



Ingenieur fakultät Bau Geo Umwelt  
Lehrstuhl für Methodik der Fernerkundung

# **Fusion of Hyperspectral Images and Digital Surface Models for Urban Object Extraction**

*Janja Avbelj*

Vollständiger Abdruck  
der von der Ingenieur fakultät Bau Geo Umwelt  
der Technischen Universität München  
zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs (Dr.-Ing.)  
genehmigten Dissertation.

Vorsitzende:

Univ.-Prof. Dr.-Ing. habil. Xiaoxiang Zhu

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. habil. Richard H. G. Bamler
2. Hon.-Prof. Dr.-Ing. Peter Reinartz  
Universität Osnabrück
3. Prof. Dr.-Ing. Markus Gerke  
University of Twente, Niederlande

Die Dissertation wurde am 16.09.2015 bei der Technischen Universität München eingereicht und durch die Ingenieur fakultät Bau Geo Umwelt am 20.11.2015 angenommen.

Die Dissertation ist erstveröffentlicht als Forschungsbericht des Deutschen Zentrums für Luft- und Raumfahrt unter: Avbelj, Janja: Fusion of Hyperspectral Images and Digital Surface Models for Urban Object Extraction, DLR-Forschungsbericht 2016, Köln, ISRN DLR-FB-2016-07 (zugleich Dissertation Technische Universität München 2015) und parallel online freigeschaltet unter <<http://elib.dlr.de/100607/>>.

Diese Dissertation ist auf dem Server der Deutschen Geodätischen Kommission unter <<http://dgk.badw.de/>> sowie auf dem Server der Technischen Universität München unter <<http://mediatum.ub.tum.de/>> elektronisch publiziert.

## Summary

Buildings are prominent objects of the constantly changing urban environment. Accurate and up to date Building Polygons (BP) are needed for a variety of applications, e.g. 3D city visualisation, micro climate forecast, and real estate databases. The increasing number of earth observation remote sensing images enables the development of methods for building extraction. For instance, Hyperspectral Images (HSI) are a source of information about the material of the objects in the scene, whereas the Digital Surface Models (DSM) carry information about height of the surface and of objects. Thus, complementary information from multi-modal images, such as HSI and DSM, is needed to provide better understanding of the observed objects. A variation in material and height is represented by an edge in HSI and DSM, respectively. Edges in an image carry large portions of information about the geometry of the objects, because they delineate the boundaries between them. Object extraction and delineation is more reliable if information content from HSI, DSM, and edge information is jointly accounted for. The focus in this thesis is on method development for BP extraction using complementary information from HSI and DSM by accounting for edge information. Furthermore, a new quality measure, which accounts for shape differences and geometric accuracy between extracted and reference polygons, is proposed.

Object and edge detection from an image is meaningful only for some range of scales. Edge detection in scale space is motivated by showing that in the same image different edges appear at different scales. Instead of deterministic edge detection, edge probabilities are computed in a linear scale space. Bayesian fusion of edge probabilities is proposed, which employs a Gaussian mixture model. The scale, at which an edge probability is computed, is defined by a confidence probability. The impact of selecting mixing coefficients in the Gaussian mixture model according to a prior knowledge or by a fully automatic data-driven approach is investigated. Main limitations of joining the edge probabilities from different datasets are the coregistration between the datasets and the inaccuracies in the datasets.

The rectilinear BP are adjusted by means of weighted least squares, where the weights are defined on the basis of joint edge probabilities. Two mathematical models for rectilinear BP are proposed, one with a strict rectilinearity constraint and the second one, which introduces a relaxed rectilinearity constraint through weighting. The experiments on synthetic images show that the model with strict constraint gives better results, if the BP under consideration are all rectilinear. Otherwise, the relaxed rectilinearity constraint through weighting balances better between the rectilinearity assumption and fitness to the data. The approximate BP are created by a Minimum Bounding Rectangle (MBR) method. A main contribution of the proposed iterative MBR method is the automatic selection of a level of complexity of MBR through analysis of a cost function.

A metric for comparison of polygons and line segments, named PoLiS metric, is defined. It compares polygons with different number of vertices, is insensitive to the number of vertices on polygon's edges, is monotonic, and has a nearly linear response to small changes in translation, rotation, and scale. Its characteristics are discussed and compared to the commonly used measures for BP evaluation. In all experiments the BP are evaluated by computing the newly proposed PoLiS metric and quality rate.

The feasibility of joining all the proposed methods in one workflow is shown through the experiment, which is carried out on 17 HSI-DSM dataset pairs with four different ground sampling distances. The main finding of the experiment is that joining the information from multi-modal images, i.e. HSI and DSM, results in better quality of the adjusted BP. For instance, even for datasets with 4 m ground sampling distance, the completeness, correctness and quality rate values of extracted BP are better than 0.83, 0.68, and 0.60. Inaccuracies of the images, such as holes in DSM or imperfect DSM for HSI orthorectification, are influencing the accuracy and localisation of edge probabilities and consequently also the accuracy of adjusted BP.



## Zusammenfassung

Gebäude sind in einem sich stetig verändernden städtischem Raum von besonderem Interesse. Genaue und immer aktuelle Gebäudeumrisse werden in einer Vielzahl von Anwendungen benötigt, wie z.B. für 3D Städtemodelle, Mikroklima-Vorhersagen oder auch Grundstücksdatenbanken. Die Entwicklung von Methoden zur Gebäudeextraktion aus Erdbeobachtungsdaten, insbesondere aus Bilddaten, wird durch deren steigende Verfügbarkeit stetig vorangetrieben. So geben Hyperspektralbilddaten (HSI) zum Beispiel Aufschluss über das Gebäudematerial in einem Gebiet, wohingegen Oberflächenmodelle (DSM) vom gleichen Gebiet Informationen über den Geländeverlauf und die Höhe der Gebäude enthalten. Um die Gebäudeeigenschaften in einem Gebiet genauer auswerten zu können, sind sich ergänzende Informationen von multimodalen Bilddaten, wie HSI und DSM, nötig. Räumliche Änderungen von Materialien erzeugen in einem HSI eine Kante wogegen eine Kante in einem DSM durch Höhenunterschiede erzeugt wird. Kanten enthalten also Informationen über Gebäudeeigenschaften, weil sie Grenzen zwischen Objekten darstellen. Die Extraktion von Gebäuden und die Bestimmung ihrer Umrisse sind zuverlässiger, wenn der Informationsgehalt aus HSI, DSM und Kanteninformationen gemeinsam in Betracht gezogen wird. Der Schwerpunkt dieser Doktorarbeit liegt in der Entwicklung von Methoden zur Extraktion von Gebäudeumrissen aus sich ergänzenden Informationen von HSI und DSM durch Betrachtung der Kanteninformationen.

Gebäude- und Kantendetektion in einem Bild sind nur in bestimmten Skalen aussagekräftig. Kantendetektion im Skalenraum ist dadurch motiviert, dass im selben Bild unterschiedliche Kanten in unterschiedlichen Skalenräumen existieren. Anstelle von deterministischer Kantendetektion werden Kantenwahrscheinlichkeiten im linearen Skalenraum berechnet und unter Verwendung eines gaußschen Mischungsverteilungsmodells eine bayesianische Fusion von Kantenwahrscheinlichkeiten durchgeführt. Die Skala in der eine Kantenwahrscheinlichkeit berechnet wird, ist durch die Konfidenzwahrscheinlichkeit definiert. Der Einfluss einer a priori gesteuerten und einer automatisch datengesteuerten Bestimmung der Mischungsgewichte im gaußschen Mischverteilungsmodell wird untersucht. Die größte Limitierung bei der Kombination von Kantenwahrscheinlichkeiten aus verschiedenen Datensätzen ist die Ungenauigkeit der Koregistrierung zwischen den Datensätzen und die Ungenauigkeit der einzelnen Datensätze.

Rechtwinklige Gebäudepolygone werden durch die Methode der kleinsten Quadrate berechnet, wobei die Gewichtung in der Ausgleichung auf den kombinierten Kantenwahrscheinlichkeiten basiert. Zwei mathematische Modelle zur Berechnung von rechtwinkligen Gebäudepolygone werden vorgestellt. Die erste Methode beruht auf einer strengen Rechtwinkligkeit und in der zweiten Methode wird die Bedingung der Rechtwinkligkeit durch Gewichtungsfaktoren relaxiert. Untersuchungen mit synthetischen Bilddaten zeigen, dass die Methode mit streng rechtwinkliger Beschränkung bessere Ergebnisse liefert, wenn die untersuchten Gebäudestrukturen tatsächlich rechtwinklig sind. Andererseits liefert die Methode basierend auf Gewichtungsfaktoren ein ausgewogeneres Ergebnis bezüglich Rechtwinkligkeit und Passgenauigkeit der Daten. Die approximierten Gebäudepolygone werden mit der Methode der minimal umgebenden Rechtecke (MBR) erstellt. Ein Hauptbeitrag zur iterativen MBR-Methode ist die automatische Auswahl einer Komplexitätsebene der MBR durch die Analyse einer Kostenfunktion.

Es wird eine Metrik zum Vergleich von Polygonen und Liniensegmenten, die PoLiS-Metrik, vorgestellt. Die PoLiS-Metrik erlaubt Vergleiche von Polygonen mit unterschiedlicher Anzahl von Kanten, ist unempfindlich gegenüber der Eckenanzahl einer Polygonseite, ist monoton und zeigt ein fast lineares Verhalten für kleine Veränderungen in Translation, Rotation und Maßstab. Die Charakteristiken der PoLiS-Metrik werden diskutiert und mit auf diesem Forschungsgebiet anerkannten Qualitätsmaßen für Gebäudestrukturen verglichen. Die Gebäudestrukturen werden in allen Versuchen durch die vorgestellte PoLiS-Metrik und eine bekannte Qualitätsrate evaluiert.

Die Möglichkeit alle vorgestellten Methoden in einen Gesamtprozess einzubinden, wird durch einen Versuch an 17 HSI-DSM Datensatzpaaren mit vier verschiedenen Bodenauflösungen gezeigt. Eine wesentliche Erkenntnis aus diesem Versuch ist, dass die Kombination von Informationen aus multimodalen Bilddaten, z.B. HSI und DSM, ein qualitativ besseres Ergebnis der ausgeglichenen Gebäudepolygone liefert. Zum Beispiel sind für Datensätze mit 4 m Bodenauflösung die Richtig-Positiv-Rate (Vollständigkeit, completeness), der positive Vorhersagewert (correctness) und die Qualitätsrate (quality rate) der extrahierten Gebäudepolygone besser als 0.83, 0.68 und 0.60. Ungenauigkeiten der Bilder, wie z.B. Lücken im DSM oder ein ungenaues DSM für die HSI Orthorektifizierung, beeinflussen die Genauigkeit und Lokalisierung von Kantenwahrscheinlichkeiten und infolgedessen auch die Genauigkeit der ausgeglichenen Gebäudepolygone.

# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Scientific Relevance of the Topic . . . . .	11
1.2	Objectives and Focus of the Thesis . . . . .	12
1.3	Outline . . . . .	13
<b>2</b>	<b>Theoretical Background</b>	<b>15</b>
2.1	Hyperspectral Imaging . . . . .	15
2.1.1	Terminology and Basic Principles of HSI . . . . .	15
2.1.2	Distortions in HSI and Their Characteristics . . . . .	17
2.2	Digital Elevation Models . . . . .	21
2.2.1	Terminology . . . . .	21
2.2.2	DEM Generation and Their Characteristics . . . . .	21
2.3	Image Fusion in Urban Areas . . . . .	24
2.3.1	Availability of the Multi-View and Hyperspectral Images . . . . .	25
2.3.2	Potential and Challenges of the HSI and DSM Fusion . . . . .	25
<b>3</b>	<b>Advances in HSI and DSM Fusion for Building Extraction</b>	<b>29</b>
3.1	Building Extraction from Remote Sensing Data . . . . .	29
3.1.1	General Workflow of Geometrical Building Extraction . . . . .	30
3.1.2	Least Squares Adjustment of Rectilinear Building Outlines . . . . .	32
3.2	Remote Sensing and Scale Space . . . . .	33
3.3	Evaluation of Building Outline Extraction . . . . .	35
3.3.1	Wild West in Evaluation Techniques for Extracted Building Polygons . . . . .	35
3.3.2	Overview of Indices . . . . .	36
3.3.3	What is Ground Truth? . . . . .	37
3.4	Summary of Research Voids . . . . .	38
<b>4</b>	<b>Fusion of HSI and DSM</b>	<b>39</b>
4.1	Edge in an Image . . . . .	39
4.1.1	Edge Detection and Edge Probability in an Image . . . . .	42
4.1.2	Scale Space Representation . . . . .	45
4.1.3	Implementation of the Discrete Linear Scale Space and Scale Space Derivatives . . . . .	46
4.2	Bayesian Fusion of HSI and DSM Based on Edges . . . . .	48

4.2.1	Abundance and Probability of an Edge . . . . .	48
4.2.2	DSM Derived Probability of an Edge . . . . .	48
4.2.3	Gaussian Mixture Model . . . . .	49
4.3	Information Content Relation of HSI and DSM . . . . .	49
4.3.1	Weighting by Prior Knowledge . . . . .	50
4.3.2	Weighting by Data-Based Confidence Level . . . . .	50
<b>5</b>	<b>Building Outline Extraction and Adjustment</b>	<b>51</b>
5.1	Approximate Building Outline Determination . . . . .	51
5.1.1	Building Region Extraction . . . . .	51
5.1.2	Building Polygon Creation and Selection . . . . .	53
5.2	Joined use of HSI and DSM for Building Polygon Estimation . . . . .	55
5.2.1	Mathematical Model for Rectilinear Polygon . . . . .	56
5.2.2	Weights in the Adjustment . . . . .	60
5.3	A New Metric for Evaluation of Polygons and Line Segments (PoLiS) . . . . .	62
5.3.1	Measures for Quantification of Similarity . . . . .	63
5.3.2	Definition of the PoLiS Metric . . . . .	70
5.3.3	Characteristics of the PoLiS Metric . . . . .	71
<b>6</b>	<b>Case Studies</b>	<b>79</b>
6.1	Data Description and Preprocessing . . . . .	79
6.1.1	HSI Sensors and Images . . . . .	79
6.1.2	DSM from Stereo Images and LiDAR Point Clouds . . . . .	82
6.1.3	Implementation and Setting of Parameters for Proposed Methods . . . . .	82
6.2	Experiments on Scale Space for Edge Detection . . . . .	83
6.2.1	Test Dataset . . . . .	83
6.2.2	Parameter Settings . . . . .	84
6.2.3	Results and Discussion . . . . .	84
6.2.4	Summary of Experiments on Scale Space for Edge Detection . . . . .	89
6.3	Experiment on Building Polygon Selection and Adjustment . . . . .	90
6.3.1	Test Dataset . . . . .	90
6.3.2	Parameter Settings . . . . .	92
6.3.3	Results and Discussion of the Model Selection Experiment . . . . .	93
6.3.4	Results and Discussion of the Adjustment . . . . .	96

---

6.3.5	Summary of Experiment on BP Selection and Adjustment . . . . .	101
6.4	Experiments on RS Images . . . . .	102
6.4.1	Test Dataset . . . . .	102
6.4.2	Parameter Settings and Preprocessing . . . . .	104
6.4.3	Results and Discussion of the Experiment on the <i>Small Area</i> . . . . .	105
6.4.4	Results and Discussion of the Experiment on the <i>Large Area</i> . . . . .	114
6.4.5	Summary of Experiments on RS Images . . . . .	121
<b>7</b>	<b>Conclusions and Outlook</b>	<b>123</b>
7.1	Conclusions . . . . .	123
7.2	Outlook . . . . .	124
	<b>Acronyms</b>	<b>127</b>
	<b>List of Figures</b>	<b>129</b>
	<b>List of Tables</b>	<b>132</b>
	<b>Bibliography</b>	<b>133</b>
	<b>Acknowledgements</b>	<b>141</b>



# 1 Introduction

*A picture is worth a thousand words!* This proverb becomes a new dimension when a picture is a Hyperspectral Image (HSI) with dozens or even hundreds of spectral bands. How many “words” is the HSI worth? The HSI is a source of information about geometry and materials of objects in the image. If besides the HSI also a Digital Surface Model (DSM) is available, which is a source of information about geometry and heights of objects, the joined information value from both datasets is further increased. Thus, the HSI and DSM carry complementary information about the covered area. Moreover, a variation in material and height is represented by edges in HSI and DSM, respectively. More reliable discrimination and delineation of the objects is possible by extracting the knowledge about the edges from these images. The information value from HSI, DSM, and the edge information together are worth more than thousands of “words”.

The topic of this thesis is the usage of the *hyperspectral images and digital surface models for urban object extraction*. In order to detect building polygons from HSI and DSM, a method for rectilinear building polygon extraction is proposed, which accounts for edge probabilities from both datasets. A new quality measure for evaluation of the extracted building polygons, called Polygons and Line Segments (PoLiS) metric, is defined and compared to the already community accepted measures.

## 1.1 Scientific Relevance of the Topic

The variety of applications, such as 3D city visualisation, micro climate forecast and monitoring, real estate databases, require up to date building polygons or models as an input (Brédif et al., 2013; Rottensteiner et al., 2014). The City Geography Markup Language (CityGML), an international standard on city models describes representation, storage, and exchange of the 3D city models (Gröger and Plümer, 2012). The European Union also issued specifications about buildings and their properties within the Infrastructure for Spatial Information in Europe (INSPIRE) framework (INSPIRE TWG BU, 2013). The 3D geometry at different Level of Detail (LOD), semantics, and material attributes of building façades and roofs is specified in both, the CityGML standard and INSPIRE framework (Avbelj et al., 2015a; Gröger and Plümer, 2012; INSPIRE TWG BU, 2013).

Air- or space-borne Remote Sensing (RS) enable periodic acquisitions of images of larger areas. The increasing number and availability of the Earth Observation (EO) RS images enables the development of methods for building extraction, which combine several datasets (Brenner, 2005; Gamba, 2014; Hu et al., 2003). The complementary information from multi-modal RS imagery can be exploited for building extraction. Typically, one of the datasets exhibits heights of the objects, e.g. DSM and Light Detection and Ranging or Light Radar (LiDAR) point cloud, and the other one the spectral characteristics of the objects in the scene, e.g. HSI and Multispectral (MS). HSI provide, in contrast to MS images, material information about building façades and roofs, as specified in CityGML and INSPIRE framework.

The geometry of the objects in the scene can be extracted from RS images. The accuracy, with which the geometry of an object can be extracted, depends on the characteristics of an image. Spatial resolution, one of the characteristics of the RS images, limits the smallest extractable detail of an object. MS images provide higher spatial (and lower spectral) resolution in comparison to the HSI. However, some air-borne HSI sensors, e.g. HyMAP, HySpex, and AVIRIS, acquire images with not only high spectral, but also better spatial resolution. Better spatial resolution enables extraction of smaller objects common for urban areas. Also the future EO HSI missions, such as EnMAP, DESIS, HypSI, motivate further developments of the methods for automatic urban object extraction from

HSI. The building extraction on the basis of material properties in HSI is a widely addressed research topic (e.g. Roessner et al., 2001; Segl et al., 2003a). Only few researchers addressed the problem of extracting polygons, and not pixel regions, of the objects from HSI (Avbelj et al., 2013a, 2015a; Brook et al., 2010; Huertas et al., 1999).

Inaccuracies of acquisition techniques of RS sensors and processing methods influence the values in the images (Eismann, 2012; Richards and Jia, 2006). These inaccuracies are usually corrected for. Nevertheless, not only random, but also some uncorrected systematic errors may remain in the images and have to be dealt with when extracting objects and their edges. These uncorrected systematic errors are a limiting factor to the accuracy of extracted objects. Thus, the methods for extracting building polygons from HSI must be robust against these uncorrected errors.

Edges in an image carry large portions of information about the geometry of objects, because they delineate the boundaries between them. If the boundaries of the same object can be extracted from multi-modal RS images with different characteristics, such as DSM and HSI, then they can be used as the basis for image fusion. The same object, mapped to an image of coarser scale, might not be extractable. Thus, object and edge detection from an image is meaningful only for some range of scales (Lindeberg, 1998). Moreover, different edges in the image can appear at different scales (Koenderink, 1984; Lindeberg, 1994; Witkin, 1984). The edge detection in RS images at different scales has a potential to provide better results than single scale edge detector algorithms (Field and Brady, 1997; Lindeberg, 1994, 1998; Lowe, 1999; Marimont and Rubner, 1998; Perona and Malik, 1990), e.g. Canny (Canny, 1986). The analysis of the necessity of scale space edge detection from RS images has not been analysed so far. The main goal of this thesis is to extract building polygons by accounting for edge information from HSI and DSM to increase their quality (Avbelj et al., 2013b). This is achieved by adjustment process of the extracted building polygons, which accounts for fused edge probability information.

The extracted polygons have to be evaluated with regard to ground truth. A shortage of standard evaluation techniques for extracted polygons has been addressed by several authors (e.g. Awrangjeb et al., 2010; Ragia and Winter, 2000; Rottensteiner et al., 2014, 2007; Zeng et al., 2013; Zhan et al., 2005). A further goal is therefore to propose a single evaluation measure, which accounts for shape and geometric accuracies of the extracted polygon, but is at the same time insensitive to the difference in LOD between extracted and ground truth polygons (Avbelj et al., 2015b).

## 1.2 Objectives and Focus of the Thesis

Urban areas are characterised by a large number of objects on a relatively small area. The diversity of objects in urban environments can be described by the variation in size, shape, height, and material. These characteristics of urban objects are well captured by HSI and DSM images. Both datasets should be jointly considered to gain a higher information value. Edge information is used to define shape and location of the objects. The focus of this thesis is method development for building polygon extraction from HSI and DSM by accounting for edge information. The quality of extracted building shapes and the improvement of the geometric accuracy, when incorporating the edge information from HSI and DSM, have to be quantitatively evaluated. Thus, a single quality measure is needed, which accounts for shape differences and geometric accuracy between extracted and ground truth polygons. Therefore, three major and one minor objectives are to be fulfilled:

**Objective 1:** *Fusion of HSI and DSM based on edge information.*

The information content of the image edges shall be extracted by computing edge probabilities in the two modalities. The fact that edges in a single image can appear at different scales has to be accounted

for by edge probability computation in scale space. Necessity and potential of edge probability computation in scale space from RS images has to be analysed. Edge probabilities extracted from different multi-modal images have to be fused. The fusion according to prior knowledge or automatically by a fully data-driven approach has to be investigated.

**Objective 2:** *Mathematical description and adjustment of rectilinear building polygons.*

The rectilinear building polygons have to be described by a mathematical model in order to be estimated by means of Least Squares (LS). The comparison of the strict and relaxed rectilinearity constraint in the model has to be investigated. The model has to allow for incorporation of edge information. Input parameters to the adjustment are approximate building polygons, which have to well represent the building outline in the image, i.e. the approximate building polygon has to balance between the details and generalisation of the building outline.

**Objective 3:** *Definition of a new metric for evaluation of extracted building polygons.*

A new metric for comparison of building polygons is needed, which accounts for positional accuracy and shape differences between an extracted and a reference polygon. The metric shall have the following characteristics. It compares polygons with different number of vertices, is insensitive to the number of vertices on polygons' edges, is monotonic, and has a linear response to small changes in translation, rotation, and scale. It is a metric in a mathematical sense. The characteristics of the new metric shall be discussed and compared to the community accepted measures, e.g. matched rates and Root Mean Square Error (RMSE).

**Minor objective:** *Joining the proposed methods (Objectives 1, 2, 3) in one workflow.*

Building polygon extraction and adjustment from HSI and DSM with various spatial resolutions shall be carried out to demonstrate the applicability of the proposed methods. The fused edge information from both datasets has to be included in the adjustment. The preprocessing steps of the HSI and the DSM such as DSM normalisation, and material map generation using unmixing of the HSI has to be carried out. The adjusted building polygons have to be evaluated by computing the newly proposed metric and commonly used evaluation measures.

## 1.3 Outline

The thesis is structured as follows. In this Chapter 1, the introduction is given, which includes scientific relevance of the topic and objectives of this thesis. Shortly, basics and characteristics of the HSI and DSM, followed by the overview of fusion of these two datasets for urban areas is given in Chapter 2. The state of the art of building extraction and evaluation and the usage of scale space in RS images are described, and research voids are identified in Chapter 3. The following two Chapters describe methodological contributions. The edge in an image and edge probability computation in scale space are introduced and the method for Bayesian fusion of edge probabilities is proposed in Chapter 4. The method for building polygon extraction together with two adjustment models and the PoLiS metric are proposed in Chapter 5. The experiments using the methods proposed in the previous two Chapters are carried out and discussed in Chapter 6. First, the necessity of the edge probability detection in scale space is shown. Second, the proposed method for building polygon extraction is applied on synthetic images and the adjustment models are compared. Then, all proposed methods are joined in one workflow and applied on real HSI–DSM dataset pairs. The results, findings, and outlook on future works are discussed in the final Chapter 7.



## 2 Theoretical Background

The basic principles and characteristics of the two types of RS data, relevant for this thesis, HSI (Section 2.1) and DSM (Section 2.2), are introduced in this Chapter. Special focus is put on the inaccuracies of acquisition techniques and processing methods to the values in the images and DSM. Usually these inaccuracies are corrected for, nevertheless next to the random errors, some systematic errors can remain uncorrected. Therefore, image processing methods must be robust towards these shortcomings. Finally, an overview of image fusion of HSI and DSM in urban areas is given (Section 2.3).

### 2.1 Hyperspectral Imaging

Hyperspectral imaging (also called imaging spectroscopy or hyperspectral RS) is measuring electromagnetic radiation in tens or hundreds of mostly adjacent narrow spectral bands with increased spectral resolution in contrast to MS sensors with only few spectral bands (e.g.  $\leq 12$ ) of larger bandwidth and gaps between bands (Keshava, 2003; van der Meer, 2001). The abbreviation HSI is used for hyperspectral imaging and hyperspectral image. HSI sensors measure electromagnetic radiation in the optical spectral region (about  $0.4\text{-}14\ \mu\text{m}$ ). HSI used in this thesis are a result of a measured reflected sun radiation in Visible and Near (VNIR) (about  $0.4\text{-}1.1\ \mu\text{m}$ ) and/or Short Wavelength Infrared (SWIR) (about  $1.1\text{-}2.5\ \mu\text{m}$ ) spectral regions (Figure 2.1b). Therefore, the described principles and characteristics of the HSI are restricted to the passive HSI sensors acquiring data in VNIR and SWIR spectral regions. They differ in some aspects from the passive Thermal Infrared (TIR) HSI sensors measuring emitted thermal radiation (Eismann, 2012).

#### 2.1.1 Terminology and Basic Principles of HSI

HSI from RS platforms simultaneously capture material information as well as spatial information of the observed scene. The resulting image can be regarded as a Three-Dimensional (3D) dataset, named also hypercube (Figure 2.1a), and is composed of  $N$  grey scale images, where  $N$  is the number of spectral bands or channels of the HSI. The spectral signature of each spatial pixel carries information of the surface contained in it. It depends on the chemical composition of the material, vibrational and electronic resonances of molecules of the material, microscopic surface and volumetric properties (Eismann, 2012, p.6).

The term pixel is used to describe the smallest element of a digital image, which represents a resolution cell in object space. The optical sensors are usually designed in a way that the maximum spatial sampling distance is approximately the same as a resolution cell. Thus, in optical imaging community and in this thesis, a pixel in an optical RS image is considered to describe also the resolution.

#### Material Identification from HSI

Materials present in a scene and captured in a HSI can be determined by computing a similarity between the image and reference spectra, using unmixing techniques, or classification techniques. A prerequisite for material identification is a known set of reference spectra of the materials. The existing spectral libraries (Baldrige et al., 2009; Clark et al., 2007) include laboratory measured spectra with high spectral resolution. Before image spectra can be compared to the laboratory measured spectra, the HSI must be atmospherically corrected to ground reflectance values and laboratory spectra resampled to the spectral resolution of the HSI. To evade the complex atmospheric correction (Subsection 2.1.2), the reference spectra can be manually or automatically (Plaza et al., 2004) extracted from the HSI.

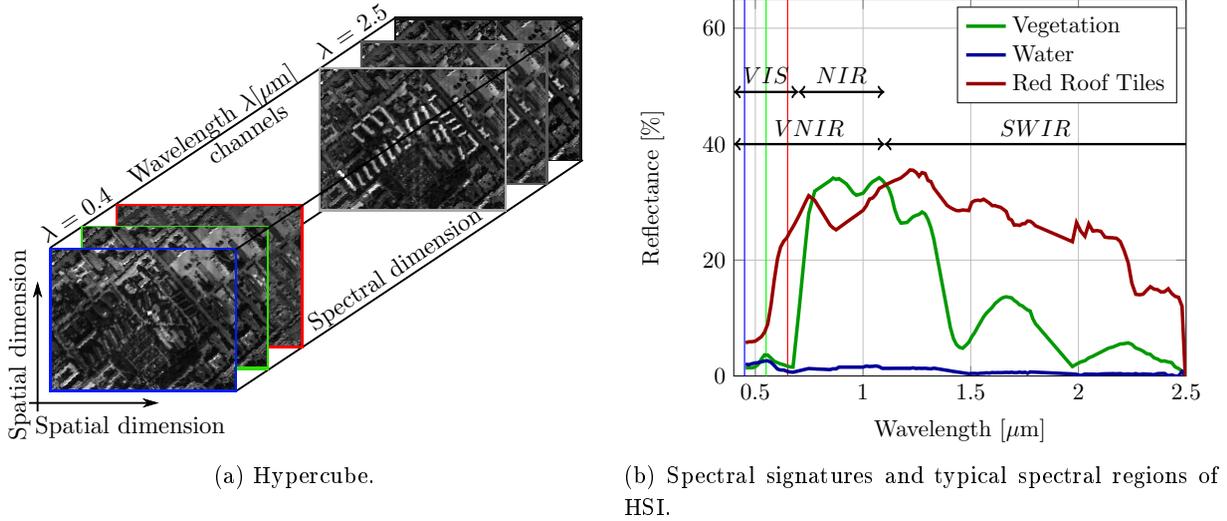


Figure 2.1: A HSI HyMAP image (Figure 2.1a) and three spectral signatures collected from it (Figure 2.1b). Six channels of HSI are shown, i.e. blue, green, and red in Visible (VIS) (a colour of a boundary of each channel is with respect to the wavelength colour), and three in SWIR portion of spectrum (grey boundary) (Figure 2.1a). Figure 2.1b shows spectral signatures of vegetation (green), water (dark blue) and red roofing tiles, which were collected from the HyMAP image. Left-right pointing arrows are showing the approximate extent of VIS, Near Infrared (NIR), VNIR, and SWIR portions of the spectrum.

The similarity between two spectra can be computed by different measures (Cerra et al., 2012; Robila and Gershman, 2005), e.g. (normalised) Euclidean distance, Mahalanobis distance, or through data compression (Cerra et al., 2011). Some of them are designed especially for HSI, for instance Spectral Angle Distance (SAD), which is also named cosine distance or spectral angle mapper (SAM) (Kruse et al., 1993),<sup>1</sup> spectral information divergence (Chang, 2000), or spectral correlation mapper (De Carvalho and Meneses, 2000).

In HSI unmixing literature the term *endmember* is used to describe spectra of so called “pure” or elementary material. The term *reference spectra* is used here instead of the term endmember because of the following two reasons. First, the concept of pure material is misleading, because materials with the same chemical composition can have different spectral response due to e.g. surface structure or reflection characteristics. Second, the definition of an endmember varies depending on the application (Bioucas-Dias et al., 2012; Ma et al., 2014) and the spatial resolution of the HSI.

Some of the pixels in HSI consist of more than just one material and are called mixed pixels. A Linear Mixing Model (LMM) assumes that any observed pixel  $\mathbf{x} \in \mathbb{R}^N$ ,  $n = 1, \dots, N$  in the HSI ( $N$  is the number of bands of the HSI) is a linear combination of the  $M$ ,  $m = 1, \dots, M$ , reference spectra vectors  $\mathbf{s}_m \in \mathbb{R}^N$  weighted by their abundances  $\mathbf{a} \in \mathbb{R}^M$  (e.g. Keshava, 2003). The  $M$  column vectors  $\mathbf{s}_m$  are arranged to a matrix with reference spectra  $\mathbf{S} \in \mathbb{R}^{N \times M}$ . LMM is physically well founded if the mixing scale is macroscopic and incident light interacts with just one material (Bioucas-Dias et al., 2012). A non-linear mixing model should be considered, when e.g. light scattered by multiple materials in a scene has a prominent influence on the measured spectra (Keshava and Mustard, 2002; Ma et al., 2014). Let a HSI be denoted as a matrix  $\mathbf{X} \in \mathbb{R}^{N \times P}$ , with columns holding the  $P$ ,  $p = 1, \dots, P$  spectral vectors  $\mathbf{x}_p \in \mathbb{R}^N$  of the  $P$  pixels, then the LMM is given by

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{N} \quad (1)$$

<sup>1</sup>A reader with computer vision background should not confuse abbreviation of the SAD as a distance measure with a sum of absolute intensity differences used in stereo matching methods.

where  $\mathbf{A} \in \mathbb{R}^{M \times P}$  is the abundance matrix, and  $\mathbf{N} \in \mathbb{R}^{N \times P}$  is a matrix of noise and modelling errors. Abundances of each material can be considered as a material map of this material. Every material map has the same size like the size of any channel of the HSI.

Hyperspectral unmixing is the process of decomposition of the mixed pixels into the fractions or abundances of the constitute reference spectra (Keshava and Mustard, 2002). Let us assume a) LMM, b) Gaussian noise, and c) known all reference spectra  $\mathbf{S}$ ,  $N > M$ . A Non-Constrained Least Squares (NCLS) solution for the abundance matrix  $\mathbf{A}$  minimises the cost function  $\|\mathbf{X} - \mathbf{SA}\|_2$

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{SA}\|_2, \quad (2)$$

where  $\|\cdot\|_2$  is the Euclidean  $L_2$  norm. The abundances should be positive to be meaningful in a physical sense, thus the cost function in Equation (2) is for the so called Non-Negative Least Squares (NNLS) *subject to* abundance non-negativity constraint

$$\mathbf{a}_{m,n} \geq 0, \quad m = 1, \dots, M, \quad n = 1, \dots, N. \quad (3)$$

Additionally to the non-negativity constraint, the cost function in Equation (2) can be also *subject to* abundance sum-to-one constraint for every column  $p = 1, \dots, P$  of the  $\mathbf{A}$

$$\sum_{m=1}^M \mathbf{a}_m = 1. \quad (4)$$

This minimisation problem is called Fully-Constrained Least Squares (FCLS). The sum-to-one constraint also has a physical meaning, because the sum of the fractional abundances for every spectral vector  $\mathbf{x}$  should be exactly one. However, if the assumption c) does not hold, and not all reference spectra in a scene are known, then the FCLS will cause overfitting of the model. Moreover, if an over-complete reference spectra matrix ( $N < M$ ) from e.g. spectral libraries is given, then the sparse unmixing methods using  $L_0$  or  $L_1$  minimisation can be applied (Iordache et al., 2010). An extensive overview about HSI unmixing approaches can be found in Bioucas-Dias et al. (2012).

Alternative to the physical foundation of the LMM is the geometric or signal processing interpretation, which assumes that the reference spectra lie on the extremities of the  $M - 1$  simplex (Keshava, 2003; Ma et al., 2014). Several methods for automatic reference spectra extraction, also called blind source separation in the signal processing community, use this assumption.

## 2.1.2 Distortions in HSI and Their Characteristics

Acquired HSI must be corrected for geometric and radiometric distortions. The geometric distortions cause incorrect location of the acquired area in an image, whereas the radiometric distortions cause distorted spectra (brightness of the pixels) of the captured area. The corrections, registration, and especially coregistration of the images are important factors within an image fusion framework. In this Subsection the main influences on the HSI are summarised.

### Spectral Distortions Due to the Atmosphere

The objective of hyperspectral imaging is to quantify the composition of the observed objects, thus the measured signal must be calibrated to physical meaningful units. The following quantities are defined for optical RS images.

◇ *Digital number* [ ] is a dimensionless quantity measured by a sensor, e.g. MS, HSI.

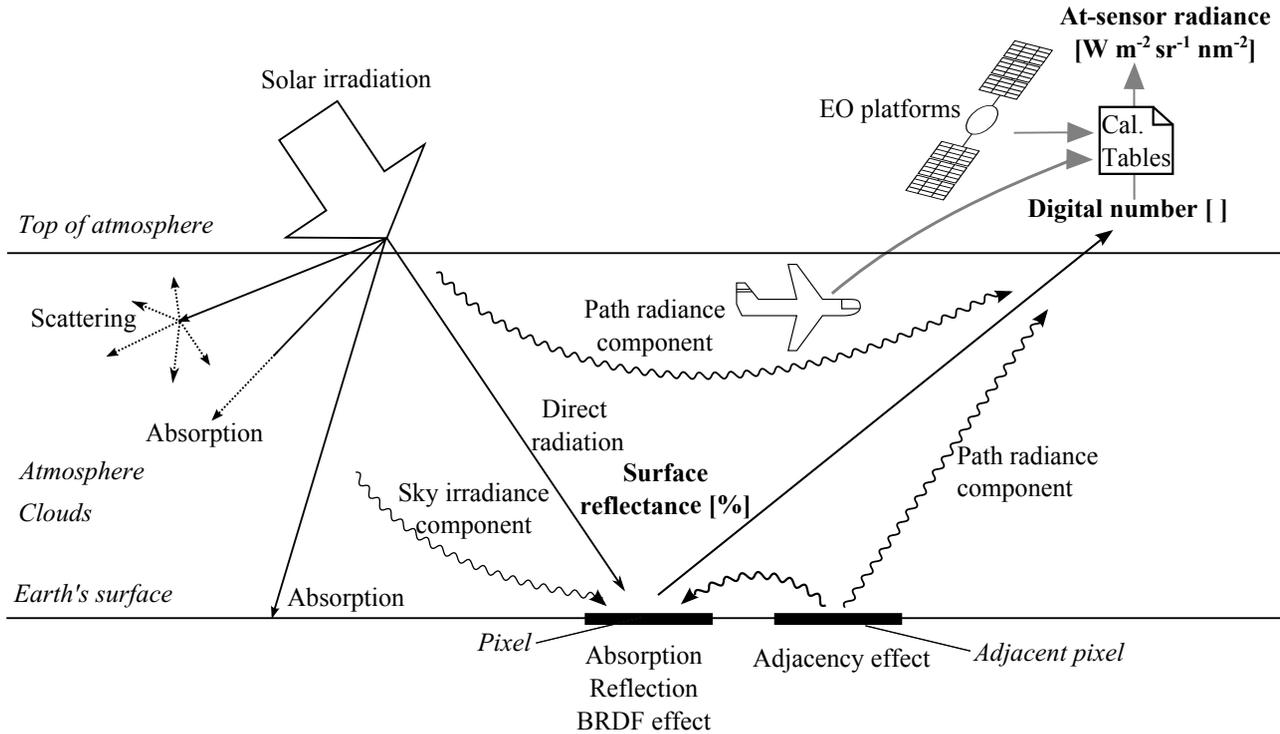


Figure 2.2: Atmospheric and other influences on the radiance measured by a passive sensor. A measured radiance of an area on ground representing an image pixel is a result of different paths through the atmosphere (wavy lines). It depends also on the incident angle, surface properties, and sensor capabilities. Actual quantity measured by a sensor is a Digital Number (DN), which can be converted to at-sensor radiance by accounting for sensor-specific systematic influences. The surface or ground reflectance are additionally corrected for the atmospheric influences. The figure is based on Richards and Jia (2006, p.28) and Eismann (2012, p.12).

- ◇ *At-sensor radiance* has physically meaningful units, energy in unit time, unit area, unit solid angle, and unit wavelength [ $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$ ]. It is computed from DN by accounting for sensor-specific systematic influences. These calibration values are given by the producer of the sensors and by carrying out laboratory and in-flight calibration.
- ◇ *Top-of-atmosphere reflectance* [%] can be converted from at-sensor radiance values, e.g. for VNIR and SWIR spectral regions the radiance is divided by the incoming solar energy.
- ◇ *Ground or surface reflectance* [%] values are corrected also for the atmospheric influences, e.g. reflection, absorption, and scattering caused by clouds, particles, and absorption of some gases in the atmosphere (Figure 2.2).

The down-welling radiation as well as reflected radiation are affected by the atmospheric influences. If the path through the atmosphere is longer, these influences are larger. Scattering by aerosols and molecules in the atmosphere is the dominant component of the radiometric distortions (Richards and Jia, 2006, p. 27–35). Figure 2.2 shows the main atmospheric influences on the measured radiance values of a single pixel.

The reflected signal is influenced also by the properties of the surface. The Bidirectional Reflectance Distribution Function (BRDF) describes the reflection of a non-Lambertian surface (opaque) under varying solar and viewing geometry. Moreover, radiance values should also be corrected for *adjacency effects*, i.e. reflections and scattering from the neighbouring areas to the target. This effect has a larger influence in the areas with high-raising objects and rugged terrain, i.e. where the elevation differences on a small area are significant. For instance, a wall of a building can cause multiple reflections of incoming solar radiation, which increases the proportion of adjacency effect in measured radiance.

## Distortions Due to the Acquisition Technique: Whisk- and Push-Broom HSI Sensors

A typical acquisition technology for collecting HSI are push-broom line scanners, which use the movement of a platform to collect the along-track spatial dimension of the image. One row of an image, collected at the same time, is approximately perpendicular to the flight direction. Two examples of HSI push-broom sensors are the air-borne HySpex (Norsk Elektro Optikk AS, 1985) and the future space-borne EnMAP (EnMAP, 2015). Another acquisition technology are whisk-broom sensors, such as the air-borne HyMap sensor, which reflect incoming light with rotating optics onto a single linear detector and collect the data across-track. This causes scan time distortion, because the data of one image line are collected from the S-shape area on the ground. The whisk-broom sensors have shorter dwell time causing lower Signal to Noise Ratio (SNR) in comparison to the push-broom sensors. Yet, the latter require extensive calibration, because each row of photo detectors of the Two-Dimensional (2D) array is effectively its own spectrometer. The limited dwell time and the moving scanning optics of the whisk-broom sensors are the main reason that the push-broom HSI sensors are preferred for space-borne missions. However, the opening angle of the optics or Field of View (FOV) is usually larger for the whisk-broom sensors, and therefore also (for the same altitude of a platform) the swath width. Another acquisition technology are frame sensors, which collect a spectral image in two spatial directions at the same time. Frame HSI sensors have lower SNR and are more commonly used on unmanned aerial platforms. Therefore, this acquisition technology is not further discussed.

For push-broom sensors, the incoming light passes through a slit and is diffracted by a grating element or a prism into the wavelength components, which fall onto the 2D array of a (photo) detector, shortly referred to as 2D array. The 2D array has a spectral and a spatial dimension. The number of detectors in spatial dimension of the 2D array is the number of pixels of an image in across-track direction, and the number of detectors in the spectral dimension is equal to the number of spectral channels.

The optical properties of an HSI sensor cause that the HSI are distorted. Therefore, the spectral and geometric (spatial) characteristics of an optical sensor must be measured. The optical properties are connected to the detector and acquisition technology (Yokoya et al., 2010), i.e. linear detector for whisk-broom sensors and 2D detector array for push-broom sensors. Spectral characteristics of detectors are described by a Spectral Response Function (SRF). Let us assume that the SRF of every detector can be described by a Gaussian. Then, the SRF of every detector is given by a central wavelength, i.e. peak sensitivity of a detector, and the Full Width at Half Maximum (FWHM). The geometric properties of the detector are described by the Line Spread Function (LSF). The LSF are different for along- and across-track directions and are characterised by the centre angles and their FWHM (Baumgartner et al., 2012; Mouroulis et al., 2000).

The response of an imaging system to a point source is described by the Point Spread Function (PSF). A simplified representation of the 2D PSF (in the zeroth order) is an ellipse. The semi-axes of the ellipse are here referred to as spatial and spectral semi-axis. Let a regular grid be assumed. If no distortions are present, the peaks of the PSF are aligned in spectral and spatial directions and also to the array. The descriptions of these misalignments and their effects are summarised below according to Gómez-Chova et al. (2008), Mouroulis et al. (2000), Yokoya et al. (2010), and Baumgartner et al. (2012).

The following two effects in MS and HSI appear in imagery acquired with whisk- and push-broom imaging sensors.

- ◊ *Smile effect* is a shift in wavelengths in the spectral domain. The smile effect influences the knowledge about the central wavelength and is different for every spectral channel.

- ◇ *Variation in the length of the spectral semi-axis of the PSF* in the spatial domain. It affects the shape of the SRF, i.e. the knowledge about FWHM.

Another two effects are peculiar for images acquired with the push-broom imaging sensors.

- ◇ *Keystone effect* is a shift between pixels in the spatial domain. For every image pixel, the peaks of the PSF in spatial direction are not aligned. This effect causes spectral contamination of the spatially adjacent pixels, and affects especially the spectral values of the pixels on the boundaries of two spatial objects, and object of a size about a pixel. The extent of this effect is dependent on the observed scene, i.e. objects in the scene.
- ◇ *Dependence of the length of the spatial semi-axes of the PSF to the wavelength* is caused by e.g. the diffraction, and influences more the channels of longer wavelengths. This effect causes, similar as the keystone effect, the spectral contamination of the spatially adjacent pixels.

Spectral contamination is wavelength dependent and causes mixtures, which have no physical meaning. For instance, given two spatially adjacent pixels with two different materials, i.e.  $A$  and  $B$ . In presence of the keystone effect, the spectra  $A$  of a pixel contaminated with the spectra  $B$  from the adjacent pixel is not the same spectral mixture as if both materials would be present in an area of one pixel (Mouroulis et al., 2000). Thus, such spectrally contaminated spectra cannot be correctly unmixed by spectral unmixing methods.

## Geometric Distortions

There are various geometric distortions influencing the acquired RS imagery with respect to a mapping frame (Müller et al., 2002). They are caused by sensor characteristics, platform motion and terrain relief (Figure 2.3a, Müller et al. (2010)). The geometric distortions in RS images are corrected in a rectification process (Figure 2.3b). Commonly, the RS images are georeferenced, i.e. transformed into a map coordinate system. The actual correction of the geometric distortions requires resampling and interpolation of the original image, which both influence the geometry as well as the spectral values of the pixels.

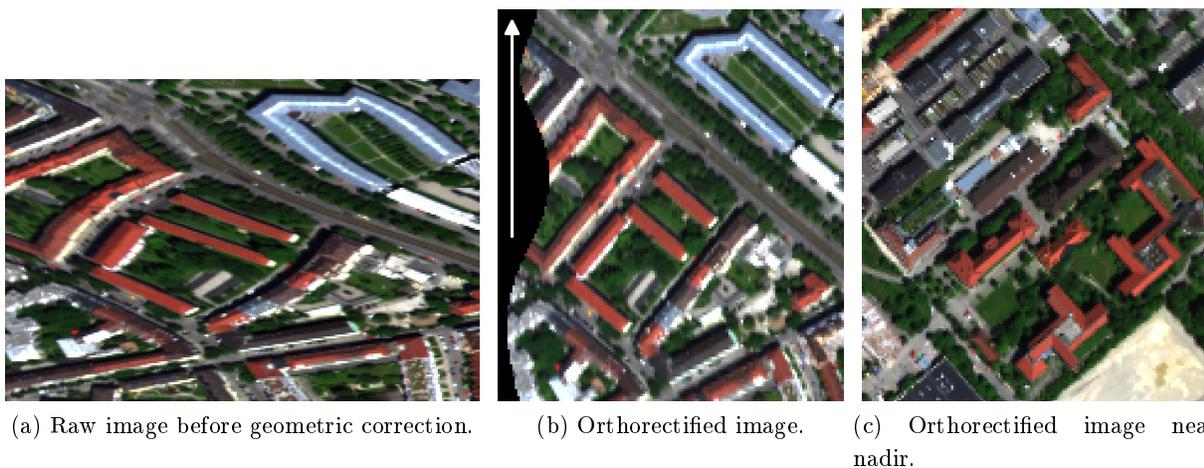


Figure 2.3: Geometric distortions of an optical RS image. An image before (Figure 2.3a) and after (Figures 2.3b and 2.3c) the orthorectification process. The image was acquired by the HySpex sensor, an air-borne push-broom HSI sensor with  $34^\circ$  FOV. The white arrow points into the flight direction. Figure 2.3b shows a geometrically corrected detail from the border of a scan line, where façades of the buildings caused by inaccurate DSM are still partially seen. This can be observed in as white areas on building boundaries. This effect is not seen in an image from the central part of the scan line, where the viewing angle is looking nearly at nadir (Figure 2.3c).

For systems with whisk- or push-broom acquisition technology, every pixel of an image in across-track direction is collected at a different viewing angle. On a rugged terrain surface two geometric distortions are difficult to correct due to this side-looking geometry and (in)accuracy of the model of the surface model required for the correction. First, named *relief displacement* is a shift in an object's position in an image due to the object's height above (or below) ground and viewing direction of the sensor. Let us assume an accurate and error free model of a terrain surface, e.g. DSM, is available for relief displacement correction. If this DSM is not precisely registered to the image, which exhibits rugged terrain, the correction will also cause relief displacement. The relief displacement is caused also by inaccurate DSM. Second, the high-rise objects are imaged from a side, and might obstruct the sensor view on other objects. The occlusion areas are interpolated in the correction process, whereas the sides of the objects can still be partially seen in a corrected image (Figure 2.3b). These two influences are larger on the extremities of the scan line and are not present at the nadir or on flat terrain. They are more prominent for higher objects and for sensors with larger FOV.

## 2.2 Digital Elevation Models

### 2.2.1 Terminology

Digital Surface Model (DSM) is a digital representation of terrain surface and all objects on it, e.g. buildings and trees, whereas Digital Terrain Model (DTM) is a digital representation of a surface without objects on it. The normalised Digital Surface Model (nDSM) exhibits only the heights of the objects above ground, and can be described by a simplification “subtraction of DTM from DSM”. The term Digital Elevation Model (DEM) is used as a generic term for all digital surface representations. The DEM defined by height values for a selected set of planar coordinates are also referred to as 2.5D DEM. The height, also called elevation is measured with respect to a high reference point or surface. They can be represented by a regular grid, i.e. raster, by an irregular grid such as triangular irregular network, or point clouds. The terms DEM, DSM, and DTM are used, when it is referred to the representation of terrain surface in a regular, equally spaced grid. Three common data types to generate DEM are stereo optical images, LiDAR point clouds (Subsection 2.2.2) and Synthetic Aperture Radar (SAR) images. The accuracy of the input data, as well as the method for DEM generation influence the accuracy of the DEM.

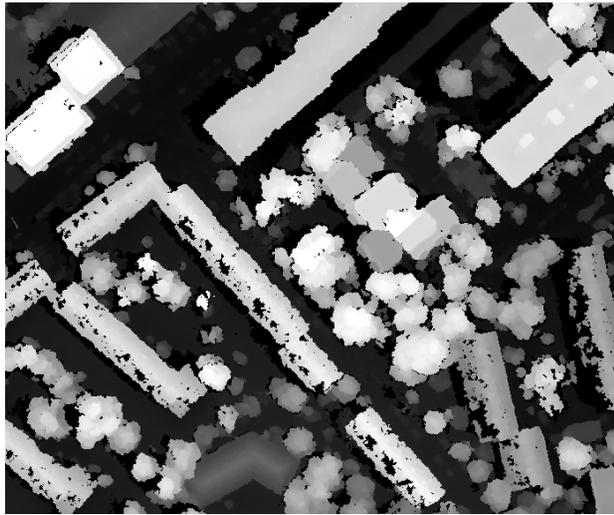
The surface model, most often a gridded DEM, is a necessary input for orthorectification of all optical RS imagery. Its accuracy and errors influence the radiometric and geometric accuracy of the orthorectified image (Section 2.1).

### 2.2.2 DEM Generation and Their Characteristics

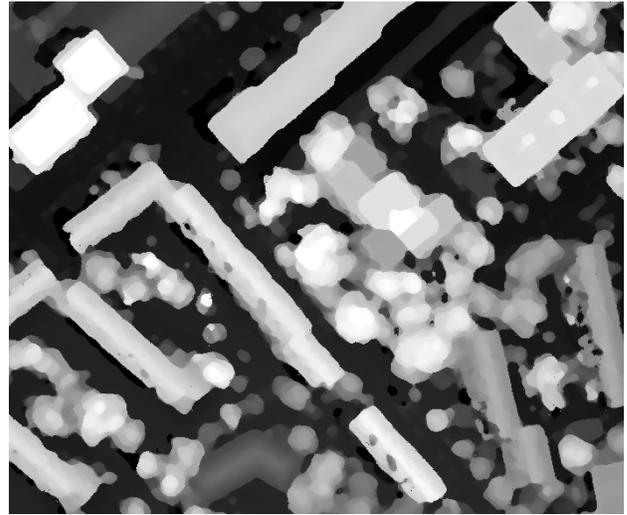
The DEM can be generated from air- or space-borne RS data or terrestrial measurements. The latter are due to the high acquisition costs seldom used for larger areas and are not further discussed. In the following Subsections, methods, acquisition principles, and possible sources of errors of the DEM from stereo optical imagery and LiDAR point cloud acquisition are discussed. Alternatively, the DSM (also of a larger coverage) can be generated by means of interferometric SAR.

### Dense-Matching

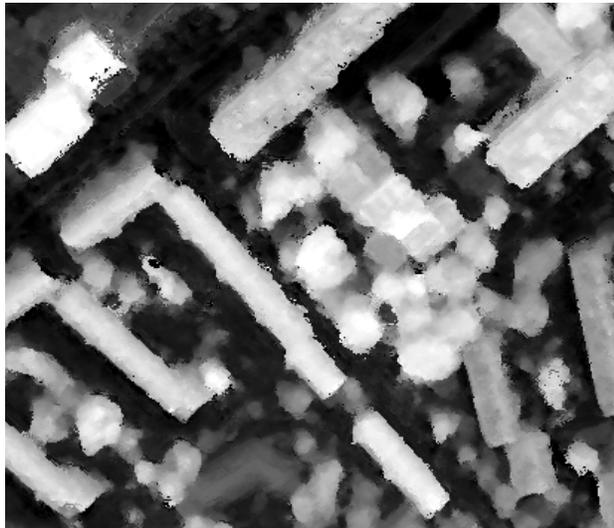
Stereo matching is one of the most active research areas in the field of computer vision and photogrammetry (Scharstein and Szeliski, 2002). The descriptions here are focused on the dense stereo matching



(a) SGM DSM from 3K images with holes.



(b) Semi-Global Matching (SGM) DSM from 3K images without holes and smoothed with median filter.



(c) SGM DSM from WV-2 images.

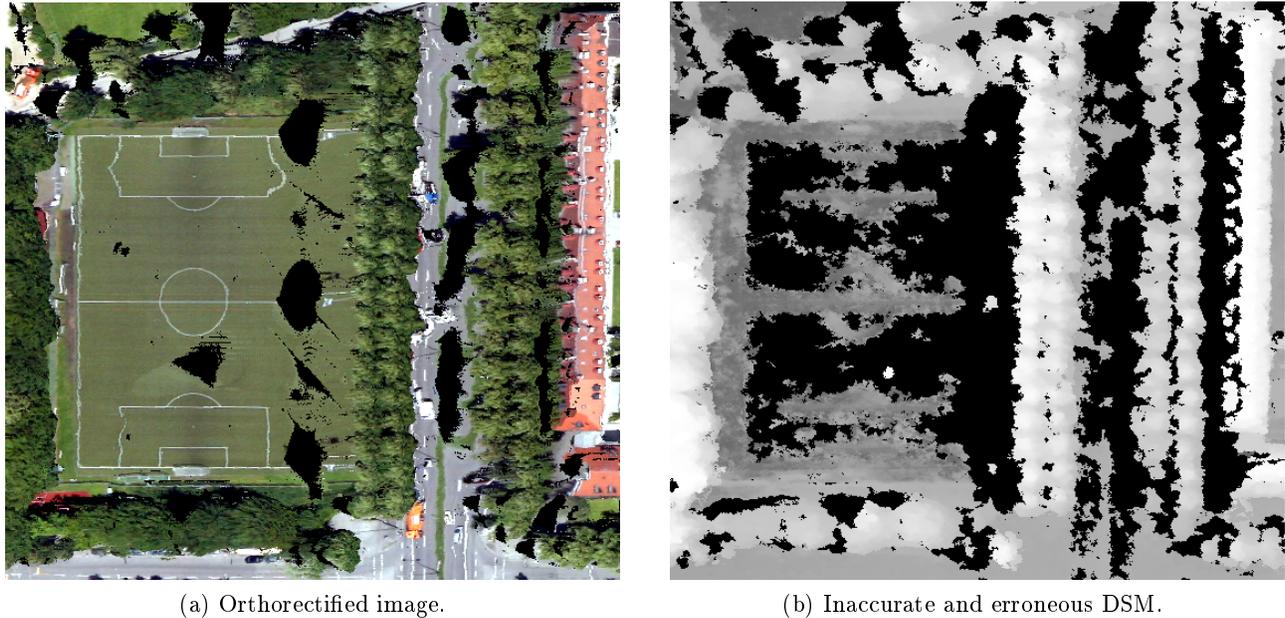


(d) RGB composite image of the urban area.

Figure 2.4: DSM of an urban area (Figure 2.4d) computed using SGM method. The SGM DSM (Figure 2.4a) is calculated from stereo images acquired by the 3K camera system (see Subsection 6.1.2, p.82 for details). Missing values (holes) are filled and the DSM is smoothed with a median filter (Figure 2.4b). Figure 2.4c shows SGM-DSM from WorldView-2 (WV-2) images. The pixel size of 3K DSM is 0.3 m (Figures 2.4a and 2.4b), and the pixel size of the WV-2 DSM is 0.5 m (Figure 2.4c). All DSM are colour coded from low (black) to high (white).

methods with known camera geometry and their application to the 3D surface model computation. The DSM from stereo imagery used in this thesis are computed by Semi-Global Matching (SGM) method described below.

A prerequisite for DEM computation are at least two images of the same scene (object) taken from different directions. First, the correspondence between the images is found for every pixel in an image based on their intensity values, i.e. area correlation methods. Alternatively, the correspondence can be based on extracted features such as Scale-Invariant Feature Transform (SIFT) features, or combination of area and feature based methods. Second, the depth map is computed, where each pixel in the depth map contains the value of the distance between the camera centre to the corresponding 3D point in the scene. Dense depth maps are a result of the area correlation methods, in contrast to the sparse depth



(a) Orthorectified image.

(b) Inaccurate and erroneous DSM.

Figure 2.5: Orthorectified view of the nadir image of the 3K camera (Figure 2.5a). It is orthorectified with erroneous DSM with holes (Figure 2.5b, black areas). The holes in DSM are due to the texture-less areas on the football field and red roofs, occlusions, etc.

maps, which are a result of matching only features (Veksler, 2001). Finally, the DEM is computed by projecting the depth map using camera geometry into the map coordinate system. Errors in dense stereo matching DEM can occur due to repetitive textures (patterns), texture-less regions, shadows and low contrast regions, specular reflection (e.g. glass, solar panels), clouds, haze, occlusions of the objects or invisible parts of scene, and moving objects, e.g. cars, people (Hirschmüller, 2005; Scharstein and Szeliski, 2002; Veksler, 2001).

The SGM algorithm is one of the dense stereo matching methods, which performs pixel-wise matching and approximation of the 2D smoothness constraint (Hirschmüller, 2005). This method is robust against illumination changes and accurate on the boundaries of an object when appropriate smoothing term for pixel-wise matching is chosen. Nevertheless, some holes in the DEM mainly due to occlusions, texture-less, and low contrast regions might still occur. In postprocessing steps, these holes can be interpolated or filled from existing DEM of the same area. Additionally, a whole DEM can be smoothed by e.g. median filter to remove possible outliers. Note that any smoothing or interpolation of the data introduces some inaccuracies to the DEM. Even though SGM performs well on the boundaries of objects, a typical SGM DEM of high-rise object areas still exhibits wave-like shaped boundaries of buildings (Figure 2.4), due to the smoothness constraint.

## LiDAR Point Cloud

The basic principle of a LiDAR system is the determination of the distance of a target by emitting pulses of a laser light and measuring the time of arrival of the reflected light. Moreover, the LiDAR sensors record also the intensity of the backscattered light. Air-Borne Laser Scanner (ALS) systems are mounted on air-borne platforms and include LiDAR system as well as other instruments, e.g. Global Navigation Satellite System (GNSS)/Inertial Navigation System (INS) unit for attitude determination. For RS applications such as topographic mapping, the LiDAR systems, which emit the pulses in NIR spectrum (usually between  $1.040\text{-}1.065\ \mu\text{m}$ ), are used (Gatzliolis and Andersen, 2008). There are two types of LiDAR sensors according to the quantification of the recorded backscattering light. First,

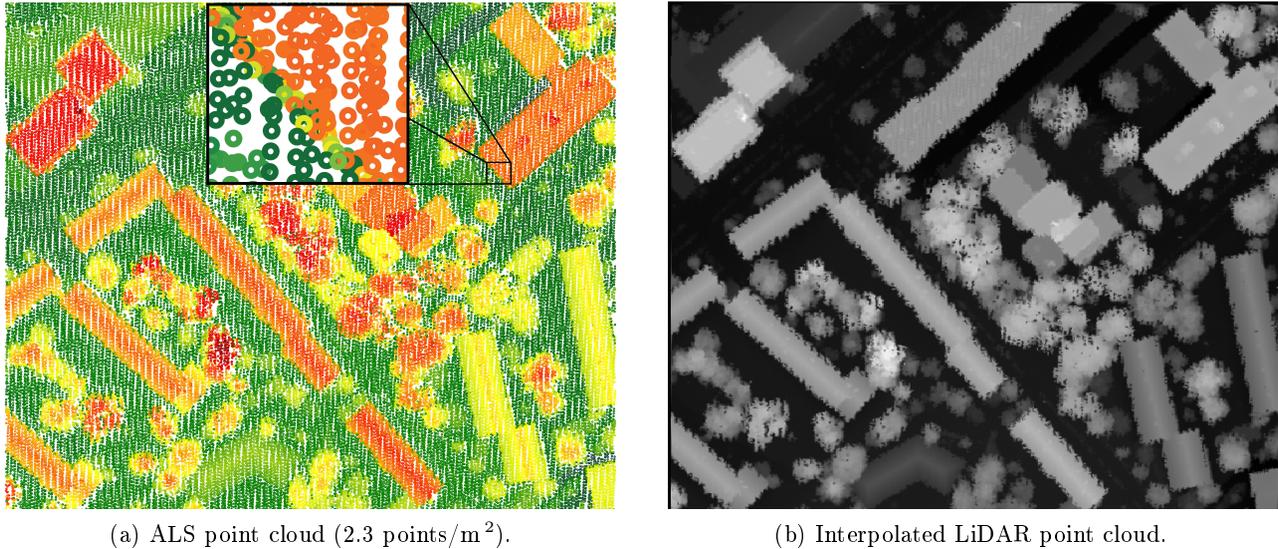


Figure 2.6: First response LiDAR point cloud (Figure 2.6a, green-low, red-high) and DSM interpolated from this LiDAR point cloud by bilinear method (Figure 2.6a). The average point density of the LiDAR point cloud is about 2.3 points/m<sup>2</sup>, and is interpolated to the 0.3 m gridded DEM. A zoomed-in point cloud demonstrates the effect of the planar error of the points on a boundary of a high-rise object. This error together with the irregular scanning pattern is propagated to the interpolated LiDAR DSM and can be observed as zig-zag pattern on the boundaries of high-rise objects (Figure 2.6b).

so called full waveform LiDAR sensors that record the reflected light nearly continuously. Second, the discrete-return LiDAR sensors that record the backscattered light at defined intervals, e.g. as a first (Figure 2.6a) and last return. The points in the LiDAR point cloud are not regularly spaced, in comparison to the DEM computed by the dense stereo matching methods. However, the LiDAR point clouds can be interpolated to a regular grid by e.g. polynomial interpolation, inverse distance weighting, kriging, and are referred to as LiDAR DEM.

Two main categories of errors of the LiDAR point cloud are elevation and horizontal errors. The measured elevation error can occur due to the sensor system and platform characteristics, e.g. analysis of the waveform, identification of the return position, pulse length, and position error measured by the GNSS/INS unit. The horizontal error has its main sources in flying height accuracy and reflections from the sloped surface (Hodgson and Bresnahan, 2004). Even if the distance from a sensor to a sloped surface is measured without an error, the planar error of such measured response introduces an apparent systematic error in the measured distance. This error has a similar behaviour like the relief displacement in optical imagery and is most prominent on the boundaries of high-rise objects. The zoomed in area in Figure 2.6b demonstrates this effect. The gridded LiDAR DEM include additionally the interpolation error.

## 2.3 Image Fusion in Urban Areas

Image fusion is a process of integrating two or more images and/or their inherent information into a single output, which includes enhanced “desired information content” compared to any of the input images. The objective of the image fusion and therefore “desired information content” of the fused result is highly dependent on the application (Chaudhuri and Kotwal, 2013, p.19–20). A prerequisite for any image fusion is a good coregistration between all the used datasets.

Urban areas exhibit large variety of objects on a relative small area, which vary in material, texture, height, shape, size, etc. Therefore, image fusion from multi-modal sensors is a necessity to enable or improve the robustness and accuracy of extracted information about urban areas (Gamba, 2014).

There are numerous ways to classify image fusion techniques, for example according to

- ◊ domain, e.g. spatial or frequency domain,
- ◊ resolution, e.g. resolution enhancement method such as spectral enhancement of MS image by using lower-spatial resolution HSI,
- ◊ modality also called “multi-sensor”, e.g. fusion of MS images, DEM, and HSI,
- ◊ acquisition time also called “multi-temporal”, e.g. images of the same area acquired at different times, often used for change detection, and
- ◊ processing level: pixel-, object-, and decision-based methods, or pixel- and feature-based methods.

The main aim of photogrammetry regarding the urban environment is to model it in 3D with all the characteristics (Brook et al., 2013), for instance as proposed by CityGML Standard (Gröger et al., 2012; Gröger and Plümer, 2012). The built urban environment comprises the 3D objects in it. Therefore, many data fusion methods modelling urban environment use height data, e.g. DEM or LiDAR point clouds, next to at least one more dataset of different modality.

### 2.3.1 Availability of the Multi-View and Hyperspectral Images

The availability and coverage of the data from stereo- and multi-view optical imagery has increased due to a larger number of air- and space-borne platforms carrying high spatial resolution cameras (Figure 2.7), e.g. WV-2, Pleiades, UltraCam. Next to these sensors, also the HSI sensors with higher spectral, but in comparison to MS imagery lower spatial resolution exist or are planned for EO missions. The main trend in EO optical sensors is going towards development of even higher spectral resolution (HSI) sensors, as well as into the direction of a larger number of lower-cost fleet of satellites, e.g. SkySat by SkyBox Imaging. Another direction of development is the area of low-cost optical systems carried on unmanned aerial platforms.

### 2.3.2 Potential and Challenges of the HSI and DSM Fusion

The main potential of HSI and DEM fusion for urban areas originates from different modality and characteristics of the images. The HSI enable e.g. material identification, the DSM the height characterisation of the objects, and both image types carry information about the objects. The fusion of these two data types includes applications for identification of the 3D objects in the scene from DSM and assignment of the material properties to these objects, detecting potentially dangerous materials, and/or change of the materials, and classification based on spectral, spatial, and/or geometric features. In view of building extraction from HSI and DSM, the influences of the geometric distortions and data characteristics are addressed in the following paragraphs.

The orthorectified HSI rarely include information of the used DSM, its accuracy, and coregistration error. Even if the accuracy of the DEM is given, it might vary significantly within the same dataset. Such an example are lines of a soccer field in Figure 2.5, which vary from nearly straight to very curvy lines, due to the erroneous DSM used for orthorectification. The marks on the soccer field, boundaries of the streets, and the building outline should be straight lines, but appear as lines with different curvatures. The wavy right boundary of the building on the right is mainly due to the wavy shape of the DSM. On the left, the building boundary is wrongly projected due to the large hole in the

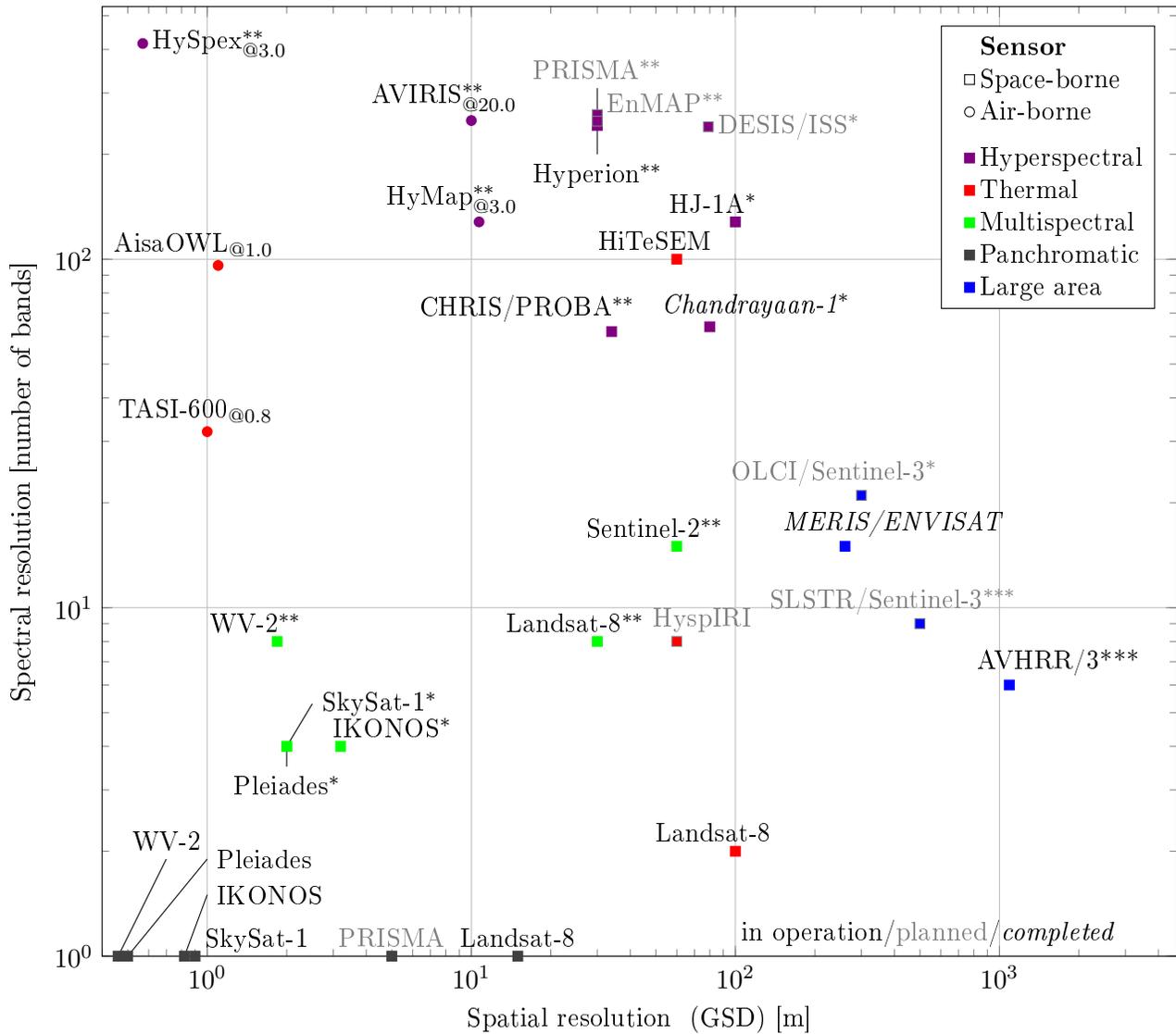


Figure 2.7: Spectral resolution as a function of spatial resolution for selected EO air- (circular marker) and space-borne (square marker) sensors. Colour of a marker denotes a type of a sensor, i.e. violet for HSI, red for TIR, green for MS, dark grey for Panchromatic (PAN), and blue for large area assessment sensors. A black text of a sensor/platform is for operational, grey text for planned, and black italic text for completed missions as of autumn 2015. The portion of observed spectrum by HSI and MS sensors is denoted by \* for VNIR, \*\* for VNIR and SWIR, and \*\*\* for VNIR, SWIR, and TIR. The spatial resolution of air-borne sensors is given for a certain flight height (@height in km) as given in producer's specifications.

DSM. A DSM for orthorectification of a HSI and a DSM used for fusion with the HSI are in general case not the same. The reason for this is that the ordered orthorectified, e.g. satellite images, are not delivered together with the DSM used for orthorectification. Thus, the inaccuracies of the HSI due to orthorectification are not dependent on or caused by the inaccuracies and errors in the DSM for orthorectification.

The decreased planar and spectral accuracy in HSI is expected on the boundaries of high-rise objects, because of the instrumental and processing influences (Subsection 2.1.2). Similarly, the decreased planar and height accuracy of DSM is expected on the boundaries of the same high-rise objects (Subsection 2.2.2). Moreover, imprecise registration of the datasets causes errors on the boundaries of these objects. Thus, these inaccuracies in HSI and DSM occur on the boundaries of the same high-rise

objects. Other partially or not corrected systematic errors are not expected on the same parts (i.e. boundaries) of these objects in the images. Let us take for example a SGM DSM with higher spatial resolution than HSI and a HSI captured by a sensor with a narrow FOV (this means the effect of relief displacement is limited). The DEM can have holes on texture-less areas such as uniform roofs, but the spectral and spatial characteristics of the same roof are still well captured by the HSI sensor. If the DSM is well modelled, the outline of a high-rise object can be better extracted from the higher spatial resolution DSM than from HSI. The fusion of HSI and DSM can support extraction of the objects and make it more robust. It is however not limited to the extraction of material and height features from HSI and DSM respectively, but has also a potential to improve the boundary extraction of these objects.



## 3 Advances in HSI and DSM Fusion for Building Extraction

Building extraction and modelling is an active research field for urban object detection from RS imagery (Rottensteiner et al., 2012), followed by road (Mena, 2003) and urban vegetation extraction (Mayer, 2008; Rottensteiner et al., 2014). Two main reasons for this are the variety of applications, which require building models or outlines as input, and lack of fully operational methods for building extraction, their transferability, and/or sufficient quality. Moreover, urban areas are continuously undergoing changes, which require cost-efficient updates of the spatial databases.

The objective of this Chapter is to give the state of the art about building extraction from RS data, which follows a common workflow. A special focus is on multi-modal datasets and 2D building outline extraction and adjustment with sub-pixel precision. A building outline is described by a polygon, which can be derived from extracted line segments and edges. In view of applicability of scale space edge detection on RS imagery, a broader overview of RS methods using scale space is given. Finally, the diversity of evaluation techniques for building extraction is summarised and the reasoning for further developments in this thesis is given.

### 3.1 Building Extraction from Remote Sensing Data

Object extraction, as understood in this thesis, is based on modelling of a 2D or 3D object by geometric description and optional extraction of additional attributes like semantic and topological information (Baltsavias, 2004). The modelling (i.e. polygon description) of objects means that the modelled object (i.e. extracted polygon) has significantly reduced number of vertices in comparison to the data points describing object's boundary in the original dataset (Wang, 2013), while maintaining its geometrical characteristics. For example, connected boundary pixels of a building region from a raster image (Vu et al., 2009) or Triangulated Irregular Network (TIN) of the building data points are not considered as geometrical building modelling, even if they are vector representations of an object.

Building extraction methods can be divided into data- and model-driven methods (Heuel and Kolbe, 2001; Tarsha-Kurdi et al., 2007; Vosselman and Maas, 2010), or hybrid versions of both approaches. The data-driven methods rely on low-level features and shapes, and combine them into a building polygon in a bottom-up manner, for instance constructing a model of a building boundary from a set of rectangles (Arefi, 2009; Avbelj et al., 2013a, 2015a; Gerke et al., 2001; Kwak and Habib, 2014). Model-driven methods fit a parametric model from a given library or set of models and select the best fitting one. The advantages of model-driven approaches are computational efficiency and robustness. Furthermore, the topological relations and constraints between the polygon edges must not be modelled, because they are included in the library of models. However, the disadvantages of such approaches are inflexibility, and inability to model complex objects not included in the library. If only few data points per object are available, i.e. the spatial resolution of the image is relatively low in comparison to the object extent, the model-driven approaches outperform the data driven ones (Henn et al., 2013).

The methods for automatic building extraction can be grouped according to the modality of the data. Some example works are give as follows:

- ◇ Optical images: PAN (Lin and Nevatia, 1998), MS (Lee et al., 2003), and HSI.
- ◇ Height data: DEM (Brédif et al., 2013; Lafarge et al., 2008; Weidner and Förstner, 1995). Prevailing amount of proposed methods for 3D building modelling is based on 3D LiDAR point clouds -

terrestrial or aerial images (Kwak and Habib, 2014; Lafarge and Mallet, 2012; Maas and Vosselman, 1999; Rottensteiner et al., 2005; Sampath and Shan, 2007).

◊ SAR imagery (Shahzad and Zhu, 2015; Thiele et al., 2007).

◊ Combination of any two or more multi-modal datasets as discussed in the subsequent paragraphs.

The focus of this work is on advances in 2D building extraction from space- or air-borne optical (particularly HSI) imagery and DSM. Parallel to these, there have been rapid developments of the methods, which use fixed terrestrial or mobile terrestrial platforms to collect images, and combine them with nadir imagery (Frueh and Zakhor, 2003; Pu and Vosselman, 2009). Moreover, oblique single-, multi-view imagery (Vanegas et al., 2010; Xiao et al., 2012), and video-sequences (Gallup et al., 2010) have also been successfully used for object extraction, as well as freely available images on the internet (Agarwal et al., 2011; Snavely et al., 2008), and mapping databases (Vanegas et al., 2010).

The increasing number of air- and space-borne RS sensors lead to a trend of methods, which combine several datasets (Gamba, 2014), as predicted by Brenner (2005) and Hu et al. (2003). Most commonly one of the datasets exhibits height information and the other one spectral information. This trend of image and data fusion is seen in the RS community. The data fusion contest, which is yearly organised by the IEEE Geoscience and Remote Sensing Society, has the goal to encourage researchers to develop new methods for RS problems. In 2013, HSI and LiDAR DSM datasets were provided, and in 2014 thermal HSI and Red-Green-Blue (RGB) images (Debes et al., 2014).

Numerous authors propose methods for building polygon extraction from height data and MS images (Awrangjeb et al., 2010; Rottensteiner et al., 2007; Sohn and Dowman, 2007). As height data, most often LiDAR point clouds are used, followed by gridded LiDAR point clouds and DEM, derived from photogrammetric stereo images or SAR images. Characteristics of height data are described in Section 2.2. From HSI with same spatial resolution as MS image, the spectral properties of buildings can be discriminated better than from MS images. The building extraction methods for MS and DSM data can also be applied for HSI and DSM data. In such case, the rich spectral information of the HSI is not fully exploited. The classification of urban surface types, including various classes of roofing materials, can be assessed from the HSI (Roessner et al., 2001; Segl et al., 2003b). Additionally, height data improve the classification of roofing material classes (Braun et al., 2011; Brook et al., 2013; Debes et al., 2014; Heldens, 2010). Nevertheless, a pixel-wise classification is only a first step to geometric building modelling. To our knowledge, Huertas et al. (1999) is one of the first works on building modelling from HSI and another dataset, i.e. PAN images. They derive a building mask from the HSI and estimate the building models on the grounds of PAN intensity values. Similarly, Brook et al. (2010, 2013) conduct HSI guided building region extraction, but the final building outline is estimated on the grounds of DSM regions. The potential of HSI and DEM fusion for geometric building boundary estimation is to jointly, and not subsequently, use of spectral and height information, respectively (Avbelj, 2012; Avbelj et al., 2015a). This joint use of HSI and DSM is the main topic of this thesis.

### 3.1.1 General Workflow of Geometrical Building Extraction

The generalised workflow of 2D building extraction includes three main steps (Figure 3.1): building region and/or building feature extraction, approximate geometric building outline creation, and building outline refinement. Some methods join two steps, or skip a step, commonly the building outline refinement.

A first step is *building region and/or building feature extraction*. Several methods make use of both, approximate building regions and geometrical feature extraction (Lafarge and Mallet, 2012; Sohn and

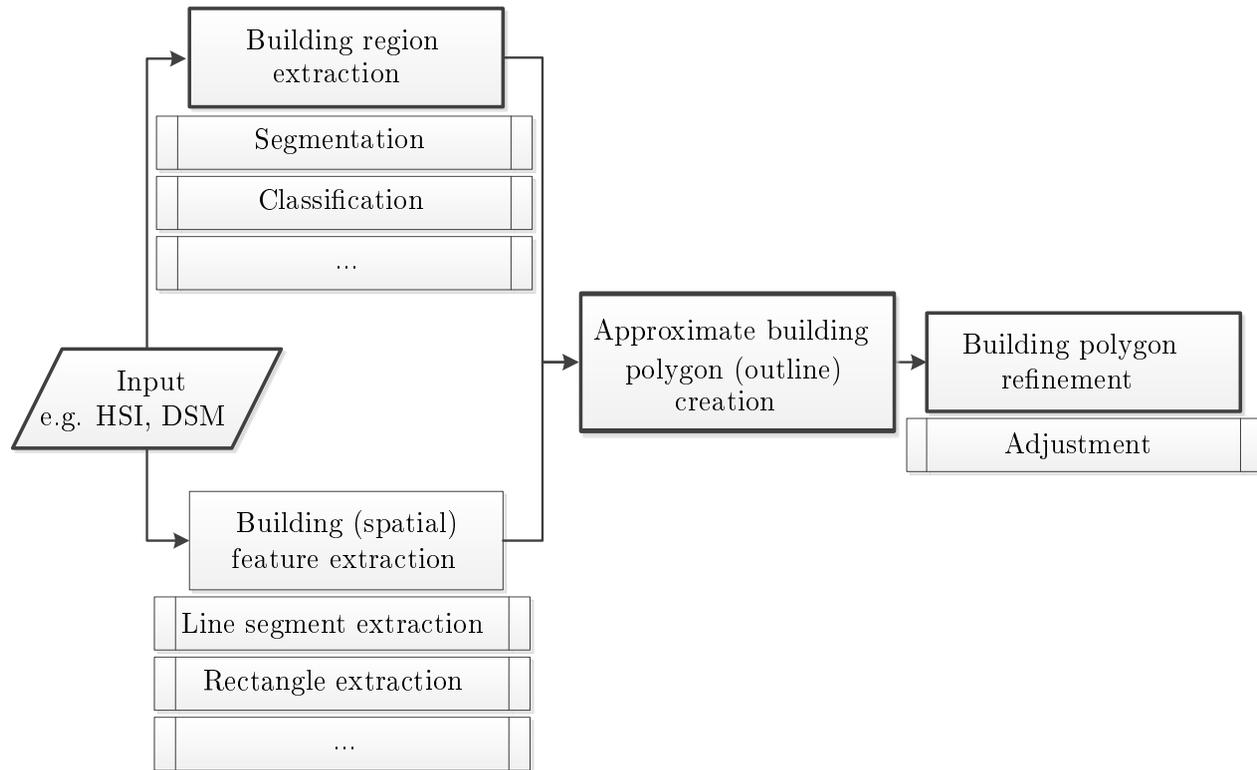


Figure 3.1: General workflow of building polygon extraction divided into three steps. These are building region and/or building feature extraction, approximate geometric building outline creation, and building outline refinement. The bold process-boxes mark a building polygon extraction steps as proposed in Chapter 5.

Dowman, 2007; Zebedin et al., 2008). Extracted building regions are a result of methods, which segment or classify images or point clouds into building and non-building regions. The output of these methods is a labelled image or a segmented point cloud. The data processing is dependent on the type of data. For RS images, the building segmentation is based on spatial, spectral, textural and/or some other features. DSM are normalised and the above-ground objects are extracted by e.g. morphological operators, filtering or (local adaptive) thresholding (Krauß et al., 2011; Sithole and Vosselman, 2004; Wack and Wimmer, 2002). In LiDAR point cloud processing, these methods are called bare-earth extraction methods (Sithole and Vosselman, 2004). High vegetation, e.g. trees, as well as buildings, are not bare-earth objects. Thus, high vegetation must be separated from the other high-rise objects. If spectral and height data are available, the high vegetation can be excluded on the basis of the spectral response, e.g. vegetation indices. If only height data are available, the high vegetation is excluded by e.g. analysing the standard deviations of surface normals or height values (Pfeifer et al., 2007). Given a LiDAR point cloud, the intensity of backscattering can be used as an pseudo-Infrared (IR) channel to support vegetation discrimination (Rottensteiner et al., 2005).

Feature extraction methods to derive geometric primitives belonging to buildings are mainly applied on single-channel images and DSM. These primitives are among others line segments (Lafarge and Mallet, 2012; Sohn and Dowman, 2007), corners, edges, contours (Ahmadi et al., 2010), and for 3D building extraction (Wang, 2013) also planes (Haala and Brenner, 1999; Verma et al., 2006) and cylinders (Lafarge and Mallet, 2012). The most common low level geometric feature is a line segment, for which several methods based on e.g. Random Sampling Consensus (RANSAC), Hough transform, or different line detectors exist. The Hough transform is also used to extract line segments perpendicular and parallel to each other (Arefi, 2009; Grigillo et al., 2012; Lee et al., 2003). Moreover, the majority of the building outlines can be well described by a rectangle (Chaudhuri and Samal, 2007), or a set of

merged rectangles as shown by Avbelj et al. (2015a); Awrangjeb et al. (2010); Brédif et al. (2013); Gerke et al. (2001); Kwak and Habib (2014); Lafarge et al. (2008). Extracted rectangles can be regarded as an approximate geometric building outline or as a primitive.

A second step is *approximate geometric building outline creation*. Extracted building regions or geometric primitives are not closed boundaries, i.e. closed polygons outlining buildings (Awrangjeb et al., 2010; Kwak and Habib, 2014). The jagged (zig-zag) boundaries, or (local) convex hulls (e.g. Sampath and Shan, 2007) of building regions must be simplified by polygon simplification or generalisation techniques (Ahmadi et al., 2010; Shan and Lee, 2005). If rectangles are extracted and regarded as already approximate geometric outline, this second step is skipped. In the other case, the rectangles (Arefi, 2009; Brédif et al., 2013) as well as line segments (Awrangjeb et al., 2010) are then connected into closed polygons. Additional constraints can be applied on adjacent polygon edges such as rectangularity, perpendicularity, and minimal/maximal intersecting angle (Lee et al., 2003; Sampath and Shan, 2007). If the subsequent edges do not meet the requirement they are merged or split. A result of this step are polygonal models of buildings with significantly reduced number of data points, i.e. vertices of a polygon.

A third step is *building outline refinement*, which is optimally fitting a building polygon model, from data- or model-driven approaches, to the data. The LS adjustment<sup>2</sup> (i.e. MLE with Gaussian noise model), or robust estimation, e.g. M-estimators, is carried out (Avbelj et al., 2015a; Kanani, 2000; Kwak and Habib, 2014). The outline refinement cannot compensate for previously wrongly selected models or erroneous data beyond the breakpoint of the estimator (Zoubir et al., 2012). Some methods skip this step, because it is computational expensive and contributes (only) to the accuracy of extraction on a sub-pixel level. In the next Subsection, the role of weights and constraints in LS adjustments for refinement of building outlines is discussed.

### 3.1.2 Least Squares Adjustment of Rectilinear Building Outlines

The building outline refinement means finding the optimal trade-off between fitting the model to the data and enforcing geometric regularity constraints, e.g. orthogonality and parallelity. The consequence of more regularity constraints in the adjustment are more unknowns, larger complexity of the estimation, and longer computational time. For the sake of simplicity, let us assume building outlines can be well described by rectilinear polygons (Kanani, 2000; Niemeier, 2008; Steadman, 2006) and the approximate outline is known. In terms of LS adjustment this means that good initial values of unknown parameters are given.

The basic idea of outline refinement can be demonstrated by fitting a line to a set of data points, which are also named observations, with optional weighting and geometric constraints. The functional model of the LS adjustment can be either LS Gauss-Markov or Gauss-Helmert (see theory by Mikhail and Ackerman, 1976; Niemeier, 2008). If the parameters of a line shall be estimated, the functional model can be written in several forms, e.g. in slope-intercept (Sampath and Shan, 2007) or normal form (Avbelj, 2012; Kwak and Habib, 2014). A rectangle (or rectilinear shape) requires more parameters as a line to be described, e.g. by position, orientation, width and height, or normal vector, distances of each line segment to the origin of a coordinate system and adjacency of polygon edges. Moreover, for building extraction some authors adjust the outlines without weighting (e.g. Awrangjeb et al., 2010; Kwak and Habib, 2014; Sampath and Shan, 2007), or derive the weights from image intensity values,

---

<sup>2</sup>In contrast to the LS method, the maximum likelihood method for parameter estimation requires an assumption on distribution of the observation vector. If the observations are normally distributed, LS and Maximum Likelihood Estimation (MLE) method lead to the identical estimator (Koch, 1999, p.161).

most commonly as a gradient of heights (Brédif et al., 2013; Sohn and Dowman, 2007), or on the basis of spectral properties of boundary pixels (Avbelj, 2012; Avbelj et al., 2015a). For multi-view imagery and under assumption of projective geometry, the stochastic properties of line segments can be used as weights for building reconstruction and establishing geometrical relations between extracted line segments (Heuel, 2004).

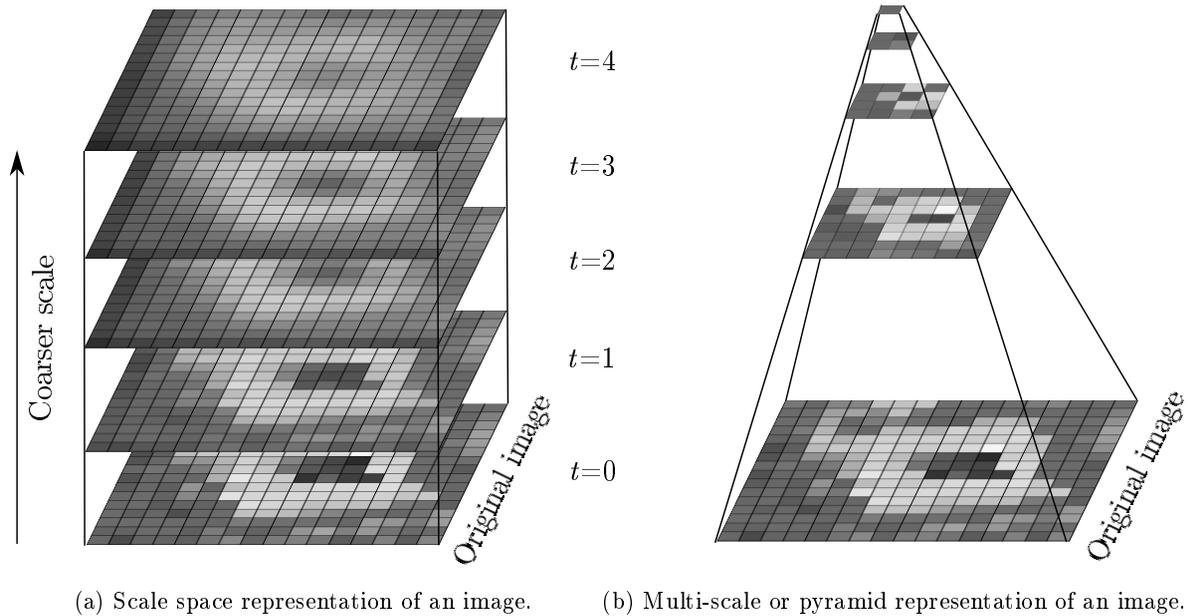
If each extracted line segment is adjusted independently, the geometric relations of rectilinear shapes are not preserved. Thus, geometric constraints shall be introduced into the LS adjustment of rectilinear building outlines by an additional set of Lagrangian multipliers (Avbelj, 2012; Mikhail and Ackerman, 1976), or as pseudo-observations (Kanani, 2000). The following examples for geometric constraints are grouped according to the simplicity of functional models, from simpler to more complex ones. One possibility is to adjust position and orientation of the prominent object orientation (Arefi, 2009) or orientation of the longest line segment (Awrangjeb et al., 2010) and then adjust other building polygon edges with respect to the first one. In a case where rectangles are extracted as geometrical primitives, another possibility is to adjust the position of each line segment of the rectangle, i.e. the size, but not the orientation of a rectangle changes after the adjustment (Brédif et al., 2013; Lafarge et al., 2008). A similar principle can be assigned for rectilinear buildings, where the largest bounding rectangle is adjusted, and only the size of smaller ones is adjusted in subsequent adjustments (Kwak and Habib, 2014). The line segments or edges of rectangles can be grouped according to geometric regularity, and then the constrained adjustment accounting for these two groups of observations is carried out (Heuel, 2004; Sampath and Shan, 2007). The most complex possibility is if all the parameters are jointly adjusted in one adjustment process, i.e. no subsequent adjustments are carried out (Avbelj, 2012; Kanani, 2000).

## Goals for Building Extraction

The developed building extraction method shall result in building polygons with sub-pixel precision of extracted edges. It shall be applicable on either DEM or HSI, or be able to join the advantages of both. A generic functional model shall be designed, which jointly adjusts all the parameters of various rectilinear shapes, i.e. L-, T-, U-shaped buildings, but also allows non-rectilinear shapes. A special focus is on exploring the weighting from both datasets in LS adjustment. The challenge is to derive weights from spectral information inherent in HSI, find a relation to the weights from DSM and join them together. To achieve these goals I propose the use of scale space.

## 3.2 Remote Sensing and Scale Space

The concept of scale space is to handle the image at different scales (Koenderink, 1984; Lindeberg, 1994; Witkin, 1984), where details in an image decrease at each coarser scale (Figure 3.2a). It is widely used in the image and signal processing communities (Bosworth and Acton, 2003), and also incorporated into some RS methods (e.g. Avbelj et al., 2013b; Dalla Mura et al., 2010; Lowe, 1999; Mayer and Steger, 1998; Pesaresi and Benediktsson, 2001). A scale space representation of an image maintains spatial the same sampling distance at all scales (Figure 3.2a) and should not be confused with a pyramid (or multi-scale) representation of an image, where the sampling distance (size) of an image changes at each level (Figure 3.2b). A multi-scale representation is a representation of an object at different fixed scales using different generalisation levels. Thus, the scale space and multi-scale representation of an object share the principle of generalisation at each coarser scale level. A prominent example of multi-scale representation in RS are five LOD for representation of buildings, as defined by the CityGML standard



(a) Scale space representation of an image.

(b) Multi-scale or pyramid representation of an image.

Figure 3.2: A scale space (Figure 3.2a) and a multi-scale or pyramid (Figure 3.2b) representation of an image. A scale space representation of an image maintains sampling distance at all scales, but the details in the image decrease with increasing scale  $t$ . A scale space representation of the image at the finest scale  $t = 0$  is the signal itself. In multi-scale or pyramid representation of an image, sampling distance (size) of an image decreases at each coarser level. Figure 3.2b shows the pyramid representation of the image with the subsampling factor of  $2^n$ ,  $n \in \mathbb{Z}$ .

(Gröger et al., 2012; Gröger and Plümer, 2012). A short overview of RS applications, which make use of a scale space, is given in this Section.

There is a conceptual difference between detecting features from an image at a single scale or multiple scales. If a single scale is used, it is assumed that all extracted details or features appear at the same abstraction level. Typical examples are classical edge detector algorithms (Canny, 1986; Clark, 1989; Fua and Leclerc, 1990; Haralick, 1984), which propose a smoothing of an image before edge detection. The principal purpose of smoothing in these algorithms is noise suppression and not a scale selection. The coarser scale of an image can be considered analogue to the human perception of abstraction, e.g. a road can be represented with its boundaries at a finer scale or as a single central line at a coarser scale (Mayer and Steger, 1998). A single scale might suffice for certain applications, thus the scale space can serve to choose the right one.

The benefit of a scale space can be fully exploited if features in an image appear and are detected across scales (Field and Brady, 1997), e.g. junctions, points, blobs, edges, shape cues (Lindeberg, 1994, 1998). Several authors showed that edge detection algorithms based on scale space perform better than classical ones on natural images (e.g. Marimont and Rubner, 1998; Perona and Malik, 1990). Moreover, a SIFT algorithm, which detects and describes local features in images, achieves the scale invariance by defining the key locations in linear scale space (Lowe, 1999). Scale space is not limited for detection of spatial features, but can also characterise spatial properties of objects in the image by a set of morphological profiles (Pesaresi and Benediktsson, 2001). They are based on the morphological scale space and are a feature vector, which can be used for e.g. supervised classification of multi-channel images (Bosworth and Acton, 2003; Dalla Mura et al., 2010; Doxani et al., 2012). Moreover, morphological scale space enables establishment of parent-child relations between objects extracted at various scales (Vu et al., 2009). These hierarchical relations can be used for generalisation or multi-scale visualisation of images in Geographic Information Systems (GIS).

Salient structures in the image are stable over several scales, i.e. they exist at a set of scales (Field and Brady, 1997; Lindeberg, 1994). On the grounds of this characteristic, a robust multi-modal image registration can be conducted, using scale structural features (Zhang et al., 2011).

In summary, RS applications mainly use two types of scale spaces, linear and morphological. Spatial feature detection prevails in linear scale space, however also non-linear, e.g. anisotropic filters perform well (Perona and Malik, 1990). The morphological scale space provides an environment in which spatial characteristics of objects over scales, their hierarchical relations, and generalisation can be carried out.

## Goals for the Use of Scale Space in RS Images

Line segments, which constitute a building polygon boundary, can be approximated from extracted edges. With a focus on edge detection from RS imagery, the potential and necessity to carry out spatial edge detection in scale space shall be researched. Emphasis shall be on scale space extraction of edges from DEM and HSI.

## 3.3 Evaluation of Building Outline Extraction

In this Section the problem of lack of standard techniques for building extraction methods is addressed, followed by an overview of evaluation indices in use, and the issues about the unavailability of appropriate ground truth are discussed.

### 3.3.1 Wild West in Evaluation Techniques for Extracted Building Polygons

Several authors tackled shortage of standard evaluation techniques for building polygon extraction by proposing, assessing, and comparing these techniques (Awrangjeb et al., 2010; Ragia and Winter, 2000; Rottensteiner et al., 2014, 2007; Rutzinger et al., 2009; Shan and Lee, 2005; Shufelt, 1999; Song and Haithcoat, 2005; Zeng et al., 2013; Zhan et al., 2005). Nevertheless, commonly accepted evaluation measures, indices, and metrics for building extraction evaluation are not yet agreed on. The closest to community acceptance are a) matched rates like: completeness, correctness, and quality rate, b) RMSE between extracted and reference vertices of polygons, and c) an agreement that these indices solely do not evaluate the quality of extracted building polygons. Moreover, the terminology of the indices for building extraction evaluation is ambiguous. For example, the same term is used for differently defined indices, and similar or even the same indices can be found under different names.

The lack of standard techniques for polygon extraction evaluation has several reasons. First, the vector data extraction from RS imagery adopted some evaluation techniques from the classical pixel based classification methods, which cannot completely assess the quality characteristics of extracted polygons (Schuster and Weidner, 2003). *Per-pixel* evaluation of detected buildings is analogous to classification accuracy assessment using *matched rates* (Zeng et al., 2013). These rates are insensitive to additional points on polygons, and respond linearly when small changes in translation, rotation, and scaling between an extracted and a reference polygon occur. The major issue with *per-pixel* evaluation for polygon comparison is that it requires rasterisation of vector data (Brédif et al., 2013; Rottensteiner et al., 2007; Shufelt, 1999) and consequently influences the accuracy. However, such low-level evaluation (Pfeifer et al., 2007) is straightforwardly applicable, requires no thresholds and can serve for a quick assessment (Schuster and Weidner, 2003; Song and Haithcoat, 2005). The matched rates can handle vector data indirectly, if instead of a number of pixels the areas of polygons are considered (Shan and Lee, 2005; Song and Haithcoat, 2005).

Second, the insufficient performance of these matched rates lead to propositions and usage of a larger number of different indices. The drawback of these numerous indices is twofold. First, some of the indices are correlated, e.g. perimeter ratio and some of the image moments. Second, they evaluate only a certain aspect of the building polygon extraction. Thus, a larger set of indices is required for complete building extraction evaluation. Song and Haithcoat (2005) list and group commonly used indices for building extraction evaluation, and give a recommendation on which indices are appropriate for a specific application. An alternative approach is proposed by Zeng et al. (2013), who decorrelates the indices and joins them into a single index by user-defined weighting.

Third, an important aspect of evaluating extracted objects is to distinguish between *per-scene* and *per-object* quality assessment. The terms per-scene and per-object evaluation are used in literature assuming different meanings, which is similar to the inconsistency in naming the indices. Per-scene evaluation is considered here as an overall evaluation of extraction. Per-object evaluation is considered whenever each building gets assigned quality indices, either as certain value, or as a binary quality measure of in-correct or dis-similar extracted object. Per-scene evaluation can always be computed by joining the per-object quality values into an overall value, or directly evaluating the whole scene. Per-object and per-scene evaluation can be based on either raster or vector data. Rutzinger et al. (2009) show that there are significant quantity differences between per-scene evaluation (in the paper referred to as per-pixel evaluation, even if vector data are extracted) and per-object evaluation methods, as well as among different per-object based indices. Furthermore, when indices are computed *per-object*, a new definition problem rises of true/false matched or detected objects and a threshold must be set (Awrangjeb et al., 2010; Ragia and Winter, 2000; Rutzinger et al., 2009). Moreover, a problem of one-to-many cardinality appears, i.e. a ground truth polygon can correspond to more than one extracted building polygons, and vice versa (Geibel and Stilla, 2000; Rutzinger et al., 2009).

Finally, the quality of extracted objects (buildings) depends highly on the object's size and complexity of the object's boundary. Generally, the larger buildings are extracted with higher quality (Rottensteiner et al., 2007). Nevertheless, because of the boundary complexity the variation of quality measures is larger for the category of larger buildings (Avbelj and Müller, 2014). Thus, the quality of extraction must be considered in relation to the spatial resolution of the imagery, smallest segment of the building polygon, and the scene under consideration. This dependence can be also regarded from a mapping point of view. For each mapping scale, the maximal planimetric deviation is standardised by the mapping authorities. Thus, if the planimetric deviation of an extracted object is within the standardised limits, it is considered correctly extracted (Freire et al., 2014).

One of the main contributions towards standardisation of extracted urban object evaluation is the ISPRS test project on urban classification and 3D building reconstruction (Rottensteiner et al., 2014, 2012). Within the project, the building extraction methods were compared and the set of indices for building extraction evaluation are defined based on Rutzinger et al. (2009) and Rottensteiner et al. (2005). Before the evaluation possible one-to-many cardinality between reference and extracted polygons are topologically clarified by split-and-merge algorithms. Then, per-scene and per-object (binary and binary weighted by the area of the buildings) matched rates can be calculated. Moreover, the geometrical accuracy is evaluated by computing the RMSE of the distances between the vertices of extracted and reference polygons, and of the distance between polygon centroids. Yet, no measures to evaluate shape similarity of extracted polygons are used.

### 3.3.2 Overview of Indices

Next to the mentioned completeness, correctness, and quality rate (Section 5.3) other *matched rates* exist, e.g. Hammoude distance and false-alarm rate. They are all based on combination (addition and

multiplication) of the common (so called overlap) and not common areas between the ground truth and extracted objects, and are related to the confusion matrix and its derivations.

*Shape similarity measures* quantify *per-object* similarity between a reference and an extracted object in raster or vector format. They can be categorised in region-based, or contour-based (also boundary-based) measures (Zhang and Lu, 2004). A choice for a certain shape similarity measure or a set of them depends on their properties and application (Veltkamp, 2001). For instance, the affine invariance is desired for some object recognition tasks, but not for evaluating building footprints, because two footprints rotated, translated, and/or scaled to each other should be recognised as different. Several shape similarity measures are in use for building polygon extraction (Ragia and Winter, 2000; Song and Haithcoat, 2005; Zeng et al., 2013), e.g. turning angle function (also named tangent space representation), geometric moments (Hu, 1962), and Fourier descriptors, which are all translation, rotation and scale invariant. In contrast, area and perimeter ratio are only translation and rotation invariant. All of the mentioned measures, except the Fourier descriptors and geometric moments, can be used to compare shapes given in vector or raster format.

In the pattern recognition community, classical metrics, e.g. Hausdorff and Chamfer metrics, are used to measure similarity between shapes (Tsai et al., 2007; Veltkamp, 2001; Veltkamp and Hagedoorn, 1999; Zhang and Lu, 2004; Zhang et al., 2004). In contrast to shape similarity measures mentioned in the previous paragraph, Hausdorff and Chamfer metrics are sensitive to translation, rotation, and scale differences. Thus, a potential of such metrics is to more objectively evaluate extracted buildings, which is further addressed in Section 5.3.

The *positional accuracy* of extracted building polygons is another category of quality measures important for geo-located data, for instance, the RMSE between reference and extracted points of a building polygon (Song and Haithcoat, 2005; Zeng et al., 2013) or Euclidean distance between centroids of the objects (Zhan et al., 2005). Both of them use vertices of extracted building footprints without accounting for the edges. Rottensteiner et al. (2014) account for the reference edges by computing the RMSE between extracted points and nearest points on the corresponding reference building polygon. If extracted objects are to be included into the GIS database, the absolute positional accuracy is necessary to quantify. Low positional accuracy indicates inaccurate object extraction, but also the coregistration or geometrical correction inaccuracies of an image and/or ground truth.

### 3.3.3 What is Ground Truth?

Ground truth or reference data are independently collected data, which are used for evaluating the accuracy of a method. The main concerns about the reference data are level of detail and availability. In relation to the spatial resolution, some details are not recognisable and therefore cannot be detected from the image. Is it better to use ground truth with all the details of the buildings or ground truth with minimal extractable details according to the spatial resolution of the image? The answer to this question is discussed in the following paragraphs together with the availability of ground truth as a limiting factor for its selection.

Ground truth data of the buildings can be manually digitised or (semi-)automatically extracted from the same dataset as extracted outlines (Brédif et al., 2013; Rottensteiner et al., 2007; Rutzinger et al., 2009), a dataset of a higher spatial resolution (Lafarge et al., 2008), mapping authorities, e.g. cadastre data (Avbelj et al., 2015a; Zeng et al., 2013), or by terrestrial measurements (Zeng et al., 2013). The latter are seldom used due to the extensive terrestrial work load and costs.

The building outlines from mapping authorities are usually accurate, complete and correct, but are collected for a different purpose (Flamanc et al., 2003), e.g. tax and real estate management. They

include also details that are not extractable from RS imagery, for instance building parts. Moreover, the building outline in a cadastre is the outline of the area of exterior walls. In orthorectified optical RS images the building outline is a horizontal projection of the roof (e.g. Taillandier, 2005), which is for roof overhang larger than the outline of exterior walls. On one hand, fully independently collected ground truth data are not directly comparable to the extracted polygons. On the other hand, ground truth data extracted from the same dataset are always available and have the similar level of detail as the extracted polygons, e.g. details of objects not extractable due to the spatial resolution of the image are not present in such ground truth.

The evaluation method should allow for different levels of detail and other variations in ground truth in comparison to the extracted data. One possible approach is to alter ground truth and/or extracted objects by e.g. boundary simplifications and merging of spatially adjacent objects (Rutzinger et al., 2009). Another approach is to choose evaluation indices, which are insensitive to these variations, e.g. accuracy of estimation (Sampath and Shan, 2007) or matched rates. The accuracy of estimation is only partially evaluating the quality, because it is a measure for fitness of the data to the model and does not evaluate the extracted object with respect to ground truth.

## Goals for Building Extraction Evaluation

Standardisation of evaluation measures for building and other object extraction, shall be considered according to the data type, i.e. raster or vector data. Moreover, the standardised measures shall be tailored to the application, which is in this case a comparison of the extracted building outlines from RS imagery to ground truth. The quality measure shall penalise a) differences due to a projection and b) errors of an extracted shape due to extraction method, noise, occlusion, and viewing geometry. The indices that are insensitive to the details and generalisation shall be selected rather than altering the extracted or reference data. The shape and positional accuracy of extracted objects should be jointly quantified. Next to accepted per-scene measures, also per-object measures shall be established, where each extracted object gets assigned a quality value, and not only a binary label of correct and incorrect.

## 3.4 Summary of Research Voids

- ◇ The potential and necessity to carry out edge detection from RS images (more specifically HSI and DSM) in scale space should be researched (Section 4.1).
- ◇ Edges in an image carry the majority of information about the objects. They can serve as a common basis between single- or multi-modal datasets, with significantly different characteristics, e.g. HSI and DSM (Sections 4.2 and 4.3).
- ◇ The emerging HSI sensors allow not only high spectral but also better spatial resolution. Thus, there is potential improvement in the rarely addressed research topic of geometric modelling of objects from HSI. In other words, the extraction of buildings from HSI should not be limited only to classification techniques (Chapter 5).
- ◇ There is a need for a single metric tailored to complete building polygon evaluation. Such a metric shall account for shape and geometrical accuracies and be insensitive to the level of detail of ground truth (Section 5.3).

## 4 Fusion of HSI and DSM

A HSI enables to identify materials of the observed area, whereas a DSM carries information about height of the earth's surface including objects on it. The complementary information from HSI and DSM provide better understanding of the observed scene. Thus, the fusion result of these two datasets has higher information content than any of the input images considered independently. Even if the HSI and the DSM carry complementary information, the fusion process requires common features or common information content between the datasets. Difference in material and height difference in HSI and DSM respectively are represented by edges. On this common feature, the fusion of the datasets is proposed. A prerequisite for this fusion method is a sufficient coregistration accuracy of HSI and DSM. In our experience, sufficient coregistration accuracy is that at least one half of an area of a pixel from an image with the higher Ground Sampling Distance (GSD) overlaps with the corresponding pixel in the other image.

In this Chapter, an alternative definition of an edge in an image is given, followed by the computation of the edge probabilities instead of detecting edges in the images. The proposed method for edge probability computation considers edges at different scales using the linear scale space. Then, the Bayesian fusion of edge probabilities is introduced, which accounts for prior knowledge about the information content relation of HSI and DSM. If no reliable prior knowledge is given, a fully data-driven fusion process is carried out.

### 4.1 Edge in an Image

Before the definition of an edge in an image and motivation for computing edge probabilities at different scales, the basic notions for differential edge description are given. Then, the conditions for a differential geometric edge detection are presented, followed by the derivation of the edge probability, which accounts also for the image noise. Finally, the linear scale space is introduced, and the implementation of the discrete scale space and the derivatives in the discrete scale space is described.

#### Basic Notions for Edge Detection in an Image

A greyscale digital image is defined as a  $f(\mathbf{x})$ ,  $\mathbf{x} = [x, y]^T$   $f : \mathbb{Z}^2 \mapsto \mathbb{R}$  and is a sampled representation of a continuous 2D function, with the spacing  $\Delta x$  and  $\Delta y$  (i.e. pixel size) in two perpendicular directions. For simplicity, equal spacing in both directions  $\Delta x = \Delta y$  is assumed. The gradient or gradient vector field  $\nabla f(\mathbf{x})$  in rectangular coordinate system is computed as

$$\nabla f(\mathbf{x}) = [f_x, f_y]^T = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]^T, \quad (5)$$

where  $f_x$  and  $f_y$  are partial derivatives of the function  $f(\mathbf{x})$  in the  $x$ - and  $y$ -direction, respectively.

The direction of the gradient vector  $\theta(f(\mathbf{x}))$  with respect to the  $x$ -axis is given by

$$\theta(f(\mathbf{x})) = \tan^{-1} \left( \frac{f_y}{f_x} \right) \quad (6)$$

and the magnitude of the gradient vector in the direction of the gradient vector  $\theta(f(\mathbf{x}))$  is

$$|\nabla f(\mathbf{x})| = \sqrt{f_x^2 + f_y^2}, \quad (7)$$

where  $|\nabla f(\mathbf{x}_0)|$  is the steepness of the slope at the location  $\mathbf{x}_0$  in the  $\theta(f(\mathbf{x}_0))$ -direction. The parameter is dropped when it is clear to which pixel it is referred to.

Let us introduce a local orthonormal coordinate system  $(u, v)$  in gradient direction at every point (pixel) of the image  $f(\mathbf{x})$ . The basis vectors of this coordinate systems are  $e_v = [\cos \theta, \sin \theta]^T$  and  $e_u = [\sin \theta, -\cos \theta]^T$ , where the  $v$ -axis points to the  $\theta$ -direction, and the  $u$ -axis is perpendicular to the  $\theta$ -direction. The directional derivative in gradient and the perpendicular direction are denoted by  $\frac{\partial f}{\partial v} = f_v$  and  $\frac{\partial f}{\partial u} = f_u$ , respectively. It follows that all directional derivatives in  $u$ -direction are equal zero, and  $f_v$  is equal to the magnitude of the gradient vector in the gradient direction  $f_v = |\nabla f|$ . The directional derivatives up to the order of three in the  $v$ -direction are

$$f_v = (f_x^2 + f_y^2)^{\frac{1}{2}}, \quad (8)$$

$$f_{vv} = \frac{f_x^2 f_{xx} + 2f_x f_y f_{xy} + f_y^2 f_{yy}}{f_x^2 + f_y^2}, \quad (9)$$

$$f_{vvv} = \frac{f_x^3 f_{xxx} + 3f_x^2 f_y f_{xxy} + 3f_x f_y^2 f_{xyy} + f_y^3 f_{yyy}}{f_x^2 + f_y^2}. \quad (10)$$

The measured images are usually corrupted by noise. In the edge detection process, the noise in the image is either reduced beforehand by convolving the signal with a smoothing filter (as the Canny edge detector) or modelled in the edge detection process. Noise in the image influences the values of the gradients. Therefore, the noise of the image and response of the gradient operators must be jointly considered by e.g. probability models (Marimont and Rubner, 1998). Our aim is to compute the probability of an edge given the image noise.

### Definition of an Edge in an Image

An edge lies on the boundary between two objects (regions) (Gonzalez and Woods, 2002) and has a length (i.e. a single point is not the edge). Definitions of an edge in an image can be grouped into two categories.

- ◇ A common way to define an edge in the image is through the definition of the discontinuity of a continuous function (e.g. Gonzalez and Woods, 2002). The grey level transition at the location of “the discontinuity” in an image is dependent on the discretisation of a continuous function.
- ◇ Another way to define the edge is directly in the discrete image. Haralick (1984) defines step edges and roof edges, where the step edge exists between pairs of the neighbouring pixels where one pixel is inside the brighter region and the other outside the region, and roof edge is at the points of change from the increasing to the decreasing grey scale values.

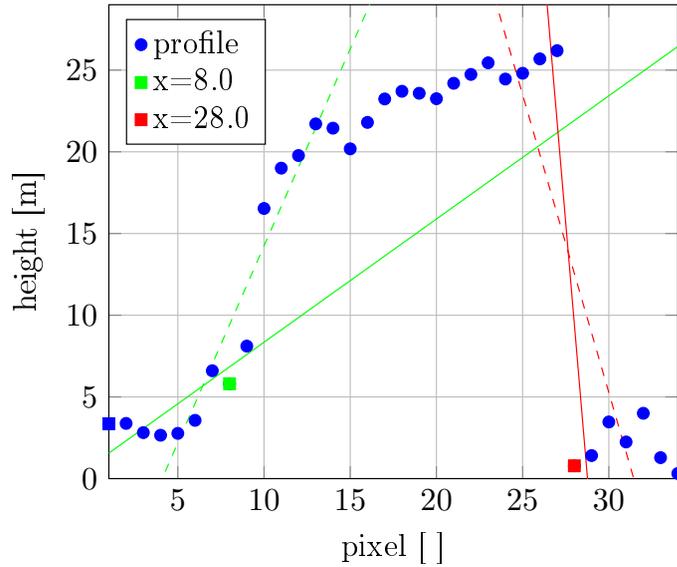
None of these definitions has a requirement on the minimal length of the edge in an image in contrast the edge detection algorithms by e.g. Canny (1986), Fua and Leclerc (1990).

Image edge detection algorithms, which rely on (some minimum) magnitude of the gradient  $|\nabla f|$  (Canny, 1986; Fua and Leclerc, 1990) have two main disadvantages. First, ramp edges are detected as thick edges requiring a thinning process for a good localisation of the edge, such as non-maximum suppression in the Canny edge detector. Second, well localised edges with low gradient magnitude are not detected (e.g. long ramp edges with low steepness). In contrast, the image edge detection algorithms, which rely on the locations of the local gradient maxima, can localise edges well (Haralick, 1984). I propose the following definition of an edge that also accounts for the local gradient maxima. An edge in an image is

- ◇ a curve with a length larger than zero, where
- ◇ the divergence of the gradient vector field is zero,  $\nabla \nabla f(\mathbf{x}) = 0$ , i.e. the vector field is source free, and
- ◇ the third derivative in a gradient direction is smaller than zero,  $f_{vvv} < 0$ .



(a) DSM.



(b) Building profile.

Figure 4.1: Challenge of scale selection for edge detection based on gradients. Blue dots are equally spaced and represent heights of a noisy building profile (blue line, 4.1a). First point of the building profile is marked with the blue square (4.1a, 4.1b). The building profile features two edges, of which the second one is steeper than the first one. The derivatives with a larger (dashed lines) and finer (solid lines) step size are computed for two points centred at each of the building edges ( $x = 8$  green and  $x = 28$  red square). The lines show the influence of the step size (scale) of the derivative operator to the estimation of an edge. Even the same DSM exhibits edges at different scales, which requires different derivative operators for their detection.

## Meaning of an Edge at Different Scales

Object and edge detection is meaningful only for a certain range of scales (Lindeberg, 1998). Edge detection algorithms, such as the Canny edge detection algorithm, are optimised for controlled image conditions, e.g. for a single distance between the sensor and the observed object. Edge detection algorithms applied on an image at a too coarse scale cannot detect the edge, whereas applied at a very detailed scale results in detection of the edges due to the noise. At a first glance edges of the objects in RS images and derived image products should all be detectable at a single scale. Reasons for this are that the RS image products are acquired in controlled image conditions and the differences of the distances between the acquisition sensor and objects in the scene are usually negligible. Yet, let us take for instance a profile of a building from a DSM (Figure 4.1a), computed from an image stereo pair with the widely used SGM method (Hirschmüller, 2005). This stereo matching method, as many others, has a data and smoothness term, which means that the resulting reconstructed surface is a trade-off between fitness to the data and smoothness of the reconstructed surface. The height discontinuities such as vertical walls of the buildings are then approximated by nearly vertical surfaces with different slopes. Even a height profile of one building can feature ramp-like edges with different slopes (Figure 4.1b). Further characteristics of the optical DSM and the LiDAR DSM are explained in Subsection 2.2.2. This example of a building profile motivates the edge detection in RS images at different scales in order to yield better performance.

Figure 4.2 shows how different smoothing factors (scale) influence edge detection and edge probability computation. In contrast to the edge probabilities, edge detector algorithms provide only information if a pixel is an edge or not. An edge in an image shorter than a sampling distance  $\Delta x$  is present only in a single pixel and is according to our definition not an edge (zero length). If the sampling distance is finer, the same edge extends over more than one pixel. For this reason and because it is aimed at

good localisation of edge probabilities, the adjacency of edge pixels in the image is not required. The Canny edge detector balances a good localisation and good detection through hysteresis thresholding and thinning. For example, the inner yard of the building on Figure 4.2 is detected by the Canny edge detector with varying  $\sigma^2$  as a different shape (4.2d-4.2f), whereas the locations of edge probabilities (4.2g-4.2i) are well localised for all  $\sigma^2$ , i.e not influenced by any thinning, tracing, and thresholding processes. I propose to compute edge probabilities for a range of scales, and define an edge probability as a maximum probability over scales. By such approach edges and even part of edges are identified at different scales (Figure 4.1).

### 4.1.1 Edge Detection and Edge Probability in an Image

In this Subsection, the computation of edge probability and three conditions for edge detection in an image are explained. First, two conditions for edge detection are given, which are needed for the computation of the probability of an edge. The third condition for an edge is a statistical test based on the distribution of the gradient derivatives.

#### Two Necessary Conditions for Edge Detection in an Image

Edges can be well localised through the locations of the local spatial gradient maxima, i.e negatively sloped zero crossing of  $f_{vv}$  (Haralick, 1984). The zero crossings of the second derivative in gradient direction  $f_{vv}$  are located exactly at the edge locations because zero crossings have no thickness. The first condition for edge detection in an image labels pixels, which contain the  $f_{vv} = 0$  and is written as

$$\left| \frac{f_{vv}}{f_{vvv}} \right| < \frac{1}{2} \Delta x(\theta), \quad (11)$$

where  $f_{vvv}$  is the steepness of the  $f_{vv}$ , and  $\Delta x(\theta)$  is the width of the sampling distance  $\Delta$  in the gradient direction defined as

$$\Delta x(\theta) = \frac{\Delta x}{\max(|\cos \theta|, |\sin \theta|)}. \quad (12)$$

The zero crossings of the  $f_{vv}$  correspond either to actual edges in an image (local gradient maxima) or so called phantom edges (local gradient minima) (Clark, 1989). Therefore, the second condition for an edge in an image checks if the zero crossing is a local gradient maximum. The zero crossing is an actual edge, if the  $f_{vvv}$  is negative

$$f_{vvv} < 0. \quad (13)$$

If the gradient  $\theta$  is zero, it is assumed there is no edge.

#### Edge Probability

An image  $f(\mathbf{x}) = s(\mathbf{x}) + n(\mathbf{x})$  can be described as the sum of the noise free image  $s(\mathbf{x})$  and an additive zero mean white noise  $n(\mathbf{x})$ . Equal noise  $\sigma_n$  in both directions is assumed. Computation of the edge probability as well as the third condition for detecting an edge in an image follows the publication by Marimont and Rubner (1998).

The probability of an edge in an image is the probability of a zero crossing  $p(zc)$  in gradient direction  $\theta$  (see Equations (11) and (12)), when no phantom edge is present (Equation (13)). It can be written as

$$p(edge) = p(zc | s_{vvv} < 0). \quad (14)$$

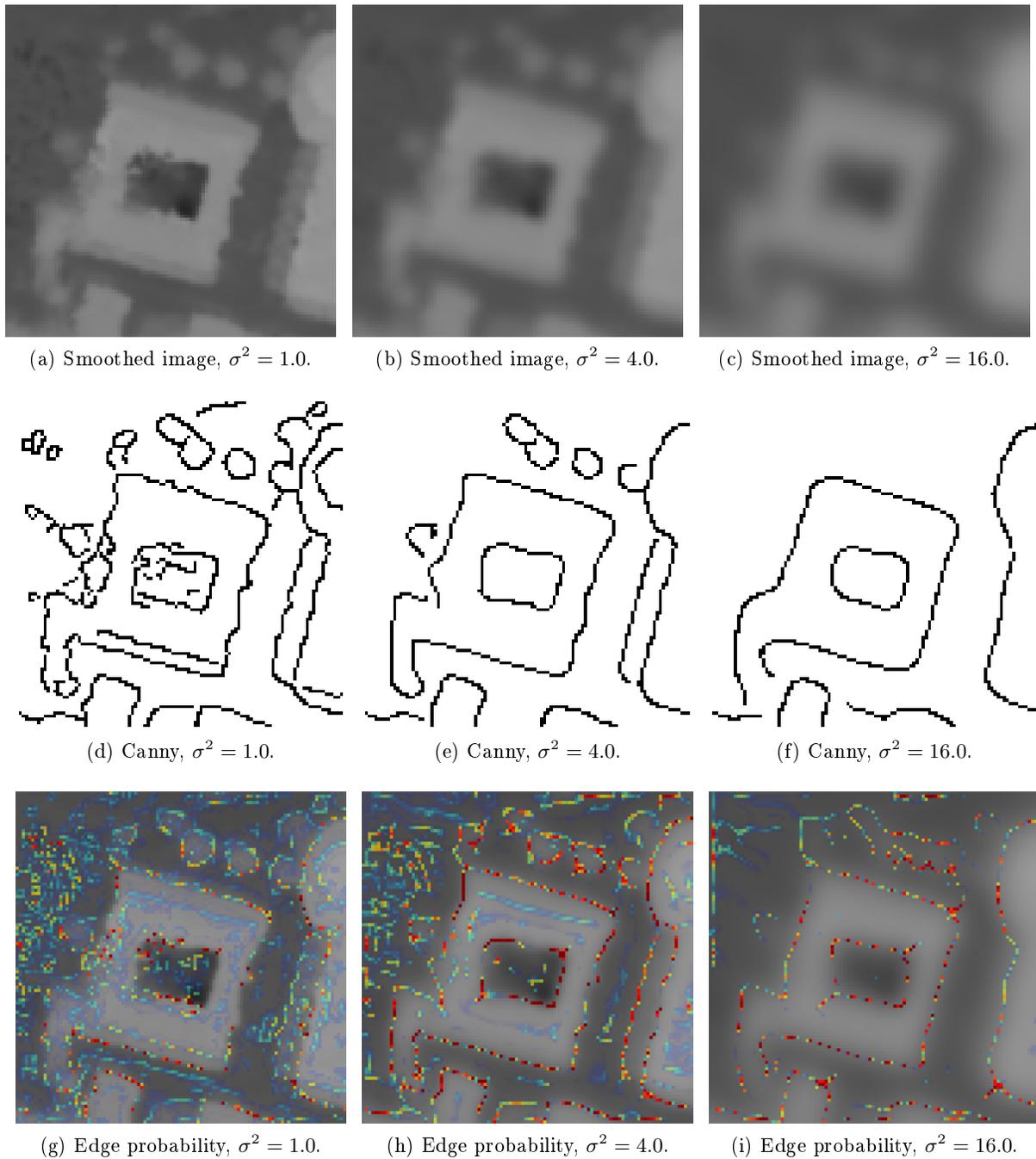


Figure 4.2: Comparison of the Canny edge detector and edge probability computation applied on a DSM (Figure 4.1a) smoothed with the Gaussian kernel,  $\sigma^2 = [1.0, 4.0, 16.0]$ . The detected edges and edge probabilities (blue-low, red-high) are scale ( $\sigma^2$ ) dependent. For instance, most of the borders of the trees at the top of the image get assigned higher edge probability at  $\sigma^2 = 4.0$  than at  $\sigma^2 = 1.0$ , and are not detected by the Canny edge detector for even larger  $\sigma^2 = 16.0$ , because their steepness is too low. For the Canny edge detector the high threshold for hysteresis thresholding is selected by analysis of the histogram of the gradient image, assuming that 30% of all pixels are edge pixels, and the low threshold is 0.4 fraction of the high threshold.

The probability of an edge in an image requires computation of the directional derivatives up to the third order and known distributions of the noise of the directional derivatives. The vectors of derivatives are denoted with two letters, i.e. **dn** for derivatives of the noise, and are given in  $(x, y)$  coordinate system. The directional derivatives are functions of derivatives in  $x$ - and  $y$ -direction (Equation (10)) and the gradient direction.

The vector of the image derivatives up to the third order  $\mathbf{df} = [f_x, f_y, f_{xx}, f_{xy}, f_{yy}, f_{xxx}, f_{xxy}, f_{xyy}, f_{yyy}]^T$  is assumed to be normally distributed with zero mean and covariance matrix  $\Sigma(\mathbf{f})$ , where  $\Sigma(\mathbf{f})$  specifies the joint probability distribution of the random variables in the vector, i.e.  $\mathbf{df} \sim \mathcal{N}(0, \Sigma(\mathbf{df}))$ . Analogically to the vector  $\mathbf{df}$ , the derivatives of the noise are jointly normally distributed,  $\mathbf{dn} \sim \mathcal{N}(0, \Sigma(\mathbf{dn}))$ , where  $\Sigma(\mathbf{dn})$  depends also on  $\sigma_n$ . The gradient direction is computed from the derivatives, thus its value is also noisy.

The probability of an edge given the vector of the partial derivatives  $\mathbf{df}$  can be written as

$$p(\text{edge}|\mathbf{df}) = \int_0^{2\pi} p(\text{edge}|\theta = \theta_s, \mathbf{df}) p(\theta = \theta_s|\mathbf{df}) d\theta_s, \quad (15)$$

where  $\theta_s$  is a simplified notation of  $\theta(s(\mathbf{x}))$  (Equation (6)). The integral in Equation (15) has no closed form solution, but can be reasonably approximated (Marimont and Rubner, 1998) by assuming that the observed gradient direction  $\theta_f$  is the correct one ( $\theta_s$ )

$$p(\text{edge} | f_{vv}, f_{vvv}, \theta_s) = p\left(\left|\frac{s_{vv}}{s_{vvv}}\right| < \Delta x(\theta_f) | f_{vv}, f_{vvv}\right). \quad (16)$$

The cumulative distribution function of a quotient of two normal random variables  $\frac{s_{vv}}{s_{vvv}}$  is computed by taking the bivariate normal distribution. The standard deviations of  $s_{vv}$ ,  $s_{vvv}$  depend on the gradient direction  $\theta_s$ , but in practice the influence of the gradient direction is small. Thus, the standard deviations of  $f_{xx}$  and  $f_{xxx}$  are taken.

### Third Necessary Condition for Edge Detection in an Image

Differentiation of a noisy signal amplifies the noise, i.e. every higher order derivative of the same signal is more noisy than lower order derivatives. Therefore, the third condition for detecting an edge in an image is needed to check the statistical reliability of the values of the second and third derivative (Equation (16)). The simplified formula for the edge probability in Equation (16) does not depend on the first derivatives ( $f_x, f_y$ ). Therefore, only higher order derivatives and their covariance matrix are needed in Equation (17).

The distribution of the vector of second and third derivatives  $\mathbf{ds} = [s_{xx}, s_{xy}, s_{yy}, s_{xxx}, s_{xxy}, s_{xyy}, s_{yyy}]^T$  is  $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ . The quadratic chi-squared distribution  $\chi^2$  with seven degrees of freedom, i.e.  $q \sim \chi_7^2$  is

$$q = (\mathbf{s} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{s} - \boldsymbol{\mu}) \quad (17)$$

where  $\boldsymbol{\mu}$  is a subvector of  $\mathbf{df}$  (i.e.  $\mathbf{df}$  without the first two rows) and  $\Sigma$  is a submatrix of  $\Sigma(\mathbf{df})$  (i.e.  $\Sigma(\mathbf{df})$  without the first two rows and first two columns).

The statistical hypothesis  $H_0 : \boldsymbol{\mu} = 0$  is rejected at the level of significance  $\alpha$  if  $q < c_m(\alpha)$ , where  $c_m(\alpha)$  is a critical value defining the upper  $\alpha$  area of the  $\chi_7^2$  distribution. If  $H_0$  is rejected, the alternative hypothesis  $H_1 : \boldsymbol{\mu} \neq 0$  is accepted, and  $q$  has a non-central  $\chi_7^2$  distribution with non-centrality  $\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$ . The value  $\alpha$  is the probability of incorrectly rejecting  $H_0$ , also called false positive rate, and is usually set as a very low value, e.g. 0.001, 0.01, 0.05.

If  $q$  has a non-central  $\chi^2$  distribution, the values of the second and third derivative under the given signal noise are considered reliable. A third condition for edge detection is then

$$q > c_m(\alpha). \quad (18)$$

The pixels whose corresponding  $q$  does not fulfil the condition in Equation (18) are not used for the fusion process of the HSI and the DSM. Additionally to the hard decision on  $q$  (Equation (18)) at every pixel in the image, the confidence probability can be computed. The confidence probability is defined as the power of the test  $\beta$ , where  $\beta(\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})$  is a probability of correctly detecting that  $q$  has a non-central  $\chi^2$  distribution, when  $q > c_m(\alpha)$ .

### 4.1.2 Scale Space Representation

A scale space representation of signals is a methodology that can handle image structures at different scales (Koenderink, 1984; Lindeberg, 1994; Witkin, 1984). The scale space methodology is explained on 2D continuous signals, but can be expanded to any arbitrary dimension ( $N$ -Dimensional (ND)). The denotation  $\tilde{\cdot}$  is used when referred to continuous signals. The continuous 2D signal  $\tilde{f} : \mathbb{R}^2 \rightarrow \mathbb{R}$  is embedded into the one-parameter scale space family  $\tilde{L} : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$  of the smoothed signals, where the parameter  $t$  is named scale parameter. The linear scale space representation of the  $\tilde{f}(x, y)$  is defined as

$$\tilde{L}(x, y; t) = \begin{cases} \tilde{f}(x, y), & t = 0 \\ \tilde{f}(x, y) * \tilde{g}(x, y; t), & t > 0. \end{cases} \quad (19)$$

Thus, the (linear) scale space representation of the signal at the finest, also called zero scale, is the signal itself, and at all coarser scales it is given by the convolution of the signal with the Gaussian kernel  $\tilde{g} : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$  defined as

$$\tilde{g}(x, y; t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}}, \quad (20)$$

with  $t = \sigma^2$  the variance of the Gaussian kernel. The Gaussian kernel  $\tilde{g}$  is isotropic (circularly symmetric), normalised to integral one, and separable, which means that for any dimension higher than one, e.g. for the  $N$ -th dimension, it can be described as a product of  $N$  One-Dimensional (1D) Gaussian kernels.

The  $\tilde{g}$  is a unique kernel for generating the linear scale space, which can be shown through the solution of the heat equation (Lindeberg, 1994). The linear scale space is a solution of the partial differential heat equation (Koenderink, 1984)

$$\partial_t \tilde{L}(\cdot; t) = \frac{1}{2} \nabla^2 \tilde{L}(\cdot; t) \quad (21)$$

with the initial condition  $\tilde{L}(x, y; 0) = \tilde{f}(x, y)$ .

Linear scale space derivatives are defined as

$$\partial_{x^\xi y^\tau} \tilde{L}(x, y; t) = \partial_{x^\xi y^\tau} (\tilde{f}(x, y; t) * \tilde{g}(x, y; t)) \quad (22)$$

where  $(\xi, \tau)$  denotes the order of differentiation. The operation of differentiation and convolution commute, thus the following holds

$$\partial_{x^\xi y^\tau} (\tilde{f}(x, y; t) * \tilde{g}(x, y; t)) = \partial_{x^\xi y^\tau} \tilde{g}(x, y; t) * \tilde{f}(x, y; t). \quad (23)$$

Moreover, for any  $\xi = \tau$ -order derivative of  $\tilde{L}$ ,  $\partial_{x^\xi y^\tau} \tilde{L}(x, y; t)$  also satisfies the diffusion equation given in Equation (21).

The linear scale space representation and linear scale space derivatives have two crucial properties that are used for edge probability computation in an image.

- ◇ *Cascade smoothing property.* A representation of a signal at a coarser scale  $t_2$  can be computed from a representation at a finer scale  $t_1 < t_2$  by convolving the signal with the Gaussian function with the parameter  $t_2 - t_1$

$$\tilde{L}(\cdot; t_2) = \tilde{g}(\cdot; t_2 - t_1) * \tilde{L}(\cdot; t_1). \quad (24)$$

This property is inherited from the semi-group structure of the Gaussian function  $\tilde{g}(\cdot; t_1) * \tilde{g}(\cdot; t_2) = \tilde{g}(\cdot; t_1 + t_2)$ .

- ◇ *Non-enhancement of local extrema.* If at some scale  $t_0 \in \mathbb{R}_+$  a point  $(x_0, y_0) \in \mathbb{R}$  is
- a local maximum for  $(x, y) \rightarrow \tilde{L}(x, y; t_0)$ , then  $\partial_t \tilde{L}(x_0, y_0; t_0) < 0$
  - a local minimum for  $(x, y) \rightarrow \tilde{L}(x, y; t_0)$ , then  $\partial_t \tilde{L}(x_0, y_0; t_0) > 0$ .

The edges in an image and their probabilities are defined through the local extrema. Therefore, it is necessary that the smoothing operation does not introduce any new zero crossings at higher levels.

For an exhaustive and complete scale space theory and axioms an interested reader is referred to the book by Lindeberg (1994).

### 4.1.3 Implementation of the Discrete Linear Scale Space and Scale Space Derivatives

To compute a scale space family of a discrete signal  $f(x, y)$ , a discrete analogue  $g(x, y)$  of the Gaussian function  $\tilde{g}(x, y)$  (Equation (20)) and a discrete derivative operator  $\mathcal{D}$  are needed. All the properties of the linear scale, e.g. non-enhancement of local extrema space must be kept.

#### Discrete Analogue of the Gaussian Function

The Gaussian kernel  $\tilde{g}(x, y)$  is separable so it is sufficient to find the 1D discrete analogue of the  $\tilde{g}(x; t)$ . The 1D Gaussian function  $\tilde{g}(x)$  can be discretised in several ways (e.g. Equations (25) and (26)). The widely used sampled Gaussian kernel  $g_{sampled}$  is defined as

$$g_{sampled}(n; t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{n^2}{2t}} \quad (25)$$

and is not a scale space kernel, because for convolution of a signal with the  $g_{sampled}$  kernel violates the property of the non-enhancement of the local extrema between two arbitrary levels. For the special case where  $t_1 = 0$  or  $t_2/t_1$  is an odd integer, the transformation  $f(\cdot; t) * g_{sampled}(n; t)$  from a level  $t_1 \geq 0$  to an arbitrary level  $t_2 > t_1$  represents a scale space transformation (Lindeberg, 1994). However, convolution of the discrete signal  $f(x)$  with the discrete Gaussian kernel  $g(n; t)$  preserves the scale space properties (Lindeberg, 1994) without special settings of the scale parameter. The discrete Gaussian kernel  $g_{discrete}(n; t)$ , shortly  $g(n; t)$ , is defined as

$$g(n; t) = e^{-t} I_n(t), \quad (26)$$

where  $I_n$  is the modified Bessel function of integer order  $n$ . Figure 4.3 shows  $g_{discrete}$  and  $g_{sampled}$  for three different values of  $t$ , and a difference between them for  $t \in [0.5, 5.0]$ .

The linear scale space representation of a discrete 1D signal  $f(x)$  is analogous to continuous signals (Equation (19))

$$L(x; t) = \begin{cases} f(x), & t = 0 \\ \sum_{n=-\infty}^{\infty} f(x-n) * G(n; t) = \sum_{n=-\infty}^{\infty} f(x-n) * e^{-t} I_n(t), & t > 0. \end{cases} \quad (27)$$

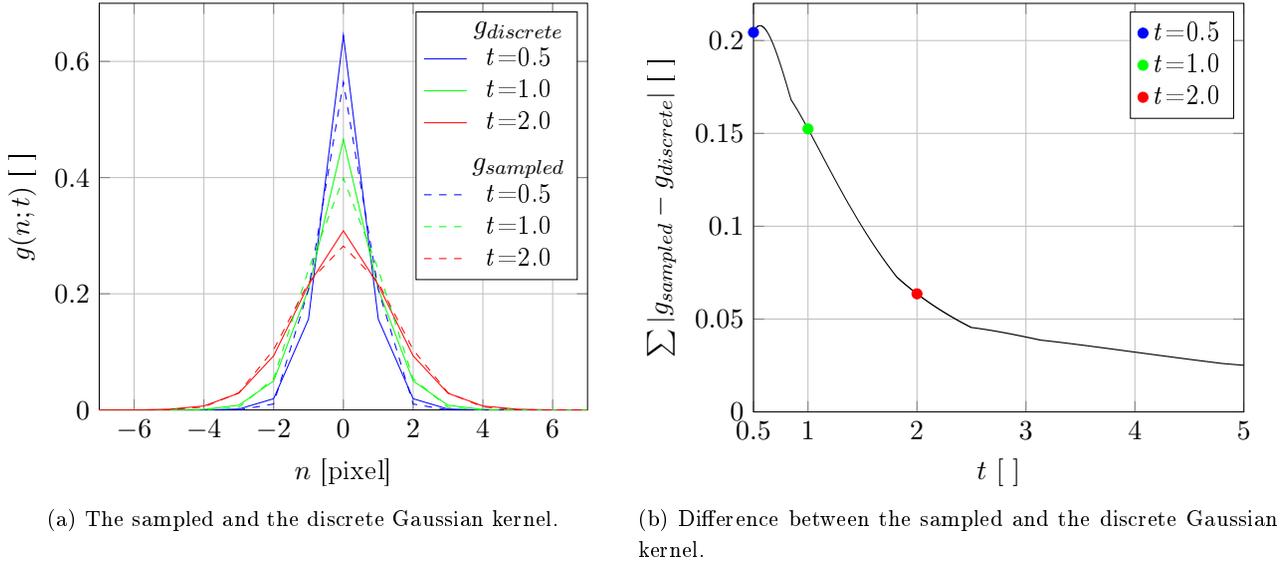


Figure 4.3: Comparison of the discrete  $g_{discrete}$  and the sampled Gaussian  $g_{sampled}$  kernel for three different values of the parameter  $t$  and their difference on an interval  $t \in [0.5, 5.0]$ . The difference is in general larger for smaller scales and decreases gradually for larger scale parameters. Note that around  $t = 0.5$  the difference function is not monotone. All the kernels are normalised to integral one.

For implementation, the discretised Gaussian function  $g$  is truncated

$$L(x; t) = \sum_{n=-M}^M f(x-n) * g(n; t) = \sum_{n=-M}^M f(x-n) * e^{-t} I_n(t), \quad (28)$$

where  $2M + 1$  is the size of the truncated discrete Gaussian kernel. Typical values for truncation of  $g$  are  $3\sqrt{t} + 1 \leq M \leq 6\sqrt{t} + 1$ . The discrete Gaussian kernel (as well as the sampled Gaussian kernel) is separable, so the linear scale space of the 1D discrete signal can be generalised to  $ND$  discrete signal.

## Discrete Scale Space Derivatives

Derivatives of the digital image are needed to compute the probabilities of edges in the linear scale space (Subsection 4.1.1). Thus, a discrete correspondence to the derivative operator  $\partial_{x\xi}$  denoted  $\mathcal{D}_{x\xi}$  is needed. As proposed by Lindeberg (1994), the  $\mathcal{D}_{x^\gamma}$  up to the third order are defined as

$$(\mathcal{D}_x L)(x; t) = \frac{1}{2}(L(x+1; t) - L(x-1; t)) \quad (29)$$

$$(\mathcal{D}_{xx} L)(x; t) = L(x+1; t) - 2L(x; t) + L(x-1; t) \quad (30)$$

$$(\mathcal{D}_{xxx} L)(x; t) = \frac{1}{2}((\mathcal{D}_{xx} L)(x+1; t) - (\mathcal{D}_{xx} L)(x-1; t)). \quad (31)$$

These numerical discrete derivatives (Equations (29)-(31)) preserve the properties of the scale space representation (see Subsection 4.1.3).

Smoothing and discrete derivative operations  $\mathcal{D}_{x^\gamma}$  commute. To yield better computational efficiency, the following order of operations should be applied. First, the smoothing of the image is computed by convolving it in two perpendicular directions by 1D  $g$  kernel with a larger support (large  $M$ ). Second, the discrete derivatives (Equations (29)-(31)) are computed in two perpendicular directions. Third, the gradient direction and the directional derivatives are computed for every pixel in the image.

## 4.2 Bayesian Fusion of HSI and DSM Based on Edges

An object in space is captured in an image as a region and exhibits a boundary. Edges in two (or more) images of the same modality are expected to be on the boundary of the same object. This is also expected for the edges of the object in multi-modal images with the same characteristic, e.g. LiDAR and stereo-matching DSM both represent object's height. Yet, in multi-modal images with considerable different characteristics, like HSI and DSM, not all the objects are detectable in both of the images. The image fusion should increase the reliability of extracted edges when applied on single-modality images. When applied on multi-modal images, the fusion should enable identification of the boundaries of the regions that are undetectable from one of the images. I propose an image fusion method with such characteristics, based on edge probabilities (Avbelj et al., 2013b). Edges, their probabilities and confidence probabilities can be extracted from any grey image using the methodology described in the previous Section 4.1. In this Section, I first concisely explain the applicability of edge probability computation (Section 4.1) to HSI (Subsection 4.2.1) and DSM (Subsection 4.2.2) images, and then propose the Bayesian fusion based on edge probabilities (Subsection 4.2.3). For simplicity, the HSI and DSM data fusion approach is explained for datasets of same spatial resolution. Nevertheless, using averaging or interpolation techniques, the fusion can be applied to datasets of different spatial resolution.

### 4.2.1 Abundance and Probability of an Edge

Edge probabilities of the objects detectable in HSI cannot be directly computed from HSI given the framework in the previous Subsection 4.1.1. My aim is to compute the probability of an edge, which is indicated by the difference of material detectable in HSI. Thus, I propose how to extract meaningful data from HSI, from which in the next step the edge probabilities can directly be computed.

Let us take all abundances of one material computed from HSI by FCLS unmixing, where the complete spectral library is given. The LMM is assumed, where every abundance is greater or equal zero and all abundances sum up to one. Thus, the abundances are on the closed interval  $[0, 1]$ . Both, abundances and probabilities are bounded on the same interval  $[0, 1]$ , and they also both sum up to one. The abundance map of the material, from which the object of interest is build, is appropriate for estimating the edge probability. However, the abundance is not directly the probability of an edge. The greater the difference of abundances in adjacent pixels is the higher is the computed edge probability. If there are no mixed pixels in HSI, the probability of an edge computed from an abundance map is either 0 or 1, i.e. only step edges are expected (Subsection 4.1.1) or the material is not present in the scene. In most cases, the HSI consists of mixed pixels, and the abundance maps derived from it are grey images. Moreover, when the mixed pixels are spanning over several pixels in the gradient direction  $\theta_g$ , the scale space representation is advantageous to compute the probabilities of the difference of material.

### 4.2.2 DSM Derived Probability of an Edge

The edges extracted from DSM or LiDAR DSM are in contrast to edges from HSI straightforward to extract, because the input datasets are already grey images. The edges in DSM are discontinuities in height or their approximations. Similarly to the edges of objects in HSI, the objects in DSM can also span over several pixels (Figure 4.1). The reasons for these have different sources, which are described in detail in Subsection 2.2.2 about characteristics of DSM.

### 4.2.3 Gaussian Mixture Model

Two normally distributed edge probabilities of the same object are given, which are computed from two different and independently collected images. The given edge probabilities are maximum edge probabilities considering all scales  $t = 1, \dots, T$  in the scale space family (see Subsection 4.1.3). The question is, how can the two given probabilities be joined? According to the product rule of probability, the joint probability  $p(A, B)$  of two independent events  $A$  and  $B$  is computed by multiplying both marginal probabilities, i.e.  $p(A, B) = p(A)p(B)$ . However, the images are collected independently, but the given edge probabilities are *of the same object*. In this case, the product rule of probability is not applicable, because the two edge probabilities are two measurements of the same event, i.e. of the same object boundary, measured by two different acquisition systems. Both edge probabilities are normally distributed, but with different parameters, which depend on e.g. image resolution, sensor precision, image noise, characteristics of images (Subsections 2.1.2 and 2.2.2). Thus, the two Gaussian distributions linearly superimposed on each other can be formulated as the Gaussian mixture distribution, also named Gaussian Mixture Model (GMM).

The mathematical formulation of GMM is as follows. The marginal density of  $K$  normally distributed Gaussian densities  $\mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$  each parametrised by  $\mu_k$  mean and  $\Sigma_k$  covariance is

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k), \quad (32)$$

where  $\mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$  is a  $k$ -th component of the mixture of the Gaussian functions, and  $\pi_k$  is the  $k$ -th mixing coefficient (Bishop, 2007, p.110-113, 430-432). Given that  $p(x)$  and all the components are normalised, the mixing coefficient  $\pi_k$  must fulfil two conditions. First, the sum of mixing coefficients equals to one  $\sum_{k=1}^K \pi_k = 1$ , and second they are non-negative  $0 \leq \pi_k \leq 1$ , for every  $k = 1, \dots, K$ . Hence, the mixing coefficients satisfy the requirements to be probabilities. Thus,  $\pi_k$  is equal to the prior probability  $p(k)$  of picking  $k$ -th component, and  $\mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) = p(\mathbf{x}|k)$ , i.e. density is equal to the conditional probability of  $\mathbf{x}$  given component  $k$ . Thus, given the normally distributed edge probabilities (Subsection 4.1.1, Equation (32)) can be rewritten as

$$p(\mathbf{x}) = \sum_{k=1}^K p(k)p(\mathbf{x}|k) \quad (33)$$

$$p(\text{edge}) = \sum_{k=1}^K p(k)p(\text{edge}|k), \quad (34)$$

where the probability of an edge has to be computed for every pixel separately. Setting and choosing the value of the mixing coefficient  $\pi_k$  is discussed in Section 4.3. I focus on fusion of the edges from HSI and DSM, i.e. number of components  $k = 2$ , however the proposed framework allows fusion of any finite number of images.

## 4.3 Information Content Relation of HSI and DSM

Up to now the theory of edge probabilities on different scales for 2D discrete signals was given, and I proposed the GMM to join this edge probabilities of HSI and DSM. In this Section I address the problem of selecting the mixing coefficients  $\pi_k$  of the components of the GMM by using prior knowledge (Subsection 4.3.1) or by a fully data-driven approach (Subsection 4.3.2).

### 4.3.1 Weighting by Prior Knowledge

Prior probabilities are used, when an expert has good knowledge about the relation of the information content of the datasets. For instance, two datasets are acquired under the same conditions with the same air-borne system at different altitudes. The dataset acquired at the lower altitude has a better resolution and consequently should get higher weight, i.e. higher value of  $\pi_k$  than the dataset acquired at the higher flight altitude. The data with larger weights are considered more reliable. When dealing with the images of different modalities, the relation between information content is not straightforward. If no decision can be met on which dataset is more reliable, one possibility is to set the weights for all datasets to be fused to the same value  $\pi_k = \frac{1}{K}$ , where  $K$  is the number of datasets. Another possibility is to rely solely on the data as I propose in the following Subsection 4.3.2.

### 4.3.2 Weighting by Data-Based Confidence Level

The relation between the information content of two multi-modal datasets such as HSI and DSM is difficult to define. Due to the inaccuracies and characteristics of the datasets, this relation might not be the same for the whole dataset (see Figure 4.1). I propose an alternative fully data-driven approach for fusion of edge probabilities (Subsection 4.1.1).

Each pixel in every dataset gets assigned the edge probability at  $t_{max}$  denoted  $p(edge; t_{max})$  and the associated confidence probability  $\beta(t_{max})$  of the maximum edge probability (Subsection 4.1.1). The scale  $t_{max}$  is a scale among all scales, at which the probability of an edge  $p(edge; t)$  is the maximal edge probability, i.e.  $t_{max} = \arg \max_t (p(edge; t))$ . I propose to use the ratio between confidence probabilities of the pixels presenting the same spatial areas to steer the mixing coefficient  $\pi_k$ . Mixing coefficients are computed for every pixel independently as

$$\pi_k = \frac{\beta_k(t_{max})}{\sum_{k=1}^K \beta_k(t_{max})}, \quad (35)$$

where subscript  $k$  denotes the values derived from the  $k$ -th dataset. The mixing coefficients for every fused pixel obviously sum up to 1.

## Summary and Outlook of This Chapter

A method for fusion of edge probabilities from multi-modal datasets is proposed. The method detects edge probabilities at different scales and joins them according to the GMM. Two modalities of the images are considered in detail, one with spectral information and the other one with information about height, HSI, and DSM, respectively. The joint edge probabilities can be further used to support accurate extraction of objects in the images, e.g. polygonal objects (Chapter 5). More precisely, the joint edge probabilities are used as weights in adjustment of the objects' boundaries.

## 5 Building Outline Extraction and Adjustment

Building region extraction, polygon creation, and adjustment follows the workflow for geometric Building Polygon (BP) extraction as described in Subsection 3.1.1. They can be applied on either single or multiple images of different modalities. The approximate outline of a building (Section 5.1) is determined by extracting the building region (Subsection 5.1.1) and then creating and selecting an approximate building polygon (Subsection 5.1.2). The essential focus of the method is on the adjustment of this approximate building polygon to the data, while keeping its geometrical properties (Section 5.2). Different image modalities require alternations in two parts of the proposed method. First, the building region extraction procedure, and second, weight determination for the adjustment of the building polygon (proposed in previous Section 4.2).

In Section 5.3, a new metric for evaluation of extracted polygons, i.e. PoLiS metric, is introduced. Its characteristics are discussed and compared to commonly used evaluation measures.

### 5.1 Approximate Building Outline Determination

Determination of approximate building outlines consists of two steps, building region extraction (Subsection 5.1.1) and approximate building polygon creation (Subsection 5.1.2), which is based on iteratively fitting a Minimum Bounding Rectangle (MBR).

#### 5.1.1 Building Region Extraction

One approach for building region extraction is described in this Subsection. Initial building regions could be extracted by other methods, e.g. classification techniques (see Subsection 3.1.1). The building region extraction depends on the modality of the input data. Three cases are distinguished, these are HSI and DSM, only HSI, and only DSM. If buildings in an image are distinguishable from the surrounding, also images with other modalities could be used. Missed building regions (i.e. false negatives) cannot be recovered in the following processing steps, i.e. building polygon creation and selection (Subsection 5.1.2), and adjustment (Subsection 5.2). Thus, high completeness values of building region extraction step are required.

Let both datasets, i.e. HSI and DSM, be available. The building regions are extracted as follows:

- ◊ First, the DTM is computed from DSM as a mean value of a moving minimum. The window (i.e. filter) size and percentage of outliers in DSM must be set.

Let us assume that the lowest point in a certain neighbourhood is lying on a terrain (Wack and Wimmer, 2002). A DSM (also DTM and nDSM) can be regarded as a matrix **DSM** with elements representing surface heights. For every pixel in **DSM** (Figure 5.1, solid orange lines), a terrain height  $\hat{h}_{m,n}$  (solid grey lines) is estimated as a minimum height in a window of a size  $W \times W$

$$\hat{h}_{m,n} = \min_{h_{i,j}} \left\{ h_{i,j}; i = m - \frac{W-1}{2}, \dots, m + \frac{W-1}{2}, j = n - \frac{W-1}{2}, \dots, n + \frac{W-1}{2} \right\}. \quad (36)$$

The value of  $W$  (red solid line) should be larger than the largest above-ground object in the scene, and is an odd integer number. A LiDAR point cloud contains outliers, e.g. high outliers due to reflection from particles in an atmosphere or birds, and low outliers due to multiple reflections. Similarly, height outliers can be present in DSM due to false disparity estimation. Therefore, a predefined percentage of values in the  $W \times W$  window is excluded, before computing the minimum

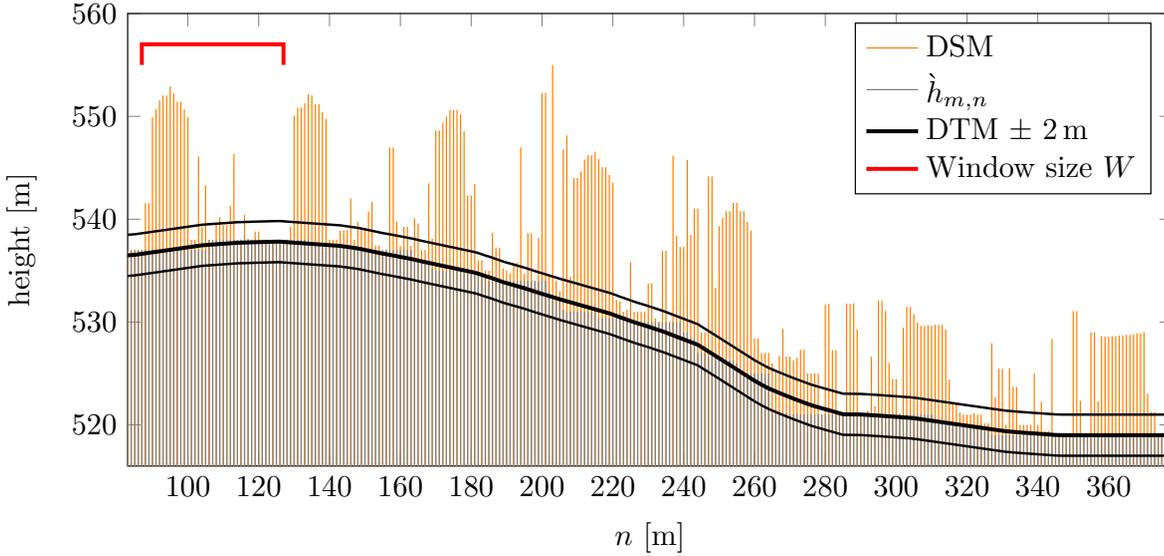


Figure 5.1: Computation of a DTM from a DSM on the 1D example. A height profile of the WV-2 DSM with the added height variation is shown by orange vertical lines with a spacing of 1 m. This hilly height variation is one of the height variations, which are used by Krauß et al. (2011) for comparison of methods for nDSM generation. The estimated height  $\hat{h}_{m=\text{constant}, n=x}$  (grey solid lines) are plotted over the DSM. The value  $m$  is constant for 1D height profile, and the  $n$  is the pixel size of the WV-2, which is resampled to 1 m. The thick black line shows the estimated DSM, and the thin black lines an optional  $\pm 2$  m threshold for above- and under-ground objects. The windows size is  $W = 41$  m (red solid line), and percentage of outliers is set to 10%.

value in Equation (36). A terrain height for the  $m, n$ -th pixel of **DTM** (thick black solid line) is computed as a mean value of terrain heights (Equation (36)) in a window of a size  $W \times W$

$$DTM_{m,n} = \text{mean}_{\hat{h}_{i,j}} \left\{ \hat{h}_{i,j}; i = m - \frac{W-1}{2}, \dots, m + \frac{W-1}{2}, j = n - \frac{W-1}{2}, \dots, n + \frac{W-1}{2} \right\}. \quad (37)$$

A **DTM** (thick black solid line) is produced by computing Equation (37) for every pixel in **DSM**.

- ◇ Second, the nDSM is computed by subtracting **DTM** from **DSM**.
- ◇ Third, building regions are defined by thresholding nDSM and removing high vegetation. Typical height for defining above-ground objects, i.e. thresholding nDSM, is 2–3 m (Figure 5.1, thin solid black lines). Height accuracy and spatial resolution of the DSM must be sufficient that the nDSM can be computed. Vegetation index values (e.g. Normalised Difference Vegetation Index (NDVI)) are computed from the selected spectral channels of the HSI. The vegetation areas are removed from the building regions by accounting for vegetation index values.

If only HSI is given, the building regions are gained by segmentation or thresholding the material maps of roofing materials using a spectral library (Section 2.1). A typical value for thresholding material maps is  $0.7 \pm 0.1$  (see Avbelj et al., 2015a).

If only DSM is available, high vegetation is removed by analysing surface normals (e.g. Pfeifer et al., 2007).

From the mathematical (estimation) point of view, the smallest building that can be estimated is of size 3 pixel  $\times$  3 pixel. Every polygon edge must be at least three pixels long, because three observations (boundary points) are needed to estimate the parameters of a line and their accuracy (Subsection 5.2.1). Two vertices of each polygon edge, i.e. begin and end pixel of each polygon edge, span usually only through a part of a pixel. Therefore, the end pixels of the edge are unreliable and the minimal length of a polygon edge is five pixels. Any building region smaller than 9 pixel<sup>2</sup> (or 25 pixel<sup>2</sup>, if end pixels are counted as unreliable) is not further considered.

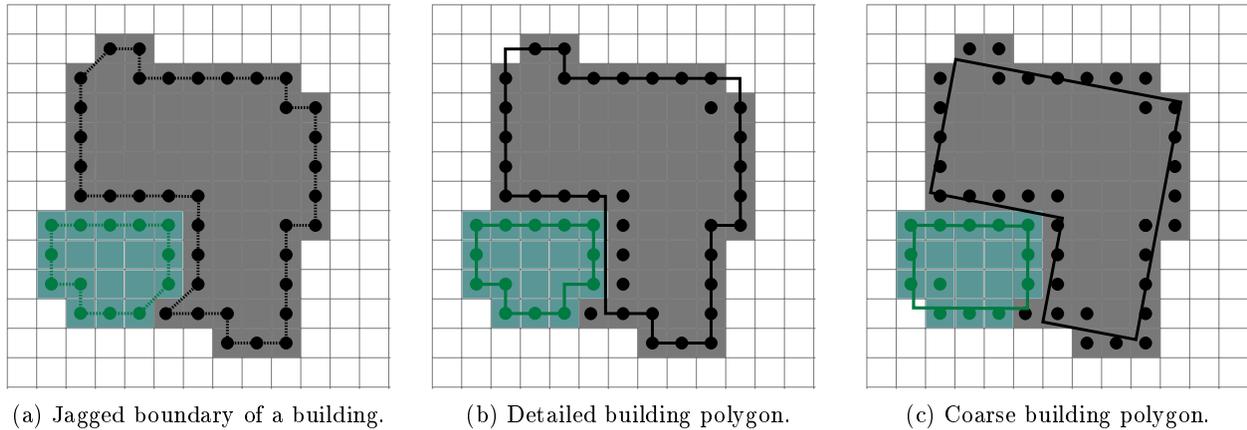


Figure 5.2: Jagged boundary (Figure 5.2a, finely dashed lines), detailed polygon (Figure 5.2b, solid lines), and coarse polygon (Figure 5.2c, solid lines) of two building regions (grey and green). The jagged boundaries, i.e. connection of adjacent boundary pixels is not referred to as a polygon representation (Figure 5.2a). The two detected building regions are represented by grey and green pixels, and their boundaries by black and green dots, respectively. The same building outline can be represented by different BP. For instance, according to the level of detail a BP can be more detailed (Figure 5.2b) or coarser (Figure 5.2c).

### 5.1.2 Building Polygon Creation and Selection

2D representation of a building outline is a BP (Figures 5.2b and 5.2c). In contrast, a jagged (zig-zag) boundary of a building (Figure 5.2a), which can be extracted from building regions, is not referred to as a polygon representation. One building outline can be represented by infinite number of BP. In this Subsection a model-driven approach for creation of a set of BP is presented. Then, the most appropriate among them is automatically selected.

Extracted building boundary points are noisy to some extent. Noisy building boundaries are one reason why a model-driven approach is more suitable than a data-driven one for BP creation. Another reason is that model-driven approaches perform better if only few points per building edge are available. However, the largest limitation of model-driven approaches is that only building shapes included in a model library can be extracted. Instead of creating a model library with complex shapes, I propose to create an approximate BP by using a combination of one primitive. The majority of building outlines can be well represented by rectangles, i.e. they have rectilinear shape (Steadman (2006), >80% according to Kanani (2000) and Niemeier (2008)). Thus, the rectangle primitive is suitable to create a set of BP.

Every building region is modelled by the method of iterative Minimum Bounding Rectangle (iMBR). The principal behind this hierarchical approach is iteratively adding and subtracting MBR as shown in Figure 5.3. A MBR of a region or a set of boundary points is a rectangle with a minimum area among all possible bounding rectangles. A BP of a region (Figure 5.3a, grey area) at a first hierarchical level  $BP_{k=1}$ ,  $k = 1, \dots, K$ , is its MBR, i.e.  $BP_{k=1} = MBR_1$  (Figure 5.3a, grey rectangle). Then, the not overlapping areas between the  $BP_1$  and the building region are defined (Figure 5.3b, green areas). At a second level  $k = 2$ , the MBR of the remaining areas (green) are identified ( $MBR_2$ ). The minimal area to which a MBR is fitted is limited by the requirements of the adjustment method. For example on Figure 5.3b, only one remaining area (green region with green delineated MBR) meets this requirement. In a general case, any  $MBR_k$ ,  $k > 1$  can constitute of more than one rectangle. The building polygon at the second level  $BP_2$  is defined by subtracting  $BP_2$  from  $BP_1$  (Figure 5.3c, black

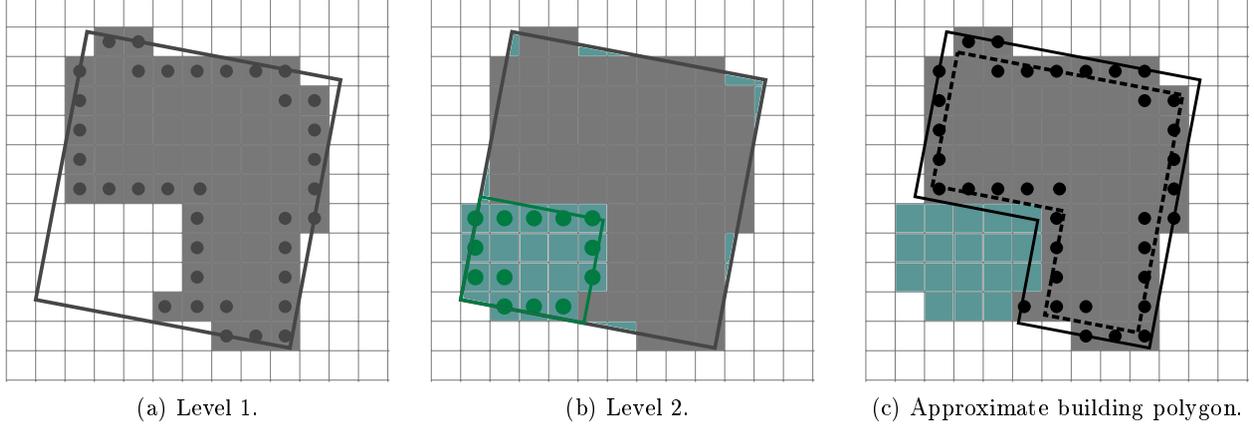


Figure 5.3: Approximate building polygon creation by recursive adding and subtracting MBR. Figure 5.3a shows level 1 MBR (grey lines). Grey dots represent the boundary of the initial building region. Figure 5.3a) shows the level 2 MBR (green lines). Green points are the boundary of the region, that is subtracted from the first level region. The approximate building polygon is represented by solid black lines (Figure 5.3c). The dashed black lines represent the adjusted polygon (Subsection 5.2.1).

line). The iterations are repeated until no more regions are left. The building polygon at the  $k$ -th level  $BP_k$  is given by

$$BP_k = \sum_{k=1}^{K=k} (-1)^{k+1} MBR_k. \quad (38)$$

The optimal BP shall be selected from the set of building polygons  $BP_k$ ,  $k = 1, \dots, K$ . The cost function is defined, which penalises complexity of a polygon and awards how well a BP fits to the data (Avbelj et al., 2013a). For  $k$ -th level of  $BP$ , the cost function is defined as

$$\text{cost}(BP_k) = \sqrt{k} \text{RMSE}_{\text{line}}(B, MBR_k), \quad (39)$$

where  $B$  is a set of boundary points (Figure 5.3, grey points) of a building region, and  $\text{RMSE}_{\text{line}}(B, MBR_k)$  is RMSE of Euclidean distances between  $B$  and  $MBR_k$  (see Equation (68)). The best approximate BP is given by

$$BP_{k=k_{\min}} = \min_{k=1, \dots, K} \{\text{cost}(BP_k)\}, \quad (40)$$

where  $k_{\min}$  is the hierarchical level of the  $BP$  with the minimal cost (Equation (39)). In the following Sections, only the selected  $BP_{k=k_{\min}}$  is considered and is concisely denoted by  $BP$ .

An optional restriction can be set on the orientation of the  $MBR_k$ . All  $MBR_k$ , with  $k > 1$  are aligned to the  $MBR_1$  or any other predefined orientation, and have the same orientation (Figure 5.3c). A result of this restriction is that all BP are exactly rectilinear. The drawback of this restriction is that non-rectilinear buildings are falsely approximated or approximated by a too complex polygon.

## Comparison of the Proposed iMBR Method to Related Methods from the Literature

Compared to the related methods by Gerke et al. (2001), Arefi (2009), and Kwak and Habib (2014), the following can be noted. All these authors proposed iMBR-like methods for BP creation. All methods define the BP in the coordinate system given by the data, except the method of Arefi (2009),

which projects the binary image of the building region into the rotated coordinate system of the main orientation. This causes changes to the boundary of building regions and possible inaccuracies. According to the dimensionality of the building polygons, Gerke et al. (2001) and Arefi (2009) deal with 2D, whereas Kwak and Habib (2014) with 3D building polygons.

(1) All authors address the problem of orientation of the MBR. Gerke et al. (2001) defines the main orientation and the MBR by using image invariant moments. It is stressed in the paper that the orientation defined in such a way may be inaccurate for complex shapes and numerically unstable for quadratic and circular shapes. Circular shaped buildings are rare, whereas quadratic shaped ones are common in suburban areas with one-family houses. Arefi (2009) showed that the main orientation of the  $MBR_1$  is less robust than the main orientation defined by a Hough transform. The former approach requires an additional threshold to select the main orientation(s) in a Hough space. Kwak and Habib (2014) define  $MBR_1$  as a minimum area MBR, which has also a considerable overlap between the MBR and boundary points. Such MBR are more reliable than a minimum area MBR. However, this additional criterion also requires a threshold. The proposed iMBR framework allows a predefined orientation of a MBR, but does not specify the method how to define it. All the discussed methods show satisfactory results on the test areas regarding the selection of the main orientation. Thus, any of the aforementioned approaches can be used in iMBR.

(2) Gerke et al. (2001), Arefi (2009), and Kwak and Habib (2014) all propose to select the highest level of BP. This means, the iterative MBR procedure is limited by the area of the remaining regions. For example, Kwak and Habib (2014) present 3D iMBR-like roof modelling, and exclude all planar roof regions smaller than  $200 \text{ m}^2$ . According to the tested datasets, this corresponds to about 175 and 1300 LiDAR points per planar region. Specific information about the minimum area is not provided for the other two methods. A main contribution of the iMBR method is the automatic selection of a building polygon through analysis of a cost function, instead of a predefined area threshold.

There are several difficulties in objectively comparing these methods and their performance. They were all tested on different datasets with varying spatial resolutions. Kwak and Habib (2014) applied the method on two datasets, one with a mean point density of  $0.88 \text{ points/m}^2$  (from ISPRS benchmark project on urban classification and 3D building reconstruction (Rottensteiner et al., 2014, 2012)) and the other one with a median point density of  $6.7 \text{ points/m}^2$ . Arefi (2009) used a LiDAR point cloud with mean point density of  $4.8 \text{ points/m}^2$  resampled to a  $0.5 \text{ m}^2 \times 0.5 \text{ m}^2$  grid. Gerke et al. (2001) used DSM and MS ortho images with spatial resolution of  $0.10 \text{ m}$ . Moreover, quality assessment of the results has not been carried out using the same quality indices. The lack thereof is discussed in Section 3.3. Finally, the method performance is dependent on the spatial resolution (point density), the dimensions, and the complexity of the objects in the scene. Thus, the performance of BP extraction methods should be considered in relation with the length of a building edge. A restriction on minimal length of a straight edge (building area) of a 2D polygon is similar to the restriction on the minimal area of a planar surface (volume) of a building for 3D roof polygon estimation. The method by Kwak and Habib (2014) requires a minimal area of a roof region to be estimated. Thus, it is the only method among the mentioned ones, which accounts for a dimension of a building part.

## 5.2 Joined use of HSI and DSM for Building Polygon Estimation

A goal is to find the best fitting rectilinear BP in the sense of LS method by refining approximate BP (Subsection 5.1.2). This is a minimisation problem, where the best set of parameters according

to the given data and one or more constraints must be found. The relation between the parameters of a rectilinear polygon and observations (boundary pixels) is defined in the functional model. This model is based on minimisation of the distances between the rectilinear polygon and boundary pixels. In the Cartesian coordinate system the distance between a point and a rectilinear polygon cannot be given by a single equation, in contrast to e.g. a line or conic section. Thus, two different functional models for adjustment of rectilinear polygon are proposed (Subsection 5.2.1). A LS method is used to estimate the unknown parameters. The joint edge probabilities are included in the weighting matrix as proposed in Subsection 5.2.2.

In this Section it is assumed that the following quantities are given

- ◊ boundary pixels of a building region (Subsection 5.1.1),
- ◊ approximate building polygon with known approximate scale, position, and orientation (Subsection 5.1.2), and
- ◊ their edge probabilities (Section 4.2), which carry joined information about edges from HSI and DSM (i.e. images considered as input datasets).

### 5.2.1 Mathematical Model for Rectilinear Polygon

Two mathematical models for rectilinear building polygon are proposed (Figure 5.4). The parameters are estimated by means of LS. The first one is the Gauss-Markov (GM) model with derived- and pseudo-observations. The second one is the constrained Gauss-Helmert (GH) model. The stochastic models of both adjustment models are discussed in the following Subsection 5.2.2.

#### Least Squares Method

Let the observations, which contain the information about the unknown parameters  $\boldsymbol{\beta}$  be collected in a random vector  $\mathbf{y}$ . The vector  $\mathbf{y}$  has a positive definite covariance (dispersion) matrix  $\boldsymbol{\Sigma} = D(\mathbf{y})$ . The linear relations between the parameters and the observations are assumed to be known and given by a set of functions  $F(\cdot)$ . These functions are collected in a vector  $\mathbf{F} = [F_1(\cdot), F_2(\cdot), \dots]^T$ . Furthermore, let  $s(\mathbf{y})$  be estimates and  $E(\mathbf{y})$  the expectation values of  $\mathbf{y}$ . The estimates of the expected values of  $\mathbf{y}$  are denoted  $s[E(\mathbf{y})]$ . The LS estimate minimises the quadratic form

$$\omega = (\mathbf{y} - s[E(\mathbf{y})])^T \boldsymbol{\Sigma} (\mathbf{y} - s[E(\mathbf{y})]). \quad (41)$$

The LS estimates are concisely denoted by a hat symbol, i.e.  $\hat{\boldsymbol{\beta}}$ . The functional model of LS adjustment is given by a set of linear functions  $\mathbf{F}$ , and stochastic model by expectation value  $E(\mathbf{y})$  and dispersion  $D(\mathbf{y})$  of observations.

In a case of non-linear relations between parameters and observations, approximate values of parameters, denoted  $\boldsymbol{\beta}^0$ , must be known. The functions  $\mathbf{F}(\cdot)$  are linearised by taking first order of the Taylor series expansion around approximate values of unknown parameters  $\boldsymbol{\beta}^0$ . In other words, first order partial derivatives of all functions with respect to the all unknown parameters are collected in a Jacobian matrix  $\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \boldsymbol{\beta}}$ , and evaluated at  $\boldsymbol{\beta}^0$ . Then, the LS estimation is iterated and in every iteration the  $\boldsymbol{\beta}^0$  is updated.

In contrast to the LS method, the maximum likelihood method for parameter estimation requires an assumption on distribution of the observation vector. Given normally distributed  $\mathbf{y}$ , both methods lead to the identical estimator for the unknown parameters (Koch, 1999, p.161). Further details about LS adjustment and parameter estimation can be found in the textbooks by e.g. Koch (1999), Niemeier (2008), and Mikhail and Ackerman (1976).

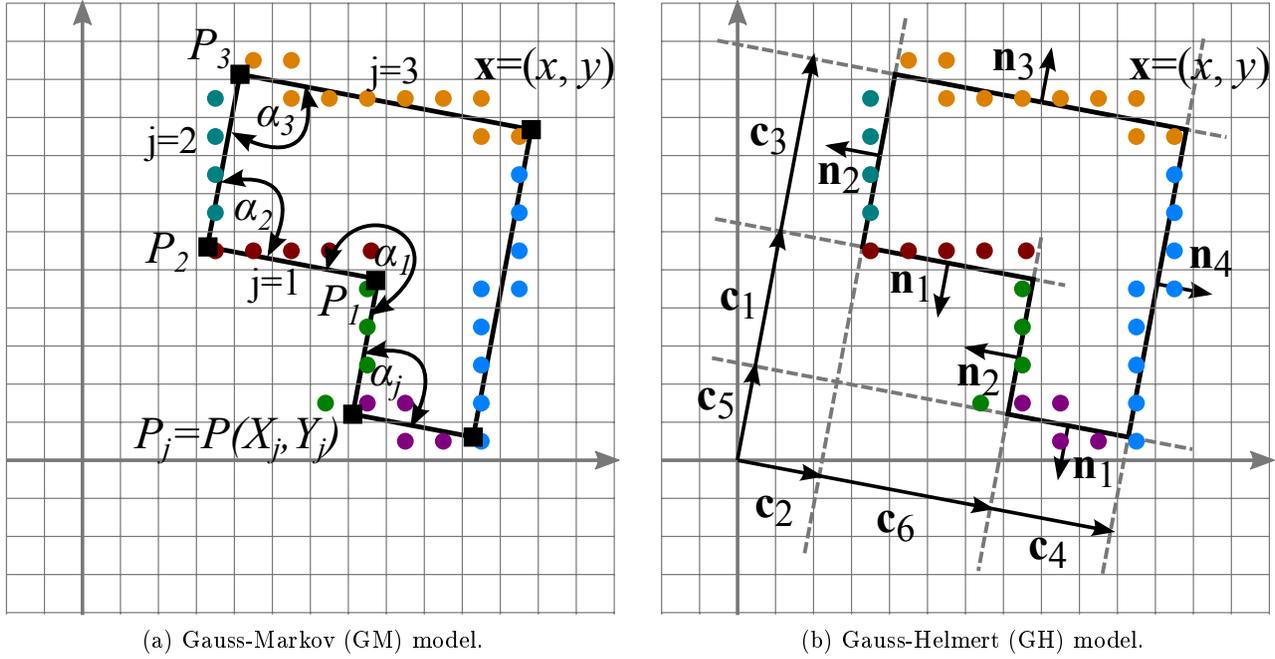


Figure 5.4: Functional model for LS adjustment of a rectilinear building. The approximate building polygon (black), and boundary points  $\mathbf{x}_i$  with assignment to the polygon edges are given (dots in various colours). For a Gauss-Markov (GM) model (Figure 5.4a), coordinates of vertices  $P_j$  (black squares) are unknown parameters  $\beta$ . The vector of observations  $\mathbf{y}$  includes squared distances between boundary points and polygon edges  $d^2(i, j)$  (derived observations), cosines of angles  $\alpha_j$  and vertices of the polygons  $P_j$  (both pseudo-observations). For a Gauss-Helmert (GH) model (Figure 5.4b), the unknown parameters are distances from the origin of a coordinate system to the polygon edges  $c_j$  and a normal vector of polygon edges  $\mathbf{n}$ .  $\mathbf{n}$  is s.t. constraint  $\mathbf{n}^2 = 1$ . The observations are boundary points  $\mathbf{x}_i$ . To enhance readability of Figure 5.4b, lengths of normal vectors  $\mathbf{n}$  are intentionally longer than a unit indicated by a grid of coordinate system.

### Gauss-Markov Model

The Gauss-Markov model (GM, also spelled Gauss-Markoff and known as adjustment of observations) estimates linear functions of fixed parameters by linear functions of the observations, i.e. observation equations have a form  $F(\hat{\mathbf{y}}) = 0$ . Given observations  $\mathbf{y}$ , design (coefficients) matrix  $\mathbf{A}$ , and weight matrix  $\hat{\mathbf{P}}$ , the GM model is

$$\mathbf{A}\beta = \mathbf{y} + \hat{\mathbf{e}} \quad \text{with} \quad E(\hat{\mathbf{e}}) = \mathbf{0}, \quad D(\hat{\mathbf{e}}) = \sigma_0^2 \hat{\mathbf{P}}^{-1}, \quad (42)$$

where  $\hat{\mathbf{e}}$  is random error vector of the observations (or vector of corrections) and  $\sigma_0^2$  is a known constant.

Let the 2D coordinates of boundary points  $\mathbf{x}_i = [x_i, y_i]^T$ ,  $i = 1, \dots, I$  of a building region and their assignment to the approximate polygon edges be given (Figure 5.4a). The building polygon can be described by its vertices  $P_j = P(X_j, Y_j)$ , which are unknown parameters in the model  $\beta = [P_j]^T$ ,  $j = 1, \dots, J$ . The observation equations can be set in such a manner that the rectangularity between subsequent polygon edges is included as pseudo-observation and not as an additional constraint. Three types of equations are set, these are observation equations for squared distances between each boundary point and the corresponding edge of a polygon  $F_d$ , observation equations for right (and  $\frac{3\pi}{2}$ ) angles  $\alpha$  denoted  $F_\alpha$ , and observation equations for the polygon vertices  $F_{XY}$  (Figure 5.4a). Since the actual observations are coordinates of boundary points  $\mathbf{x} = [x, y]^T$ , the squared distances  $d^2$  are derived observations and the cosines of angles  $\alpha$  and polygon vertices  $P$  are both pseudo-observations.

The distance between the  $i$ -th boundary point and corresponding  $j$ -th polygon edge is given by

$$d(i, j) = \frac{|(X_{j+1} - X_j)(Y_j - y_i) + (X_j - x_i)(Y_{j+1} - Y_j)|}{\sqrt{(X_{j+1} - X_j)^2 + (Y_{j+1} - Y_j)^2}}. \quad (43)$$

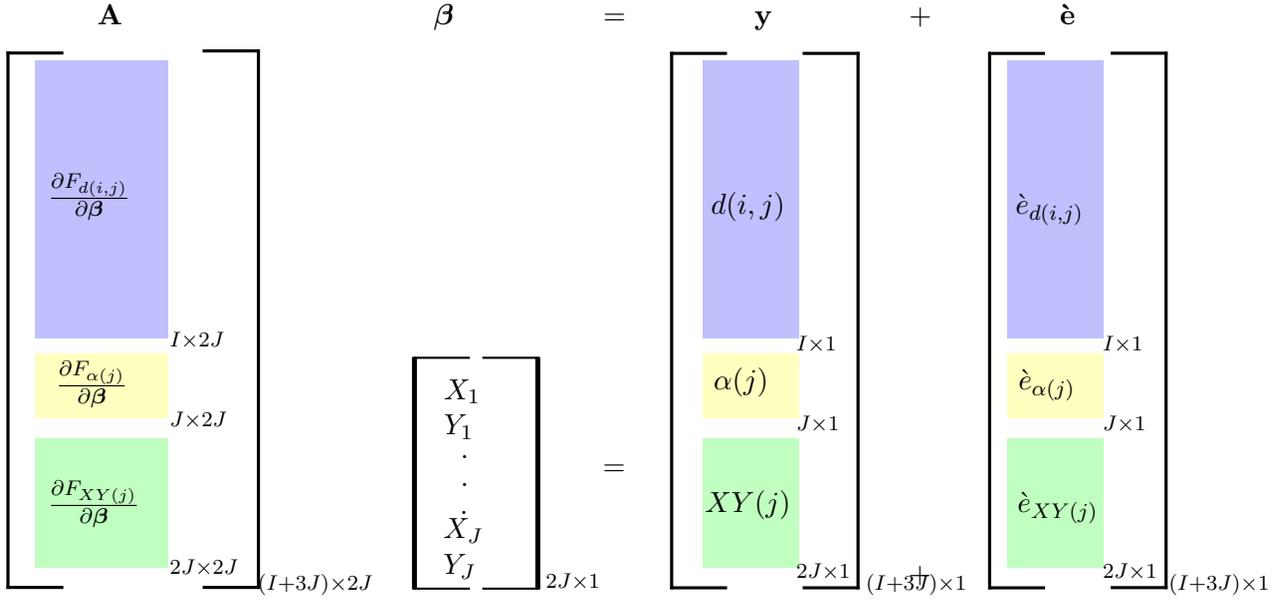


Figure 5.5: Schematic representation of the Gauss-Markov model (Equation (42)) for rectilinear BP estimation. Three types of observation equations are set, for squared distances between the boundary points and building polygon (purple), for cosines of angles (yellow), and for polygon vertices (green). The relation between observations and unknown parameters are non-linear, thus the design matrix  $\mathbf{A}$  is linearised.

The denominator of a function  $d(i, j)$  is the Euclidean distance between two polygon vertices  $d_e(P_j, P_{j+1})$  and is always positive, if  $X_j \neq X_{j+1}$ . This is a plausible assumption, because two subsequent vertices of a building polygon are not the same. If the boundary point and the polygon vertex are exactly the same, the numerator of Equation (43) is equal zero. In such a case, function  $d(i, j)$  is not differentiable, because the nominator is an absolute value function. To bypass this problem, the first set of observation equations is transformed to the squared distance  $d^2$  between  $i$ -th boundary point and corresponding  $j$ -th polygon edge

$$F_{d(i,j)} = d^2(i, j), \quad (44)$$

where  $d^2(i, j)$  is differentiable on the full domain (Figure 5.5, purple).

All interior angles  $\alpha$  of a rectilinear polygon are either right angles  $\frac{\pi}{2}$  or  $\frac{3\pi}{2}$ . A second set of observation equations are cosines of the interior angles  $\cos(\alpha)$  (Figure 5.5, yellow), because  $\cos((2k+1)\frac{\pi}{2}) = 0$ , for  $k \in \mathbb{Z}$ . The observation equation for the cosine of the  $j$ -th angle is

$$F_{\alpha(j)} = \frac{(X_j - X_{j-1})(X_{j+1} - X_j) + (Y_j - Y_{j-1})(Y_{j+1} - Y_j)}{\sqrt{(X_j - X_{j-1})^2 + (Y_j - Y_{j-1})^2} \sqrt{(X_{j+1} - X_j)^2 + (Y_{j+1} - Y_j)^2}}. \quad (45)$$

This third set of observation equations are coordinates of  $j$  polygon's vertices (Figure 5.5, green). One pair of observation equations for the  $j$ -th polygon vertex is

$$F_{X(j)} = X_j - X_j^0 \quad (46)$$

$$F_{Y(j)} = Y_j - Y_j^0 \quad (47)$$

and is concisely denoted  $F_{XY(j)}$ . This third set of observation equations could be excluded from the functional model, because the polygon vertices are unknown parameters. However, the position of the

$$\begin{array}{c}
\mathbf{A} \quad \boldsymbol{\beta} \quad + \quad \mathbf{B} \quad \mathbf{e} \quad = \quad \mathbf{y} \\
\\
\left[ \begin{array}{c} \frac{\partial F_i}{\partial \boldsymbol{\beta}} \\ \vdots \\ \frac{\partial F_i}{\partial \boldsymbol{\beta}} \end{array} \right]_{I \times (2+J)} \left[ \begin{array}{c} n_x \\ n_y \\ c_1 \\ \vdots \\ c_J \end{array} \right]_{(2+J) \times 1} + \left[ \begin{array}{c} \frac{\partial F_i}{\partial \mathbf{y}} \\ \vdots \\ \frac{\partial F_i}{\partial \mathbf{y}} \end{array} \right]_{I \times 2I} \left[ \begin{array}{c} e_{x_1} \\ e_{y_1} \\ \vdots \\ e_{x_I} \\ e_{y_I} \end{array} \right]_{2I \times 1} = \left[ \begin{array}{c} \mathbf{y}_i \\ \vdots \\ \mathbf{y}_i \end{array} \right]_{I \times 1}
\end{array}$$

Figure 5.6: Schematic representation of the Gauss-Helmert (GH) model (Equation (48)) for rectilinear BP. One type of conditional equations is set (red), in contrast to GM model (Figure 5.5), where three types of observation equations are set.

adjusted polygon vertices should be kept close to the approximate polygon vertices (Kanani, 2000). If  $F_{XY(j)}$  are kept in the functional model as observations their accuracy can be steered through weights.

The observation equations are non-linear, thus the coefficient of the design matrix  $\mathbf{A}$  are defined through linearisation of functions  $\mathbf{F}$  and evaluated at approximate values of unknown parameters  $\mathbf{A} = \frac{\partial F}{\partial \boldsymbol{\beta}} |_{\boldsymbol{\beta}^0}$  (Figure 5.5).

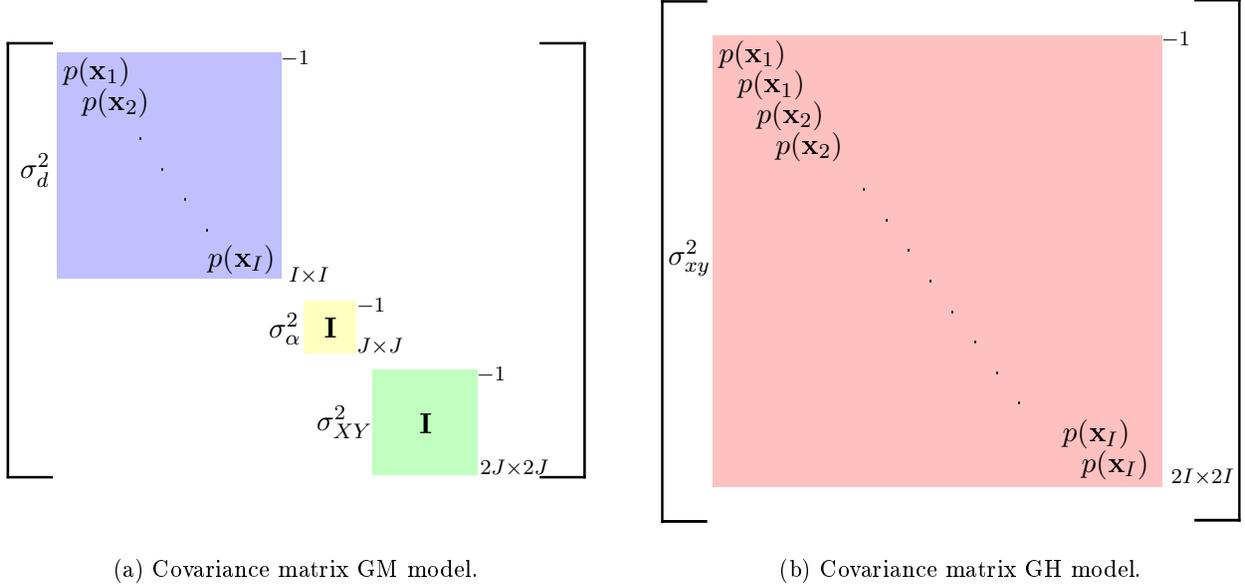
## Gauss-Helmert Model

Gauss-Helmert model (GH, also known as a mixed model, general model, or a model with condition equations and parameters) treats unknown parameters  $\boldsymbol{\beta}$  and observations  $\mathbf{y}$  as random variables. The condition equations have a form  $F(\hat{\mathbf{y}}, \hat{\boldsymbol{\beta}}) = 0$ . In contrast to GH model, the GM model treats only  $\boldsymbol{\beta}$  as fixed quantities. The so-called total least squares model approach also jointly estimates parameters and observations, but it was shown that this is only a variation of the GH model (Neitzel and Petrovic, 2008). Given  $\boldsymbol{\beta}$ , design matrices  $\mathbf{A}$  and  $\mathbf{B}$ , and weight matrix  $\mathbf{P}$ , the GH model is given by

$$\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{e} = \mathbf{y} \quad \text{with} \quad E(\mathbf{e}) = \mathbf{0}, \quad D(\mathbf{e}) = \sigma_0^2 \mathbf{P}^{-1} \quad (48)$$

The GH model can be transformed into the GM model, by substituting  $\hat{\mathbf{e}}$  (Equation (42)) with the linear combination  $-\mathbf{B}\mathbf{e}$ . It can be shown that the stochastic model is then given by  $\mathbf{e} = \mathbf{0}$ , and  $D(\mathbf{e}) = \sigma_0^2 \mathbf{B}^{-1} \hat{\mathbf{P}} \mathbf{B}^T$ .

Similar as in GM model, the boundary points  $\mathbf{x}$  and their assignment to polygon edges is known. However, in GH model the boundary points  $\mathbf{x}$  are the only observations and are denoted by  $\mathbf{y}$ . Each polygon edge can be described by a vector pair  $(\mathbf{n}, \mathbf{c})$ , where  $\mathbf{n} = [n_x, n_y]^T$  is the normal vector and  $\mathbf{c}$  is the vector from the origin of a coordinate system to the polygon edge (Figure 5.4b). Only the length of a vector  $\|\mathbf{c}\| = c$  must be estimated, because the orientation of a rectilinear polygon is described



(a) Covariance matrix GM model.

(b) Covariance matrix GH model.

Figure 5.7: Covariance matrix  $\Sigma$  for rectilinear building for GM model (Figure 5.7a) and GH model (Figure 5.7b). For GM model, the variances of three groups of observations must be set. These are  $\sigma_d^2$  for square distances between polygon edge and boundary points (purple),  $\sigma_\alpha^2$  for cosines of angles (yellow), and  $\sigma_{XY}^2$  for vertices of the building polygon (green). GH model has only one group of observation, i.e. vertices of boundary points (pink). Before multiplication with the corresponding variances, these weight matrices must be all inverted, which is schematically shown by  $-1$ . The relative relation between variances of groups of observations is relevant, thus the variance  $\sigma_{xy}^2$  can be set to one. The variances of the square distances (Figure 5.4a) and the boundary points (Figure 5.4b) are both inverse-proportional to the edge probabilities. The coloured squares represent weight matrices of groups of observations for GM model and observations for GH model.

by  $\mathbf{n}$ . The unknown parameters are then  $\boldsymbol{\beta} = [n_x, n_y, d_1, \dots, d_J]^T$ , where  $j = 1, \dots, J$  is a number of polygon edges. For each boundary point  $\mathbf{x}$  a condition equation is set, which has the form

$$F_i = \mathbf{n}_i \mathbf{x} - c_j, \quad (49)$$

where  $\mathbf{n}_i$  is one of the four normal vectors perpendicular or parallel to each other (Figure 5.4b), i.e. given  $\mathbf{n}_1 = [n_x, n_y]^T$ , then  $\mathbf{n}_2 = [-n_y, n_x]^T$ ,  $n_3 = -n_1$ , and  $n_4 = -n_2$ . The normal vector  $\mathbf{n}$  must have length one, thus the Equations (49) are *subject to* constraint

$$n_x^2 + n_y^2 = 1. \quad (50)$$

The condition equations are non-linear, thus the coefficient of the design matrices are determined through linearisation at approximate values of unknown parameters for  $\mathbf{A} = \frac{\partial F}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}^0}$  and at values of observations for  $\mathbf{B} = \frac{\partial F}{\partial \mathbf{y}}|_{\mathbf{y}^0}$  (Figure 5.6).

## 5.2.2 Weights in the Adjustment

A covariance matrix of observations  $\Sigma$  is a symmetric positive semi-definite matrix, with variances of observations on the diagonal, and covariances on the off-diagonal entries. The relation between  $\Sigma$  and weight matrix  $P$  is given by

$$\Sigma = \sigma_0^2 P^{-1}, \quad (51)$$

where  $\sigma_0^2$  is an a priori known constant, which influences neither the adjustment nor the weighting. Therefore, it is set to  $\sigma_0^2 = 1$ . To set the weights for the observations in the adjustment means the definition of a covariance matrix of observations  $\Sigma$ .

	Gauss-Markov model	Gauss-Helmert model
General formula	$\mathbf{A}\boldsymbol{\beta} = \mathbf{y} + \boldsymbol{\acute{e}}$	$\mathbf{A}\boldsymbol{\beta} + \mathbf{B}\mathbf{e} = \mathbf{y}$
Stochastic model	$E(\boldsymbol{\acute{e}}) = \mathbf{0}, D(\boldsymbol{\acute{e}}) = \sigma_0^2 \mathbf{P}^{-1}$	$E(\mathbf{e}) = \mathbf{0}, D(\mathbf{e}) = \sigma_0^2 \mathbf{P}^{-1}$
Variance matrix $\Sigma =$	$\sigma_d^2 (\propto p_{\mathbf{x}_i}^2), \sigma_\alpha^2, \sigma_{XY}^2$	$\sigma_{\mathbf{x}_i}^2 (\propto p_{\mathbf{x}_i}^2)$
$D(\cdot)$ components		
Equations $F$	observation equations	condition equations
	$F_{d^2(i,j)}, F_{\alpha(j)}, F_{XY(j)}$	$F_i$ s.t. $\mathbf{n}^2 = 1$
Unknown parameters $\boldsymbol{\beta}$	$P_j = [X_j, Y_j]$	$\mathbf{n} = [n_x, n_y]^T, \mathbf{c}_j$
Observations $\mathbf{y}$	$d^2(i, j), \alpha(j), P_j = [X_j, Y_j],$	$\mathbf{x}_i = [x_i, y_i]^T$
Right angle	weighting of pseudo-observations $\sigma_\alpha^2$	condition equations and constraint

Table 1: General characteristics of the GM model (Figure 5.5) and GH model (Figure 5.6) for rectilinear polygon. In GH model strict rectilinearity is given through the functional model, whereas in GM model the rectilinearity constraint is introduced through weighting in variance matrix. This means that GM model allows some deviation from the right-angle, i.e. in other words the rectilinearity constraint is relaxed, which depends on the  $\sigma_\alpha^2$ . The weights of boundary points are for both models proportional to the edge probabilities  $p_{\mathbf{x}_i}^2$  (Section 4.2). However, the GM model constitutes of three groups of derived- and pseudo-observations, i.e. squared distances between boundary points and polygon edges  $d^2(i, j)$ , cosines of angles  $\alpha(j)$ , and coordinates of the polygon's vertices  $P_j(X, Y)$ . For each group of observations, a variance must be set.

Let us assume that no correlation between observations  $\mathbf{y}$  exists. Then,  $\Sigma$  is a diagonal matrix,  $\Sigma = \text{diag}[\sigma_1^2, \dots, \sigma_{no}^2]$ , where  $no$  is a number of observations. In this Subsection, weighting for GM model and GH model for a rectilinear polygon is described. Both stochastic models use edge probabilities  $p(\text{edge}) = p(\mathbf{x}_i)$  (Subsection 4.2.3, Equations (34) and (33)) to define the covariance matrices for observations (Figure 5.7).

## Weighting Groups of Observations in GM Model

The GM model constitutes of three groups of derived- and pseudo-observations, i.e. squared distances between boundary points and polygon edges  $d^2(i, j)$ , cosines of angles  $\cos(\alpha_j)$ , and coordinates of the polygon's vertices  $P_j = P(X_j, Y_j)$ . The relations between these three groups of observations must be set by defining the variances of groups of observations. These variances are  $\sigma_d^2$  for  $d^2(i, j)$ ,  $\sigma_\alpha^2$  for  $\cos(\alpha)$ , and  $\sigma_{XY}^2 = \sigma_X^2 = \sigma_Y^2$  for  $(X_j, Y_j)$ . A challenge of setting variances of groups of observations is twofold. One,  $d^2(i, j)$  are observations for GM model, but are derived from actual observations, which are boundary points  $\mathbf{x}$  with given edge probabilities. Two, pseudo-observations  $\alpha_j$  and  $(X_j, Y_j)$  have a different role in a functional model. The  $\alpha_j$  should ensure rectilinearity, i.e. they shall all be  $\frac{\pi}{2}$  or  $\frac{3\pi}{2}$ . The position of the polygon's vertices should be close to the approximate position, which is bounded by the spatial distribution of the boundary points. Therefore, the weights of angles should be much larger than the weights of vertices and boundary points.

Propagation of the given uncertainty  $\sigma_d^2$  of the distance between boundary point and polygon edge to the squared distance  $d^2$  is  $\sigma_d^2 = 2d^2(i, j)\sigma_d^2$ . The boundary points of an object are extracted from a raster image with an equal spacing. In an ideal case the actual polygon boundary is  $d(i, j) \leq \frac{\sqrt{2}}{2}$  pixel away from the boundary point. It follows that the upper boundary is  $\sigma_d^2 = \sigma_d^2$  in units pixel<sup>4</sup>. The standard deviation of  $(i, j)$ -th squared distance between polygon edge and boundary point is

$$\sigma_{d_i}^2 = \sigma_d^2 \frac{1}{p(\mathbf{x}_i)}, \quad (52)$$

for all boundary points  $\mathbf{x}_i$  with corresponding weights  $p(\mathbf{x}_i)$ ,  $i = 1, \dots, I$ .

To ensure rectilinearity, the standard deviation of the angles  $\alpha$  should be small, e.g.  $\sigma_{\alpha} = 1^{\circ}$ . The observation equations for angles are given as  $\cos(\alpha_j)$  (Equation (45)). It follows from the propagating the uncertainty of the angles that the standard deviations of  $\cos(\alpha)$  denoted  $\sigma_{\alpha}^2$  are  $\sin(\alpha)$ . The error is largest at the  $\frac{\pi}{2}$  ( $\frac{3\pi}{2}$ ) angles. The rectilinear buildings are being estimated, therefore the approximation of error by its upper bound is reasonable  $\sigma_{\alpha}^2 = \sigma_{\alpha}^2$ .

The variance matrix  $\Sigma$  of GM model is composed of three sub-matrices and has a size  $I + 3J \times I + 3J$  (Figure 5.7a). These three sub-matrices correspond to the three groups of observations.

### Weighting Groups of Observations in GH Model

The observations in the GH model are boundary points  $\mathbf{x}_i$  with given edge probabilities  $p(\mathbf{x}_i)$ . The variances of a boundary point  $\sigma_{\mathbf{x}_i}^2$  are inverse-proportional to the edge probabilities  $p(\mathbf{x}_i)$

$$\sigma_{\mathbf{x}_i}^2 = \sigma_{xy}^2 \frac{1}{p(\mathbf{x}_i)}, \quad (53)$$

where  $\sigma_{xy}^2 \in \mathbb{R}^+$  for all boundary points  $i = 1, \dots, I$ . Both coordinates of every boundary point have the same standard deviation,  $\sigma_{\mathbf{x}_i}^2 = \sigma_{x_i}^2 = \sigma_{y_i}^2$ . The variance matrix of GM model has a size of  $2I \times 2I$  (Figure 5.7b).

## 5.3 A New Metric for Evaluation of Polygons and Line Segments (PoLiS)

The increasing spatial resolution of satellite and aerial imagery (Figure 2.7) together with the ongoing development of methods (Section 3.1) enable accurate and (semi-)automatic object detection from RS imagery. The extracted objects are in vector or raster format. The latter is usually a result of classification methods in which each pixel is labelled, whereas objects in vector format are represented by e.g. points, lines, and polygons. Evaluation of performance of these methods should be carried out with regard to the format of the result. The current status about evaluation measures is summarised in Section 3.3.

The results of different methods can be compared to each other or to the reference, i.e. ground truth data. Focusing on building extraction and not on general classification of RS images, several authors aim at obtaining 2D or 3D building polygons, which are in vector format. The measures to evaluate similarity between polygons should consider vector data without any alternations (see Subsection 3.3.3). For instance, building outlines in cadastre, which can be used as a reference, are very detailed. These details are not extractable from RS imagery. One way to handle the different level of extracted and reference outlines is to simplify the reference.

A measure is needed to compare reference and extracted polygons, which fulfils the following requirements

- ◇ compares polygons, not only point sets, with different number of vertices,
- ◇ is insensitive to additional points on polygons' edges,
- ◇ is monotonic and also has linear response to small changes in translation, rotation, and scale,
- ◇ is a metric in the mathematical sense.

A level of similarity between two polygons is quantified by using different measures. Several terms are used in the literature to describe these measures, e.g. index, rate, metric. The RS community

has not yet agreed to the consistent usage of these terms for quality measures. In this thesis, the term “measure” is used as a generic term for index, rate, and metric. The terms “rate” and “index” are used as synonyms, but always consistently with the same kind of measure. The terms “metric” or “distance” are used for measures, which fulfil the mathematical requirements for the metric, and the term “directed metric” or “directed distance” is used for measures, which are an intermediate step in the metric computation.

In this Section, commonly used measures for building polygon evaluation are defined and their characteristics are discussed (Subsection 5.3.1). Then a metric for comparison of **Polygons** and **Line Segments** (PoLiS) is proposed (Subsection 5.3.2) and compared to these other measures. The PoLiS metric was first proposed in the journal publication Avbelj et al. (2015b). Its content is paraphrased and extended through additional figures, discussion, and comparisons to other measures than Hausdorff and Chamfer metrics (Subsection 5.3.3).

### 5.3.1 Measures for Quantification of Similarity

In this Subsection, first the term metric is defined and the notation of a point set and a polygon describing a building footprint is given. Then, commonly used measures for building polygon evaluation are defined. They are grouped in three categories (Zeng et al., 2013), i.e. shape similarity measures (Hausdorff, Chamfer, PoLiS), matched rates (completeness, correctness, quality measure), and positional accuracy measures (RMSE, distance between centroids). An emphasise is put on:

- ◊ usage of these common measures with regard to the input data (polygon or point set),
- ◊ fulfilment of the requirements for a metric,
- ◊ codomain and units of the measure,
- ◊ possible applicability to raster data.

All the considered measures compare polygons or point sets with a different number of vertices. The other requirements for a measure to compare reference and extracted polygons (i.e. insensitivity to the additional points on polygons’ edges and monotonic and linear response to small changes in translation, rotation, and scale) are analysed in the subsequent Subsection 5.3.3.

### Metric or Distance Function

A metric  $d$  on a set  $S$  is a distance function  $d : S \times S \mapsto \mathbb{R}$ . For all  $s_1, s_2, s_3 \in S$ , the function  $d$  must satisfy the following conditions (e.g. Copson, 1988; Veltkamp and Hagedoorn, 1999)

$$d(s_1, s_2) \geq 0 \text{ and } d(s_1, s_2) = 0 \Leftrightarrow s_1 = s_2, \quad \text{positive definiteness} \quad (54)$$

$$d(s_1, s_2) = d(s_2, s_1), \quad \text{symmetry} \quad (55)$$

$$d(s_1, s_2) + d(s_1, s_3) \geq d(s_2, s_3). \quad \text{triangle inequality} \quad (56)$$

Non-negativity of a function  $d(s_1, s_2) \geq 0$  can be derived from the coincidence axiom  $d(s_1, s_2) = 0 \Leftrightarrow s_1 = s_2$  (a part of Equation (54)) and the other two conditions in Equations (55) and (56). All the distance functions are denoted with a letter  $d$ , where a subscript specifies the metric. E.g.  $d_e$  denotes Euclidean metric as defined in Equation (57). The arrow  $\vec{\cdot}$  above the distance function is used to denote a directed distance  $\vec{d}$ , which is an intermediate step in some distance function definitions. Directed distance  $\vec{d}$  is in a general case not a metric.

One of the commonly used metrics is the Euclidean distance  $d_e : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}$ . If  $N$ D vectors  $\mathbf{a}$ ,  $\mathbf{b} \in \mathbb{R}^N$ ,  $n = 1, \dots, N$  are two points in Euclidean space, then the Euclidean distance between  $\mathbf{a}$  and  $\mathbf{b}$  is defined as

$$d_e(\mathbf{a}, \mathbf{b}) = \left( \sum_{n=1}^N (a_n - b_n)^2 \right)^{\frac{1}{2}} \quad (57)$$

and denoted also by  $\|\cdot\|$ , i.e.  $d_e(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|$ . The Euclidean distance always meets all the conditions for a metric (Equations (54)–(56)). Thus, the arrow sign  $\vec{\cdot}$  is unnecessary, i.e.  $\vec{d}_e(\mathbf{a}, \mathbf{b}) = \vec{d}_e(\mathbf{b}, \mathbf{a}) = d_e(\mathbf{a}, \mathbf{b})$ .

## Point Set and Polygon

Let us start by defining  $A$  and  $B$  as two sets of 2D points, with elements  $\mathbf{a}_j \in A$ ,  $j = 1, \dots, J$  and  $\mathbf{b}_k \in B$ ,  $k = 1, \dots, K$ , respectively. The coordinates of a 2D point of a point set  $A$  are denoted by  $\mathbf{a}_j = [ax_j, ay_j]^T$ . If points  $\mathbf{a}_j$  of a set  $A$  represent salient points of a shape, e.g. a building footprint, then they can be connected in a closed polygon in  $\mathbb{R}^2$  (Figure 5.8) consisting of straight line segments. A point set and the polygon with the same points and vertices, respectively, are denoted with the same capital letter, e.g.  $A$ . Thus, the points  $\mathbf{a}_j$ ,  $j = 1, \dots, J$  of the set  $A$  represent the vertices of the closed polygon  $A$ , where the first and the last vertex coincide  $\mathbf{a}_1 = \mathbf{a}_{J+1}$ . A boundary  $\partial A$  consists of  $J + 1$  vertices  $\mathbf{a}_j$  of the closed polygon  $A$ ,  $J$  edges, and points that lie on the boundary. Analogically to the point set  $A$ , the point set  $B$  can be considered as a closed polygon  $B$  with  $\mathbf{b}_k$ ,  $k = 1, \dots, K + 1$  vertices. Moreover, if not explicitly written otherwise, the term polygon is used for any closed, piecewise linear, and not self-intersecting polygon. Note that such a polygon can have holes.

A point of a polygon, which has defined coordinates, is referred to as a vertex, even if it is not a corner point of a polygon and lies on the polygon's boundary. Any point, e.g.  $a \in A$ , without subscript can be either a vertex or a point without explicitly defined coordinates. Two vertices are equal, if all coordinates of each dimension are equal. For a 2D case,  $\mathbf{a}_j = \mathbf{b}_k$ , if  $ax_j = bx_k$  and  $ay_j = by_k$ . However, if two point sets are equal this does not imply that polygons with the same vertices are equal too. For example, solid and dotted blue polygon edges (Figure 5.9), connect the same set of vertices in two different polygons. The order of vertices of the solid and the dotted blue polygons is not equal.

Two polygons are considered as equal if all the vertices and their order are equal (circular permutations of the vertices are allowed). All equal vertices in one polygon and vertices lying on the polygon's edges are not accounted for the comparison. For example, a polygon with added vertex exactly on the polygon's boundary and the same polygon without the added vertex are equal. In such cases, two polygons with different number of vertices are considered equal. Another example of two polygons, which are considered equal, are the polygons with the same vertices, but with a different first vertex.

## Shape Similarity Measures: Hausdorff $d_h$ and Chamfer $d_c$ Metrics

Many shape similarity measures are based on distances between points (Veltkamp, 2001). The Euclidean distance  $d_e(\mathbf{a}_j, \mathbf{b}_k)$  between any two points of the sets  $A$  and  $B$  is defined, however it does not define correspondences between the points. Thus, other distances are needed to evaluate shape similarity. Such distances are e.g. bottleneck or taxicab (Manhattan) distance, which require equal size of the sets  $A$  and  $B$  ( $J = K$ ) and one-to-one correspondence between the points. For applications like stereo-matching or comparing generalised to more detailed shapes, distances are needed, which take into account different sizes of the sets  $J \neq K$ , e.g. Hausdorff and Chamfer distances.

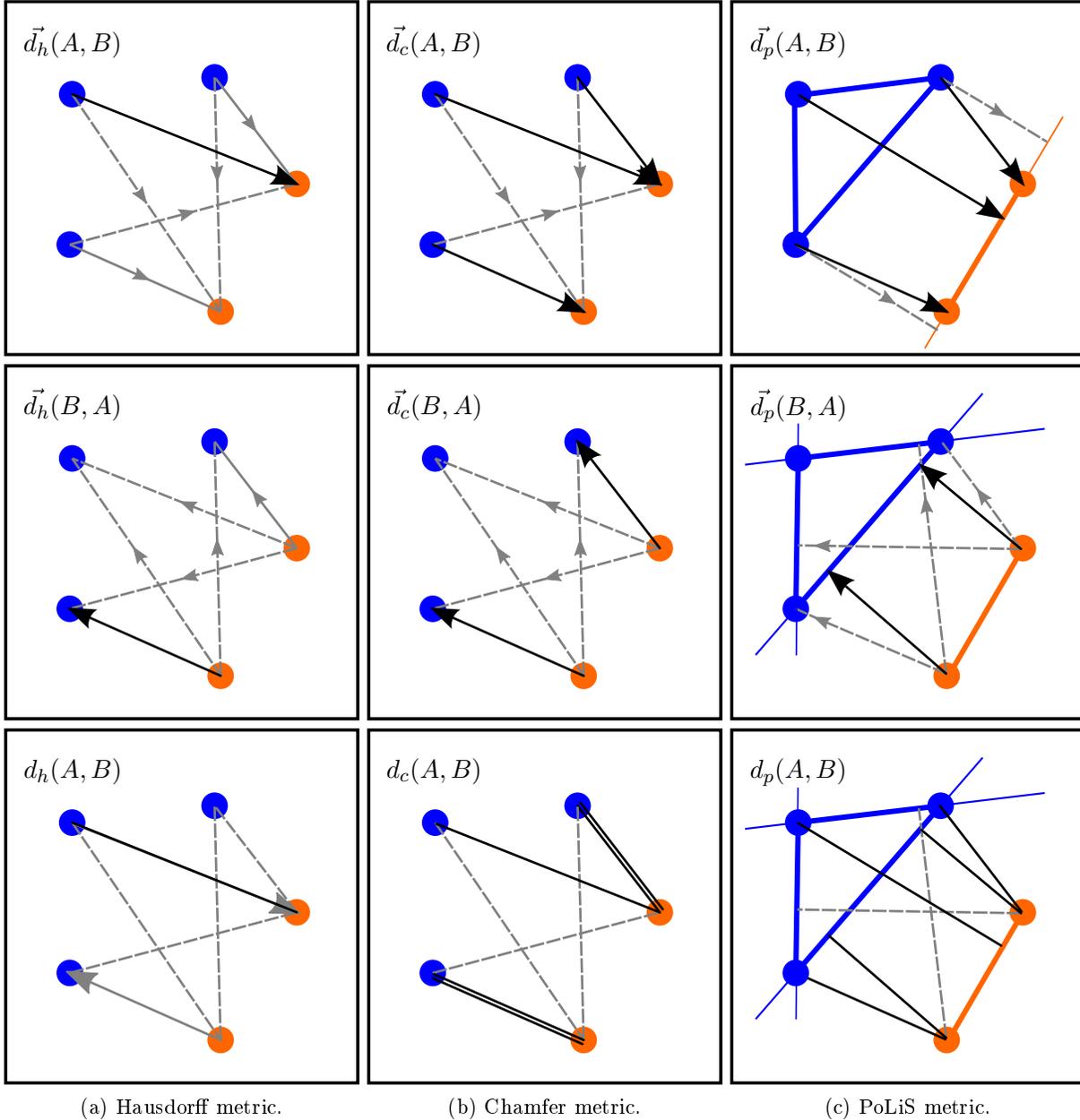


Figure 5.8: The Hausdorff, the Chamfer and the PoLiS metric on an example of two point sets. Distances (black solid lines) between two sets of points, i.e.  $A$  (blue) and  $B$  (orange). Directed  $\vec{d}(A, B)$  (first row),  $\vec{d}(B, A)$  (second row), and symmetric Hausdorff  $d_h$  (Figure 5.8a), symmetric Chamfer  $d_c$  (Figure 5.8b) and symmetric PoLiS  $d_p$  metric (Figure 5.8c) between  $A$  and  $B$  is shown. Arrows represent direction in which the distance is computed, grey (solid or dashed) connections between points show an intermediate step in computing a distance, i.e. an underlying Euclidean distance  $d_e$  between points. For example, for Hausdorff metric, grey solid arrows show the  $\vec{d}_h(A, B)$  and  $\vec{d}_h(B, A)$ , grey (dashed, solid) lines with an arrow all possible Euclidean distances between point sets, and solid grey lines with an arrow the minimum distance from a point to the nearest point in the other point set. The PoLiS metric is defined for polygons and not for point sets, thus connections between the points (blue, orange lines) must be known (Figure 5.8c). This figure is based on a figure in Avbelj et al. (2015b).

A directed Hausdorff distance  $\vec{d}_h(A, B)$  between the set  $A$  and the set  $B$  is defined as the maximum distance (Figure 5.8a, first row) between each point  $\mathbf{a}_j \in A$  and its closest point  $\mathbf{b}_k \in B$

$$\vec{d}_h(A, B) = \max_{\mathbf{a}_j \in A} \min_{\mathbf{b}_k \in B} \|\mathbf{a}_j - \mathbf{b}_k\|. \quad (58)$$

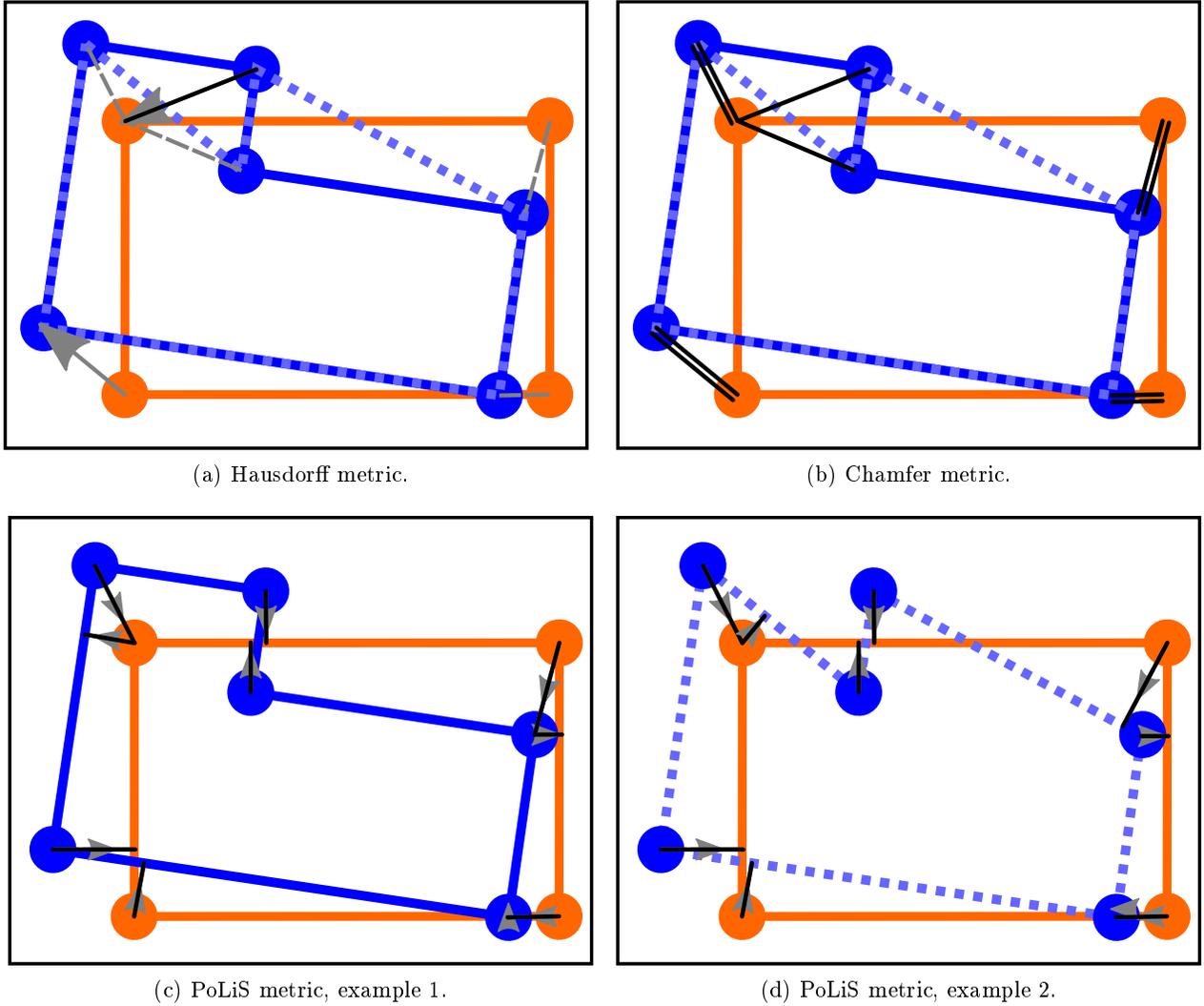


Figure 5.9: A same point set is connected into two different polygons and the influence on the Hausdorff, the Chamfer, and the PoLiS metric computation. PoLiS metric accounts for polygons and not only point sets like the Hausdorff and the Chamfer metrics. Distance (solid black lines) between extracted building footprint  $A$  (orange) and a reference building footprint  $B$  (blue), marked with solid black lines. The PoLiS metric (Figures 5.9c and 5.9d) accounts for the shape, whereas the Hausdorff (Figure 5.9a) and the Chamfer metric (Figure 5.9b) are independent of the connections between the point sets (solid and dotted blue lines). The dotted light blue lines demonstrate one alternative way to connect the point set  $B$  into a regular polygon. This figure is based on a figure in Avbelj et al. (2015b). The meaning of the colours used in this Figure is analogue to the meaning of the colours described in the caption of the Figure 5.8.

A directed Chamfer distance  $\vec{d}_c(A, B)$  between the sets  $A$  and  $B$  is defined as the sum of the distances (Figure 5.8b, first row) between each point  $\mathbf{a}_j \in A$  and its closest point  $\mathbf{b}_k \in B$

$$\vec{d}_c(A, B) = \sum_{\mathbf{a}_j \in A} \min_{\mathbf{b}_k \in B} \|\mathbf{a}_j - \mathbf{b}_k\|. \quad (59)$$

Both, directed Hausdorff  $\vec{d}_h$  and directed Chamfer distance  $\vec{d}_c$ , fail to fulfil the condition of the symmetry (Equation (55)) and are therefore not a metric in the mathematical sense. To fulfil requirements in Equations (54)–(56),  $\vec{d}_h$  is made symmetric by computing the maximum of directed Hausdorff distances (e.g. Veltkamp, 2001)

$$d_h(A, B) = \max\{\vec{d}_h(A, B), \vec{d}_h(B, A)\} \quad (60)$$

or an average of directed Hausdorff distances  $\frac{1}{2}(\vec{h}(A, B) + \vec{h}(B, A))$  (e.g. Tsai et al., 2007; Zhang et al., 2004). The  $d_h$  as defined in Equation (60) is the largest distance of all possible minimal distances between two point sets (Figure 5.8a, third row).

Chamfer distance  $\vec{d}_c$  is made symmetric in analogy with  $\vec{d}_h$  (Equation (60)) by summing the normalised directed Chamfer distances

$$d_c(A, B) = \frac{1}{2q}\vec{d}_c(A, B) + \frac{1}{2r}\vec{d}_c(B, A), \quad (61)$$

or by computing the average between  $\vec{d}_c(A, B)$  and  $\vec{d}_c(B, A)$  (Tsai et al., 2007). The normalisation factors  $\frac{1}{2q}$  and  $\frac{1}{2r}$  account for the number of points of each point set. The  $d_c$  as defined in Equation (61) is an average of the average distances between two point sets (Figure 5.8b, third row).

The Chamfer distance is also defined for the raster binary images, where it is computed with the aid of the distance transform of an image. The distance transform of a binary image assigns to each pixel of the image a value that is a distance between this pixel and the nearest non-zero pixel. The Chamfer distance between the binary reference (also named template) image and another binary image of the same size is computed as a dot product between the distance transform of the reference image and the other image.

The Hausdorff and the Chamfer distances in Equations (60) and (61), respectively are defined with an underlying Euclidean distance. However, any other distance could be underlain. The values of the  $d_h$ ,  $d_c$  are on an interval  $[0, \infty)$  and have the same units as the units of the coordinates of the input datasets. These two distances compare point sets and not polygons, thus they are independent of the connections between the points. Figures 5.9a and 5.9b illustrate an example of how the same point set can be connected into two different simple polygons (solid and dotted blue lines), which does not influence the values of  $d_h$  and  $d_c$  (shown as black lines).

### Matched Rates: Completeness, Correctness, and Quality Rate

Let polygon  $A$  represent an extracted polygon (Figure 5.10a, orange) and polygon  $B$  represent a reference polygon (Figure 5.10a, blue), with the areas  $\text{ar}(A)$  and  $\text{ar}(B)$ , respectively.  $\text{ar}(\cdot)$  denotes the area of a polygon and is computed as

$$\text{ar}(A) = \frac{1}{2} \left| \sum_{j=1}^J ax_j ay_{j+1} - ax_{j+1} ay_j \right|. \quad (62)$$

The match rates, i.e. completeness, correctness, and quality rate, are defined through true positive  $\text{ar}(TP) = \text{ar}(A \cap B)$  (Figure 5.10b, green), false negative  $\text{ar}(FN) = \text{ar}(B) - \text{ar}(A \cap B)$  (Figure 5.10b, red), and false positive  $\text{ar}(FP) = \text{ar}(A) - \text{ar}(A \cap B)$  (Figure 5.10b, purple) detected areas.

Completeness *comp* also named producer's accuracy, recall, detection rate, true positive rate, or matched overlay (Rutzinger et al., 2009; Song and Haithcoat, 2005) is defined as

$$\text{comp} = \frac{\text{ar}(TP)}{\text{ar}(TP) + \text{ar}(FN)}. \quad (63)$$

Correctness *corr* also named user's accuracy, or precision, (Rutzinger et al., 2009) is defined as

$$\text{corr} = \frac{\text{ar}(TP)}{\text{ar}(TP) + \text{ar}(FP)}. \quad (64)$$

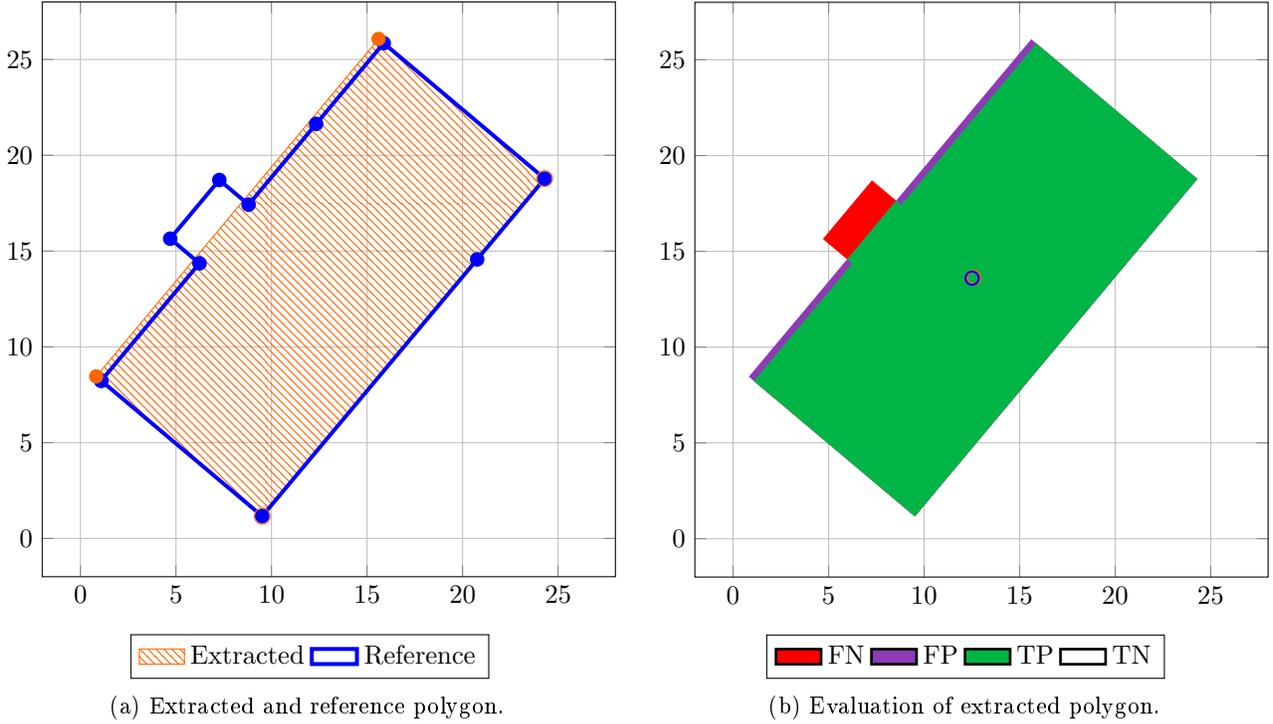


Figure 5.10: Extracted (orange) and reference polygon (blue) of a building outline (Figure 5.10a). The extracted polygon is a generalised representation of the reference polygon, i.e. without division in two units and additional structure. The false negative ( $FN$ ), false positive ( $FP$ ), true positive ( $TP$ ), and true negative ( $TN$ ) detected areas (Figure 5.10b). The centres of the circles are the centroids of the extracted and the reference polygon. This figure is partly based on the figure in Avbelj et al. (2015b).

Quality rate  $qual$  is defined as

$$qual = \frac{ar(TP)}{ar(TP) + ar(FP) + ar(FN)} \quad (65)$$

$$= \frac{comp\ corr}{comp + corr - comp\ corr}. \quad (66)$$

The matched rates are based on the areas and intersection area (or pixel or object count, Section 3.3) of the extracted and the reference polygon. Therefore, different connections between the vertices of a polygon can change the area of the polygon and consequently the values of the matched rates. The matched rates account for polygons and not only point sets in unlike e.g. Hausdorff and the Chamfer distances.

The codomain of the three above defined matched rates (Equations (63)-(66)) is on an interval  $[0, 1]$ , where 0 is the lowest and 1 is the highest value of the matched rate. All the matched rates are dimensionless quantities.

For two equal polygons  $A = B$ , the matched rates are equal one  $comp(A, B) = corr(A, B) = qual(A, B) = 1$ . If the reference and extracted polygon are swapped, the value of correctness is the value of completeness, and vice versa. The  $comp$  and  $corr$  are not symmetric rates, but their combination into the quality rate  $qual$  is symmetric (Equation (66)). The way the  $comp$  and  $corr$  are made symmetric is analogue to combining the directed Hausdorff (Equation (58)) distances or directed Chamfer (Equation (59)) distances to a symmetric metric (Equations (60) or (61), respectively). None of the matched rates fulfils the positive definiteness (Equation (54)) and triangular inequality (Equation (56)) condition for a metric.

The matched rates are also used for pixel based evaluation, where the Equations (63)-(66) are simplified by taking the number of pixels instead of the area  $\text{ar}(\cdot)$ . The matched rates can be computed per-object or per-scene, both using the same equations above. However, some authors compute per-scene matched rates by taking the object counts (Song and Haithcoat, 2005), which requires setting a threshold for the minimum required overlap between a reference and extracted object (Section 3.3).

### Geometrical Quality Measures: $RMSE_{line}$ , $RMSE_{point}$ , and Distance between Centroids

Geometrical quality of the extracted polygons is characterised by e.g. the RMSE of the distances between polygons' vertices or distance between centroids  $dc$  of the extracted and reference polygon (Rutzinger et al., 2009; Song and Haithcoat, 2005; Zeng et al., 2013).

For building polygon comparison, two ways to compute RMSE of the distances between two polygons can be found in the literature. One computational way accounts for only point sets  $RMSE_{point}$  and the other accounts for polygons  $RMSE_{line}$ . The more commonly used and computationally simpler measure is  $RMSE_{point}$ .  $RMSE_{point}(A,B)$  is computed as a square root of the ratio between a sum of the squared minimal Euclidean distances between the vertices of the extracted  $A$  and reference  $B$  polygon and the number of the vertices  $q$  of extracted polygon

$$RMSE_{point}(A, B) = \sqrt{\frac{\sum_{j=1}^J (\min_{\mathbf{b}_k \in B} \|\mathbf{a}_j - \mathbf{b}_k\|)^2}{J}}. \quad (67)$$

Computation of the  $RMSE_{point}$  is similar to the computation of the directed Chamfer distance  $\vec{d}_c$  (Equation (59)), because they both include the sum of the minimal distances, or minimal square distances from one point set to another.

The  $RMSE_{line}(A, B)$  differs from  $RMSE_{point}(A, B)$  in computation of the distances between the vertices of the extracted  $A$  and the reference polygon  $B$  by taking the distance to the nearest point of the polygon  $B(\mathbf{b} \in \partial B)$  and not the nearest vertex of the polygon  $B(\mathbf{b}_k \in B)$ . It is defined as

$$RMSE_{line}(A, B) = \sqrt{\frac{\sum_{j=1}^J (\min_{\mathbf{b} \in \partial B} \|\mathbf{a}_j - \mathbf{b}\|)^2}{J}}. \quad (68)$$

Both,  $RMSE_{point}$  and  $RMSE_{line}$ , can be computed from extracted to reference polygon, and vice versa. This is analogue to the computation of the directed metrics  $\vec{d}_h$  and  $\vec{d}_c$ . None of them is a symmetric measure, thus they are not a metric in mathematical sense. Some authors (e.g. Rottensteiner et al., 2012) exclude all the distances  $\|\mathbf{a}_j - \mathbf{b}_k\|$ ,  $\mathbf{a}_j \in A$ ,  $\mathbf{b} \in B$  (or  $\mathbf{b} \in \partial B$ ) exceeding a predefined threshold from computation of the  $RMSE_{point}$  (or  $RMSE_{line}$ ) between the polygons. This threshold compensates for possible gross errors, inaccurate extraction, and difference in the level of detail of extracted and reference polygons.

The centroid of a polygon  $A$  denoted  $\mathbf{c}(A) = [cx_A, cy_A]^T$  (Figure 5.10a, orange circle) is computed as

$$cx_A = \frac{1}{6 \text{ar}(A)} \sum_{j=1}^q (ax_j + ax_{j+1})(ax_j ay_{j+1} - ax_{j+1} ay_j) \quad (69)$$

$$cy_A = \frac{1}{6 \text{ar}(A)} \sum_{j=1}^q (ay_j + ay_{j+1})(ax_j ay_{j+1} - ax_{j+1} ay_j), \quad (70)$$

```

1: procedure POLISMETRIC( $A, B$ )
2:    $p1, p2 \leftarrow 0$ 
3:   for  $j = 1, \dots, J$  do ▷ for every point  $\mathbf{a}_j \in A$ 
4:     for  $k = 1, \dots, K$  do ▷  $\mathbf{b}_k, \mathbf{b}_{k+1} \in B$ 
5:        $p1 \leftarrow p1 + \text{DISTPT2LNSEG}(\mathbf{a}_j, \mathbf{b}_k, \mathbf{b}_{k+1})$ 
6:     end for
7:   end for
8:   for  $k = 1, \dots, K$  do ▷ for every point  $b_k \in B$ 
9:     for  $j = 1, \dots, J$  do ▷  $\mathbf{a}_j, \mathbf{a}_{j+1} \in A$ 
10:       $p2 \leftarrow p2 + \text{DISTPT2LNSEG}(\mathbf{b}_k, \mathbf{a}_j, \mathbf{a}_{j+1})$ 
11:    end for
12:  end for
13:   $p \leftarrow \frac{p1}{2J} + \frac{p2}{2K}$ 
14:  return  $p$  ▷ PoLiS distance value
15: end procedure

```

Figure 5.11: Pseudo-code for computing PoLiS metric between two closed polygons,  $A$  and  $B$ . The procedure `DISTPT2LNSEG` computes the shortest distance between a point and a line segment given by two points. For a 3D case, the procedure `DISTPT2LNSEG` is replaced by a procedure, which computes a distance between a 3D point a polyhedron.

where  $\text{ar}(A)$  is the area of the polygon  $A$  (Equation (62)). Then, the Euclidean distance between centroid of the polygon  $A$  and  $B$ ,  $d_e(\mathbf{c}(A), \mathbf{c}(B))$ , is concisely denoted  $dc(A, B)$ .

The values of the  $RMSE_{point}$ ,  $RMSE_{line}$ , and  $dc$  are all on an interval  $[0, \infty)$  and their units are the same as the units of the coordinates of the vertices of the polygons. All three above defined geometrical quality measures can be computed for each coordinate separately, i.e.  $RMSE_x(A, B)$ ,  $RMSE_y(A, B)$  (*point* or *line*),  $dc_x(A, B)$ , and  $dc_y(A, B)$ .

For two equal polygons,  $A = B$ ,  $RMSE(A, B) = 0$  and  $dc(A, B) = 0$ . The  $dc$  is in contrast to  $RMSE_{point}$  and  $RMSE_{line}$  a symmetric measure, but it does not fulfil the positive definiteness requirement for a metric (Equation (54)).

### 5.3.2 Definition of the PoLiS Metric

A directed PoLiS distance  $\vec{d}_p(A, B)$  between polygons  $A$  and  $B$  (Figure 5.8c, first row) is defined as the average of the distances between each vertex  $\mathbf{a}_j \in A$ ,  $j = 1, \dots, J$  of  $A$  and its closest point  $\mathbf{b} \in \partial B$  (not necessary a vertex) on the polygon  $B$

$$\vec{d}_p(A, B) = \frac{1}{J} \sum_{\mathbf{a}_j \in A} \min_{\mathbf{b} \in \partial B} \|\mathbf{a}_j - \mathbf{b}\|. \quad (71)$$

The directed PoLiS distance  $\vec{d}_p$  is made symmetric, similar like  $\vec{d}_c$  (Equations (59) and (61)), by summing and normalising the directed distances

$$d_p(A, B) = \frac{1}{2J} \sum_{\mathbf{a}_j \in A} \min_{\mathbf{b} \in \partial B} \|\mathbf{a}_j - \mathbf{b}\| + \frac{1}{2K} \sum_{\mathbf{b}_k \in B} \min_{\mathbf{a} \in \partial A} \|\mathbf{b}_k - \mathbf{a}\|. \quad (72)$$

The pseudo-code for PoLiS metric computation is given in Figure 5.11. The first summand in Equation (72) is the average of the distances between each vertex of  $\mathbf{a}_j \in A$ ,  $j = 1, \dots, J$  and its closest

point  $\mathbf{b} \in \partial B$  on the polygon  $B$ . The second summand in Equation (72) is computed vice versa. Both summands are normalised by  $\frac{1}{2}$  to quantify the overall average dissimilarity per point. This normalisation factor is the same like for the Chamfer distance in Equation (61). The PoLiS distance is on the interval  $[0, \infty)$  and has the same units as units of the polygon vertices.

The PoLiS metric combines some properties of the Chamfer distance and  $RMSE_{line}$ . It is a metric in mathematical sense, like Chamfer distance, but accounts, like  $RMSE_{line}$ , for polygons and not only point sets (Figures 5.9c and 5.9d). Other characteristics of the PoLiS metric are discussed in the following Subsection 5.3.3.

### 5.3.3 Characteristics of the PoLiS Metric

In this Subsection, characteristics of the proposed PoLiS metric are discussed and compared to the measures defined in Subsection 5.3.1. Moreover, it is explained through examples how well the PoLiS metric fulfils the requirements for a measure to compare reference and extracted polygons. These requirements are listed in the introduction of Section 5.3. In addition, the consistency of a visual perception of quantification of similarity between polygons and quantification by the above defined shape similarity measures (Subsection 5.3.1) is discussed.

The main focus is on the insensitivity of the PoLiS metric to additional points on polygons' edges and response to the small changes in translation, rotation, and scale. The examples are given on the 2D building polygons. However, the PoLiS metric may also have broader applications in the field of shape similarity and comparison of polygons.

#### Comparison of Polygons with Different Number of Vertices

The PoLiS metric and  $RMSE_{line}$  consider shapes of the polygons by computing distances to the polygon edges and not only between the point sets like e.g.  $RMSE_{point}$ , Hausdorff and Chamfer distances (Figure 5.9). This means that the point sets must be arranged and connected with edges into open or closed polygons (Figure 5.8c). For application in this thesis, the polygons are always closed polygons representing building outlines. The number of vertices of polygons to be compared can differ  $J \neq K$ .

If one of the polygons has much larger number of vertices than the other, the numerical value of the PoLiS metric underestimates the actual dissimilarity, because of normalisation factors. Nevertheless, under the assumption of small translation, rotation and scale between the polygons, the relations between values of PoLiS distances are consistent relative to each other in contrast to the values of Hausdorff and Chamfer distances.

The matched rates are independent of the number of the polygons' vertices, because their computation is based on the areas and intersection area of the extracted and the reference polygon.

#### Insensitivity to Additional Points on Polygons' Edges

The PoLiS metric is robust towards partitioning of the polygon, i.e. adding vertices on the polygon edges. The values of the  $d_p$  (green),  $d_c$  (blue), and  $d_h$  (red) are compared (Figure 5.13), when additional points on the polygon's edges, considered as vertices, are added (Figure 5.13). The examples are given through two simple geometrical shapes, a line segment (Figure 5.12) and a square (not graphically shown). The vertices are added randomly (Figure 5.12a), from one side, i.e. the added vertices have a constant distance to the previously added vertex (Figure 5.12b), and as equally distributed, i.e. the distance between subsequent vertices is equal (Figure 5.12c).

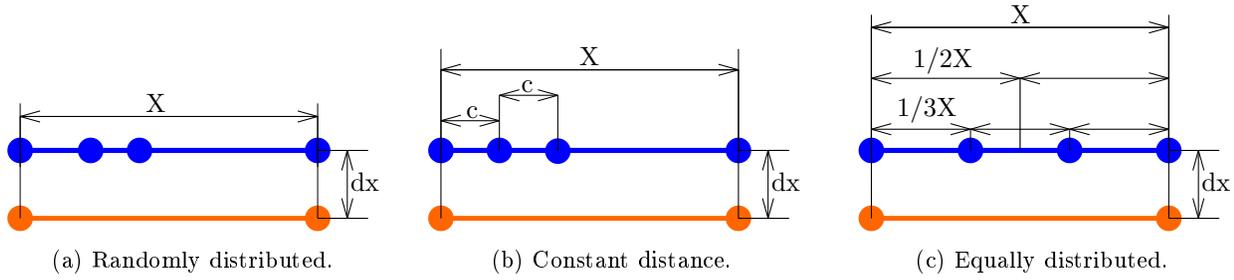


Figure 5.12: The additional vertices on a line segment to test sensitivity of the Hausdorff, the Chamfer and the PoLiS metrics. The extracted line segment (orange) has two vertices, whereas to the reference (blue) line segment the vertices are added randomly (Figure 5.12a), with a constant distance from one side (Figure 5.12b), or equally distributed (Figure 5.12c). The  $dx$  is translation between extracted and reference polygon and is introduced only to enhance visualisation.  $X$  is the length of both line segments and  $c$  is a constant distance between vertices (Figure 5.12b).

There are two main findings about metrics behaviour when adding the vertices, which lie on polygon edges. First, the PoLiS metric is independent of the manner how the vertices are added, unlike Hausdorff and Chamfer metrics. This is seen by comparing the graphs on Figures 5.13a–5.13c, or by comparing the definitions of all metrics (Equations (60), (61), and (72)). The graph of a  $d_p$  metric has the same shape, whereas graphs of the  $d_h$  and the  $d_c$  metrics vary depending on how the vertices are added. Second, the value of PoLiS metric is independent of the number of vertices added, i.e. the graph is nearly constant. The value of  $d_p$  varies slightly, because the compared line segments are translated for  $dx$ . This translation between line segments (and squares) is introduced for visualisation reasons only. The values of the  $d_h$  and  $d_c$  are dependent on the number of vertices added.

The matched rates are independent of the number of the added vertices on polygons' edges, because their computation is based on the areas and intersection area of the extracted and the reference polygon.

### Nearly Linear Response to Small Changes in Translation, Rotation, and Scale

To compare extracted polygons to the reference polygons, both data (or images from, which polygons are extracted), must be coregistered. Then, the expected differences in translation, rotation, and scale between the two datasets (images) are normally much smaller than e.g.  $\pm 3$  m,  $\pm 5^\circ$  and  $1 \pm 0.1$ . The difference of the value of the PoLiS metric to small translation, rotation, and scale differences is nearly linear.

Let us take two polygons of equal area, the reference polygon (Figure 5.10a, blue) and the extracted polygon (Figure 5.10a, orange). The reference polygon has a small structure that is not captured in the extracted polygon, and additional two vertices on the polygon's edges. This is a typical building polygon extraction scenario, where two additional vertices dividing a building into two building units are not detectable from RS images and the small structure is not distinguishable due to the spatial resolution of the image. The extracted building is translated, rotated, and scaled according to the centroid of the building polygon. Then, the  $d_h$ ,  $d_c$ ,  $d_p$  (Figure 5.14), the matched rates,  $RMSE_{point}$ ,  $RMSE_{line}$  values, and distance between centroids  $dc$  (Figure 5.15) are computed.

The Chamfer and the PoLiS distances have a minimum value at an initial position of the extracted polygon. The impact of the variations in translation, rotation, and scale to the values of  $d_c$  and  $d_p$  can be well approximated by a linear function. Different minimum values and slopes are a consequence of the definitions of metrics. Thus, the values of the metrics should be compared relative to each other. The Hausdorff distance does not have a minimum value at an initial position, it is not monotonic

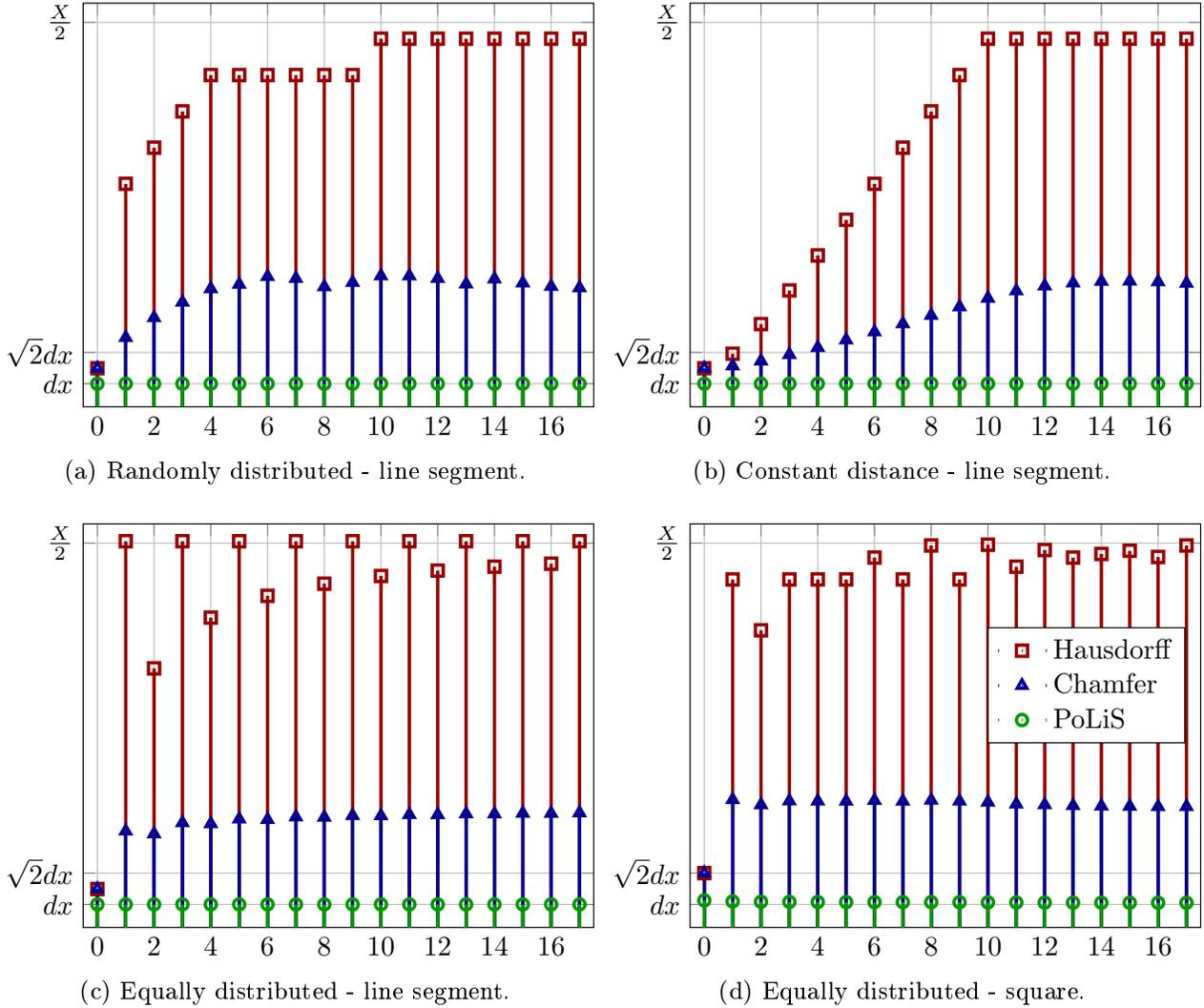


Figure 5.13: Sensitivity of the Hausdorff (red), the Chamfer (blue), and the PoLiS metrics (green) to the additional points on a line segment (Figure 5.12) and a square. The PoLiS metric is insensitive to the number and a way the vertices are added to the line segment or the square, unlike Hausdorff and Chamfer metrics. The added vertices to the reference line segment are added randomly (Figure 5.12a), with a constant distance from one side (Figure 5.12b), or equally distributed (Figure 5.12c). Translation  $dx$  between the reference and the extracted square with a side  $X$  is equal in both direction. Equally distributed points are added to the square polygon (Figure 5.13d). The general behaviour of the values of  $d_h$ ,  $d_c$ , and  $d_p$  is the same for added vertices to the line segment with a length  $X$ , or to the square with the side  $X$ . This figure is based on a figure in Avbejl et al. (2015b).

with respect to the initial position of a polygon, and is therefore not an appropriate measure for quantification of similarity between two polygons. The graphs in the first column of Figure 5.14 are zoomed-in parts of the graphs (grey dashed lines) in the second column. It can be observed that  $d_c$  and  $d_p$  are not monotonic for larger variation of translation and rotation (Figures 5.14b and 5.14d outside of the grey dashed lines). This is a break point of a measure, and occurs when vertices of the reference polygon are assigned to the wrong vertices ( $d_h$ ,  $d_c$ ) or wrong edges ( $d_p$ ) of the extracted polygon, or vice versa.

The matched rates (Figures 5.15a, 5.15c and 5.15e), completeness  $comp$  (cyan), correctness  $corr$  (yellow dashed line), and quality rate  $qual$  (magenta dashed line) are computed as 1 minus the matched rate for easier visual comparison to the graphs in Figure 5.14. The values  $1 - comp$  and  $1 - corr$  are the same for variations in translation and rotation, because the  $ar(A) = ar(B)$ , and the rotation is around

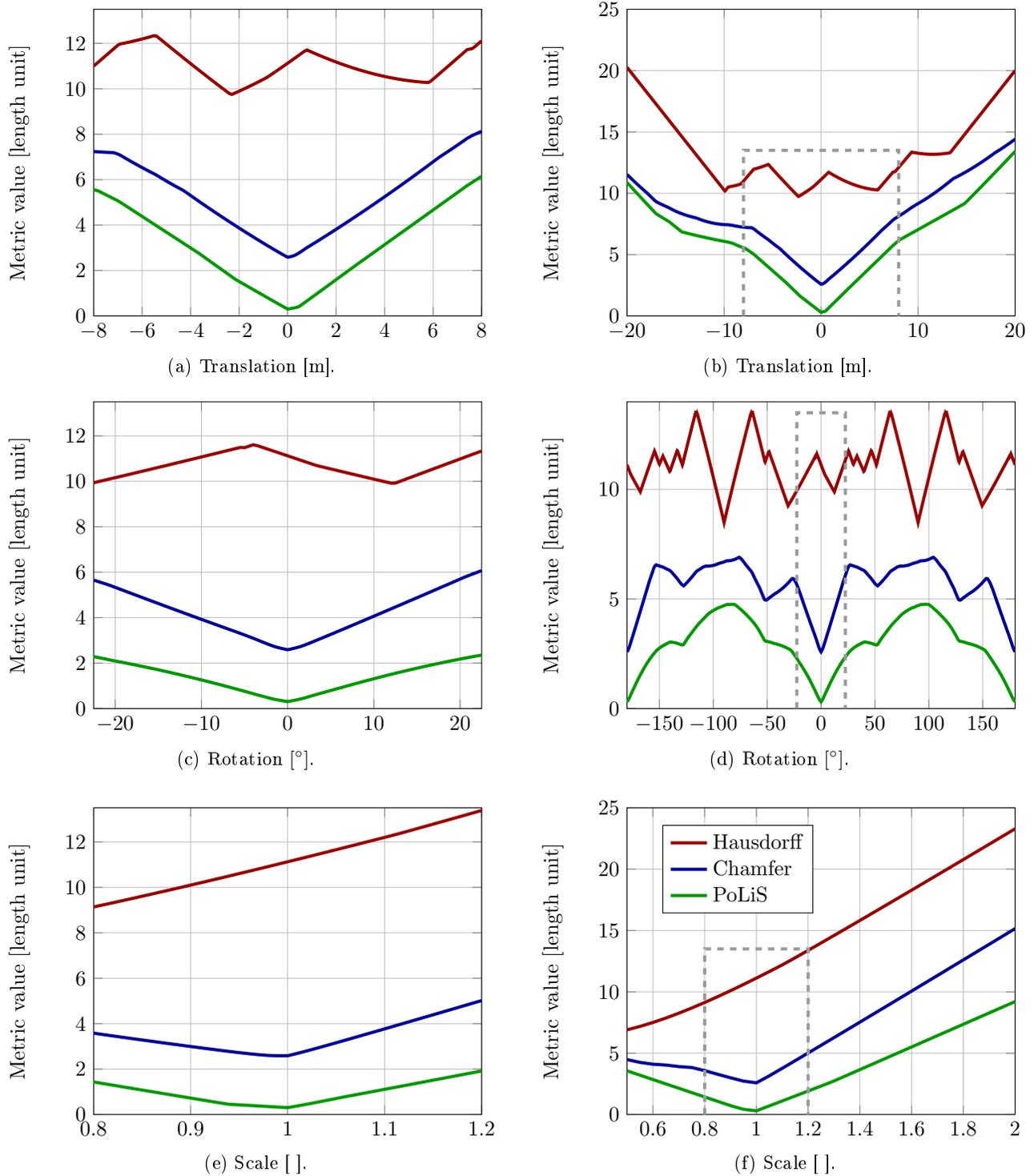


Figure 5.14: The Hausdorff (red), the Chamfer (blue), and the PoLiS (green) metric as a function of translation (Figures 5.14a and 5.14b), rotation (Figures 5.14c and 5.14d), and scale (Figures 5.14e and 5.14f). Figure 5.10a shows the initial position of the reference and the extracted polygon. The graphs in the first column show zoomed in graphs (grey dashed lines) of the graphs in the second column. It can be observed that all considered metrics have a break point. This break point is, when vertices of the reference polygon are assigned to the wrong vertices ( $d_h$ ,  $d_c$ ) or wrong edges ( $d_p$ ) of the extracted polygon, or vice versa. This figure is taken from Avbelj et al. (2015b).

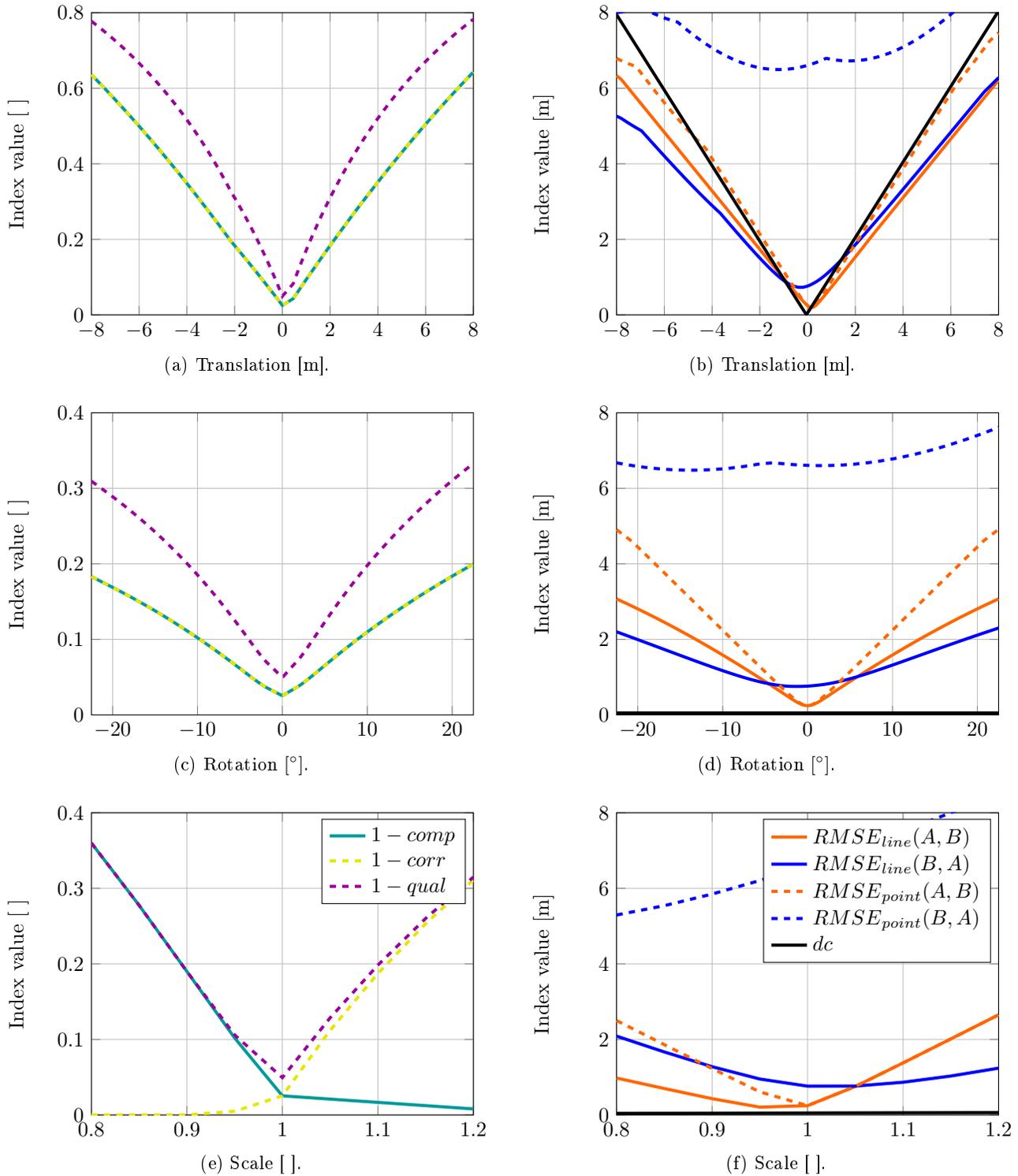


Figure 5.15: The matched rates, completeness *comp* (cyan), correctness *corr* (yellow dashed line), and quality rate *qual* (magenta dashed line) as a function of translation (Figure 5.15a), rotation (Figure 5.15c), and scale (Figure 5.15e). All matched rates are dimensionless quantities. For easier visual comparison to the graphs in Figure 5.14, all the matched rates computes as 1 minus the matched rate. Geometrical quality measures, i.e.  $RMSE_{point}(A, B)$  (dashed orange line),  $RMSE_{point}(B, A)$  (dashed blue line),  $RMSE_{line}(A, B)$  (solid orange line),  $RMSE_{line}(B, A)$  (solid blue line), and distance between centroids  $dc$  as a function of translation (Figure 5.15b), rotation (Figure 5.15d), and scale (Figure 5.15f). All geometrical quality measures are in [m].

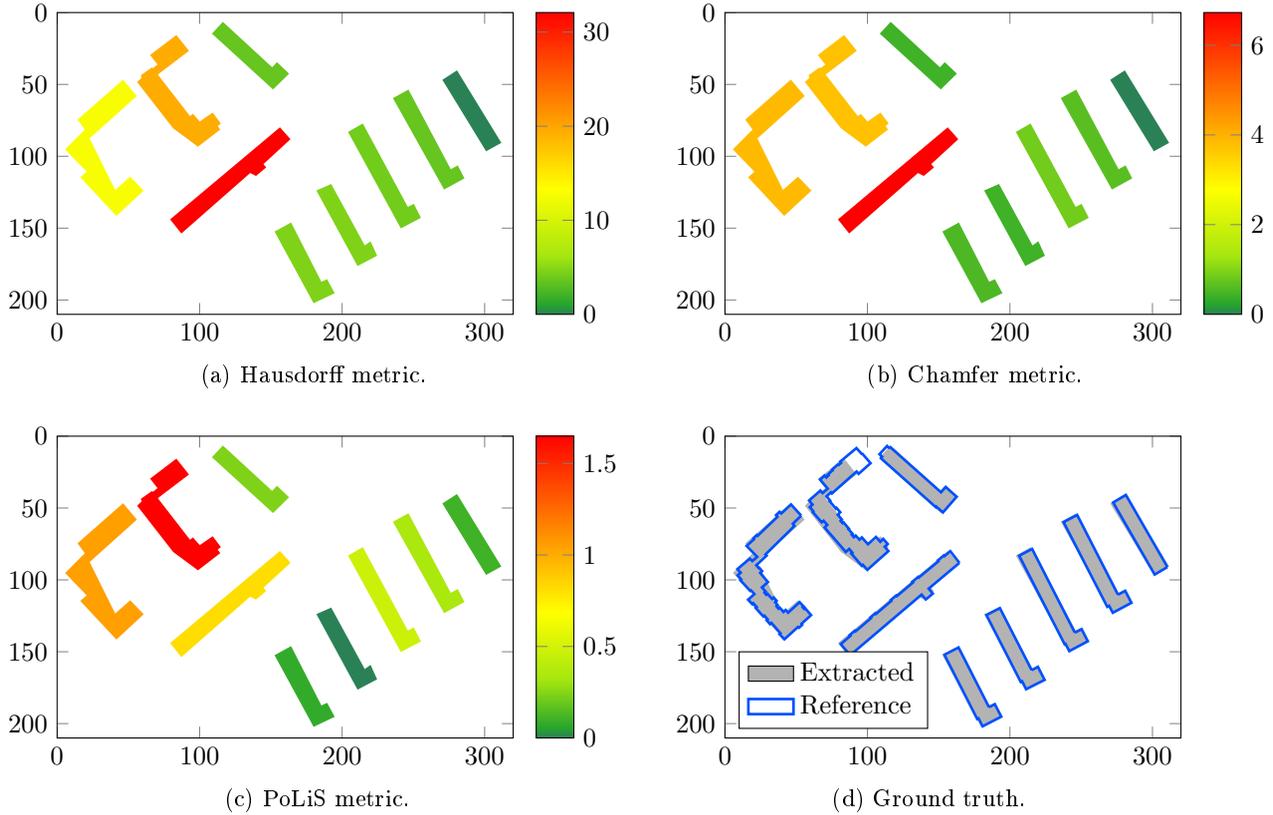


Figure 5.16: Evaluation of building polygon extraction by the Hausdorff (Figure 5.16a), the Chamfer (Figure 5.16b), and the PoLiS (Figure 5.16c) metric. All units are in [m]. For easier visual interpretation, all the colour bars are scaled from the worst (red) over middle (yellow) to the best (dark green) extracted building polygon. Figure 5.16d shows the extracted (gray) and reference (blue) building polygons. This figure is taken from Avbelj et al. (2015b).

the centroid, respectively. In a general case, values of  $comp$  and  $corr$  differ. All the matched rates have monotonic, and nearly linear response to the small translations, rotations and scale. Likewise, responses by  $RMSE_{line}$  in both directions and  $RMSE_{point}(A, B)$  (Figures 5.15b, 5.15d, and 5.15f) can be well approximated by a line, even if the response for small variations in translation and rotation is parabola-like. The distance between centroids relies only on a single point of each polygons. Thus, it is not appropriate as a single measure for quantification of polygon similarity.

### The PoLiS Metric is a Metric in Mathematical Sense

The PoLiS metric is, like the Hausdorff and the Chamfer metrics, a metric in a mathematical sense, i.e. it meets all the conditions in Equations (54)-(56) (Subsection 5.3.1). In contrast, the  $RMSE_{point}$  and  $RMSE_{line}$  are not metric.

### Consistency with a Visual Perception

Figure 5.16 shows an example of a real dataset and comparison of the  $d_h$ ,  $d_c$ , and  $d_p$  metrics. The building polygons are extracted from the DSM (Figure 5.16d, grey), and the reference polygons (Figure 5.16d, blue) are detailed cadastral data. For easier visual interpretation, all the colour bars are scaled from the worst (red) over middle (yellow) to the best (dark green) extracted building polygon (Figures 5.16a–5.16c), i.e. from the minimal to the maximal value of each metric. The rectangular and elongated L-shape buildings are better extracted, than more complex-shaped building. The values of all metrics for simpler-shaped buildings is low (green). However, the worst extracted building

according to the  $d_h$  and  $d_c$  is the elongated building with several small structures (Figures 5.16a and 5.16b). According to  $d_p$ , the worst estimated building is the one of which a part was not extracted due to the vegetation on a roof part. Thus, the PoLiS metric penalises  $FN$  detected areas more than generalisation of the boundary. This is consistent with the visual interpretation and corresponds to the application at hand.

## Summary and Outlook of This Chapter

A generic workflow for building region extraction is proposed, with the focus on HSI and DSM datasets. Then, a model driven method is introduced for creating building polygons by an iterative MBR approach. The most appropriate building polygon is automatically selected as a trade-off between complexity and fitness to the data. Finally, two mathematical models are proposed to adjust the building model. They both use edge probabilities, which are computed from DSM and HSI datasets, as weights in the adjustments.

The development of a new Polygons and Line Segments (PoLiS) metric is one of the main contributions of this thesis. The PoLiS metric quantifies overall average dissimilarity between polygons. An extensive comparison to shape similarity measures and matched rates is carried out. The PoLiS metric is symmetric, and has nearly linear response to small variations in translation, rotation, and scale. The same is true for quality rate, which is estimated on the basis of polygon areas. However, the PoLiS metric has an advantage on estimating the dissimilarity directly from polygon vertices and edges. It is a combined measure which accounts for positional and shape differences. These are crucial characteristics for comparison of polygon objects extracted from RS imagery. Moreover, it can be straightforwardly extended to a 3D PoLiS metric.



## 6 Case Studies

The methods introduced in the previous two Chapters are tested on HSI and DSM datasets (Figure 6.1). The sensor characteristics and the test datasets are presented and the preprocessing steps are described (Section 6.1). The tests are carried out in the same sequence as they are introduced in the Chapters 4 and 5. First, the necessity of the edge probability detection in scale space is analysed on HSI (Section 6.2). Second, the iMBR method for creation and selection of BP is examined and both LS adjustment models are tested on synthetic images and compared to each other (Section 6.3). Finally, the whole workflow is applied on RS imagery (Section 6.4).

### 6.1 Data Description and Preprocessing

#### 6.1.1 HSI Sensors and Images

The developed methods are applied on HSI acquired by two air-borne sensors, namely HyMAP and HySpex. Different types of detectors are used to measure the reflected radiation in VNIR and SWIR spectral regions. Thus, the HSI sensors consist of more than one sensor or several spectral modules. The name of these sensors and modules is related to the names of the spectral regions in which they measure reflected radiation. However, the spectral ranges of sensors and spectral modules might not completely overlap with the spectral regions as defined in Section 2.1 and Figure 2.1b.

**HyMAP** was manufactured by Integrated Spectronics Pty Ltd, Australia, and is operated by HyVista Corporation. It consists of four spectral modules, each containing 32 bands. The average spectral sampling distance differs for each module, specifically it is 13 nm for SWIR1, 15 nm for VIS and NIR, and 17 nm for SWIR2 spectral module. Selected specifications of the HyMAP HSI sensor are listed in Table 2, and further ones can be found in HyVista Corporation (1999). For typical air-borne flight altitudes of 1-5 km the GSD<sup>3</sup> of the HyMAP HSI is between 2.0-10.0 m. The HyMAP dataset used for experiments has a GSD of 4 m.

**HySpex** was manufactured by Norsk Elektro Optikk AS (NEO) and was purchased by Deutsches Zentrum für Luft- und Raumfahrt (DLR) in 2011. The sensor was calibrated at the DLR's Calibration home base, i.e. radiometric, geometric, and spectral properties were characterised. HySpex comprises of two spectral sensors, these are VNIR and SWIR, containing 160 and 256 bands, respectively. The average spectral sampling distance is 3.7 nm for VNIR and 6.0 nm for the SWIR sensor. The VNIR and the SWIR HySpex images can be joined in a postprocessing procedure. Selected specifications of the HySpex HSI sensor are listed in Table 2 and further ones can be found in Norsk Elektro Optikk AS (1985). For typical air-borne flight altitudes of 1-5 km the GSD<sup>3</sup> of the VNIR and the SWIR HSI is between 0.19-0.95 m and 0.75-3.75 m, respectively

The HySpex HSI used in this thesis were all acquired on 16.06.2012, while the sensor was still in a test phase. Thus, the keystone and smile effect were corrected by using preliminary calibration tables. Two areas over the city of Munich were acquired at three different altitudes. All the stripes have roughly north-south direction. The field expander (also referred to as FOV expander) was used, thus the swath width of all the stripes is extended by a factor of two (see \* in Table 2). Consequently, the actual pixels size is larger for the same altitude. For one flight altitude, both VNIR and SWIR images are resampled to the same GSD. The characteristics of the VNIR and the SWIR HySpex sensors differ, and therefore also the characteristics of the VNIR and the SWIR HySpex images. Due to these

<sup>3</sup>The GSD values are given for across-track Instantaneous Field of View (IFOV) for the pixel at nadir. The pixels at the sides of swath are larger, as well as they can differ in across-track direction.

	Spectral Bands	Spectral range [ $\mu\text{m}$ ]	FOV [ $^\circ$ ] (pixel)	IFOV across-/along- track [mrad]
<b>HyMAP</b>	128	0.45-2.48		
VNIR	32	0.45-0.89		
NIR	32	0.89-1.35		
SWIR1	32	1.40-1.80	61.3 (512)	2.0 / 2.5
SWIR2	32	1.95-2.48		
<b>HySpex</b>	416	0.4-2.5		
VNIR-1600	160	0.41-0.99	17* (1600)	0.18 / 0.36*
SWIR-320m-e	256	0.97-2.50	13.5* (320)	0.75/ 0.75*

\*can be doubled with field expander

Table 2: Selected sensor specifications of whisk-broom HyMAP and push-broom HySpex sensor. Both HSI sensors operate in a similar spectral range, these are VNIR and SWIR spectral regions. The spectral modules and sensors constituting the HyMAP and the HySpex, respectively have the same names as the spectral regions in which they acquire reflected radiation. The spectral range of these modules and sensors does not completely overlap with the spectral ranges defined in Section 2.1. Further characteristics of the sensors can be found on the producers' web pages (HyVista Corporation, 1999; Norsk Elektro Optikk AS, 1985).

different characteristics and different resampling scale, the VNIR and the SWIR images are considered separately. Three VNIR and three SWIR HySpex images are used for experiments, with GSD, 0.3 m, 1.0 m, and 2.0 m. The images are denoted according to the HySpex sensor name and the corresponding GSD, e.g. VNIR03 for VNIR image with 0.3 m GSD, and SWIR10 for SWIR image with 1.0 m GSD.

During this flight campaign the HySpex sensor was not mounted on a drift frame (stabilising sensor mount). A drift frame stabilises the drift and compensates for angular motions, i.e. roll and pitch angles. The influence of the angular motions to the acquired data is corrected by carrying out geometric corrections. Among the three angular motions, the roll motion has the most prominent influence to the location of the pixels. For the given HySpex HSI, the roll effect could not be fully correct for. This effect is more prominent in SWIR images than in VNIR images due to the different resampling rate.

## Preprocessing of HSI Images

All used HSI are geometrically and radiometrically corrected and orthorectified (see Subsection 2.1.2). The noisy bands caused by, e.g. atmospheric absorption regions, are removed. The applied methods require as an input material maps of materials under consideration. Thus, reference spectra of at least these materials are collected manually from the HSI. To suppress the noise, the reference spectra is an average of typically between 20–30 samples (pixels) of one material. In order to compensate for illumination effects, the spectra of the HSI and of the reference spectra are normalised before unmixing. The complete set of reference spectra present in a scene is not given, therefore the NNLS unmixing is applied on the HSI, which results in a *material map* for each reference material.

The *building regions*, detected from material maps, are all the regions with the abundances of a roofing material under consideration larger than 0.7 (Avbelj et al., 2015a).

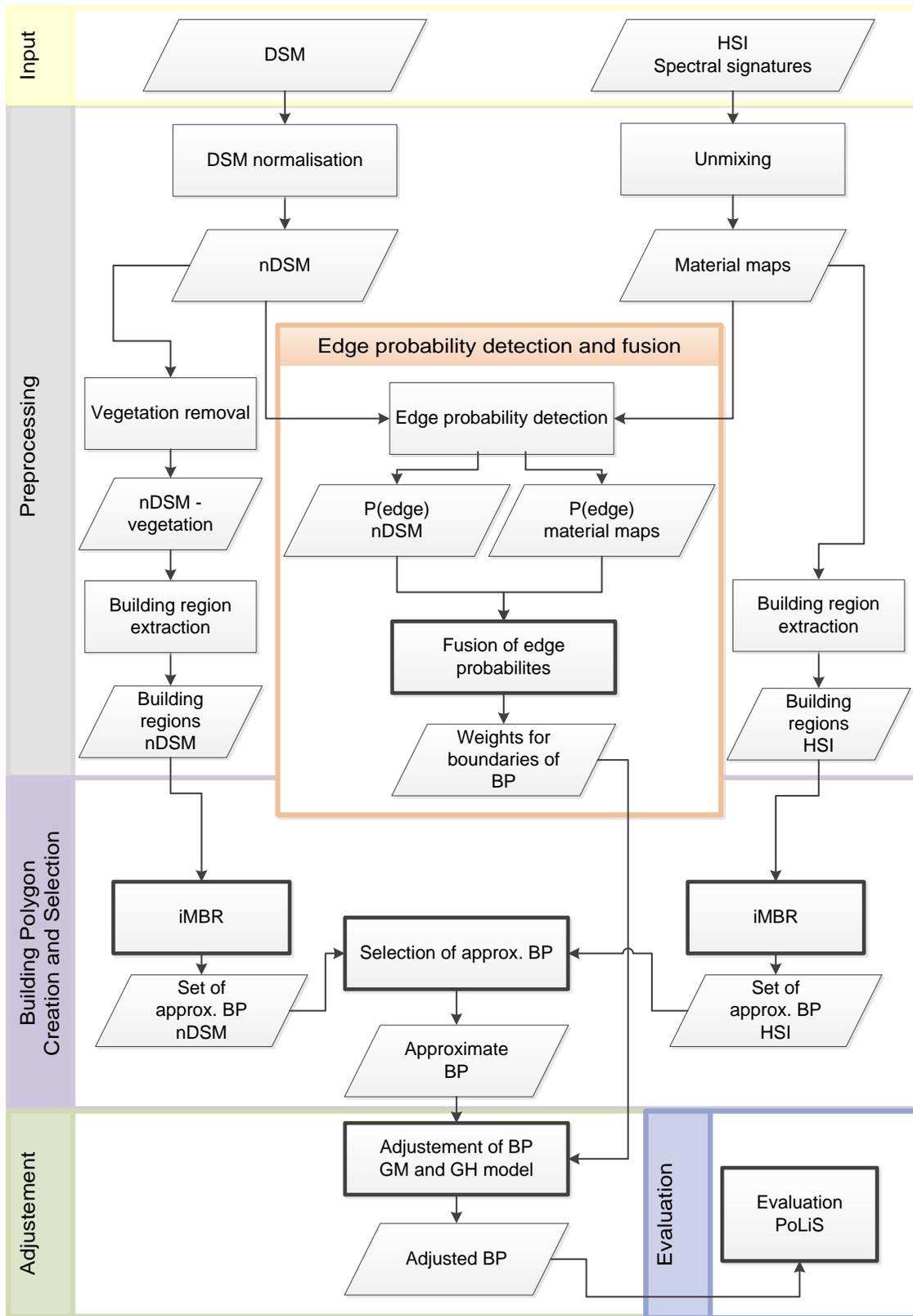


Figure 6.1: Workflow for extraction of BP with sub-pixel precision from HSI and DSM. Arrows point into the direction of subsequent steps and connect data (parallelogram box) to the processes or methods (rectangle box). The input datasets are HSI and DSM of the same area, and the spectral signatures of the roofing materials (yellow box). First, some preprocessing of the HSI and the DSM is required. Then, the edge probability detection and fusion (orange box) is carried out, and iMBR method is applied on the preprocessed images to create and select the approximate BP (violet box). The approximate BP with corresponding edge probabilities are adjusted by the LS adjustment (green box). Finally, the adjusted BP are evaluated (blue box). The methods, which are a main contribution to this thesis are marked by bold rectangle boxes, i.e. fusion of edge probabilities, iMBR method for BP extraction and automatic selection of approximate BP, two adjustment models, and PoLiS metric for comparison of polygon data.

### 6.1.2 DSM from Stereo Images and LiDAR Point Clouds

Two DSM used for experiments are both computed from stereo images by the SGM method (Section 2.2). The original SGM DSM output is used for experiments, i.e. holes are not filled and no smoothing of the DSM is carried out. Additionally, a LiDAR DSM is used for the edge probability fusion experiment.

The **3K** optical images were acquired by the air-borne 3K camera system on 17.06.2012. The system consists of three non-metric cameras, of which one is nadir looking, and two are oblique sideways looking. The 3K system was developed at DLR (Kurz et al., 2012). The DSM computed from 3K images is referred to as 3K DSM. The original sampling of the 3K DSM is 0.3 m.

**WV-2** is a MS EO satellite owned by DigitalGlobe. It provides MS imagery in eight spectral channels and one PAN channel with 1.85 m and 0.46 m GSD, respectively. The DSM computed from WV-2 images is referred to as WV-2 DSM. The original sampling of the WV-2 DSM is 0.5 m.

The **LiDAR** last-pulse point cloud with an average density of 1.69 point/m<sup>2</sup> is resampled to a 1 m grid by bilinear interpolation.

#### Preprocessing of DSM

For details on preprocessing steps and typical values of parameters see Subsection 5.1.1. For the DSM normalisation procedure a square window of the size  $W$  pixel  $\times$   $W$  pixel is chosen, where  $W$  is an odd integer number. The outlier level is chosen as 0.15 for 3K and WV-2 DSM, and 0.10 for LiDAR DSM. The higher level for outliers was chosen for both SGM DSM, because they include more errors due to the false or missing matches between the stereo images.

The building regions in the nDSM are non-vegetation above-ground objects, higher than 2.5 m. Vegetation areas are removed from building regions by accounting only for the areas, where the corresponding material map of vegetation is lower than 0.3.

### 6.1.3 Implementation and Setting of Parameters for Proposed Methods

In this Subsection the implementation and application of the proposed a) edge probability detection in scale space and fusion of edge probabilities and b) BP model creation and selection from the images are summarised. The required parameters are listed below.

a) Choice of parameters for edge probability detection in scale space (Subsection 4.1.1) and fusion of edge probabilities (Section 4.2):

- ◇ The set of scales  $t$  is chosen as a geometric sequence (e.g.  $t \in \{0.5, 1.0, 2.0\}$ ) with a common ratio 0.5, as used by Lindeberg (1994) and Marimont and Rubner (1998).
- ◇ The discrete Gaussian kernel has a size of  $6\sqrt{t}$  rounded to the nearest odd integer number  $M$  (Equation (28)) and is normalised to integral one (Subsection 4.1.3).
- ◇ Level of significance of statistical test is  $\alpha = 0.01$  for all edge probability computations.
- ◇ The noise in an image is assumed to be zero-mean additive Gaussian white noise and is denoted  $\sigma_n$ . The value of  $\sigma_n$  is a scaling factor of the covariance matrix  $\Sigma(\mathbf{dn})$  of the vector of the derivatives of the noise  $\mathbf{dn}$  (Subsection 4.1.1). For DSM, it is given as the standard deviation of the flat region in DSM, e.g. 0.88 m for 3K DSM and 0.97 m for WV-2 DSM. For each material map,  $\sigma_n$  is given as a standard deviation of the the region consisting of the same material. For considered material maps  $\sigma_n$  varies between 0.04 and 0.07.

- ◇ The scale  $t_{max} = \arg \max_t (p(\text{edge}; t))$  is selected for every pixel in an image (Subsection 4.3.2).
- ◇ The edge probability fusion is carried out using the GMM model (Subsection 4.2.3). The mixing coefficients  $\pi_k$  are defined as the ratio between confidence probabilities at the scales  $t_{max}$  (Subsection 4.3.2, Equation (35)).
- b) Choice of parameters for BP model creation and selection (iMBR method, Subsection 5.1.2):
  - ◇ The width of a *MBR* at any level  $k$  shall not be smaller than 2 pixel.
  - ◇ The orientation of all *MBR* is the same.
  - ◇ The boundary points are pixels with the highest edge probability values within the buffer zone of the approximate BP. More precisely, line segments perpendicular to the approximate BP are created, which have a length of a buffer width. The sampling distance of these line segments is one pixel. For every line segment, the pixel through which the line segment passes and has the maximum edge probability is chosen as the boundary point.
  - ◇ The boundary points are assigned to the nearest edge of the approximate BP. No re-assignment of boundary points during the LS adjustment is carried out.

## 6.2 Experiments on Scale Space for Edge Detection

In Section 4.1, the influence of the scale to the detection of an edge is demonstrated on an example of the building profile from SGM DSM (Figure 4.1). This profile exhibits edges at different scales, which motivates edge detection in scale space.

The objective of this experiment is to compare edge probabilities detected at single scales to the edge probabilities detected in scale space. These edge probabilities are evaluated by comparing them to reference polygons of the objects under consideration. The scale selection for each pixel in an image is based on the maximal edge probability (Subsection 4.1.1) and is compared to the scale selection on the basis of so called “optimal scale” for each pixel (Marimont and Rubner, 1998).

### 6.2.1 Test Dataset

In this experiment, the edge probabilities are detected from HSI. A DSM is influenced by the matching inaccuracies and errors, which may occur in the SGM method. This is the main reason to choose HSI for a test dataset instead of the SGM DSM.

The test dataset is VNIR20 HySpex image, which exhibits several tennis courts located north of the Olympiapark, Munich, Germany and is hereafter referred to as the *tennis court* dataset (Figure 6.2). The tennis courts are selected for the object of consideration, because they consist of one distinguishable material and have well-defined rectangular-shaped boundaries. Unlike building rooftops, they lie on the terrain’s surface. Thus, inaccurate geometric corrections due to height displacements are negligible.

Each of the tennis court areas consists of several tennis courts, which are divided by either paths, markings, or nets. Between some tennis courts no division is visible. All the tennis courts consist of clay material, having white marks, nets and some smaller objects on them. Thus, only the outer boundaries of the tennis court areas are considered for experiments. Moreover, all the tennis courts are not consisting of the same type of clay material or the material aged longer. However, their reference spectra are similar and one reference spectra for all was sufficient for unmixing of the HSI.

Reference tennis court polygons (yellow) are collected manually from the 10-times enlarged *tennis court* dataset using bilinear interpolation. The boundaries are digitised from the RGB composite of



Figure 6.2: The *tennis court* test dataset with superimposed reference polygons. The RGB composite of the HySpex HSI is shown. The outer boundaries of the six clay tennis courts were manually digitised and are superimposed over the image (yellow). The most northern tennis court is excluded from the evaluation, because the clay material is spread around the tennis court. Consequently, the material boundary could not be reliably digitised.

the enlarged dataset and are all generalised to rectangular shapes. The most northern tennis court is excluded from the evaluation, because the material boundary could not be reliably digitised. The reason for this is that the clay material is spread over a larger area around the tennis court. Six outer boundaries of the tennis court areas with a total length of about 1.25 km are used as a reference.

## 6.2.2 Parameter Settings

The material map of the clay tennis courts is computed by NNLS unmixing. Then, the edge probabilities  $p(\text{edge}; t)$  are computed for scales  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$ . For every pixel in an image, the scale  $t_{max} = \arg \max_t (p(\text{edge}; t))$  is selected, where the probability of an edge  $p(\text{edge}; t)$  is the maximal edge probability. In cases where more than one scale has the same and also maximal edge probability, the minimum scale among them is defined as  $t_{max}$ . Furthermore, the confidence probability  $\beta(t_{max})$  is computed for every pixel and the level of significance of the statistical test is set to  $\alpha = 0.01$ .

## 6.2.3 Results and Discussion

Figure 6.3 shows the selected scale  $t_{max}$  for the *tennis court* dataset, from the finest scale ( $t = 0.5$ , dark red), over middle rang scales ( $t = 4.0$ , yellow) to the coarsest scale ( $t = 16.0$ , white). If the statistical test  $H_0$  is accepted, the values of the second and third derivatives are considered unreliable, and the edge probabilities are set to zero. Thus, no scale can be selected for these pixels and they are labelled as “no value” (grey). Moreover, pixels with very low confidence probability ( $\beta(t_{max}) \leq 0.0001$ ), are also labelled “no value”.

### Visual Interpretation

Figure 6.4 shows the edge probabilities  $p(\text{edge}; t_{max})$ , and Figure 6.6 the confidence probability  $\beta(t_{max})$  for the selected scales of the *tennis court* dataset and clay material. Border effects can be observed on the boundaries of both figures, which are the result of a convolution. The edge probabilities

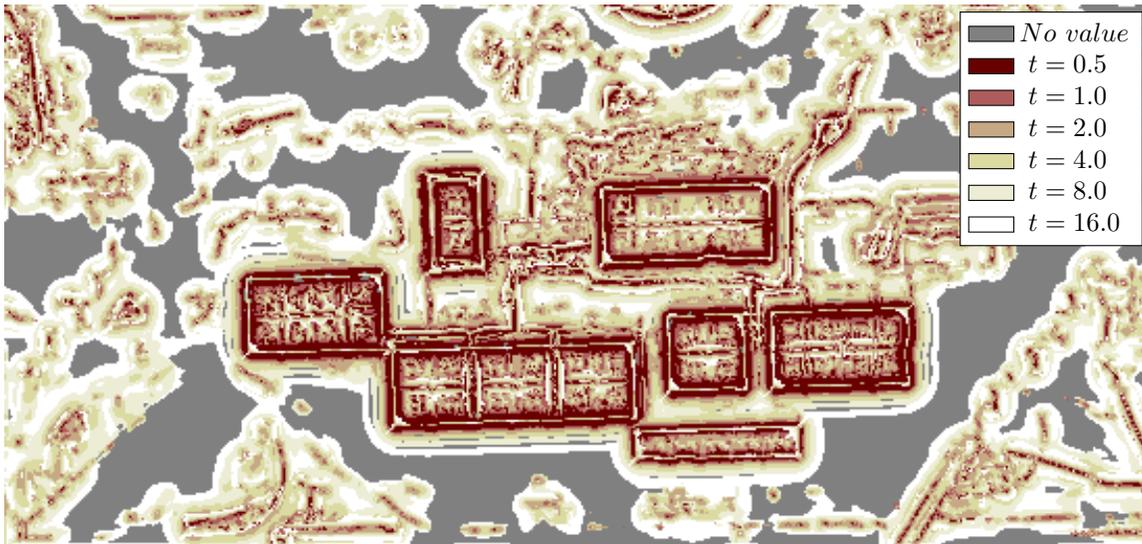


Figure 6.3: Selected scale  $t_{max} \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$  on the basis of maximal edge probability for *tennis court* dataset. The scales span from the finest scale ( $t = 0.5$ , dark red), over middle rang scales ( $t = 4.0$ , yellow) to the coarsest scale ( $t = 16.0$ , white). The edge probabilities are computed from the derivatives. If the derivatives are considered unreliable (Equation (18)), the edge probability is set to zero. Thus, these pixels have no corresponding selected scale and are labelled as “no value” (grey).

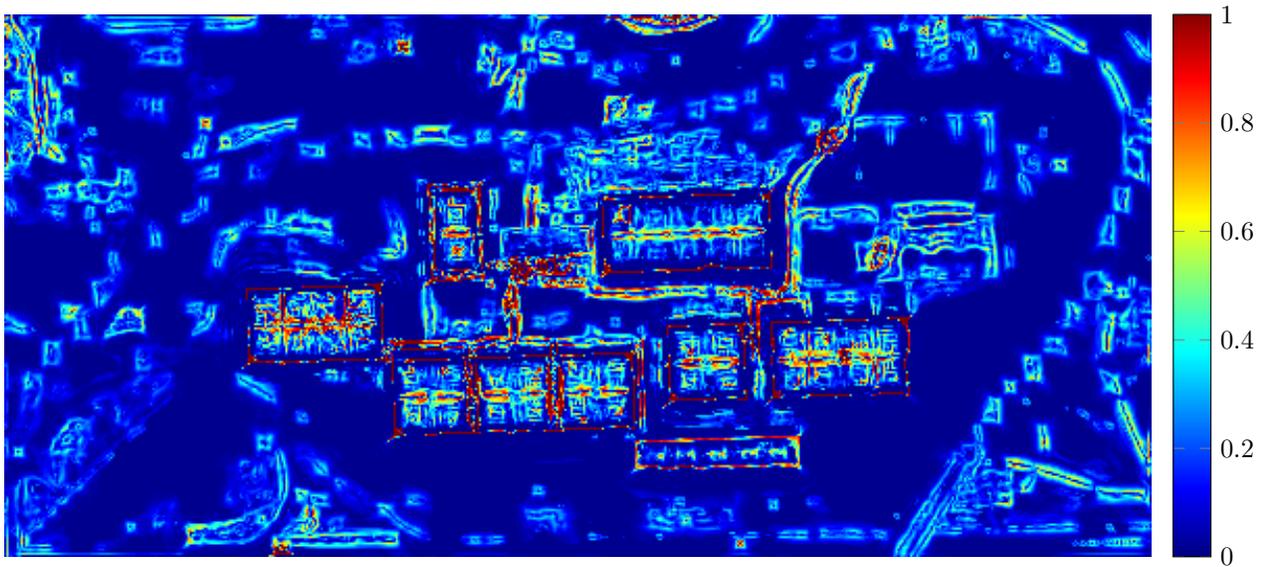


Figure 6.4: Edge probability of clay material for selected scale  $t_{max}$  (Figure 6.3) of the *tennis court* dataset. The edge probabilities are ranging from zero (blue) to one (red). The values are the highest around the boundaries of the clay material objects. These objects are mainly the tennis courts. Higher edge probabilities can be observed also on the boundaries of other objects, i.e. street and vegetation. Due to the convolution, boundary effects are present. See also Figure 6.5.

$p(\text{edge}; t_{max}) > 0.1$  are superimposed on a red channel ( $\lambda = 0.64 \mu\text{m}$ ) of the corresponding HySpex test dataset for easier visual comparison of locations of edge probabilities (Figure 6.5).

Both, edge probability and confidence probability have a range spanning from zero (blue) to one (red). The edge probabilities, as well as the confidence probabilities, have the highest values on the boundary of the clay material objects. Higher edge and confidence probabilities are present on the boundaries of some other objects, e.g. street and vegetation. For homogeneous areas, i.e. areas consisting of the same material, the edge probabilities are low or are set to zero, because they are unreliable. The nets

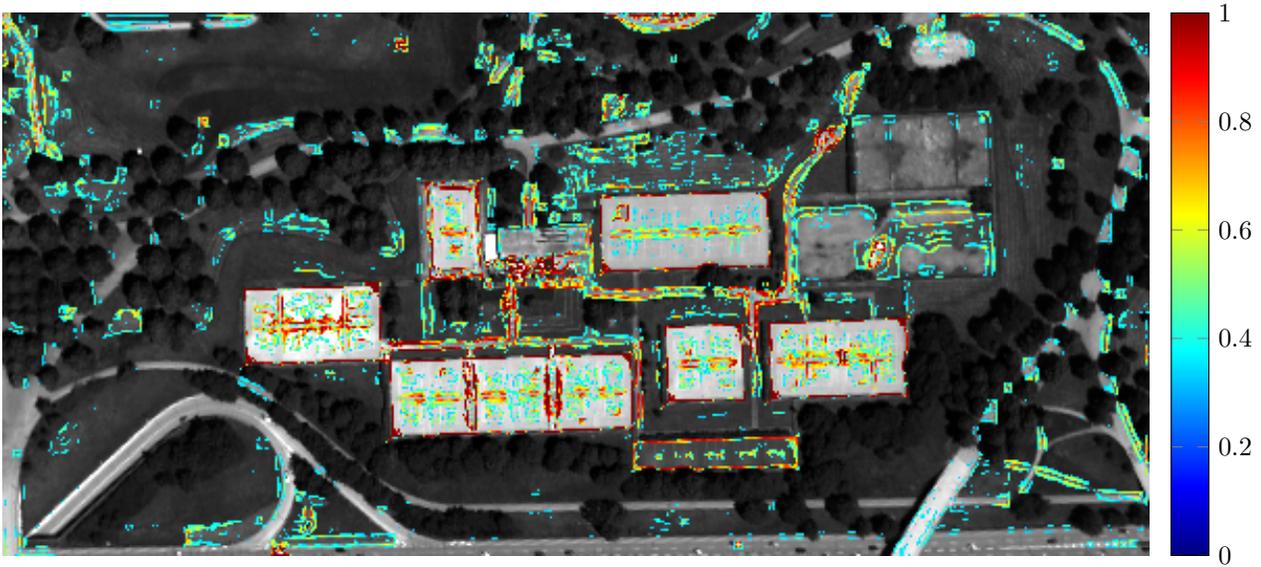


Figure 6.5: Edge probability of clay material for selected scale  $t_{max}$  (Figure 6.3) superimposed on the HySpex red channel ( $\lambda = 0.64 \mu\text{m}$ ) of the *tennis court* dataset. For the clarity, only the edge probabilities larger than 0.1 are shown, i.e.  $p(\text{edge}; t_{max}) > 0.1$ . Figure 6.4 shows all the edge probabilities.

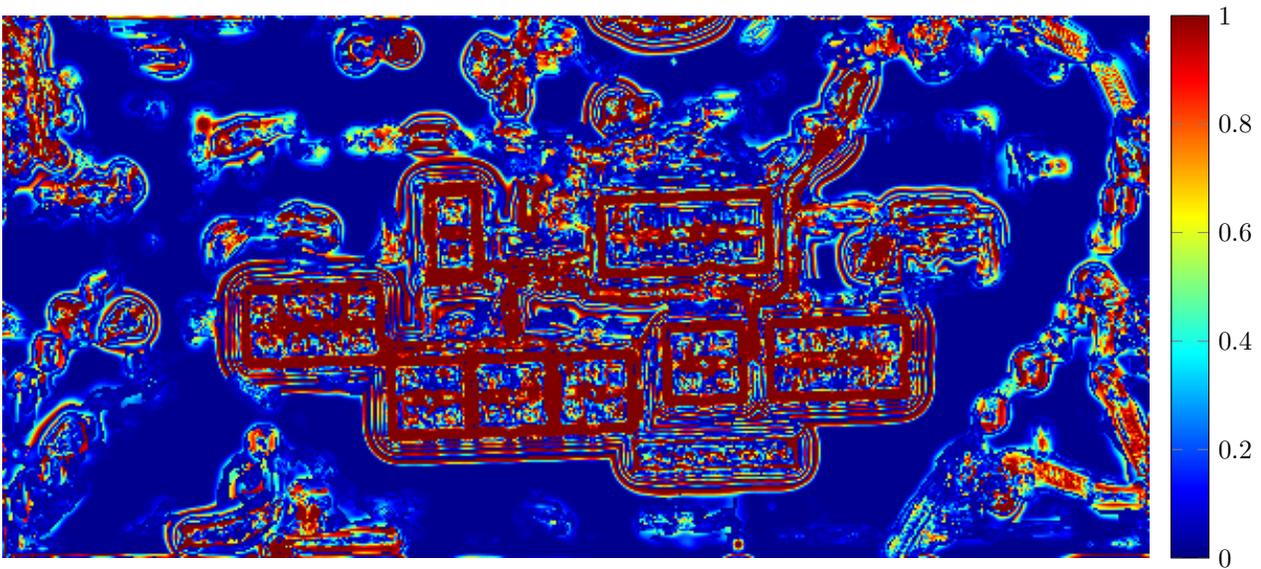


Figure 6.6: Confidence probability  $\beta(t_{max})$  of the edge probabilities for clay material (Figure 6.4). The edge probabilities and corresponding confidence probability are computed for the selected scale  $t_{max}$  (Figure 6.3). The confidence probability ranges from zero (blue) to one (red). The confidence probability has the highest values around tennis courts, therefore these edge probabilities are considered reliable.

on tennis courts are visible, but difficult to distinguish (Figure 6.2, approximately east-west direction). Nevertheless, they are detected, similar like white markings, as high edge probabilities (Figure 6.4). The most prominent example are high values of edge probabilities on the boundaries of the two paths, which divide the largest tennis court area into three parts. Both paths consist of non-clay material, but some of the clay material is spread over them.

### Quantitative Evaluation

The edge probabilities of the clay material for selected scale  $p(\text{edge}; t_{max})$  are compared to the edge probabilities at six single scales  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$ . The boundaries of tennis courts,

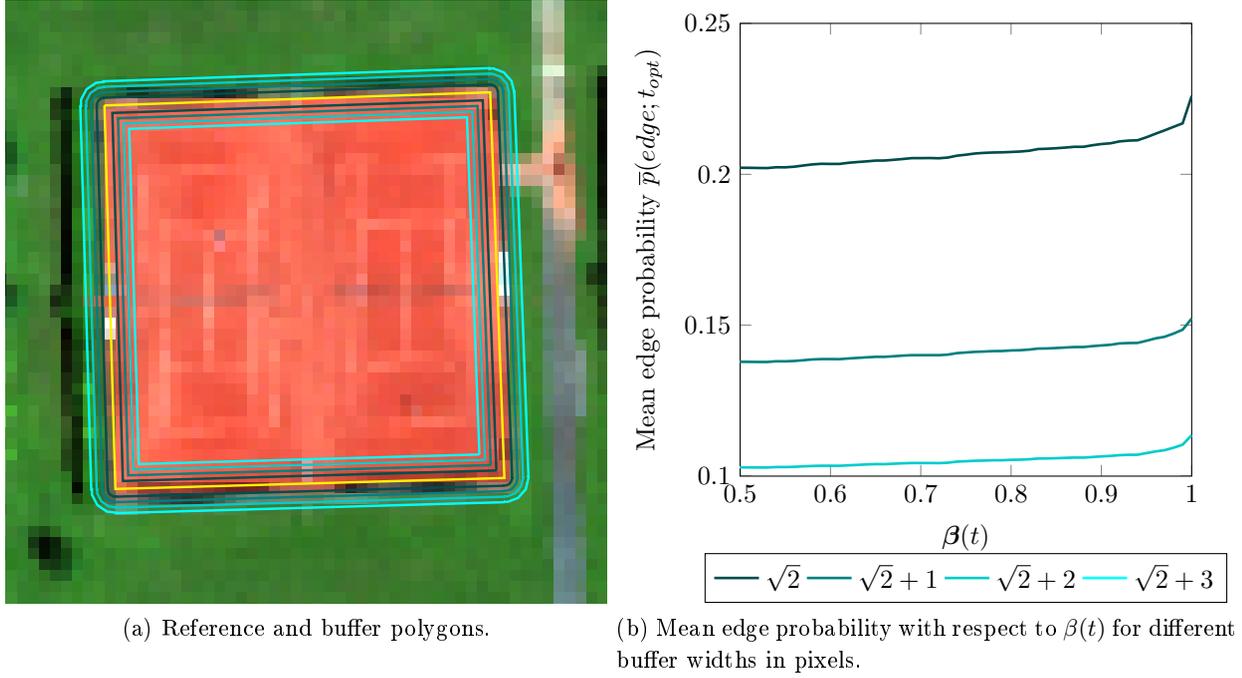


Figure 6.7: Detail of the *tennis court* dataset with superimposed buffer zones (Figure 6.7a) and mean edge probability with respect to  $\beta(t)$  for different buffer widths (Figure 6.7b). Buffer polygons are defined around the reference polygon (yellow). The width of the buffer polygons ranges from  $\sqrt{2}$  pixel (dark cyan) to  $\sqrt{2} + 3$  pixel (light cyan).

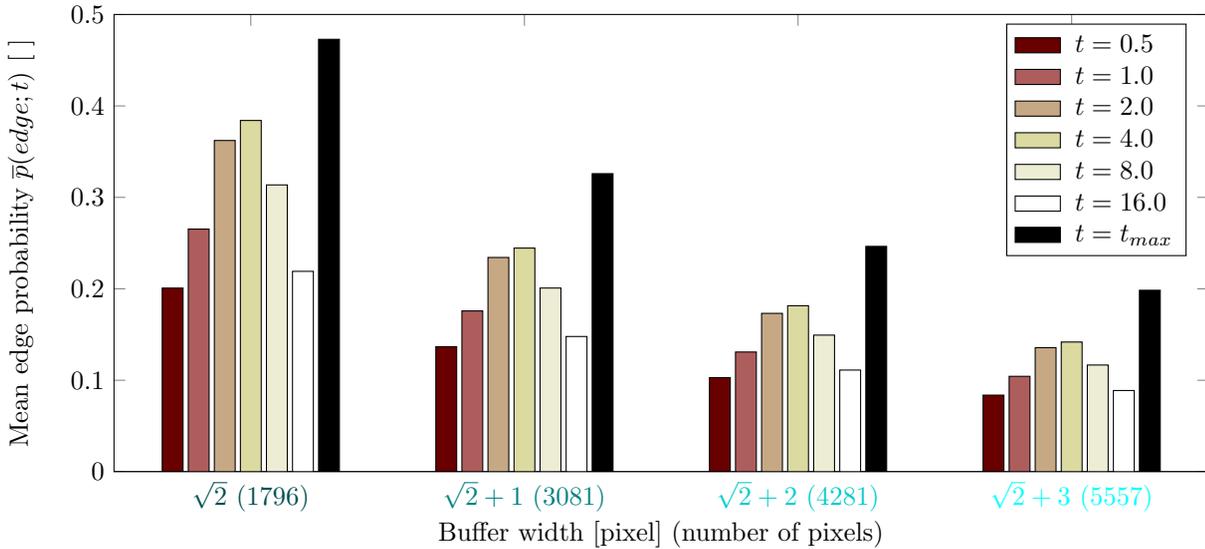


Figure 6.8: Mean edge probability for single scales  $\bar{p}(edge; t)$ ,  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$  and selected scale  $\bar{p}(edge; t_{max})$  for four different buffer widths around the reference boundary. The colour of the bars of single scales corresponds to the colours in Figure 6.3. The bars corresponding to the  $\bar{p}(edge; t_{max})$  are black. The X-labels denote the buffer width and number of pixels for every buffer width (Figure 6.7a). The total length of the reference boundaries is about 2.4 km. Computing edge probabilities at several scales gives better results than detecting at a single scale, i.e.  $\bar{p}(edge; t_{max}) > \bar{p}(edge; t)$ ,  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$ .

which are composed of clay material, are vector data, whereas the edge probabilities are raster data. Therefore, the buffer zones (Figure 6.7a, cyan) around reference polygons (Figures 6.2 and 6.7a yellow) are constructed. If a centre of a pixel lies in the buffer zone, it is considered as the edge pixel of the tennis court.

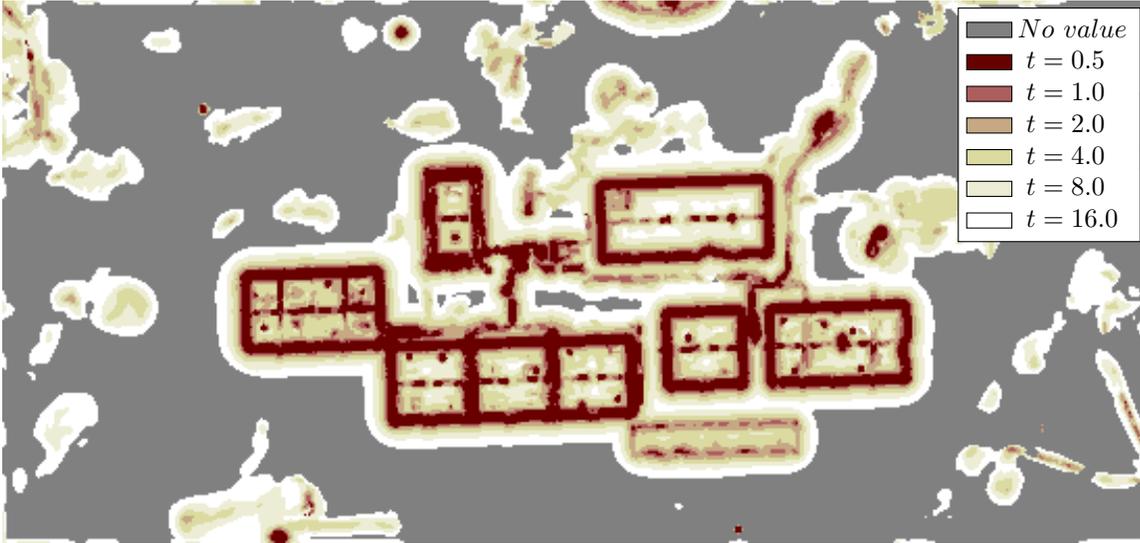


Figure 6.9: “Optimal scale”  $t_{opt}$  selected on the basis of confidence probability  $\beta(t) \geq 0.95$ ,  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$  for *tennis court* dataset. The scales span from the finest scale ( $t = 0.5$ , dark red), over middle rang scales ( $t = 4.0$ , yellow) to the coarsest scale ( $t = 16.0$ , white) same as in the Figure 6.3, where the scale is selected as  $t_{max}$ . All the pixels whose  $\beta(t) \leq 0.95$  for all  $t$  are marked as “no value”.

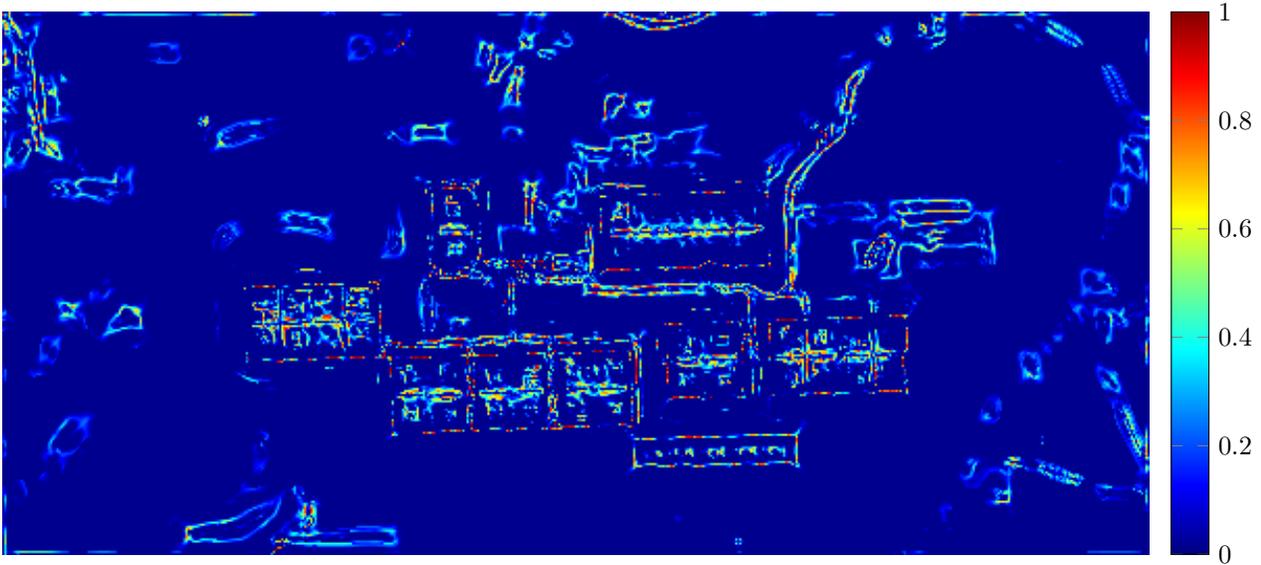


Figure 6.10: Edge probability of clay material for selected scale  $t_{opt}$ ,  $\beta(t) = 0.95$ ,  $\forall t$  (Figure 6.9) of the *tennis court* dataset. The edge probabilities are ranging from zero (blue) to one (red). The  $\bar{p}(edge; t_{opt})$  can be compared to the edge probabilities at differently selected scale  $t_{max}$ , which are shown on Figure 6.4.

The mean value of the edge probabilities  $\bar{p}(edge)$  in the buffer zone is computed for  $t_{max}$  and all  $t$ . Four different widths of buffer zones with the widths  $\{\sqrt{2}, \sqrt{2} + 1, \sqrt{2} + 2, \sqrt{2} + 3\}$  [pixel] are considered (Figure 6.7a, the buffer polygons with smaller width are in dark cyan colour, and with the larger width in light cyan colour). These buffer widths correspond to the  $\{1, 2, 3, 4\}$  [pixel] buffer widths around a line aligned with one of the axes of a coordinate system. For example, the same line segment aligned with one of the axis of the coordinate system or rotated for  $\frac{\pi}{2}$  has a different representation in a raster image. To account for the same neighbourhood in a square raster independently of the orientation of the object, the buffer widths include the length of a diagonal of the square raster.

The mean edge probability of tennis court material with respect to all considered buffer widths and scales is computed (Figure 6.8). The colours of bars, which represent mean edge probabilities for a

single scale, correspond to the colours of scales in Figure 6.3. The bars representing the mean edge probabilities for selected scale  $t_{max}$  are black. X-label text is coloured according to the buffer width (dark to light cyan), and denotes the buffer width and number of pixels in the buffer. The mean edge probability for selected scale is larger than for any single scale, i.e.  $\bar{p}(edge; t_{max}) > \bar{p}(edge; t)$ ,  $t \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$ , for all buffer widths. The value of the mean edge probability for  $t_{max}$  decreases with increasing buffer width. A larger buffer includes more pixels with low edge probability and consequently, the mean edge probability values are lower. However, the ratio between the mean edge probabilities for different scales and the  $t_{max}$  is approximately the same for different buffer widths.

I proposed to select the scale on the basis of the maximum values of the edge probabilities ( $t_{max}$ ). However, Marimont and Rubner (1998) select so called “optimal scale” for each pixel. The optimal scale  $t_{opt} \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$  is defined as a minimum scale, at which the confidence probability  $\beta(t_{opt})$  is larger than a predefined threshold. Figure 6.9 shows the optimal scale  $t_{opt}$  for the *tennis court* dataset, for  $\beta(t_{opt}) \geq 0.95$ . All the pixels, whose  $\beta(t) < 0.95$ ,  $\forall t$  are marked as “no value”. Therefore, Figure 6.9 exhibits larger number of “no value” pixels in comparison to the selected scale  $t_{max}$  (Figure 6.3). Inspecting the boundary regions of the tennis courts (clay materials), it can be seen that these edges are detected at the finest scale  $t = 0.5$ , for the confidence probability larger or equal than 0.95.

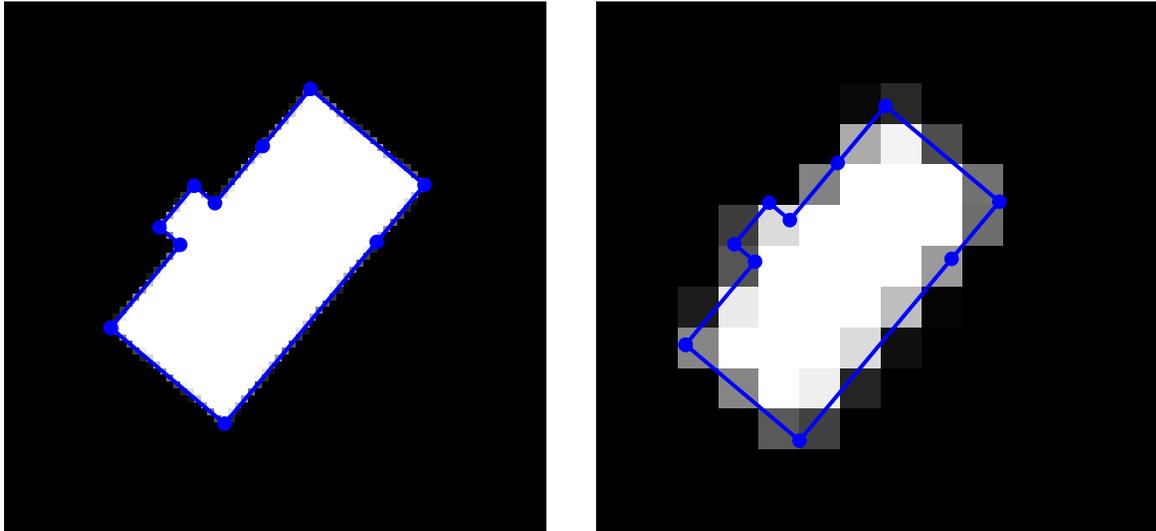
The mean edge probability is computed for the edge probabilities at the optimal scale  $t_{opt}$ , denoted  $\bar{p}(edge, t_{opt})$ , for the four buffer widths. Figure 6.7b shows the  $\bar{p}(edge, t_{opt})$  with respect to the threshold on confidence probability  $\beta(t)$ . This is analogue to the black bar plots in Figure 6.8, which represent the  $\bar{p}(edge, t_{max})$ . For all buffer widths  $\bar{p}(edge, t_{max}) > \bar{p}(edge, t_{opt})$ ,  $t_{max}, t_{opt} \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$ . With regard to the buffer width, the value of the mean edge probability for  $t_{opt}$  shows the same trend as for the  $t_{max}$ , i.e. it decreases with increasing buffer width.

Figure 6.10 shows the edge probability for the *tennis court* dataset at the scale  $t_{opt}$ ,  $\beta(t) \leq 0.95$  (Figure 6.9). The edge probabilities  $p(edge; t_{opt})$  on the boundaries of the clay tennis courts have the highest values. However, in comparison to the  $p(edge; t_{max})$ , they have on an average lower values and larger number very low values. This observation corresponds to the graph of the mean edge probability for  $t_{opt}$  (Figure 6.7b), where the highest values of the  $\bar{p}(edge; t_{opt})$  are at  $\beta(t) = 1.00$  are still lower than the the values of  $\bar{p}(edge; t_{max})$  (Figure 6.8, black bars).

## 6.2.4 Summary of Experiments on Scale Space for Edge Detection

The following conclusions can be summarised from the experiments on scale space for edge detection:

- ◇ Edge probability computation in HSI is possible on the basis of material differences. The edge probabilities are computed for each material from the corresponding material map. The same method is applicable on one channel images or DEM. A prerequisite for edge probability computation in HSI are known spectral signatures of materials under consideration.
- ◇ The edge probabilities are computed for the whole test image, and are high also on the boundaries of some objects, which are not from material under consideration. A spatial limit could be set with regard to the values of unmixing, i.e. values in material maps. Limiting the areas for computation positively influences the computational effort.
- ◇ Edge probability computation benefits from analysing RS images at different scales.
- ◇ A scale selection on the basis of the maximal edge probability over all scales showed better results than scale selection on the basis of thresholding confidence probability at a high value. Moreover, the former requires no threshold for the minimal confidence probability.



(a) Rasterised reference polygon with 0.5 m grid size. (b) Rasterised reference polygon with 3.0 m grid size.

Figure 6.11: Rasterisation of the reference building polygon (blue) by  $GSD \in \{0.2, 0.3, \dots, 5.0\}$  [m]. Examples for the GSD 0.5 m (Figure 6.11a) and 3.0 m (Figure 6.11b) are shown. Pixels get assigned values, which correspond to the fraction of the pixel lying inside of the reference polygon. The small structure of the building, which is represented by the reference polygon, is not distinguishable in the image with lower GSD.

## 6.3 Experiment on Building Polygon Selection and Adjustment

The experiments on BP selection and adjustment are carried out on synthetic building polygons (Figure 5.10a). The first objective of these experiments is to investigate the iMBR method (Subsection 5.1.2) under varying GSD. The BP creation and selection of the sampled reference BP is carried out. The reference BP is sampled at varying sampling distances. The second objective aims at comparison of the GM and the GH functional models, where the approximate BP serves as a basis for initial approximate values for the unknowns in the GM and the GH models. Accuracy assessment of the approximate and adjusted BP is carried out by comparing them to the reference polygon. The accuracy is quantified using the PoLiS metric and the quality rate ( $1 - qual$ ). The quality rate is, like PoLiS metric, a symmetric measure and the trend of both of the measures is similar (Section 5.3). The quality rate is computed, because it is an established quality measure in contrast to the newly proposed PoLiS metric. Furthermore, the main orientation of the reference and adjusted polygons is compared and discussed.

### 6.3.1 Test Dataset

To assure that no data quality and processing influences are present in the test dataset, the experiments on the model selection and adjustment are carried out on a synthetic dataset. The same building polygon as the reference polygon for the PoLiS metric experiments (Section 5.3, Figure 5.10a, blue) is used in this experiment. This reference rectilinear polygon represents a building of a size  $23\text{ m} \times 11\text{ m}$  with a small structure (e.g. jutting roof over the entrance area) of a size  $4\text{ m} \times 2\text{ m}$ . It has two additional vertices, located exactly on the polygon's edges, which represent the division of the building in two units. This division is not detectable and plays no role in BP creation, selection, and adjustment. The main orientation of the reference BP with respect to the horizontal axis of the coordinate system is  $50^\circ$ , and its area is  $261\text{ m}^2$ . This BP is used as ground truth (reference) for the accuracy assessment

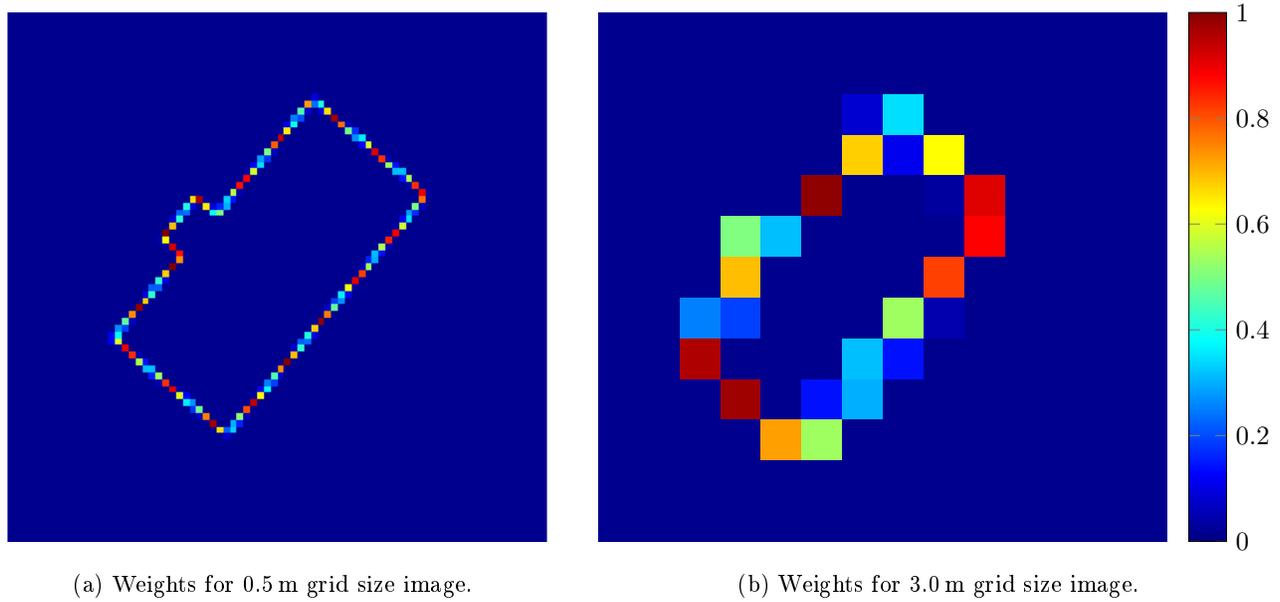


Figure 6.12: Weights for the building polygon rasterised at different GSD. Examples for the GSD 0.5 m (Figure 6.12a) and 3.0 m (Figure 6.12b) are shown. A weight is derived for every pixel in a raster image and is scaled between 0 and 1. These weights are used in LS adjustment. Pixels with the 0.50 fraction inside the reference polygon get assigned the highest weights (1). The values of weights linearly drop with respect to the increasing and decreasing fraction values.

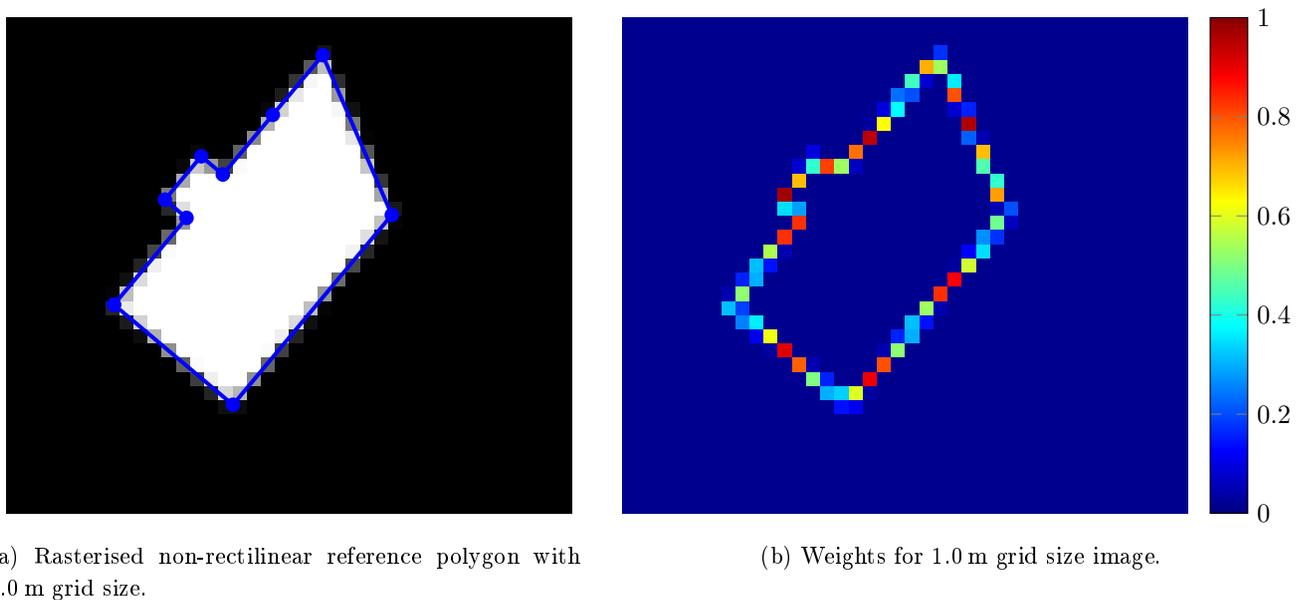


Figure 6.13: Rasterised non-rectilinear reference polygon (Figure 6.13a) and its weights (Figure 6.13b). An example for the GSD 1.0 m are shown. A weight is derived for every pixel in a raster image and is scaled between 0.0 and 1.0. These weights are used in LS adjustment. The non-rectilinear BP are used for comparison of the GM and GH adjustment models.

and to establish a set of synthetic raster images. From these raster images the approximate polygons are estimated. The weights of the boundary points are defined on the basis of the raster images.

A set of *raster images* is created by rasterisation of the reference polygon at different sampling distances, i.e. at different GSD (Figure 6.11). The GSD of 1.0 m means that the distance between two pixel centres projected on the earth's surface is 1.0 m. A total of 49 raster images are created, with the

GM model	$\sigma_d^2$ [pixel]	$\sigma_\alpha^2$ [°]	$\sigma_{XY}^2$ [pixel]
Setting 1	1	1	3
Setting 2	1	10	3

Table 3: Two settings of variances of the three groups of observations in GM model are defined. These variances define the weights between the groups of observations. They are required for the LS adjustment with GM model of rectilinear (Figure 6.11) and non-rectilinear (Figure 6.13a) BP. The relation between three groups of observations is defined by these values. The angles  $\alpha_j$  are pseudo-observations in a functional model of GM model. Their role is to ensure rectilinearity of adjusted BP. The requirement of rectilinearity is relaxed by setting the value  $\sigma_\alpha^2$  higher (Setting 2), i.e. the weights are lower. The variances of the other two groups of observations, i.e. squared distance between a polygon edge and a boundary point  $d^2(i, j)$ , and coordinates of the polygon's vertices  $P_j = [X_j, Y_j]$ , are the same in both settings. Thus, the Setting 1 and 2 are referred to as GM model with  $\sigma_\alpha^2 = 1^\circ$  and GM model with  $\sigma_\alpha^2 = 10^\circ$ , respectively.

$GSD \in \{0.2, 0.3, \dots, 5.0\}$  [m]. Pixels get assigned values which correspond to the fraction of the pixel lying inside of the reference polygon. For instance, pixels that lie completely outside or inside the reference polygon, get assigned values 1 and 0, respectively.

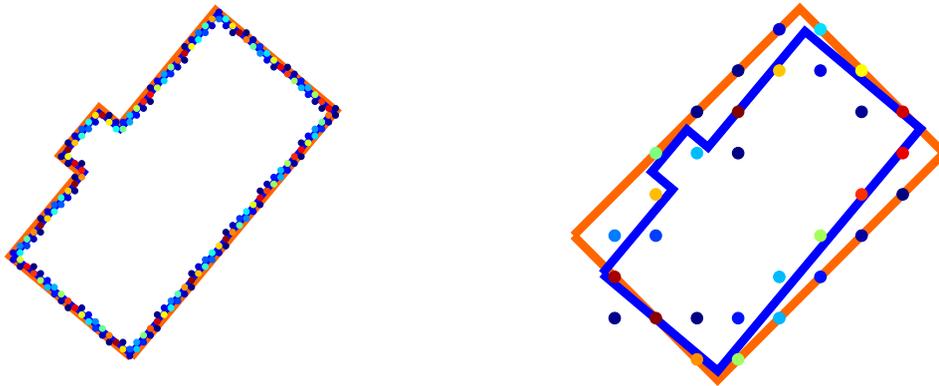
For every point in the raster image a *weight* is derived. The weights are scaled from 0 to 1, the same like edge probabilities. Boundary pixels with the 0.50 fraction inside the reference polygon get assigned the highest weights (1.0). The values of weights linearly drop with respect to the increasing and decreasing fraction values. For example, the fraction of 1 and 0 both correspond to the weight of 0, and the fraction of 0.80 and 0.30 both correspond to the weight of 0.6.

For the comparison of the GM and the GH adjustment models the set of the approximate BP is altered in a manner that they are not rectilinear (Figure 6.13). From each of the approximate BP, one of the vertices is removed. Then, the missing vertex is added to the approximate and altered BP exactly at the location of one of the vertices, which are located on the edge of the reference polygon (Figure 6.13a). The approximate and altered BP is hereafter shortly referred to as non-rectilinear BP. The weights of the boundary points are computed from the set of raster images (Figure 6.13b).

### 6.3.2 Parameter Settings

A building mask is derived from every raster image by thresholding the raster image at level 0.1. A minimum area of a building part to be estimated is set to 9 pixel. The main orientation of the BP is defined as the orientation of the  $MRB_1$ . All the  $MRB_k$ ,  $k > 1$ , are aligned to the  $MRB_1$ . This means, that all the resulting BP are strictly rectilinear.

The GH model requires no settings for the groups of observations. In contrast, the GM model for estimation of BP requires weighting of groups of observations, which are given in Table 3. GM model for BP consists of three groups of observations. The first group of observations are the angles  $\alpha_j$  between subsequent polygon edges. The  $\alpha_j$  are pseudo-observations in the GM functional model and their role is to ensure rectilinearity of adjusted BP. The requirement of rectilinearity is relaxed by setting the value  $\sigma_\alpha^2$  higher (Setting 2), i.e. the weight of this group of observation is lower. The second and the third group of observations are squared distance between a polygon edge and boundary point  $d^2(i, j)$ , and coordinates of the polygon's vertices  $P_j = [X_j, Y_j]$ , respectively. The variances of these two groups of observations are the same in both settings, i.e.  $\sigma_d^2 = 1$  pixel and  $\sigma_{XY}^2 = 3$  pixel and are chosen as an upper values as proposed by Kanani (2000) ( $\sigma_d^2 < 1$  pixel and  $\sigma_{XY}^2 = 2 - 3$  pixel). Setting 1 and 2 are referred to as GM model with  $\sigma_\alpha^2 = 1^\circ$  and GM model with  $\sigma_\alpha^2 = 10^\circ$ , respectively.



(a) Approximate and reference BP for 0.5 m grid size. (b) Approximate and reference BP for 3.0 m grid size.

Figure 6.14: Comparison of approximate BP (orange) created by iMBR method and reference BP (blue) for 0.5 m (Figure 6.14a) and 3.0 m (Figure 6.14b) GSD. Points represent the boundary pixels of the BP, and their colour represent the weight of each pixels (see Figure 6.12). The approximate BP is very close to the reference polygon for 0.5 m GSD, and they both have the same shape (Figure 6.14a). For 3.0 m GSD, the difference in orientation between the reference and the approximate BP can be observed (Figure 6.14b). Moreover, the approximate BP is a rectangle, because small structure of the building is not distinguishable from an image with 3.0 m GSD (see Figure 6.11b).

### 6.3.3 Results and Discussion of the Model Selection Experiment

For each of 49 raster images, which represent the same reference building at different GSD, a set of BP is created by iMBR method and the best one, according to the cost function (Equation (39)), among them is selected. For increasing GSD, the same building is represented by fewer pixels. Consequently, the selected approximate BP is simpler (Figure 6.14, orange). The approximate BP is required in the LS adjustment to compute the approximate values of the unknowns. The reference polygon under consideration is the level 2 BP, denoted  $BP_{k_{min}=2}$ .

Figure 6.14 shows an example of an approximate BP by iMBR method for 0.5 m and 3.0 m GSD raster images. The approximate BP for 0.5 m GSD is very close to the reference polygon and they both have the same shape, i.e.  $BP_{k_{min}=2}$  (Figure 6.14a). For 3.0 m GSD, the difference in orientation between the reference and the approximate BP can be observed (Figure 6.14b). Moreover, the approximate BP is a rectangle ( $BP_{k_{min}=1}$ ), because the small structure of the building is not distinguishable for the image with 3.0 m GSD (see Figure 6.11b).

#### Relation between a Level of Selected BP, Minimal Area of a Building Part, and GSD

A parameter in iMBR method is the minimal area of a building part, which defines the smallest part of an object to be considered. The selection of the level of BP is influenced by this parameter. The minimal area of a building part is set to 9 pixel<sup>2</sup>. For this setting, the approximate BP is a rectangle ( $BP_{k_{min}=1}$ ) for all  $GSD \geq 1.1$  m, and  $BP_{k_{min}=2}$  for all  $GSD < 1.1$  m. This can be observed in the jump of the quality measures, i.e. PoLiS (Figure 6.16) and  $1 - qual$  (Figure 6.18), around the  $GSD = 1.0$  m. If the minimal area of a building part to be estimated is set to 16 pixel<sup>2</sup> and 25 pixel<sup>2</sup>, the approximate BP is a rectangle at  $GSD \leq 1.0$  m, and  $GSD \leq 0.9$  m, respectively. For an increasing minimal area parameter, a simpler BP is selected for a smaller value of GSD. Figure 6.15 shows this dependence of

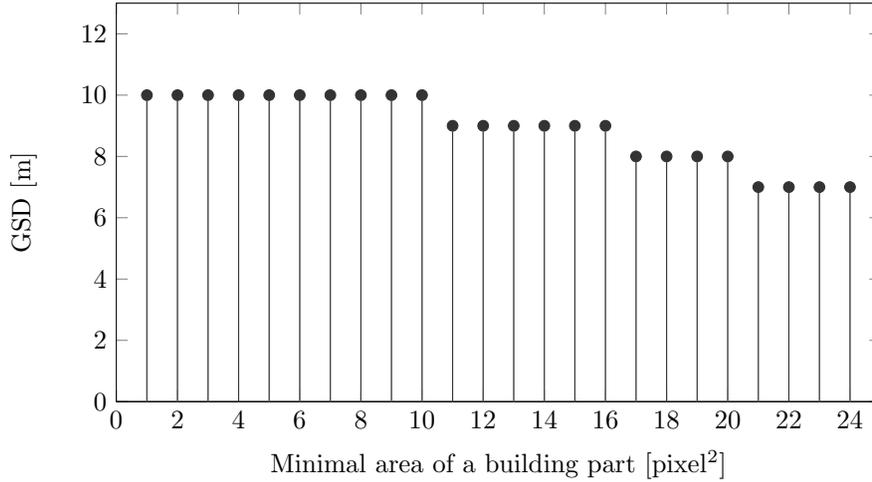


Figure 6.15: GSD [m] for which a Level 2 building polygon ( $BP_{k_{min}=2}$ ) is selected as a function of the minimal area of a building part [pixel<sup>2</sup>]. The reference polygon under consideration (Figure 6.11, blue) is a level 2 BP. For larger GSD, a simpler BP (Level 1,  $BP_{k_{min}=1}$ ) is preferred over the more complex one (Level 2,  $BP_{k_{min}=2}$ ). For increasing value of the minimal area of a building part, a simpler BP is selected for smaller GSD.

the selected BP, to the GSD and a value of the minimal area of a building part. Moreover, the selected level of BP is dependent on the relation between dimensions of an object or object parts and GSD, and also on the complexity of the object's boundary.

### Quality Assessment of Selected BP by iMRB Method

The quality of the selected approximate BP is evaluated by two measures, i.e. PoLiS metric (Figure 6.16 and 6.17, orange) and quality rate (Figure 6.18, orange). Additionally the deviation of the main orientation between the reference polygon and selected BP is computed (Figure 6.19, orange). The same measures are used to assess the quality of the adjusted BP. Non-rectilinear reference polygon is not a result of the iMRB method. Thus, the quality parameters for non-rectilinear BP are discussed in the following Subsection. The PoLiS metric is given in [m] and in [pixel], where the values of the PoLiS metric in [m] are normalised by GSD to get PoLiS metric in [pixel]. For easier visual interpretation, the values of the quality rate  $qual$  are plotted as  $1 - qual$ . The trend of the PoLiS metric in [m] (Figure 6.16) and  $1 - qual$  (Figure 6.18) with respect to the GSD is similar. For GSD between 0.5 m and 3.0 m mean, standard deviation (std), and median values of the quality parameters are computed (Table 4). These values are not given for larger GSD ( $GSD > 3.0$  m), because the shorter of both main building edges is too short (i.e.  $< 4$  pixel).

For  $GSD \leq 1.0$  m (i.e.  $BP_{k_{min}=2}$  is correctly chosen as level 2 BP) the values of the PoLiS metric [m] and  $1 - qual$  increase about linearly with the increasing GSD (Figures 6.16 and 6.18). The accuracy of approximate BP is nearly constant regarding the GSD and is better than 0.5 pixel (Figure 6.17). The main orientation of the approximate BP is close to the orientation of the reference polygon, i.e. the absolute deviation of the orientation of the BP is  $\ll 0.5^\circ$

For  $GSD > 1.0$  m (i.e.  $BP_{k_{min}=1}$  is chosen as a simplified level 1 BP) the values of the PoLiS metric [m] and  $1 - qual$  [ ] are in general much larger in comparison to the values for  $GSD \leq 1.0$  m and larger variations among the quality values for the similar GSD can be observed. For e.g.  $GSD = \{3.3, 3.4, 3.5, 4.6, 4.7\}$  [m] the PoLiS (and  $1 - qual$ ) values are significantly lower than for those of similar GSD. For e.g.  $GSD = \{3.2, 4.3, 4.4\}$  [m] the values are higher than for those of similar GSD. These larger variations can be partially explained by the inaccurate main orientation (Figure 6.19) and the inaccurate area of the polygon region. The polygon region is defined by thresholding the

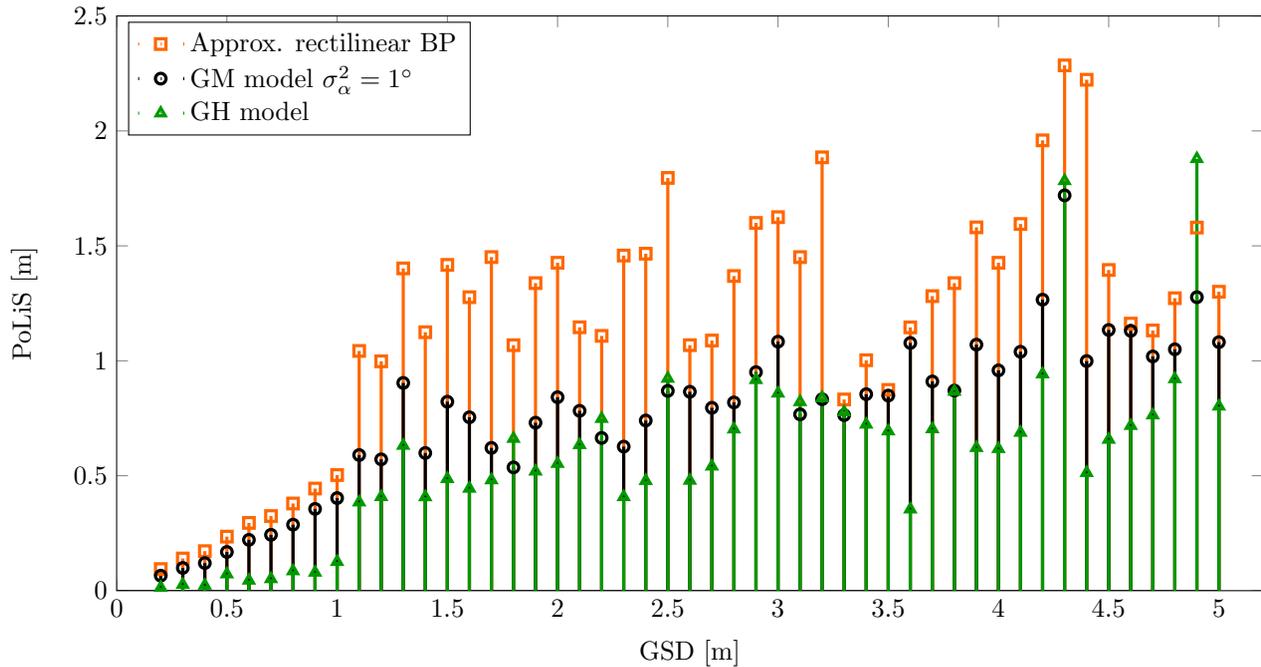


Figure 6.16: PoLiS metric between the reference polygon and three different extracted BP as a function of GSD. (1) the approximate rectilinear BP is compared to the reference BP (orange). Then, this approximate BP is adjusted by the LS method. The adjusted BP by (2) the GM (black) and (3) the GH (green) functional model are compared to the reference polygon. The lower value of PoLiS metric indicates higher similarity to the reference polygon. The units of a PoLiS metric are [m]. (See also Figure 6.17 with normalised PoLiS metric values.)

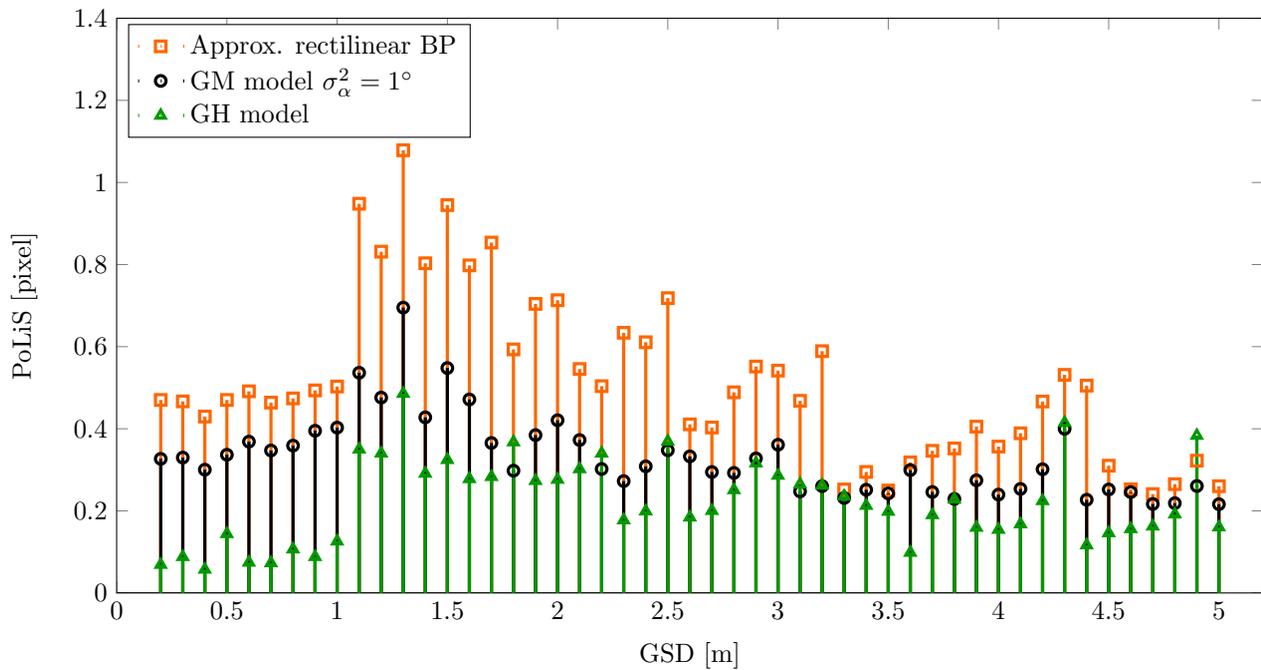


Figure 6.17: PoLiS metric between the reference polygon and three different extracted BP as a function of GSD. These extracted BP are, approximate rectilinear BP, BP adjusted by LS method using GM (black) and the GH (green) functional model. The lower value of PoLiS metric indicates better fitness to the reference polygon. The values of the PoLiS metric (Figure 6.16) are normalised by the GSD and are in [pixel].

		Approximate BP	GM model $\sigma_\alpha^2 = 1^\circ$	GM model $\sigma_\alpha^2 = 10^\circ$	GH model
PoLiS [m]	mean	0.99	0.59	0.47	0.42
	std	0.52	0.29	0.30	0.29
	median	1.11	0.63	0.49	0.48
PoLiS [pixel]	mean	0.62	0.38	0.27	0.23
	std	0.18	0.09	0.14	0.11
	median	0.55	0.36	0.25	0.27
1-quality rate [ ]	mean	0.16	0.10	0.08	0.08
	std	0.08	0.05	0.05	0.05
	median	0.19	0.11	0.09	0.09
Main orientation deviation [°]	mean	2.35	2.14	1.65	1.30
	std	2.43	2.27	1.85	1.64
	median	0.60	0.39	0.25	0.83

Table 4: Quality parameters, i.e. PoLiS [m], PoLiS [pixel],  $1 - qual$  [ ], and main orientation deviation [°], for the test rectilinear building polygon (Figures 6.11b and 6.13a). Values are computed for  $0.5 \text{ m} \leq GSD \leq 3.0 \text{ m}$ . Two sets of parameters are used for the GM model (Table 3), which differ in the weights for rectilinear angles defined through  $\sigma_\alpha^2$ . No settings are required for the GH model. The mean value for all quality parameters and main orientation deviation indicating the worst performance of the models are marked red, and the best ones green. For rectilinear test building, the GH model outperforms the GM models with both of the settings. However, the GM model with higher weights for rectilinearity of the angles  $\sigma_\alpha^2 = 1^\circ$  performs better than the one with the lower weights ( $\sigma_\alpha^2 = 10^\circ$ ).

synthetic images at 0.1. Consequently, the polygon region is increasingly too large for increasing GSD. For instance, the areas of selected BP for  $GSD = \{0.5, 3.0\}$  [m] are  $274.75 \text{ m}^2$  and  $351.00 \text{ m}^2$ , which corresponds to 55 pixel and 39 pixel too large in comparison to the reference polygon ( $261.00 \text{ m}^2$ ), respectively.

For many  $GSD > 1.2 \text{ m}$ , the deviation of the main orientation for the approximate rectilinear BP is exactly  $5^\circ$  (Figure 6.19). This value is a result of rasterisation, i.e. GSD of the image, size and orientation of the building. The main orientation of the selected BP has a significant influence on its quality. Moreover, inaccurate or wrong orientation of the Level 1,  $BP_1$  can result in falsely too high level of selected BP or poorly approximation of the building region.

For  $GSD \leq 3.0 \text{ m}$  (Table 4) the mean value of PoLiS metric is much better than one pixel ( $0.62 \text{ pixel} \pm 0.18 \text{ pixel}$ ) and the quality rate is better than 0.80 ( $1 - qual = 0.16 \pm 0.08$ ,  $qual = 0.84 \pm 0.08$ ). The deviation of the main orientation is  $2.35^\circ \pm 2.43^\circ$ . In most cases, the main orientation is very well defined (i.e.  $\gg 1^\circ$ ), or is  $5^\circ$  (Figure 6.19). Thus, the standard deviation of the main orientation is high.

### 6.3.4 Results and Discussion of the Adjustment

The selected approximate BP are used to define initial values of unknowns in LS adjustment models. The unknown values are adjusted by using the GM and the GH models. The performance of each of the models is evaluated by comparing the quality parameter values (i.e. PoLiS metric and quality rate) before and after adjustment. Then, the quality parameter values after adjustment of both of the models are compared. Figures 6.16–6.19 show the quality parameters and the deviation of main orientation for approximate (orange) and adjusted BP. The values corresponding to BP adjusted by GM and GH models are in black and green colour, respectively. The adjustment is carried out for BP with

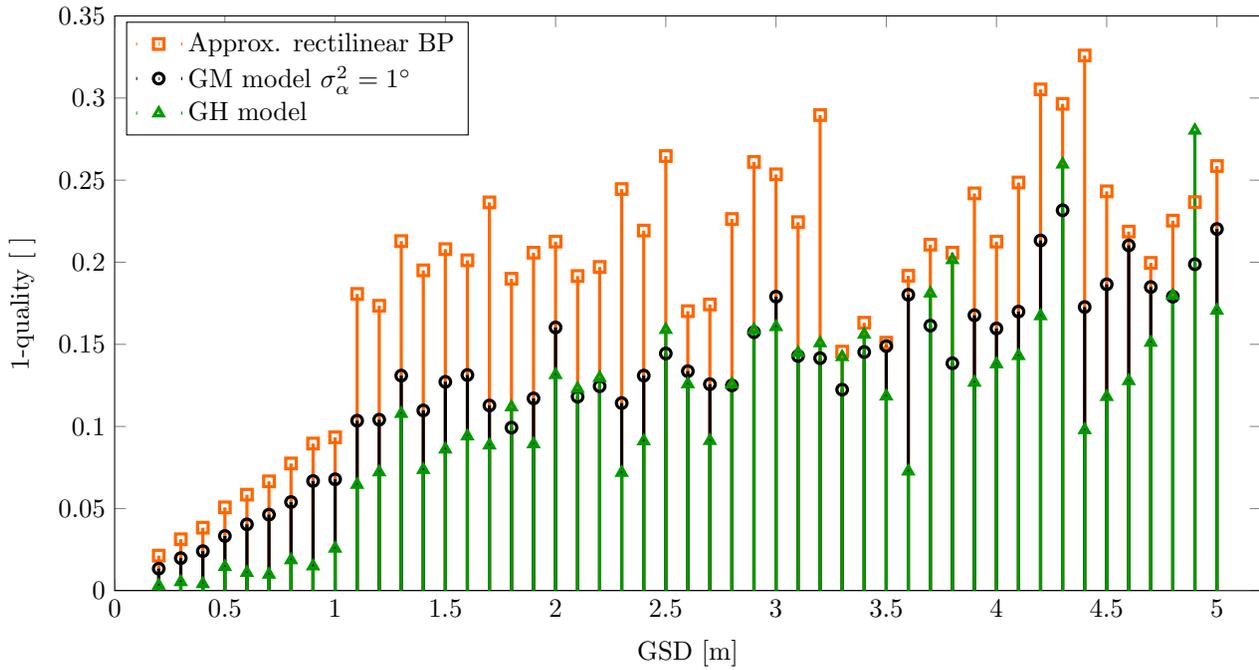


Figure 6.18: Quality rate [ ] for evaluation of the three different extracted BP as a function of GSD. These extracted BP are, approximate rectilinear BP, BP adjusted by LS method using GM (black) and the GH (green) functional model. For easier visual comparison with the PoLiS metric values, 1-quality rate values are shown (denoted  $1 - qual$ ). Lower values indicate better quality of extracted BP. The trend of the  $1 - qual$  values is similar to the PoLiS metric values (Figure 6.16).

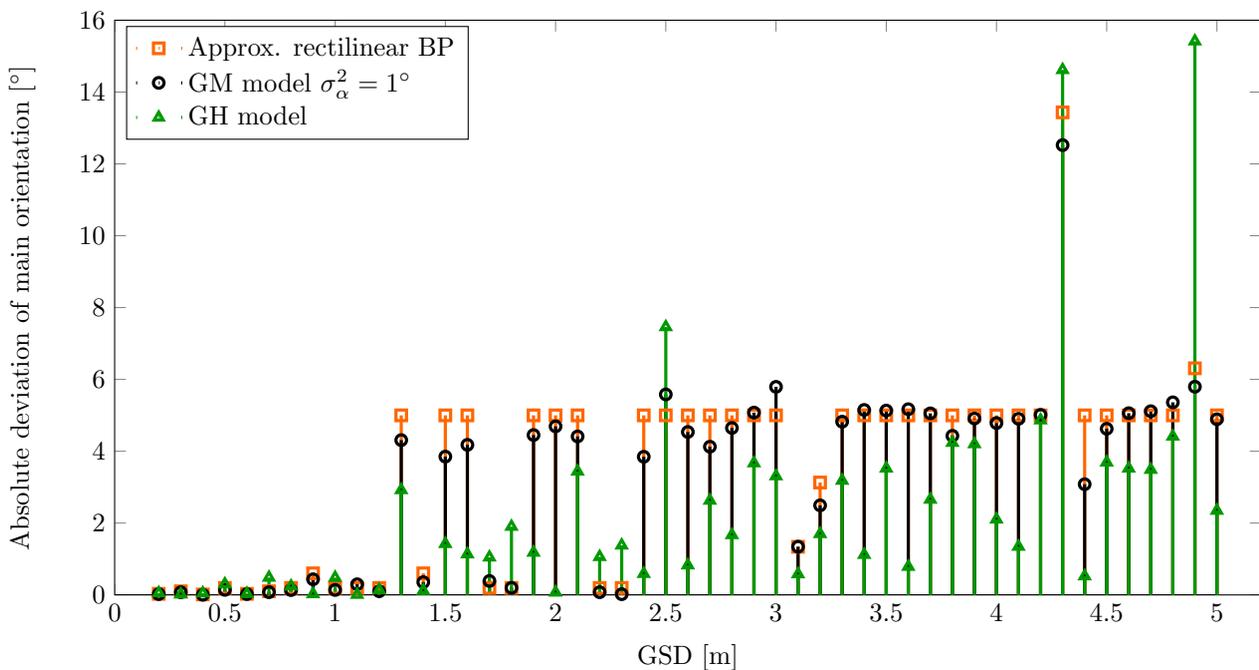


Figure 6.19: Absolute deviation of the main orientation [°] between the BP and the reference polygon. For smaller GSD, the main orientation is well estimated for approximate (orange), as well as the adjusted BP (black for GM and green for GM model), i.e. the difference in orientation is  $< 1^\circ$ . The main orientation of the approximate BP is the orientation of the MBR. For many  $GSD > 1.2$  m, the deviation of the main orientation for the approximate rectilinear BP is exactly  $5^\circ$ . This value is due to the rasterisation, i.e. size, orientation of the building and the GSD of the building.

$GSD = \{0.5, 0.6, \dots, 5.0\}$  [m] and for rectilinear and non-rectilinear BP (Figure 6.13). Two settings of weights for GM model are used (Table 3). The quality parameters and main orientation deviation values for all  $GSD > 3.0$  m and both models are summarised by their mean, standard deviation and median values (Tables 4 and 5). The mean value, indicating the worst performance of the models and their settings are marked red, and the best ones green.

The improvement of quality parameters for adjusted BP in comparison to the approximate BP can be small, even if the adjusted BP fits well to the image of the building outline under consideration. An example for such a case is, when the approximate BP, and consequently also the approximate values of the unknowns, are already very close to the optimal values of unknowns in the sense of LS (Figure 6.12a). If the approximate values of unknowns are not close enough to the optimal values, the adjustment does not converge or converges to a local minima in the sum of squares.

The main orientation of the BP is one of the crucial issues regarding the approximate and/or adjusted BP (see Subsection 5.1.2). Both proposed functional models allow for changes of the main orientation of the BP in the adjustment process. However, the actual change of orientation depends on the boundary point distribution, their assignment to the edges of the BP, and also the relation between the size of the building under consideration and GSD.

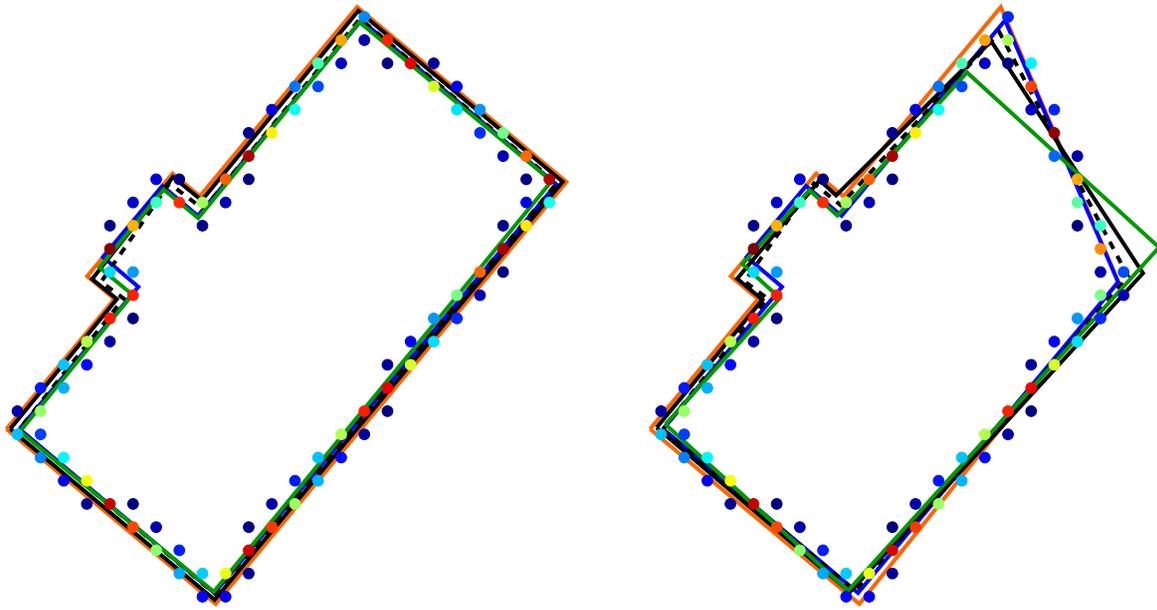
## Rectilinear Building

The approximate BP are improved through the adjustment (Figures 6.16–6.18, Table 4). This can be observed by comparing the PoLiS metric and  $1 - qual$  values of the approximate BP (orange) to the adjusted BP (black, green). The quality parameter values after adjustment are improved in an average between  $1/3$  and  $1/2$  (Table 4). The rate of improvement indicates the limitation of the iMBR algorithm for BP creation and selection, rather than the evaluation of the proposed LS models. Thus, the actual values of quality parameter values should be discussed. For example, the mean value of PoLiS metric for  $GSD \leq 3.0$  m before adjustment is 0.99 m and after adjustment it is 0.59 m (GM model  $\sigma_\alpha^2 = 1^\circ$ ), 0.47 m (GM model  $\sigma_\alpha^2 = 10^\circ$ ), or 0.42 m (GH model). The quality of the adjusted BP is below 0.5 pixel and the quality rate above 0.90.

The main orientation deviation of the adjusted BP for  $GSD \leq 3.0$  m is in an average  $2.14^\circ \pm 2.27^\circ$  for GM  $\sigma_\alpha^2 = 1^\circ$ ,  $1.65^\circ \pm 1.85^\circ$  for GM  $\sigma_\alpha^2 = 10^\circ$ , and  $1.30^\circ \pm 1.64^\circ$  for GH model (Table 4). The main orientation is not improved for all GSD after LS adjustment (Figure 6.19). For  $GSD \leq 1.2$  m, no significant improvement or worsening of the main orientation can be observed, because the approximate BP (orange) and their main orientation are already close to the reference polygon (e.g. Figure 6.14a, blue). This is also true for some  $GSD > 1.2$  m, e.g.  $GSD = \{1.4, 2.2, 2.3\}$  [m]. For most of the  $GSD > 1.2$  m, the main orientation is improved after the adjustment, if GM  $\sigma_\alpha^2 = 10^\circ$  or GH model (Figure 6.19, green) is applied.

Yet, for e.g.  $GSD = \{4.3, 4.9\}$  [m] the main orientation of the BP is worsen after the adjustment with the GH model (green). This also resembles in the poorer quality parameters for these GSD. For  $GSD = 2.5$  m, the the main orientation of the BP is also worsen after the adjustment with the GH model, but this does not resemble in the quality parameters. The reasons for this are a) the main orientation of the approximate BP is not as wrongly estimated as for  $GSD = \{4.3, 4.9\}$  [m], and b) the values of the quality parameters are compensated by the correct area and position of the adjusted BP (for quality parameter), and close position of the BP's vertices to the reference polygon (for PoLiS).

The improvement of the main orientation when using the GM  $\sigma_\alpha^2 = 10^\circ$  is insignificant. The setting  $\sigma_\alpha^2 = 10^\circ$  relaxes the rectilinearity constraint. Thus, the main orientation of the adjusted BP is computed from the longest edge of the BP without the entrance (i.e. the most right of the edges of the



(a) Adjusted rectilinear BP with 1.0 m grid size.

(b) Adjusted non-rectilinear BP with 1.0 m grid size.

Figure 6.20: Adjusted rectilinear (Figure 6.20a) and non-rectilinear (Figure 6.20b) BP by GM  $\sigma_\alpha^2 = 1^\circ$  (solid black), GM  $\sigma_\alpha^2 = 10^\circ$  (dashed black), and GH (green) model. Reference (blue) and approximate BP (orange) are also shown. Points represent the boundary pixels of the BP, and their colour represents a weight of each pixel (see Figure 6.20a). Both, GM and GH model assume rectilinearity of the BP. The rectilinearity is strictly enforced by the GH model, thus the adjusted BP is rectilinear (green). For GH model, the requirement of rectilinearity can be relaxed by setting the parameter  $\sigma_\alpha^2$  higher, i.e. weights of the angles are lower. Thus, the BP adjusted by the GM model  $\sigma_\alpha^2 = 10^\circ$  (dashed black) is nearly rectilinear. This can be best observed on the edges of the BP, which represent the small structure (entrance to the building).

BP, Figure 6.11a, blue). This longest edge is not exactly parallel or perpendicular to the other edges. If the building under consideration is rectilinear, then the relaxation of the rectilinearity constraint negatively influences the accuracy of the main orientation. Yet, it might improve the values of quality parameters of the adjusted BP.

The quality parameter values for GM  $\sigma_\alpha^2 = 10^\circ$  are better than for GM  $\sigma_\alpha^2 = 1^\circ$ . The GM model does inaccurately estimate the main orientation in some cases ( $GSD = \{2.5, 3.9\}$  [m]), even if the approximate orientation is of sufficient accuracy. This is because the LS adjustment converges to another local minima in the sum of squares. However, in general the GH model (green) outperforms the GM model (black) with either of the settings, i.e. it estimates the main orientation better and also the quality rate parameters are better (Table 4, Figure 6.20a).

### Non-Rectilinear Building

In this experiment the non-rectilinear BP (Figure 6.13) is adjusted by the GM and the GH functional models, which both assume rectilinearity. In an ideal case, only exact rectilinear buildings shall be adjusted by such models. Even if most of the buildings are rectilinear (Chaudhuri and Samal, 2007; Kanani, 2000; Niemeier, 2008; Steadman, 2006), few ones have different shape. Without a prior check, all the approximated BP are adjusted by the same LS adjustment model. The goal of this experiment is to investigate how do the proposed GM and GH models perform, when the BP under consideration is non-rectilinear. A performance of the GM with relaxed rectilinearity constraint (Setting 2,  $\sigma_\alpha^2 = 10^\circ$ ) is also discussed. The experiment is carried out in an analogue way as for the rectilinear BP. The

		Approximate BP	GM model $\sigma_\alpha^2 = 1^\circ$	GM model $\sigma_\alpha^2 = 10^\circ$	GH model
PoLiS [m]	mean	0.84	0.81	0.66	0.88
	std	0.44	0.41	0.40	0.26
	median	0.96	0.99	0.81	0.99
PoLiS [pixel]	mean	0.53	0.50	0.37	0.72
	std	0.17	0.13	0.17	0.44
	median	0.44	0.47	0.39	0.59
1-quality rate [ ]	mean	0.14	0.13	0.11	0.14
	std	0.07	0.06	0.07	0.03
	median	0.17	0.16	0.14	0.15
Main orientation deviation [°]	mean	2.51	3.91	3.13	3.95
	std	2.30	2.96	3.03	1.54
	median	1.04	2.90	1.68	3.67

Table 5: Quality parameters, i.e. PoLiS [m], PoLiS [pixel],  $1 - qual$  [ ], and main orientation deviation [°], for the test non-rectilinear building polygon (Figure 6.13). Values are computed for  $0.5 \text{ m} \leq GSD \leq 3.0 \text{ m}$ . Two sets of parameters are used for the GM model (Table 3), which differ in the weights for rectilinear angles defined through  $\sigma_\alpha^2$ . No settings are required for the GH model. The mean value for each quality parameter and main orientation deviation indicating the worst performance of the models are marked red, and the best ones green. For non-rectilinear test building, the GM model with  $\sigma_\alpha^2 = 10^\circ$  outperforms the GM model, which more strictly enforces the rectilinearity, and also GH model.

same quality parameters and the deviation of the main orientation are computed to assess the quality of adjustments (Table 5).

Figure 6.20b shows adjusted non-rectilinear BP by GM  $\sigma_\alpha^2 = 1^\circ$  (solid black), GM  $\sigma_\alpha^2 = 10^\circ$  (dashed black), and GH (green) model. The approximate BP (orange) is needed prior to adjustment. The quality of adjusted BP is compared to the reference BP (blue). The GH model strictly enforces rectilinearity through a constraint to the normal vectors. Thus the resulting adjusted BP is rectilinear (green), even if one edge of the building is obviously not perpendicular to the neighbouring two. The adjusted BP by GM model (black solid and dashed) fit this edge of the building better to the boundary points. Yet, the other edges are only nearly rectilinear, which can be best observed on the edges of the BP, which represent the small structure (entrance to the building).

In contrast to the adjustments of the rectilinear BP, the adjustments of non-rectilinear BP do not always improve the quality parameters. In general, the quality parameter values are higher for adjusted non-rectilinear BP than for adjusted rectilinear BP. The mean value for PoLiS and  $1 - qual$  remain about the same for the approximate BP and adjusted BP by the GM  $\sigma_\alpha^2 = 1^\circ$  model. The GH model performs worse than GM model with either of both settings (Table 5, red numbers). The GM  $\sigma_\alpha^2 = 10^\circ$  model, which does not strictly enforce rectilinearity, performs the best (Table 5, green numbers).

The main orientation of the non-rectilinear BP is worse than before any of the adjustments. The GM model with relaxed rectilinearity constraint can compensate for false assumption of rectilinearity of the BP under consideration. However, the main orientation can be improved only if the rectilinearity assumption holds and large enough number of boundary points per building and building edge are available. The deviation of the main orientation remains below  $4.00^\circ$  for both considered adjustment models and their settings.

### 6.3.5 Summary of Experiment on BP Selection and Adjustment

The proposed iMBR method establishes, for every building region, a set of BP. Each of the BP in the set has different level of details, where a higher level of BP is more complex than a lower one. The experiment is carried out on an example of one building, but with varying GSD. One BP is automatically selected from the set of BP as the best trade-off between complexity and fitness to the data. For finer GSD, more complex, i.e. higher level of BP, is selected, and for larger GSD the selected BP is more generalised representation of the building outline.

The following conclusion can be summarised from the experiments on BP selection and adjustment:

- ◇ The selected level of BP depends on the relation between the dimensions of an objects or object parts and the GSD, and also on the complexity of the object's boundary. Thus, the quality of extracted boundaries of different objects in one scene shall be evaluated on per-object basis.
- ◇ For larger GSD a simpler BP (lower level) is selected over a more complex one. For larger value of the minimal area of a building part and the same GSD, also a simpler BP is selected.
- ◇ The main orientation of the BP is one of the crucial issues for both, approximate BP, and adjusted BP. The improvement of the main orientation of the BP in the adjustment depends on the boundary point distribution, their assignment to the edges of the BP and also the relation between the size of the building and GSD of the image. An adaptive re-assignment of the boundary points to the BP edges in every iteration of the LS adjustment could additionally improve the main orientation. The consequence of such re-assignment is a change of the structure of the design matrices in every iteration. There is an increased risk the adjustment does not converge or converges to a local minima.
- ◇ The PoLiS ( $1 - qual$ ) of approximate BP defined by iMBR method is in average 0.62 pixel (0.16). After the adjustment, the PoLiS metric value improves to 0.38 pixel (0.10) for the GM model  $\sigma_\alpha^2 = 1^\circ$ , and 0.23 pixel (0.08) for the GH model. In the brackets are the values for  $1 - qual$ .
- ◇ The quality of adjusted BP by GM model is improved by relaxing the rectilinearity constraint, i.e. PoLiS and quality rate values are lower. The relaxation causes that the adjusted BP fit better to the boundary points (data). Better fitness of the model to the data does not necessarily mean that the adjusted BP are better representation of objects under consideration. The main drawback of relaxing the rectilinearity constraint is that the adjusted BP are not strictly rectilinear. Relaxation of the rectilinearity constraint is a good choice, when no assumption about the rectilinearity of the buildings under consideration can be met.
- ◇ The proposed GH model for rectilinear BP performs better than the GM model and requires no setting of the weights for groups of observations. However, if the BP under consideration is not rectilinear the GM model with relaxed rectilinearity constraint performs better. The GM model allows for more general shapes of the considered BP, but the result has lower quality. The GM model is more flexible than the GH model with respect to the assumption on rectilinearity.
- ◇ If all buildings under consideration are rectilinear, the GH model shall be applied for BP adjustment. If no such assumption can be met, the GM model with  $\sigma_\alpha^2 = 1^\circ$  (or similar) shall be chosen, because it balances between the rectilinearity assumption and fitness of the model to the boundary points.

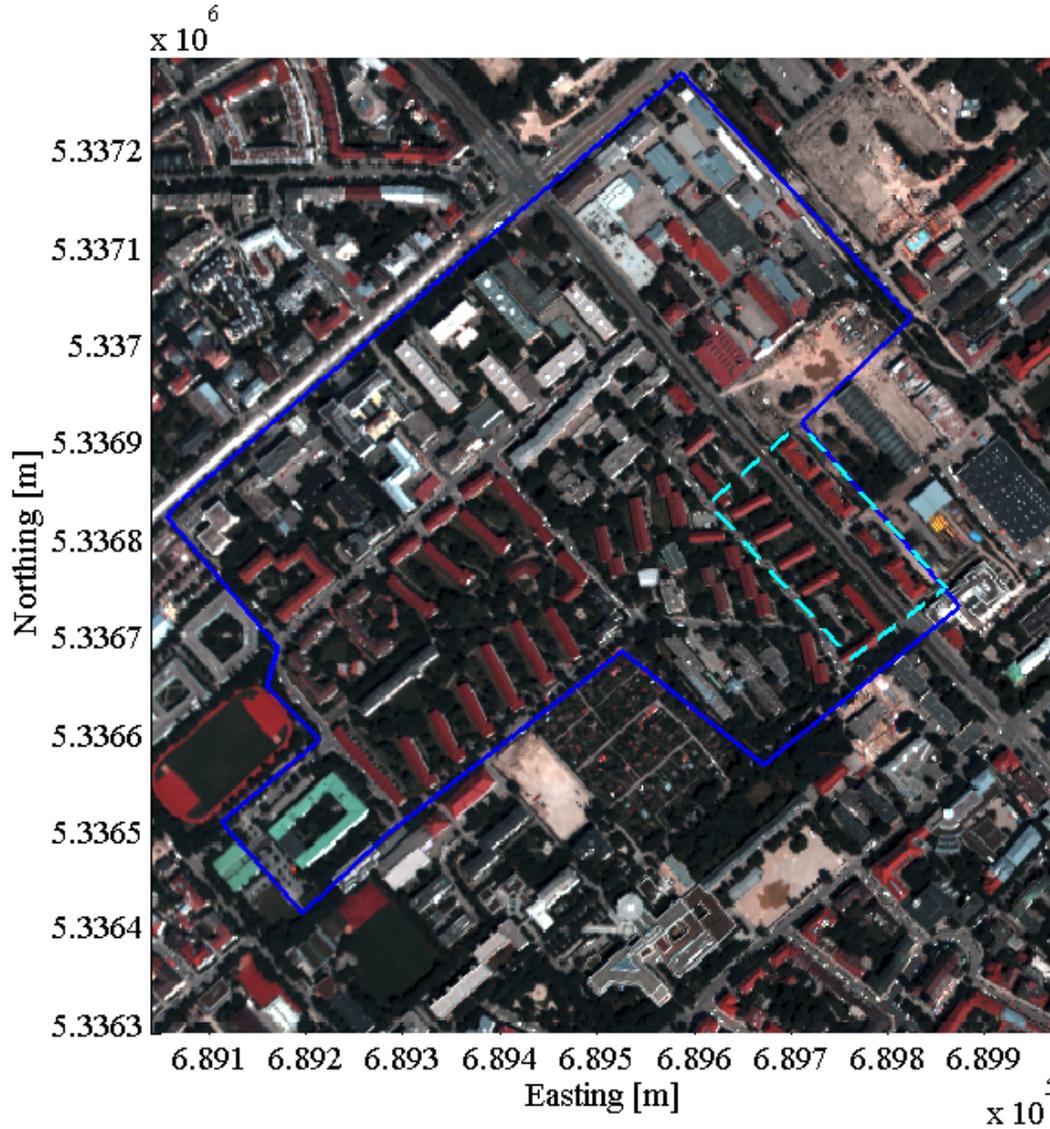


Figure 6.21: Test area (blue line), referred to as the *large area*, is showing an urban scene of a part of a city of Munich. RGB-composite of the VNIR20 dataset is shown. All proposed methods are connected into one workflow (Figure 6.1, bold rectangles) and applied on the real RS images. An extensive investigation is carried out for all datasets for the *small area* (cyan dashed line).

## 6.4 Experiments on RS Images

The proposed methods, i.e. edge probability detection and fusion, BP creation and selection, adjustment of approximate BP, and finally the evaluation of results are joined into one workflow. The main objective of experiments on RS images is join the proposed methods (Figure 6.1, bold rectangles) into one workflow (Figure 6.1). The methods are applied on three pairs of real RS images with different GSD. Additionally, an investigation of edge probabilities is carried out on 17 HSI–DSM dataset pairs with different characteristics and GSD.

### 6.4.1 Test Dataset

For all experiments a pair of HSI and DSM is required as an input. All test images show the same urban area of the city of Munich, Germany (Figure 6.21). The test area of a size of about  $0.8 \text{ km} \times 0.6 \text{ km}$  is

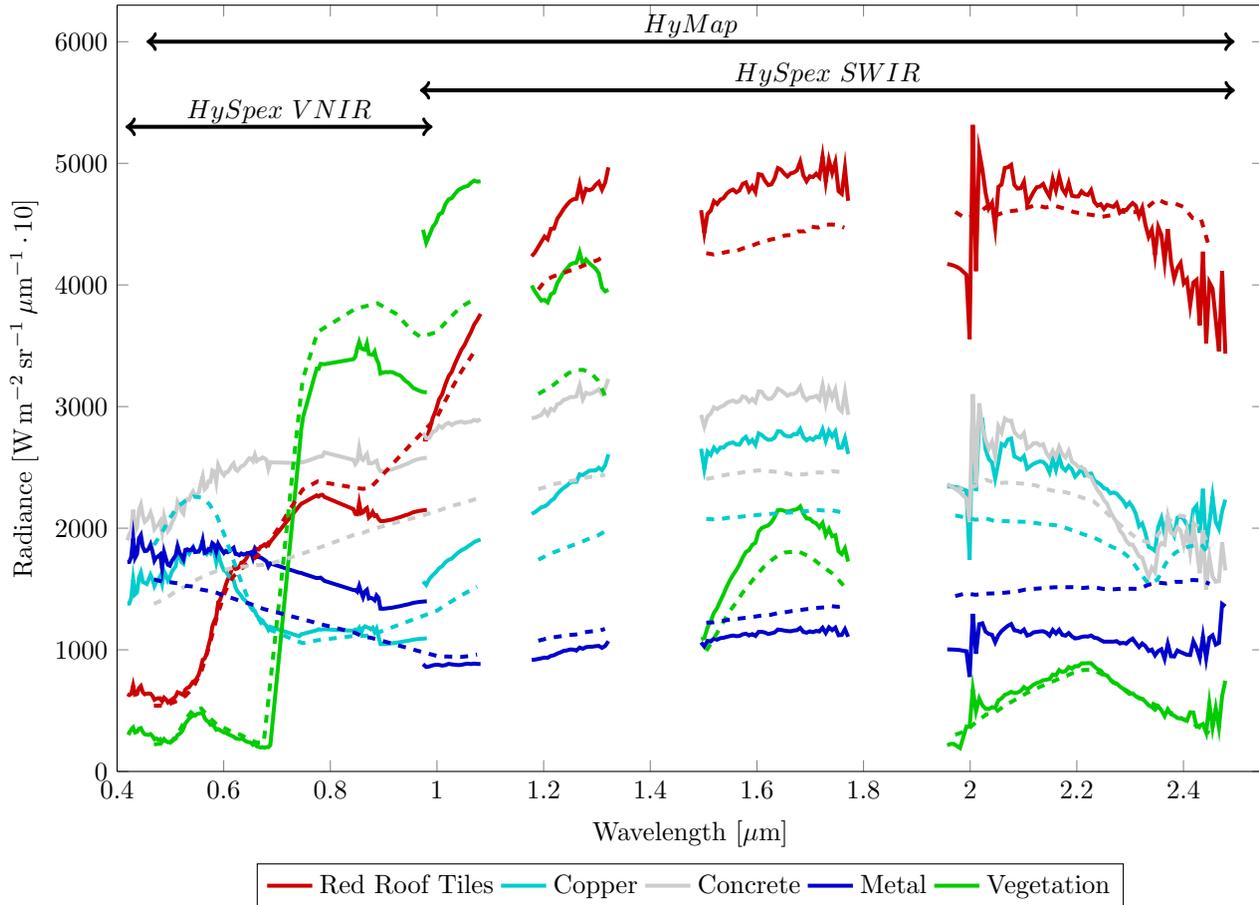


Figure 6.22: Spectral signatures of five selected materials from HyMAP (dashed line) and HySpex (solid line) HSI. Due to the high signal noise and absorption regions, some of the channels were removed. The arrows denote the spectral interval, on which each of the sensors, i.e. HyMAP, HySpex VNIR and SWIR, acquires data. These spectral signatures are used as a reference spectra for NNLS unmixing, which results in the material maps of corresponding materials. The high frequency noise can be observed in the spectral signatures collected from the HySpex sensors. The main source of this noise can be explained by uncorrected systematic errors.

marked by a blue line and referred to as the *large area*. An extensive investigation of the fusion of edge probabilities is carried out on a subset of the test area (cyan dashed line), which is referred to as the *small area*. The *small area* exhibits eight buildings. It was chosen for detailed experiments, because it shows buildings with the same type of roofing material (red roof tiles), similar complexity level of the BP, and similar size.

Seven HSI are used in the experiments with four different GSD (Subsection 6.1.1), acquired by HyMAP and HySpex sensors, and two DSM models, computed from the stereo 3K and WV-2 images. Additionally, the methods are applied also on the LiDAR DSM for comparison of the edge probability fusion results.

Vegetation (green) and four roofing materials (Figure 6.22), these are red roof tiles (red), copper (cyan), concrete (grey), and metal (blue) are selected as reference materials. Due to the low SNR and atmospheric absorption regions, some of the channels were not considered for further processing. Gaps represent the removed channels. Spectral signatures of the materials selected from HyMAP are represented by dashed lines and from HySpex by solid lines. A high frequency noise can be observed in the spectral signatures collected from the HySpex sensors. The main source of this noise can be explained by uncorrected systematic errors. The material maps are obtained by NNLS unmixing of the HSI using these five reference materials.

The building footprints from the municipal of Munich are used as ground truth for evaluation of adjusted BP. The ground truth BP are very detailed, consisting of many short line segments. The municipal data about buildings are collected for a different purpose, e.g. as a part of a real estate database. Therefore, the given ground truth BP are divided into the building parts. For our purpose, the building parts are irrelevant and were removed from the ground truth BP. Moreover, the ground truth BP represent the outline of the outer walls of buildings and not of the roofs as seen from nadir imagery. For instance, a gable roof has usually an overhang, which is not a part of the ground truth BP. For this reason and because most of the buildings in the test area have roofs with overhangs of about 1 m ground truth BP appear to be smaller than the buildings in the RS images. The evaluation results are not compared to results of other authors due to two main reasons. First, no HSI–DSM benchmark dataset pair for BP extraction is available. Second, comparison of per-scene matched rates among each other is always dependent on the quality of the ground truth (see Section 3.3).

In our experience, a prerequisite for applying the proposed methods is at least half a pixel accurate registration between HSI and DSM (Chapter 4). For evaluation, also the ground truth BP shall be registered to the HSI and DSM.

## 6.4.2 Parameter Settings and Preprocessing

The *DSM normalisation* is carried out for each DSM (Subsection 5.1.1). For the DSM normalisation of the large area, a window size  $W$  in pixels corresponds to 150 m. For the *small area*, the window size  $W$  corresponds to 50 m, because the *small area* only includes smaller buildings. The window size is always rounded to the nearest odd integer number.

The *vegetation removal* from nDSM is carried out independently for each dataset. For the *large area*, the high vegetation is removed on the basis of the VNIR20 material map of vegetation, because VNIR03 and VNIR10 images do not cover the whole area. Thus, the vegetation material map is up-sampled using bilinear interpolation, which negatively influences the building region extraction and consequently the BP creation. For the *small area*, the high vegetation is removed on the basis of the original or down-sampled VNIR03 vegetation map.

For all experiments, HSI–DSM dataset pairs have the same GSD. The nDSM with removed high vegetation are down-sampled to the GSD of the HSI by nearest-neighbour interpolation.

The *building regions* are gained from the DSM with removed high vegetation, and from the material maps. Buildings smaller than  $25\text{ m}^2$  are not considered, pixels that have exactly one 8-connected neighbour, and H-connected pixels are also removed from building regions.

The approximate BP are created from each of the datasets, i.e. from all HSI and all DSM, by the iMBR method, i.e. the result are two sets of approximate BP representing the same building outlines. Thus, only one set of approximate BP has to be chosen. In this experiment, the selected set of approximate BP consists of BP from both dataset. The task of automatic *BP selection* raises a problem of correspondence and cardinality. This problem is related to the so called one-to-many cardinality problem, which has been addressed by many authors in relation to BP evaluation (see Section 3.3 and Geibel and Stilla (2000); Rutzinger et al. (2009)). Let  $Q, q = 1, \dots, Q$ , BP extracted from a HSI correspond to  $P, p = 1, \dots, P$ , BP extracted from a DSM. Their costs are computed as given in Equation (39) and are denoted  $\text{cost}(BP_p)$  for DSM BP and  $\text{cost}(BP_q)$  for HSI BP. The joined cost of all  $P$  BP from the DSM, denoted  $\text{cost}(BP_{DSM})$ , which corresponds to  $Q$  BP from HSI, has to be

computed, and vice versa. The joined cost shall prefer simpler BP and a smaller number of the BP. The joined costs, i.e.  $\text{cost}(BP_{HSI})$  and  $\text{cost}(BP_{DSM})$ , are computed as

$$\text{cost}(BP_{HSI}) = P \sum_{q=1}^Q \text{cost}(BP_q) \frac{\text{ar}(BP_q)}{\sum_{q=1}^Q BP_q} \quad (73)$$

$$\text{cost}(BP_{DSM}) = Q \sum_{p=1}^P \text{cost}(BP_p) \frac{\text{ar}(BP_p)}{\sum_{p=1}^P BP_p} \quad (74)$$

where  $\text{cost}(BP_{HSI})$  is the joined cost of all  $Q$ ,  $q = 1, \dots, Q$ , BP, which correspond to  $P$ ,  $p = 1, \dots, P$ , BP extracted from the DSM,  $P$  is the number of the corresponding BP from DSM, and the  $\text{ar}(\cdot)$  denotes the area of the BP. The joined cost for BP from DSM is computed analogically (Equation (74)). The BP with the minimal joined cost is selected. If in one of the datasets the approximate BP is missing, then the BP from the other one is selected.

The selected BP are adjusted and evaluated by computing PoLiS metric and quality rate ( $1 - \text{qual}$ ) between the ground truth and the extracted BP. For the per-object evaluation, a threshold of minimal overlapping area between adjusted BP and the ground truth BP is set to 10 %, of the area of the smaller BP.

If not differently stated, the HSI are shown in figures as grey scale images representing one channel of the HSI with the following central wavelengths:  $\lambda = 0.6439 \mu\text{m}$  for VNIR,  $\lambda = 1.5253 \mu\text{m}$  for SWIR, and  $\lambda = 0.6320 \mu\text{m}$  for HyMAP images.

### 6.4.3 Results and Discussion of the Experiment on the *Small Area*

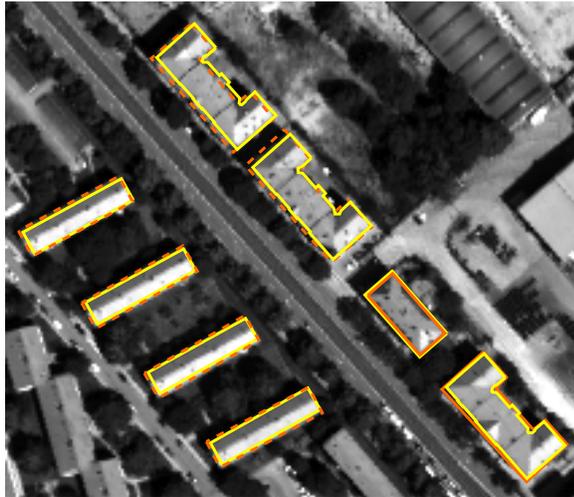
The BP extraction workflow is tested on 17 HSI–DSM dataset pairs. The HSI are VNIR and SWIR HySpex HSI with  $GSD = \{0.3, 1.0, 2.0\}$  [m], and HyMAP HSI with  $GSD = 4.0$  m. The DSM are 3K DSM, WV-2 DSM and LiDAR DSM. A prerequisite to compare the performance of the methods applied on VNIR and SWIR datasets is that the material under consideration has prominent spectral features in VNIR and SWIR regions. For this reason, the eight buildings in the *small area* characterised by the red roof tile material (see Figure 6.22), which has spectral features in VNIR and SWIR spectral regions, are chosen. The approximate BP are adjusted by the GM model ( $\sigma_\alpha^2 = 1^\circ$  and  $\sigma_\alpha^2 = 10^\circ$ , Table 3) and the GH model.

In all the figures exhibiting images, the approximate BP are orange, adjusted BP are green, and the ground truth BP are yellow. The results for all 17 tested dataset are not shown on figures, but for at least VNIR10–3K DSM dataset pair. All the (intermediate) results are shown for at least the VNIR10–3K DSM dataset pair.

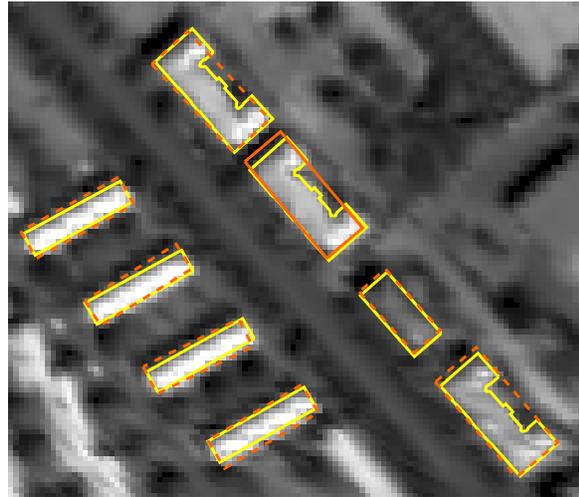
#### ***Small Area*: BP Creation and Selection**

All eight BP are extracted from the HySpex images and all DSM. From the HyMAP image, one BP is not extracted, because it is missing in the building regions due to the different spectral characteristics of the roofing material.

The *small area* exhibits eight buildings, of which five have rectangular shape (Level 1 BP), and the other three are well approximated by the Level 2 BP. According to the ground truth polygons, these three BP have more complex shape than Level 2 BP (Figure 6.23, yellow). However, the details are small, i.e. they are not recognisable and not detectable from the  $GSD = 0.3$  m images (Figure 6.28c).



(a) Approximate BP VNIR10.



(b) Approximate BP SWIR20.

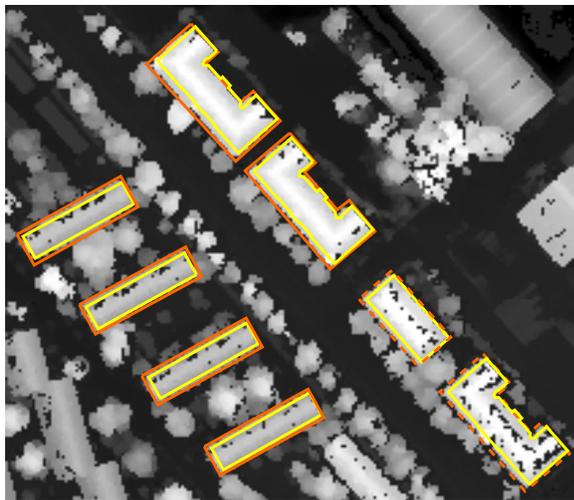
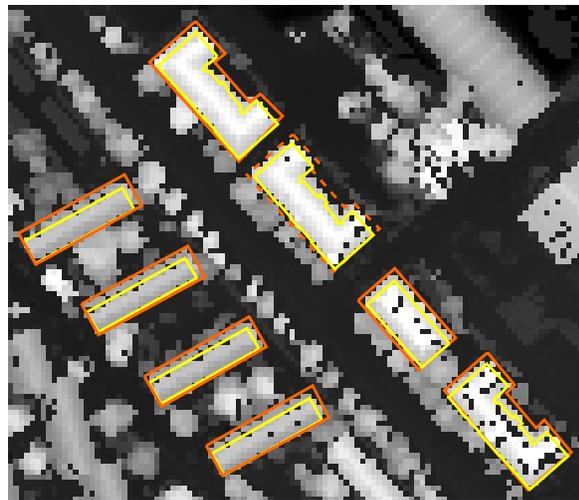
(c) Approximate BP 3K DSM  $GSD = 1.0$  m.(d) Approximate BP 3K DSM  $GSD = 2.0$  m.

Figure 6.23: Approximate BP creation and selection for VNIR10–3K DSM and SWIR20–3K DSM dataset pairs (*small area*). The approximate BP (orange) for red roof tile material from VNIR10 (Figure 6.23a), SWIR20 (Figure 6.23b), and 3K DSM resampled to  $GSD = 1.0$  m (Figure 6.23c) and  $GSD = 2.0$  m (Figure 6.23d). For each HSI–DSM dataset pair, one approximate BP is selected on the ground of the minimal cost (solid orange line). The ground truth BP are yellow. The adjusted BP for these two dataset pairs are shown on Figures 6.28a and 6.28b.

Up to a  $GSD = 1.0$  m and for all dataset pairs, five approximate BP are detected as Level 1 and three as Level 2 BP (Figure 6.25). The selected levels of BP are the same for  $GSD = 2.0$  m and VNIR dataset pairs. For  $GSD = 2.0$  m and SWIR dataset pairs, simpler BP are preferred. These results are in line with the results of the number of BP selected from DSM (Figure 6.24). For larger  $GSD$ , the cost of the BP from HSI is higher than from the DSM. A difference between VNIR–DSM and SWIR–DSM dataset pairs for  $GSD = 2.0$  m can be observed for the number of selected BP from DSM and level of the selected BP.

Figure 6.23 (orange) shows the approximate BP, which were selected from both datasets by comparing the joined cost functions (Equations (73) and (74)). A general tendency is that for larger  $GSD$ , more BP are selected from the DSM datasets. (Figure 6.24). In other words, for smaller  $GSD$  the approximate BP extraction relies more on material information from HSI than on height information from DSM.

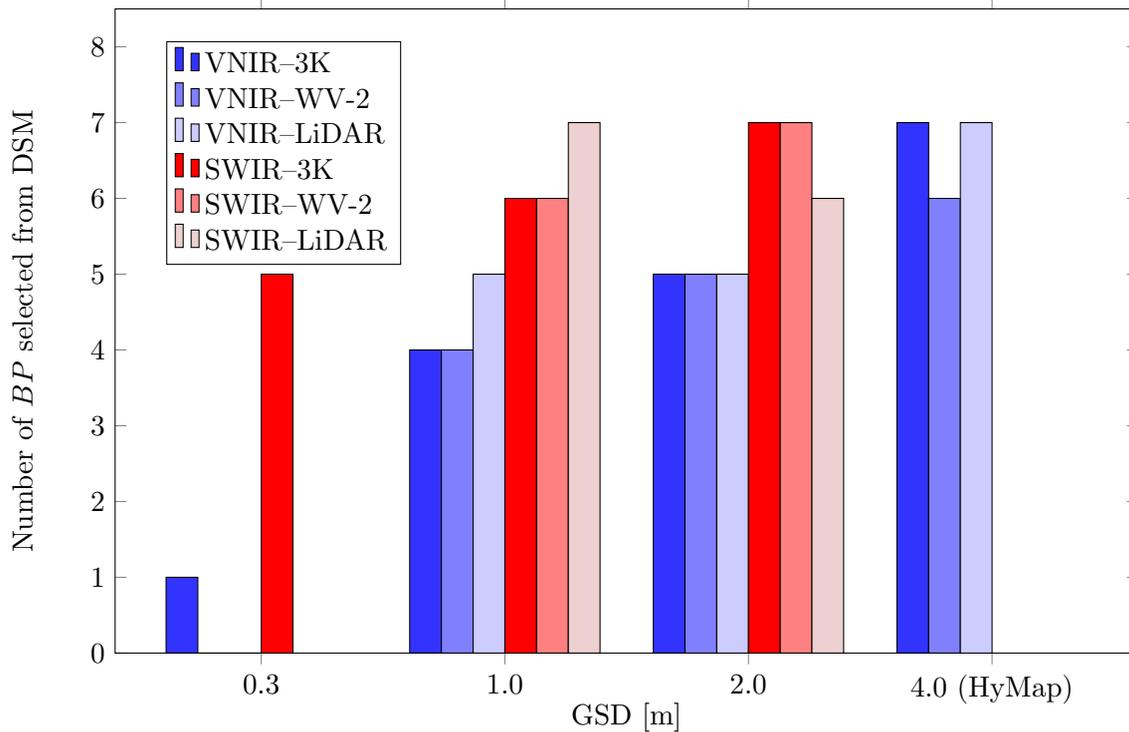


Figure 6.24: Number of approximate *BP* selected from *DSM* with respect to the *GSD* (*small area*). The *small area* exhibits eight buildings. The cost of approximate *BP* of the same building, extracted from *HSI* and *DSM* is compared. The *BP* with the smaller cost is selected as an approximate *BP*. If in one of the datasets, an approximate *BP* is missing, the *BP* from the other dataset is selected. A general tendency is that for larger *GSD*, more *BP* are selected from the *DSM* datasets. For *SWIR-DSM* dataset pairs, larger number of the *BP* is selected from *DSM* than for *VNIR-DSM* dataset pairs.

It can be concluded that the *VNIR* datasets enable better extraction of the *BP* with red roof tile material in comparison to the *SWIR* datasets with the same *GSD*. One reason for this is that *VNIR* sensor has smaller *IFOV* than *SWIR* (Table 2), but both images acquired by these sensor were resampled to the same *GSD*. The comparison of the *VNIR10* and the *SWIR10* datasets can be observed in Figures 6.27a and 6.27b.

### ***Small Area: Edge Probability Detection and Fusion***

In all experiments a *HSI-DSM* dataset pair is considered. Thus, the two mixing coefficients must be set, i.e.  $\pi_{HSI}$  and  $\pi_{DSM}$ . In Section 4.3, two possibilities to define the mixing coefficient  $\pi_k$  in the GMM for the edge probability fusion are proposed. First, the weighting by the confidence probability  $\beta(t_{max})$  is carried out (Subsection 4.3.2), where the  $\pi_k$  are defined as the ratio between confidence probabilities at the scales  $t_{max}$ . Second, the weighting of the *BP* boundary points by prior knowledge (Subsection 4.3.1, with  $\pi_{HSI} + \pi_{DSM} = 1$ ) is carried out. In this experiment, the weighting of boundary points by confidence probability ( $\pi \propto \beta(t_{max})$ ) is compared to the weighting by prior knowledge ( $\pi$  Prior).

For each *HSI-DSM* dataset pair, the weights of the boundary points in the *BP* adjustment are set 12-times. The mixing coefficients for *HSI* are set 11-times according to the prior probability, i.e.  $\pi_{HSI} = \{0.0, 0.1, 0.2, \dots, 1.0\}$  and once according to the ratio between confidence probabilities. For  $\pi_{DSM} = 0.0$ , the weights for boundary points of the *BP* are taken only from the edge probabilities

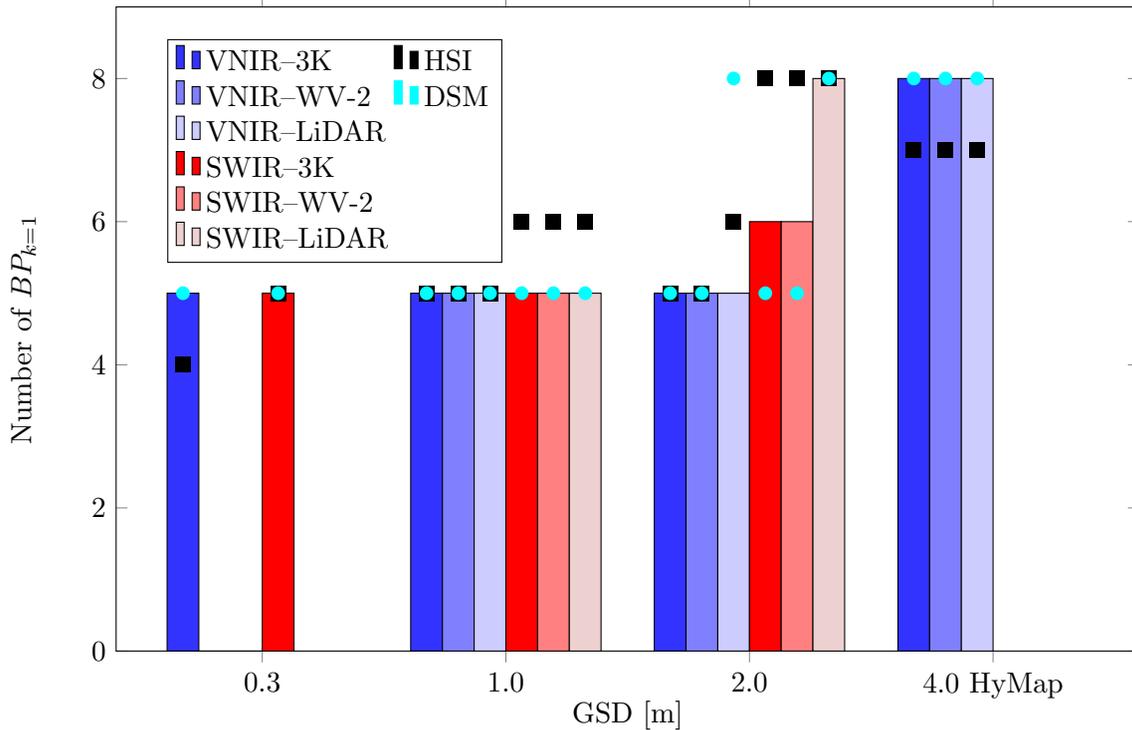


Figure 6.25: A number of Level 1 BP ( $BP_{k=1}$ ) selected from HSI-DSM dataset pairs with respect to the GSD. The *small area* exhibits eight buildings, of which five have rectangular shape (Level 1 BP), and the other three are well approximated by the Level 2 BP. Up to a  $GSD = 1.0\text{m}$  and for all dataset pairs, five approximate BP are Level 1 and three are Level 2 BP. For  $GSD = 2.0\text{m}$  and VNIR datasets the approximate BP have the same levels. However, for  $GSD = 2.0\text{m}$  and SWIR datasets, simpler BP are chosen. For  $GSD = 4.0\text{m}$ , all the BP are approximated by the Level 1 BP. For the  $GSD = 4.0\text{m}$  the HyMAP HSI are used in the HyMAP-DSM dataset pairs. The number of Level 1 BP selected from only one dataset are represented by the black squares (for HSI) and cyan circles (for DSM).

computed from the HSI, and vice versa. The joint edge probabilities according to the confidence probability  $\beta(t_{max})$  shall have higher values than the joint edge probabilities according to the prior knowledge. Moreover, values of edge probabilities extracted from only HSI or only DSM shall both carry similar amount of information, i.e. the values of  $\bar{p}(edge; t_{max})$  for  $\pi_{HSI} = 1$  and  $\pi_{DSM} = 1$  shall be comparable.

The evaluation of the experiment on edge probability detection and fusion is designed analogue to the evaluation of the experiment in Section 6.2. The mean value of the edge probabilities  $\bar{p}(edge; t_{max})$  within the buffer around the ground truth BP is considered. However, the buffer width is asymmetric, because the expected roof overhang, which is not given by the ground truth polygons, is about 1.0 m. The buffer width to the inside of the BP has a size of the roof overhang (1.0 m) in pixels plus  $\sqrt{2}/2$  pixel. The buffer width to the outside of the BP has a size of  $\sqrt{2}/2$  pixel.

Figure 6.26 shows the mean edge probabilities  $\bar{p}(edge; t_{max})$  for all dataset pairs for  $\beta(t_{max})$  (column bars), for  $\pi_{HSI} = 1.0$  (black square) and for  $\pi_{DSM} = 1.0$  (cyan circle). For the values of  $\pi_{HSI} = \{0.0, 0.1, 0.2, \dots, 1.0\}$ , the  $\bar{p}(edge; t_{max})$  can be well approximated by the linear function for all considered dataset pairs. The  $\bar{p}(edge; t_{max})$  are monotonically increasing or decreasing. Thus, if the column bar is larger than either of the marks (square, circle), the weights fused according to confidence probability  $\beta(t_{max})$  are larger than the weights fused according to the prior knowledge. The values  $\bar{p}(edge; t_{max})$  shall be compared within the datasets of the same GSD. The values of  $\bar{p}(edge; t_{max})$  are

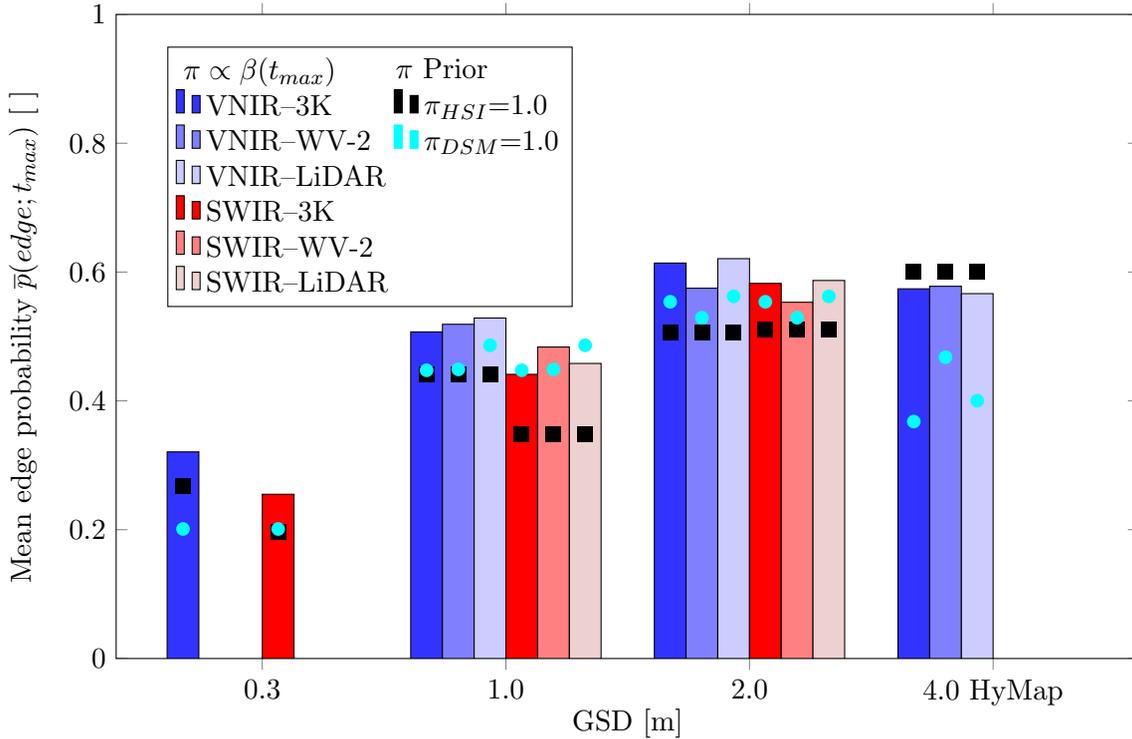
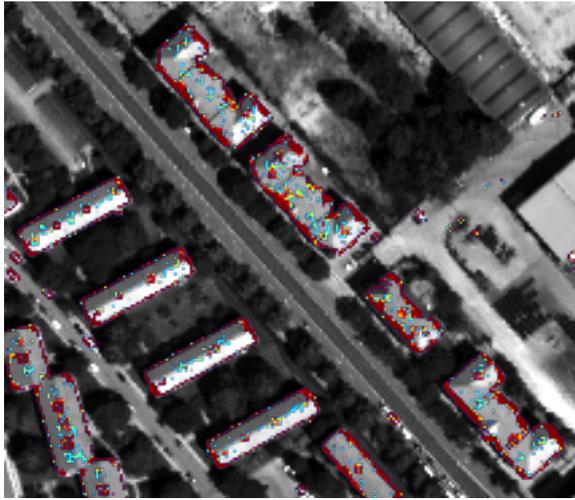


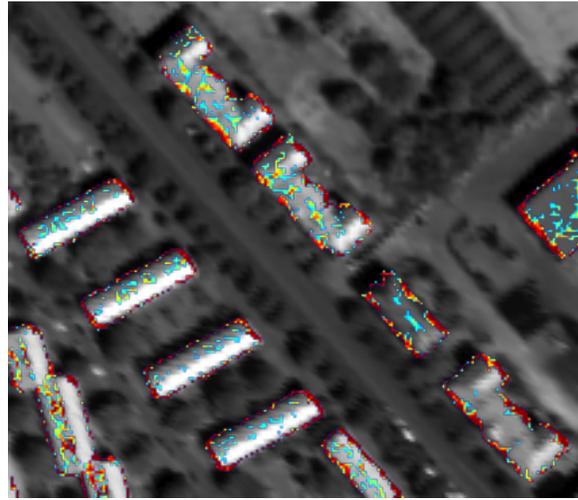
Figure 6.26: Comparison of the weighting of the BP boundary points by prior knowledge and by the confidence probability  $\beta(t_{max})$  (*small area*). The mean of the edge probabilities within the buffer around the adjusted BP  $\bar{p}(edge; t_{max})$  is computed for pairs of HSI-DSM datasets. The buffer width to the outside of the BP has a size of the roof overhang in pixels (1.0 m) plus  $\sqrt{2}/2$  pixel. The buffer width to the inside of the BP has a size of  $\sqrt{2}/2$  pixel. The height of the column bars represents the  $\bar{p}(edge; t_{max})$  for the  $\beta(t_{max})$ . The black squares ( $\pi_{HSI} = 1$ ) and the cyan circles ( $\pi_{DSM} = 1$ ) denote the  $\bar{p}(edge; t_{max})$  for the two extreme cases of the weighting by the prior knowledge, where weights are determined only from the HSI and the DSM, respectively. The values  $\bar{p}(edge; t_{max})$  shall be compared within the datasets of the same GSD, because the absolute values of the colour bars are influenced by the edge probabilities and the number of the pixels within the buffer.

influenced by the values of the edge probabilities and the number of the pixels within the buffer. For instance, for  $GSD = 0.3$  m, the buffer includes larger number of pixels than for lower  $GSD$ , therefore several pixels within the buffer have low values.

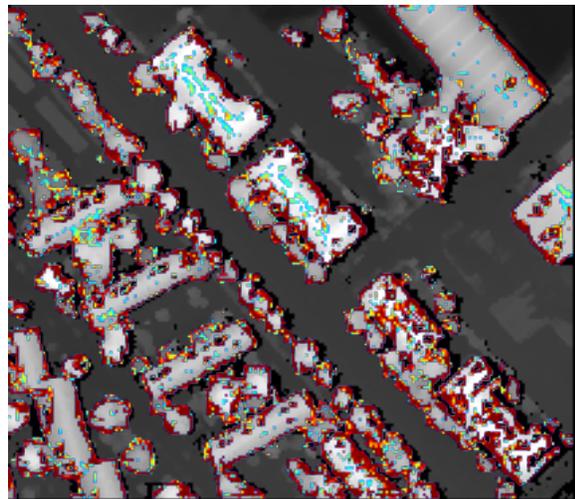
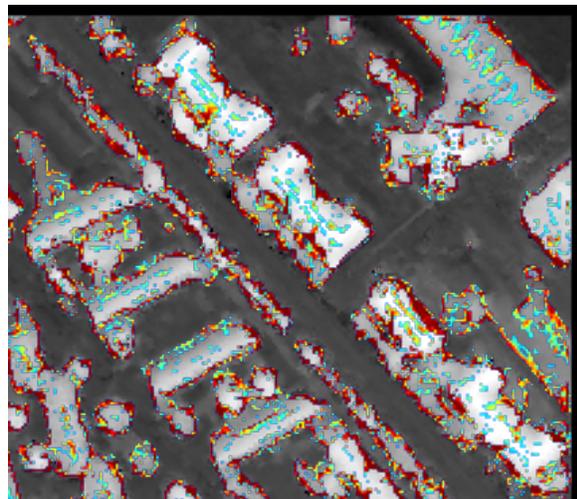
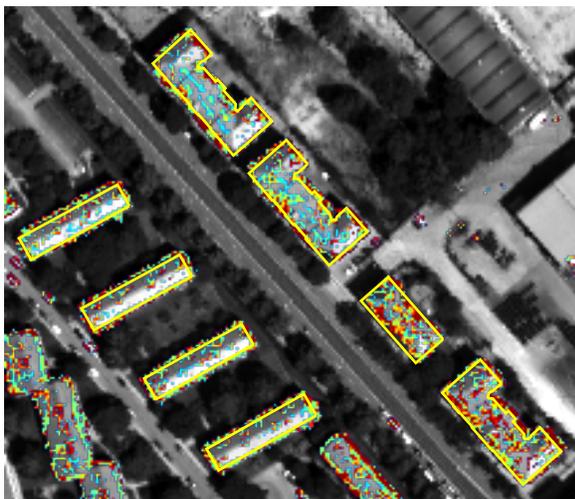
Figure 6.27 shows the edge probabilities for VNIR10 (Figure 6.27a), SWIR10 (Figure 6.27b), 3K DSM (Figure 6.27c), and WV-2 DSM (Figure 6.27d), and the joint edge probabilities according to confidence probabilities for VNIR10–3K DSM (Figure 6.27e) and SWIR10–WV-2 DSM (Figure 6.27f). The edge probabilities detected from VNIR10 are larger than the ones detected from SWIR10 datasets. In both HSI some pixels have, next to the high edge probabilities around building edges, high edge probabilities on the roofs. A reason for this can be smaller structures on the roofs and spectral changes of the red roof tile material. The edge probabilities in DSM datasets are well captured around the building edges as well as around and on the high vegetation. However, in both DSM the buildings appear larger than in corresponding HSI datasets. The holes in the 3K DSM (Figure 6.27c) are problematic for edge detection, because high edge probabilities are detected around holes. Moreover, the interpolated missing values in WV-2 DSM (Figure 6.27d) cause ramp-like edges. Therefore, the edge probabilities of these edges (due to holes or interpolated values) have poorer localisation. Different sizes of the buildings in each of the dataset pairs influences the values of the joint edge probabilities and their localisation.



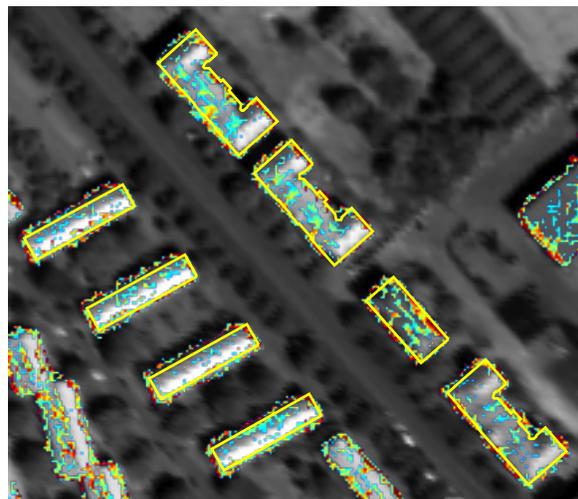
(a) Edge probability VNIR10.



(b) Edge probability SWIR10.

(c) Edge probability 3K DSM  $GSD = 1.0$  m.(d) Edge probability WV-2 DSM  $GSD = 1.0$  m.

(e) Joint edge probability VNIR10-3K DSM.

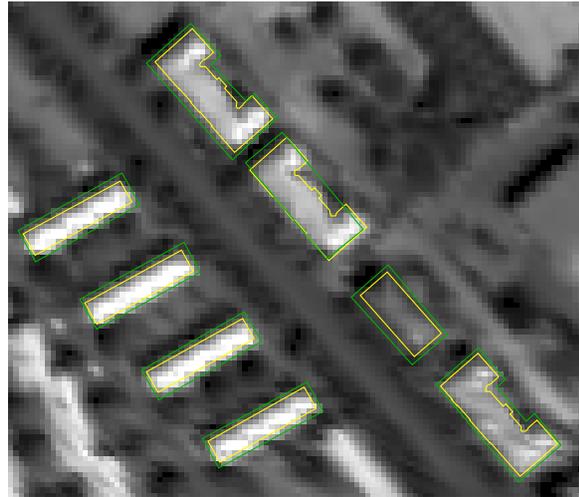


(f) Joint edge probability VNIR10-WV-2 DSM.

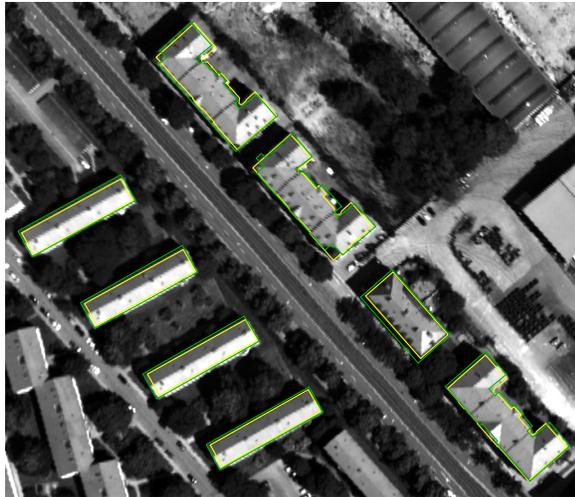
Figure 6.27: Edge probabilities for red roof tile material computed from HSI (Figures 6.27a and 6.27b), for height edges computed from DSM (Figures 6.27c and 6.27d), and joint edge probabilities (Figures 6.27e and 6.27f) according to the confidence probabilities. For visualisation reasons, only the joint edge probabilities around the objects in the *small area* with material under consideration are shown.



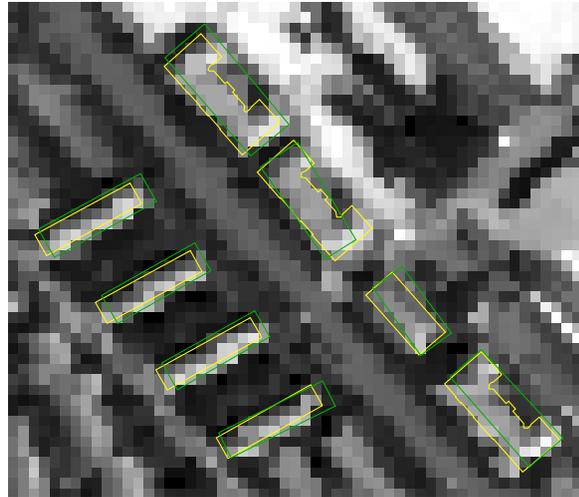
(a) Adjusted BP VNIR10.



(b) Adjusted BP SWIR20.



(c) Adjusted BP VNIR03.



(d) Adjusted BP HyMAP.

Figure 6.28: Adjusted BP for four HSI-3K DSM dataset pairs with  $GSD = \{0.3, 1.0, 2.0, 4.0\}$  [m] (*small area*). The adjusted BP are green and the ground truth BP in yellow.

For most of the dataset pairs, the values of  $\bar{p}(edge; t_{max})$  are as expected, i.e. the joint edge probabilities according to the  $\beta(t_{max})$  result in higher weights than joint edge probabilities according to the prior knowledge. Moreover, the HSI (Figure 6.26, black square) and DSM (Figure 6.26, cyan circle) edge probabilities have both a similar influence to the  $\bar{p}(edge; t_{max})$ . However, there are two exceptions to this general tendency. First, for the  $GSD = 4.0$  m, where the weights from only HSI have higher values. A possible explanation is that the DSM does not provide strong information about edge probabilities for the size of the objects under consideration. Second, for the  $GSD = 1.0$  m (but not  $GSD = 2.0$  m) and SWIR–DSM dataset pairs, where the weights from only DSM have higher values and are also significantly higher than from only HSI. The reason for the lower values of the  $\bar{p}(edge; t_{max})$  for SWIR10 and  $\pi_{HSI} = 1.0$  (e.g. Figure 6.27, right column) is related to the poorer localisation and/or detection of the edge probabilities in SWIR10 due to the resampling and remaining effect of the uncorrected roll movement. The roll movement of a platform to the acquired images causes smearing effect in across-track direction that varies between each line of the image.

On the whole, joint edge probabilities according to the confidence probabilities provide higher values for the weights than from a single dataset. They also compensate for lack of edge probability information

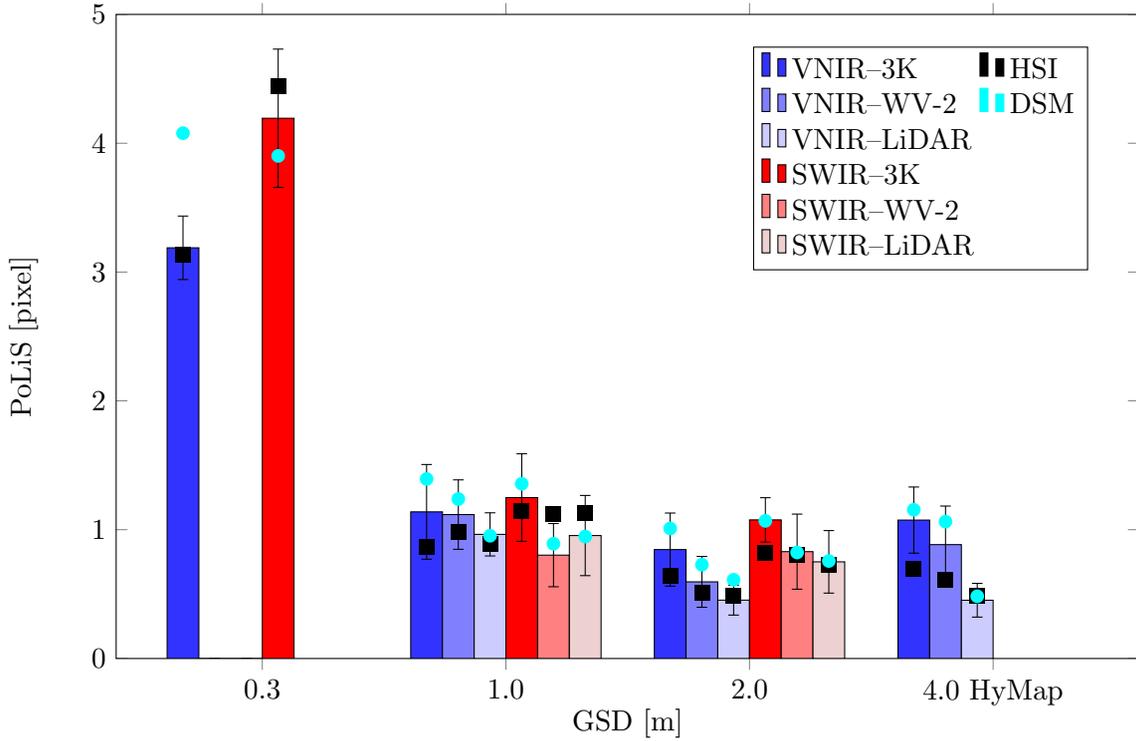


Figure 6.29: Evaluation of the BP adjustment (GM model,  $\sigma_\alpha^2 = 1^\circ$ ) for the *small area* by the PoLiS metric. Mean value of the PoLiS metric is shown. The colour of the colour bars corresponds to the HSI-DSM dataset pair, and the error bars (black) to the standard deviation of the PoLiS metric. The values can be compared to the PoLiS metric for the synthetic building example in Figure 6.17. The mean PoLiS metric for adjusted BP from a single dataset, i.e. HSI or DSM are represented by the black squares and cyan circles, respectively.

from only one dataset, i.e. none or inaccurate edge information in DSM dataset, due to e.g. a tree near the building roof and is compensated by the edge probability of a roofing material from HSI. When the ratio between the size of the object under consideration and  $GSD$  is small, the edge probability information from only HSI can provide better discrimination of the above-ground objects than from only DSM. In the case of on-ground objects, the DSM cannot provide the edge information of the objects, thus joining the edge probabilities from several datasets can be beneficial. There are mainly two concerns regarding the joint edge probabilities between multi-modal datasets. First one is the prerequisite of precise registration of the datasets. Second one is that the uncorrected errors (or the remaining uncorrected effects) in the input datasets are influencing the mapping of the objects, their edges, and localisation of the edges.

### **Small Area: Evaluation of the Adjusted BP**

All the approximate BP are adjusted by the GM model ( $\sigma_\alpha^2 = 1^\circ$  and  $\sigma_\alpha^2 = 10^\circ$ , Table 3), and the GH model. The adjusted BP are compared to the ground truth BP by computing two quality measures, i.e. PoLiS (Figure 6.29) and  $1 - qual$  (Figure 6.30). All the results are shown for the GM model  $\sigma_\alpha^2 = 1^\circ$ .

On the synthetic image (Figure 5.10a), the GH model performed best for rectilinear buildings (Section 6.2), followed by the GM model  $\sigma_\alpha^2 = 1^\circ$  and the GM model  $\sigma_\alpha^2 = 10^\circ$ . For a real RS image with only rectilinear BP, it is expected that the adjustment with the GH model will also perform best. A main reason for showing the results only for GM model  $\sigma_\alpha^2 = 1^\circ$  is that it is a trade-off between strict rectilinearity constraint (as in the GH model) and more data driven adjustment (as in the GM

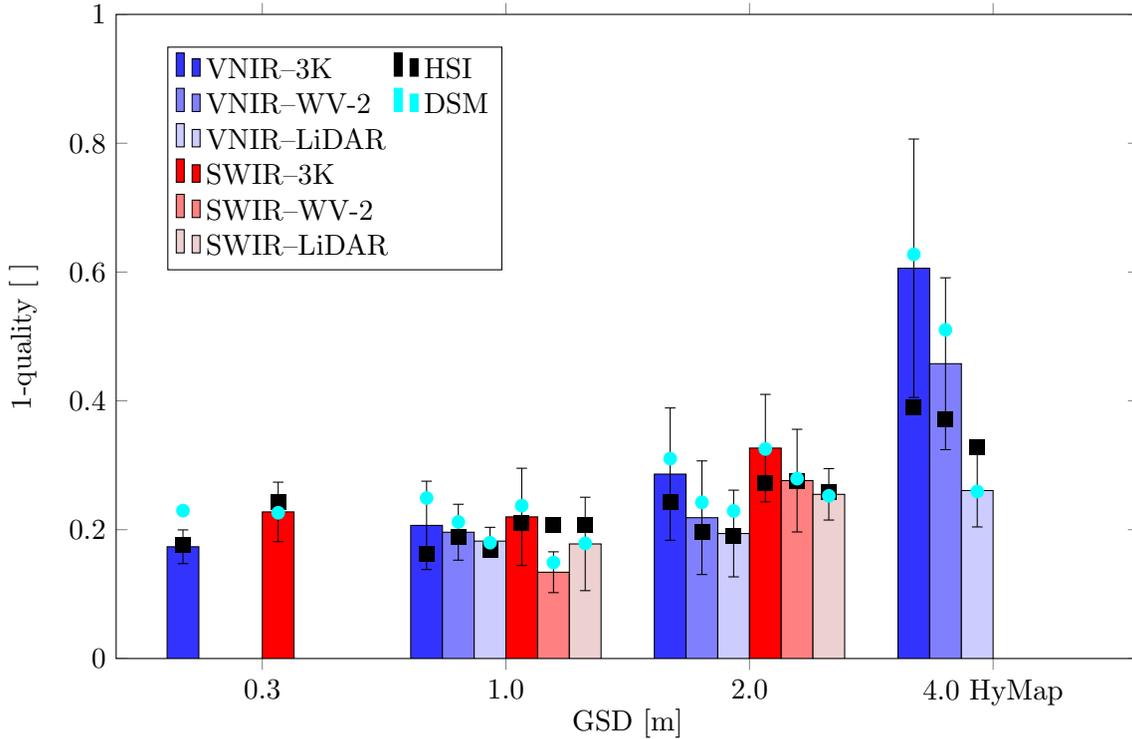


Figure 6.30: Evaluation of the BP adjustment (GM model,  $\sigma_\alpha^2 = 1^\circ$  for the *small area* by the quality rate. The mean value of the  $1 - qual$  measure is shown. The colour of the colour bars corresponds to the HSI–DSM dataset pair, and the error bars (black) to the standard deviation of the  $1 - qual$ . The values can be compared to the values of the synthetic building example in Figure 6.18. The mean  $1 - qual$  for adjusted BP from a single dataset, i.e. HSI or DSM are represented by the black squares and cyan circles, respectively.

model  $\sigma_\alpha^2 = 10^\circ$ ). Moreover, the difference between the results for both settings for the GM model is insignificant, and second, only the rectilinear BP are present in the scene of the *small area* (i.e. the  $\sigma_\alpha^2$  shall be kept small). On the synthetic image (Figure 5.10a), the GH model performed best for rectilinear buildings (Section 6.2), followed by the GM model  $\sigma_\alpha^2 = 1^\circ$  and the GM model  $\sigma_\alpha^2 = 10^\circ$ . For a real RS image with only rectilinear BP, it is expected that the adjustment with the GH model will also perform best. A main reason for showing the results only for GM model  $\sigma_\alpha^2 = 1^\circ$  is that it is a trade-off between strict rectilinearity constraint (as in the GH model) and more data driven adjustment (as in the GM model  $\sigma_\alpha^2 = 10^\circ$ ). Moreover, the difference between the results for both settings for the GM model is insignificant, and second, only the rectilinear BP are present in the scene of the *small area* (i.e. the  $\sigma_\alpha^2$  shall be kept small).

The buildings in the *small area* and the synthetic example in Section 6.3 are comparable with regard to complexity of the BP and the level of extractable details. Thus, the evaluation results for the adjustment of the BP in the *small area* (Figures 6.29 and 6.30) can be compared to the evaluation results of the synthetic image (Figures 6.17 and 6.18).

Figure 6.28 shows the adjusted (green) and ground truth (yellow) BP for four different HSI–3K DSM dataset pairs of different GSD. Due to the roof overhangs the adjusted BP are in all cases larger than the ground truth BP. The difference is more prominent for the datasets with larger GSD. For dataset pair with  $GSD = 4.0$  m, (Figure 6.28d), the false orientation of one building can be observed. For dataset pair with  $GSD = 0.3$  m, (Figure 6.28c) the adjusted BP are detailed, however all the details do not correspond to the real state.

The PoLiS metric is computed between the ground truth and adjusted BP. The mean value and standard deviations (error bars) of the PoLiS metric [pixel] (Figure 6.29) and  $1 - qual$  [] (Figure 6.30) for the BP of the *small area* are computed. The PoLiS metric varies between about 0.5 pixel and 1.1 pixel for  $GSD \geq 1.0$  m. For the same GSD it is better for the HSI–LiDAR DSM dataset pairs, than for other DSM. This corresponds to the expectation, because the quality of the LiDAR data is better in comparison to the quality of the SGM DSM.

The quality of the adjusted BP from VNIR–DSM dataset pairs is expected to be better than the quality of the ones from SWIR–DSM dataset pairs. The results show the expected tendency for  $GSD = 2.0$  m, but not for all dataset pairs with the  $GSD = 1.0$  m. For SWIR datasets with  $GSD = 2.0$  m, a larger number of simpler BP (Figure 6.25) is selected, than for the the VNIR datasets, which causes larger PoLiS values. Moreover, the imperfect coregistration between the ground truth and the adjusted BP adds to the error budget of the PoLiS metric.

For  $GSD = 0.3$  m, the PoLiS metric value is high, i.e. about 3.2 pixel and 4.2 pixel for VNIR and SWIR, respectively. These values correspond to about 1.1 and 1.4 m, which is about the same as for the dataset pairs with  $GSD = 1.0$  m. The values of about 1.0 m are expected due to the roof overhangs. Thus, the absence of roof overhang in the ground truth BP also adds to the error budget of the PoLiS metric.

Values of  $1 - qual$  and their standard deviations increase with the increasing GSD values. Exceptions are the  $1 - qual$  values for the dataset pairs with  $GSD = 0.3$  m, which are comparable to the dataset pairs with  $GSD = 1.0$  m. In fact, the  $1 - qual$  is best for  $GSD = 1.0$  m dataset pairs and is below 0.2. The quality of the BP from HSI–LiDAR DSM dataset pairs is at least as good as for HSI–SGM DSM dataset pairs. The increasing standard deviation of the  $1 - qual$  indicates the larger variation of the  $1 - qual$  for adjusted BP. This can be explained by the simpler approximate BP (all Level 1, Figure 6.25). The HyMAP–DSM dataset pairs shows insufficient quality for the purpose of BP extraction. For the buildings of the considered dimensions and level of detail, the BP extraction and adjustment from DSM shall only be carried out if combined with good quality DSM or only from DSM.

#### 6.4.4 Results and Discussion of the Experiment on the *Large Area*

The proposed workflow for BP extraction is tested on three pairs of HSI–WV-2 dataset pairs of the *large area* (Figure 6.21, blue). The HSI datasets are VNIR20 and SWIR20 HySpex HSI with  $GSD = 2.0$  m and HyMap HSI with  $GSD = 4.0$  m. These HSI are selected because they exhibit the whole *large area*. Four roofing materials are considered for the experiments on the *large area* (Figure 6.22), these are red roof tiles (red), copper (cyan), concrete (grey), and metal (blue).

The approximate BP are adjusted by the GM model with  $\sigma_\alpha^2 = 1^\circ$  setting (Table 3) and evaluated. The evaluation of the BP from only HSI (all materials, denoted hereafter shortly as all), only WV-2 DSM (denoted shortly as WV-2), and joined from both datasets (denoted hereafter shortly as joined) are compared on per-scene and per-object level. The BP from HSI datasets are excluded from joined BP, if their average height is below 2 m. In all the figures showing images, the adjusted BP are shown: in green for BP from WV-2 DSM and in a colour of the corresponding material (see Figure 6.22) for BP from HSI.

##### ***Large Area: BP Selection***

The number of extracted BP (num BP) is summarised in Table 6. The actual extractable level of BP for a given  $GSD$  can only be defined subjectively by visual inspection. Therefore, the level of selected approximate BP is not discussed for the *large area* as it is already discussed for the *small area*.

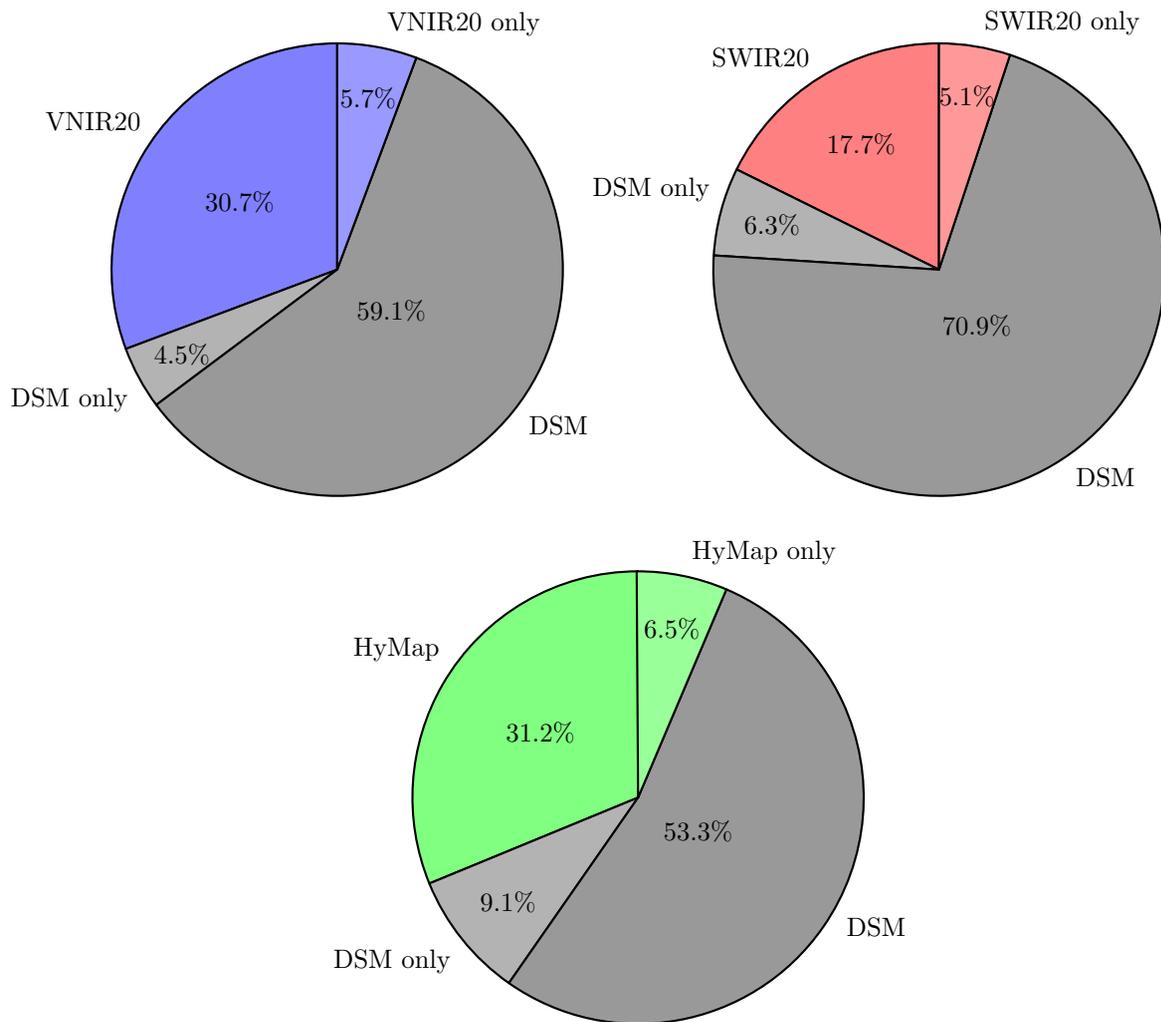


Figure 6.31: Percentage of BP selected from DSM and HSI for VNIR20–WV-2 DSM (upper left, blue), SWIR20–WV-2 DSM (upper right, red), and HyMAP–WV-2 DSM (down, green) dataset pairs. Some of the BP are extracted only from one of the datasets and are marked by the word “only”. The total number of BP is given in Table 6.

<i>Per-scene</i>	<b>VNIR20</b>		<b>SWIR20</b>		<b>WV-2</b>	<b>HyMap</b>		<b>WV-2</b>
	all	joined	all	joined	2 m	all	joined	4 m
<b>num BP</b>	148	88	109	79	72	111	77	64
<b>Completeness</b>	0.71	0.89	0.58	0.88	0.88	0.54	0.83	0.81
<b>Correctness</b>	0.77	0.76	0.76	0.75	0.75	0.60	0.68	0.68
<b>Quality rate</b>	0.59	0.69	0.49	0.68	0.68	0.40	0.60	0.59

Table 6: The number of BP and per-scene evaluation of the HSI–WV-2 DSM dataset pairs (*large area*). The HSI dataset pairs are VNIR20, SWIR20, and HyMAP. The total number (num) of the BP is given. The total number of ground truth BP is 85. The matched rates completeness, correctness, and quality rate are computed on the basis of  $TP$ ,  $FN$ ,  $FP$  detected areas (and not the number of BP). The quality rate is in all cases higher for the joined BP from both datasets, than for the BP extracted from only HSI or only DSM. For every  $GSD = \{2.0, 4.0\}$  [m] and for every computed matched rate (completeness, correctness, quality rate) its value indicating the best performance is marked green.

The total number of ground truth BP is 85 and is close to the total number of joined BP (Table 6). The number of BP from WV-2 DSM datasets is the smallest, whereas the number of BP extracted from HSI is significantly higher. The main reason for this difference is that in HSI a building regions are



Figure 6.32: Adjusted BP extracted from WV-2 DSM (green) with  $GSD = 2.0$  m (*large area*). Due to the roof overhangs, which are not included in the ground truth BP (yellow), the adjusted BP are in general larger than the ground truth BP.

defined differently than in DSM. One building region in the HSI is considered as a connected area in a material map of one roofing material and in the DSM, one building area is considered as a connected above-ground area (without high vegetation). In addition, the roofs can consist of more than a single material, or several buildings with different roofing materials are next to each other.

Figure 6.31 shows the percentage of the BP selected from HSI and corresponding WV-2 DSM dataset. The word “only” next to the name of the dataset means that some of the BP are extracted only from this dataset. For all three dataset pairs, the majority of the buildings are selected from WV-2 DSM. This result is similar to the result on the *small area* (Figure 6.24). The largest percentage of the BP selected from the WV-2 DSM is for the SWIR20–WV-2 DSM dataset. One possible explanation is that the reference spectra for metal has no significant feature in the SWIR region and the buildings with this type of metal roofing could not be extracted. For all three dataset pairs, some BP are extracted from only HSI or only DSM.

The approximate BP, selected from both datasets, are adjusted by the GM model  $\sigma_{\alpha}^2 = 1^{\circ}$  and evaluated on per-object level (PoLiS,  $1 - qual$ ) and per-scene (matched rates). The per-object evaluation measures are also shown for BP of each of four considered roofing materials, all BP (all materials together), and joined BP from HSI–DSM dataset pairs.

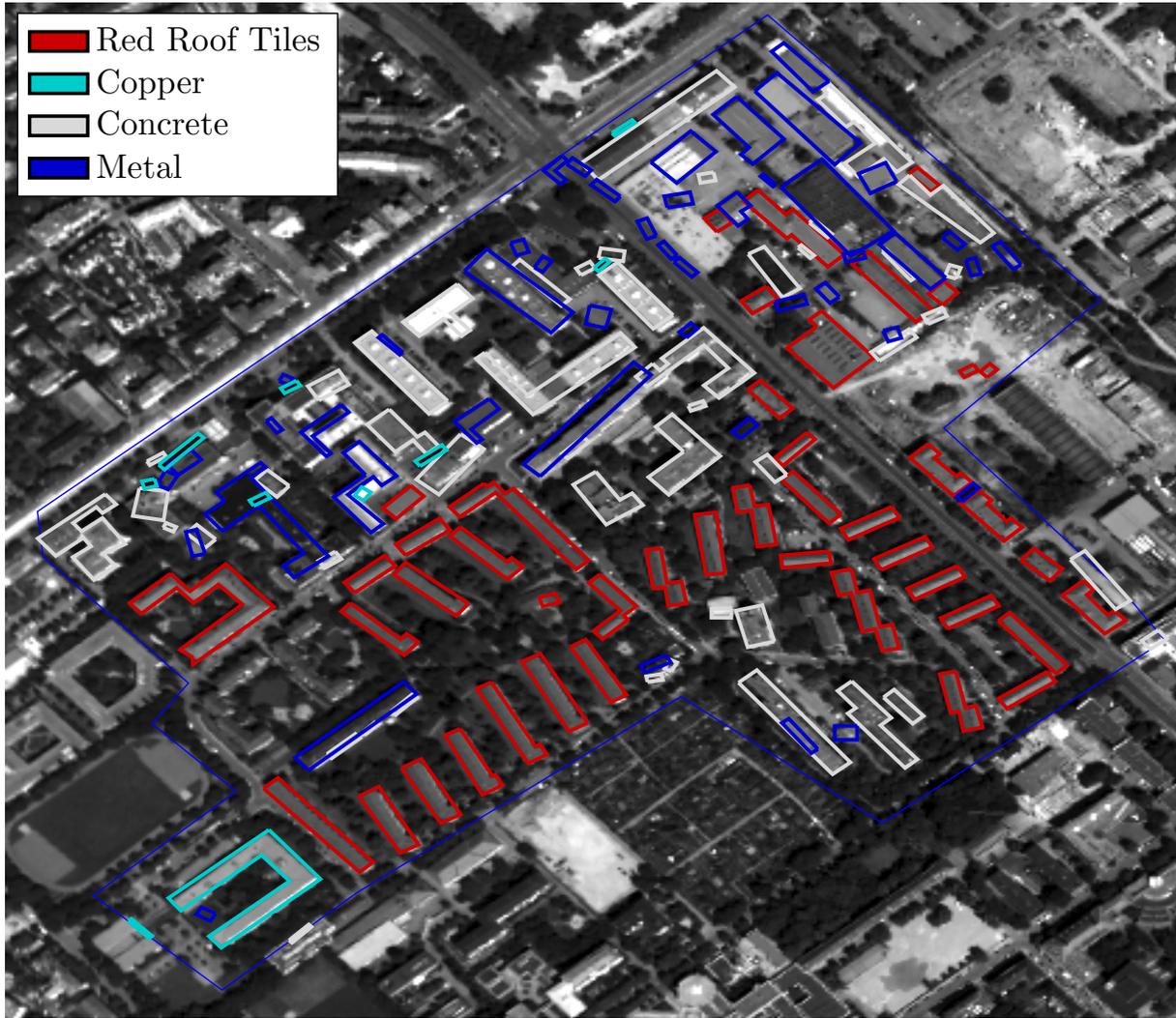


Figure 6.33: Adjusted BP from VNIR20 (*large area*) for four materials (Figure 6.22), these are red roof tiles (red), copper (cyan), concrete (grey), and metal (blue). The materials red roof tiles and copper have the most discriminant reference spectra, whereas the BP which correspond to the concrete and metal reference materials have large number of false detections.

<i>Per-object</i>		VNIR20		SWIR20		WV-2	HyMap		WV-2
		all	joined	all	joined	2 m	all	joined	4 m
PoLiS [pixel]	mean	1.45	1.39	1.73	1.38	1.53	1.38	1.14	1.08
	std	1.27	0.58	2.21	0.67	0.73	0.84	0.45	0.34
$1 - qual$ [ ]	mean	0.38	0.31	0.40	0.31	0.32	0.48	0.39	0.40
	std	0.18	0.09	0.22	0.10	0.10	0.18	0.11	0.10

Table 7: Per-object evaluation of the HSI-WV-2 DSM dataset pairs by PoLiS and  $1 - qual$  quality measures (*large area*). The HSI dataset pairs are VNIR20, SWIR20, and HyMAP. The quality measure is computed if the adjusted BP and ground truth BP overlap for at least 10% (Subsection 6.4.2). For every  $GSD = \{2.0, 4.0\}$  [m] and for every computed quality measure (PoLiS,  $1 - qual$ ) its value indicating the best performance is marked green.

### **Large Area: Evaluation of the Adjusted BP**

Figure 6.32 shows the adjusted BP from WV-2 DSM with  $GSD = 2.0$  m (green) and the ground truth BP (yellow). The per-scene evaluation (Table 6) yields the quality rate of 0.68 (best 1.00). It can be observed that some smaller buildings were not detected from the WV-2 DSM (*FN*), or some parts of

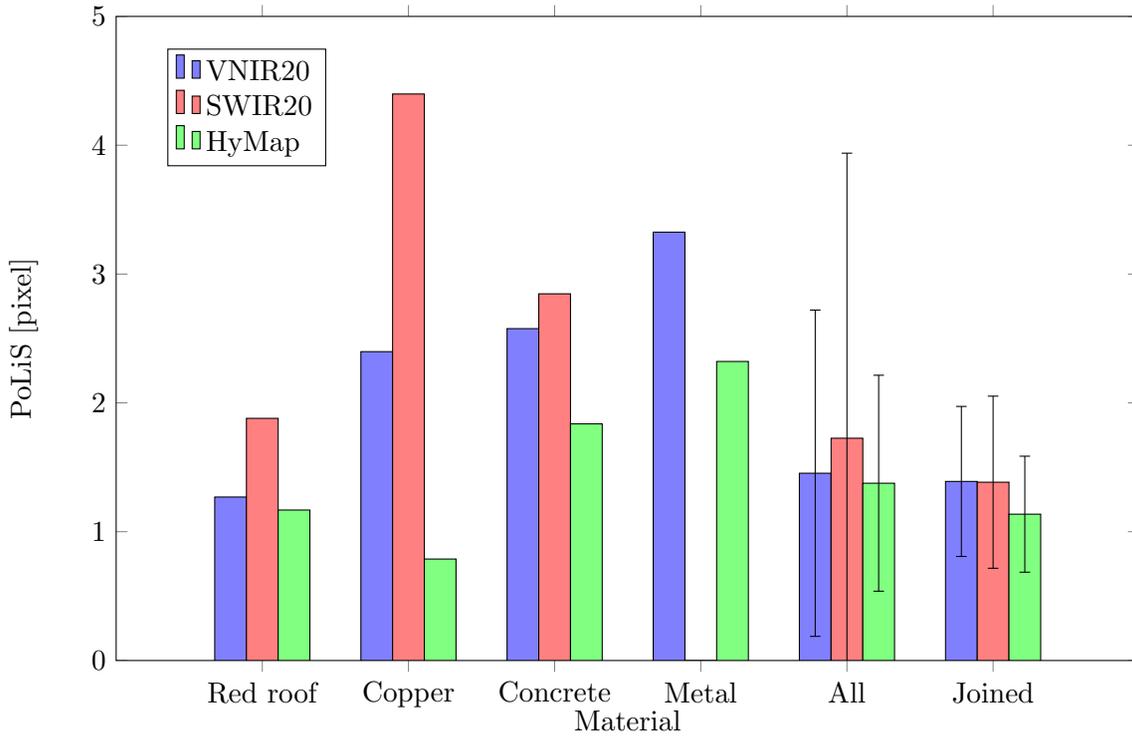


Figure 6.34: Evaluation of the BP adjustment (GM model,  $\sigma_\alpha^2 = 1^\circ$ ) for the *large area* by the PoLiS metric. Mean values of the PoLiS metric weighted by the area of the BP are shown. The evaluation is carried out for BP according to the selected material (Figure 6.22), all materials, and joined BP from HSI-WV-2 DSM dataset pair. Red roof is short for red roof tiles. The bar for metal and SWIR20 is not shown, because no metal objects were extracted. One possible explanation is that the metal reference spectra has no discriminative features in the SWIR region in comparison to other reference spectra considered (Figure 6.22). The error bars (black) represent the the standard deviation of the PoLiS metric and are shown only for all and joined BP.

the larger buildings were wrongly detected (*FP*). Due to the roof overhangs, which are not included in the ground truth BP (yellow), the adjusted BP are in general larger than the ground truth BP. The same finding is drawn from experiments on the *small area* (Figure 6.28).

Figure 6.33 shows the adjusted BP from VNIR20 for four selected roofing materials (Figure 6.22), i.e. red roof tiles (red), copper (cyan), concrete (grey), and metal (blue). The materials red roof tiles and copper have the most discriminant reference spectra, whereas the BP, which correspond to the concrete and metal reference materials, have in comparison larger number of false detections. This trend can be observed by comparing the heights of the column bars in Figures 6.34 and 6.35, showing PoLiS and  $1 - qual$  quality measures, respectively. It shall be noted that the concrete is not only roofing, but also pavement material, which contributes to the false detection. False main orientation of the BP is more prominent for metal and concrete BP than for red roof tiles BP, copper BP, and also BP from WV-2 DSM (Figure 6.32).

Per-object evaluation results are summarised in Table 7 and are also shown for all four BP considered roofing materials, all BP, and joined BP (Figures 6.34 and 6.35). For VNIR20–WV-2 DSM and SWIR–WV-2 DSM dataset pairs (both  $GSD = 2.0$  m) the PoLiS and  $1 - qual$  quality measures both show that the joined BP yield better results than from any of the single datasets (Table 7). On the contrary, for HyMAP–WV-2 DSM dataset ( $GSD = 4.0$  m) the BP from WV-2 DSM only yields better result than

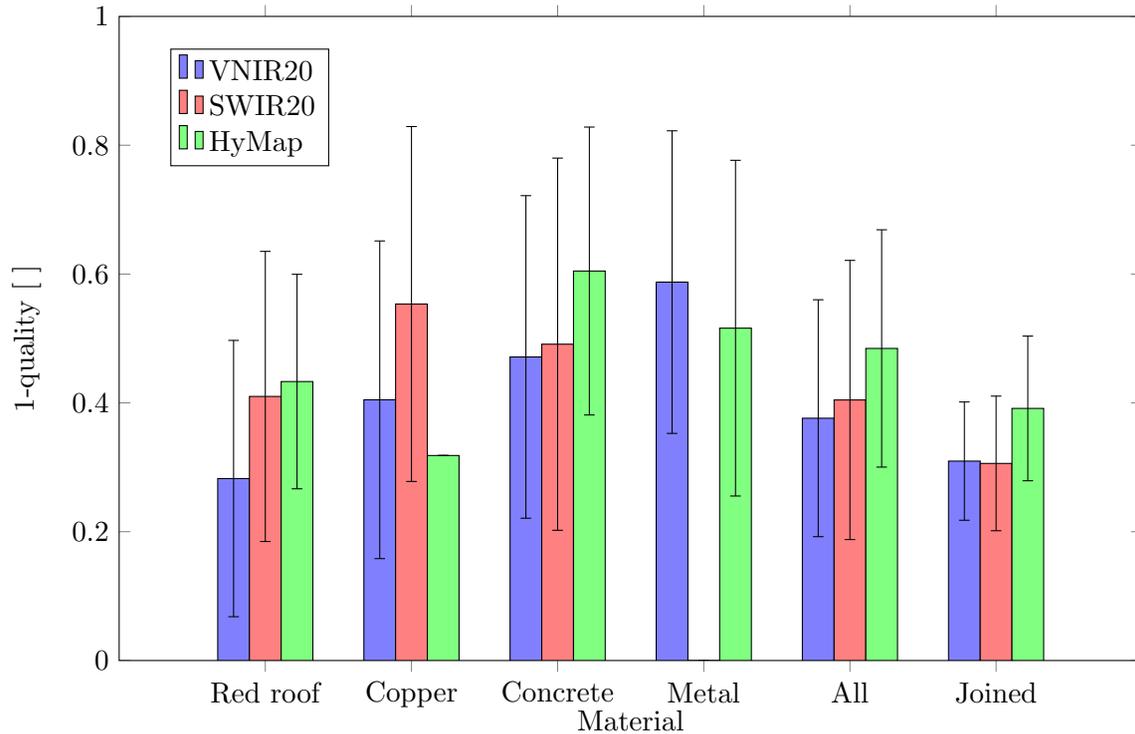


Figure 6.35: Evaluation of the BP adjustment (GM model,  $\sigma_\alpha^2 = 1^\circ$  for the *large area* by the quality rate. Mean values of the  $1 - qual$  measure, weighted by the area of the BP are shown. The evaluation is carried out for BP according to the selected material (Figure 6.22), all materials, and joined BP from HSI-WV-2 DSM dataset pair. Red roof is short for red roof tiles. The error bars (black) represent the standard deviation of the  $1 - qual$  metric. The error bar for the Copper in HyMAP dataset is not plotted, because only one Copper BP is selected and could therefore not be computed.

from joined BP. Only the HyMAP dataset is of insufficient quality for the purpose of BP extraction. This finding is in line with the finding of the test on the *small area* (Figures 6.29 and 6.30).

The values of PoLiS metric for joined BP are about 1.4 pixel for all dataset pairs. This value is expected due to the roof overhangs, which are not in the ground truth BP and the errors due to the *FN* and *FP* areas of the BP. These *FN* and *FP* areas are one reason for large standard deviations of the PoLiS metric. If the minimum required overlap between the extracted and ground truth BP would be set higher (than 10 %) the values of PoLiS and  $1 - qual$  measures would decrease as well as their standard deviations.

The quality measure of each single material can be lower for one single material, i.e. red roof tiles BP from only VNIR20 or HyMAP dataset are of better quality (compare height of colour bars in Figures 6.34 and 6.35) than from joined BP. This can be observed by comparing Figures 6.32, 6.33, and 6.36, where e.g. some of the red roof BP are selected as simpler level of BP from WV-2 DSM, rather than higher level of BP from the VNIR10. This is a consequence of automatic selection of the joined BP (Equation (74) and (73)). The imperfect automatic selection of the joined BP influences per-object and per-scene evaluation measures.

Figure 6.36 shows per-scene evaluation of the joined BP from VNIR20–WV-2 DSM dataset pair, where *TP* (green), *FN* (red), *FP* (purple), and *TN* (white) detected areas are shown. The majority of the BP are correctly extracted (green), whereas there are some larger parts of the BP detected as false positive areas (purple). This is either a consequence of poorly extracted building regions or simplification of

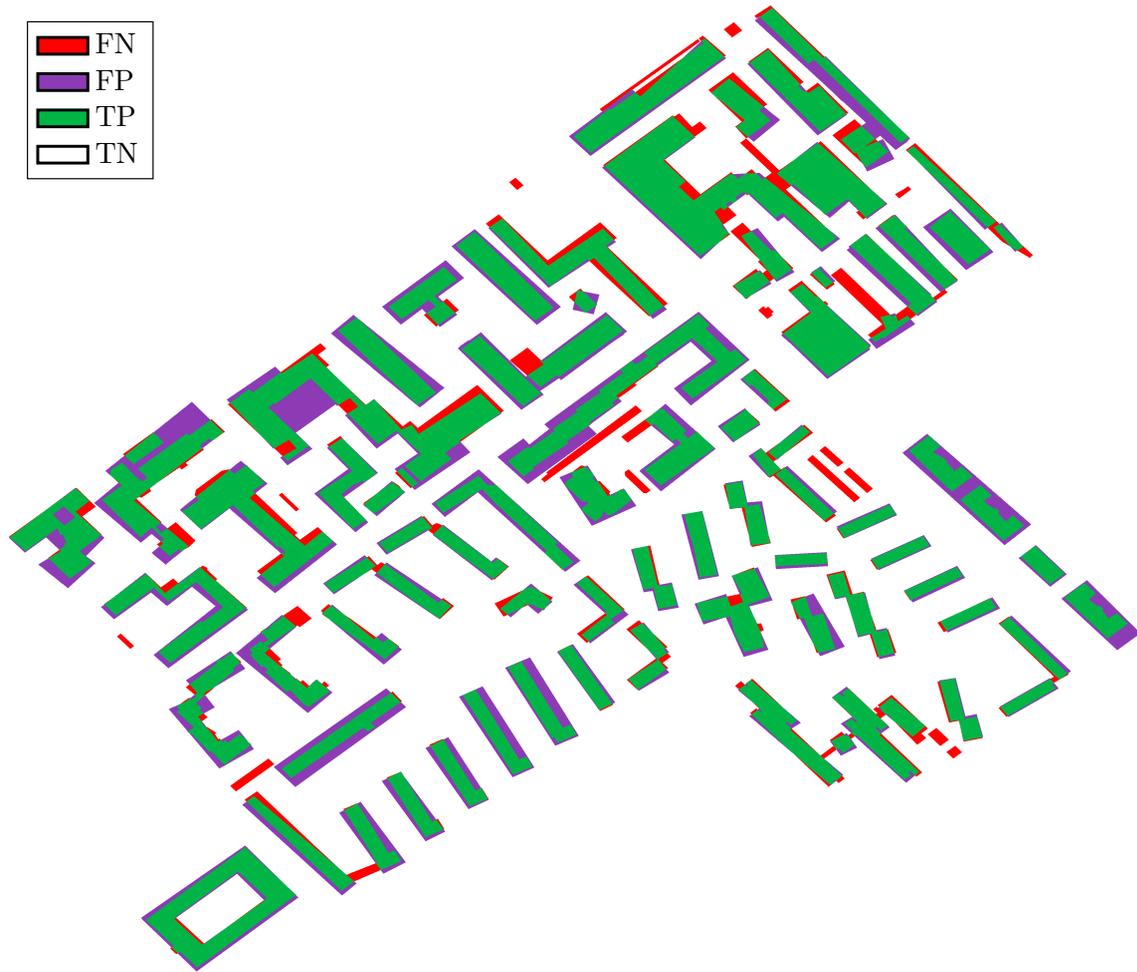


Figure 6.36: Per-scene evaluation of the joined BP from the VNIR20–WV-2 DSM dataset pair. *TP* (green), *FN* (red), *FP* (purple), and *TN* (white) detected areas. It can be observed that some buildings are not extracted (red), which is a consequence of the preprocessing steps, e.g. not present in the building regions.

the BP in the iMBR method. It can be observed that some BP are not extracted (red), which is a consequence of the preprocessing steps, e.g. not present in the building regions.

The per-scene evaluation by matched rates summarises the overall quality of the adjusted BP (Table 6). In the per-object evaluation only the BP, which overlap with the ground truth BP for at least 10% (Figure 6.36), are evaluated, whereas the per-scene evaluation accounts also for BP which were not (*FN*) or were completely falsely (*FP*) extracted. The per-object matched rates show the best results for the VNIR20–WV-2 DSM dataset pair, followed by the SWIR20–WV-2 DSM (both with  $GSD = 2.0$  m) and HyMAP–WV-2 DSM dataset pair. The completeness values are for all dataset pairs higher than correctness, because there is a relative small number of (*FN*) detected areas in comparison to the *FP* detected areas. Moreover, the quality rate values are comparably even lower, because  $ar(FN)$  and  $ar(FP)$  are both accounted for in the calculation.

The number of *FN* detected buildings is smaller for joined BP, because some BP are detected only in one dataset of the dataset pair. Consequently, the completeness for joined BP is higher (better) for all dataset pairs. As previously discussed in per-object evaluation, some BP are selected as simpler level of BP from WV-2 DSM, rather than higher level of BP from the HSI. This imperfect selection of the BP negatively influences all matched rates.

### 6.4.5 Summary of Experiments on RS Images

To join all proposed methods in one workflow is feasible and the resulting adjusted BP are of good quality. The following conclusions can be drawn from the experiments on RS images for testing the proposed methods (Figure 6.1, bold boxes) on the *small area* and the whole workflow (Figure 6.1) on the *large area*. Experiments on the *small area* are carried out for 17 HSI–DSM dataset pairs, and for the *large area* on three dataset pairs of different GSD. The conclusions are grouped according to the processing steps.

Preprocessing (*small & large area*):

- ◇ The material maps are calculated for four roofing materials, whose reference spectra are manually collected. The prominent spectral features of a roofing material are crucial for building region extraction on the basis of the material maps. Thus, the material map of e.g. red roof tiles, discriminate well the roofs consisting of this material from the background.
- ◇ The NNLS unmixing with few (five) reference spectra provides good results for material maps of roofing materials. Due to the uncorrected systematic errors of the HySpex image the selected spectra exhibit noise in the spectral domain, which influences the unmixing results. However, if available a complete set of reference materials, present in a scene shall be used for unmixing. Alternatively, building regions from HSI could be extracted by using any other classification technique.
- ◇ A building region, which is not detected in any of the datasets (nDSM, material map), cannot be recovered in later processing and negatively influences the per-scene evaluation, i.e. more *FN* detected buildings or building parts.
- ◇ The selection of BP, when one-to-many cardinality between the two sets of BP exists, is not trivial (Subsection 3.3.1). The proposed cost function for approximate BP selection from dataset pairs (Equations (73) and (74)) prefers smaller number of corresponding BP. Consequently, larger but not necessary more accurate BP are selected. This negatively influences the quality of the selected and also adjusted BP.

BP creation and selection (*small & large area*):

- ◇ For all tested dataset pairs, more BP are selected from DSM than from HSI. First, this difference is more prominent for the SWIR–DSM than from VNIR–DSM datasets. Second, for smaller GSD, the approximate BP extraction relies stronger on material information from HSI than on height information from DSM, while for larger GSD, more BP are selected from the DSM datasets.
- ◇ The main orientation of the approximate BP is for most of the buildings in the scene well extracted according to the visual interpretation. However, for some smaller buildings, the main orientation is falsely detected (see also Subsection 6.3.4).
- ◇ For larger GSD a simpler BP (lower level) is preferred over a more complex one (see also Subsection 6.3.3).

Edge probability detection and fusion (*small area*):

- ◇ The joint edge probabilities according to the confidence probability ( $\pi \propto \beta(t_{max})$ ) result in higher weights than joint edge probabilities according to the prior knowledge ( $\pi$  Prior). Thus, a better discrimination between above-ground objects is achieved.
- ◇ The joint edge probabilities compensate for lack of edge probability information from only one dataset.
- ◇ Edge probabilities computed from the HSI or from DSM have both similar influence to the mean edge probability ( $\bar{p}(edge; t_{max})$ ).

- ◇ The first main limitation of joining the edge probabilities is the imperfect coregistration between the datasets. Even if the edge probabilities are well localised in all of the imperfectly coregistered datasets, the joint edge probabilities lose to some extent their localisation accuracy. The second main limitation is that the same object is mapped differently in different datasets, i.e. the objects vary in shape and size. For instance, in SGM DSM, the buildings appear larger in an image, and their edges are slightly curved. In HSI, the edges of the buildings appear straighter, but some parts of the façades can be seen due to the slightly oblique view of the sensor for the pixels further from nadir. Moreover, remaining systematic errors have also influence on the mapping of the objects in HSI.
- ◇ The joint edge probabilities are more reliable than edge probabilities from any of the single datasets, under assumptions that the datasets are coregistered with a sub-pixel precision, the objects in the scene are accurately mapped to the images, and remaining systematic errors are negligible. If these prerequisites are not met, the localisation of the joint edge probabilities is compromised. Nevertheless, the joint edge probabilities still carry additional information.
- ◇ Edge probability fusion is recommended when the considered datasets have similar GSD and are of comparable quality. If this is not the case, the edge probability fusion shall be carried out according to the prior knowledge, and the dataset of a poorer quality and with less spatial details shall be weighted lower.

Per-object evaluation of the adjusted BP (*small & large area*) is quantified by computing PoLiS and  $1 - qual$  measures. For the *small area*, the relation of these measures with respect to the GSD is investigated. For the *large area* also per-scene matched rates are computed.

- ◇ All the adjusted BP from the images are larger than the ground truth, due to the roof overhang of about 1 m. Missing roof overhangs resembles in the values of evaluation measures. Imperfect coregistration between the dataset pairs and the ground truth adds to the error budget of the evaluation measures.
- ◇ LiDAR DSM is of better quality than the SGM DSM, thus, also the PoLiS metric and  $1 - qual$  values show better results for HSI–LiDAR DSM dataset pairs of the same GSD.
- ◇ In general, the  $1 - qual$  values increase with the increasing GSD. More detailed adjusted BP are extracted from images with smaller GSD and thus their quality is better.
- ◇ The influence of the uncorrected systematic errors, e.g. imperfect DSM used for orthorectification (see Subsections 2.1.2 and 2.2.2) limits the quality of adjusted BP ( $GSD = 0.3\text{ m}$ , *small area*). Thus, the edge information is unreliable, and the weighted adjustment does not improve the BP. The effect of systematic error is also observable by comparing the quality measures of the VNIR and SWIR datasets of the same GSD. The VNIR sensor has a smaller IFOV than the SWIR (Table 2), but both images acquired by these sensors were resampled to the same GSD. Thus, the resampling inaccuracy in the SWIR images is larger and the adjusted BP from SWIR images are of poorer quality than from VNIR images of the same GSD.
- ◇ If no high quality DSM is available, the adjusted BP from dataset pairs with  $GSD = 4.0\text{ m}$ , are of insufficient quality. The ratio between an object (or object's detail) and the GSD must be large enough, otherwise the object (or object's detail) are not or not reliably extracted and adjusted. For example, the smallest dimension of an object (or object's detail) should be at least 4 pixel.
- ◇ Per-object and per-scene evaluation (*large area*) shows that joining information from multi-modal images (i.e. from HSI and DSM) results in better quality of the adjusted BP. The joined BP are of better quality than from just one of the both datasets. Moreover, also the values of per-scene matched rates are improved.

## 7 Conclusions and Outlook

In this Chapter, the conclusions about the work presented in this thesis are drawn (Section 7.1). Possible extensions and alternative approaches to the proposed methods are outlined, and topics for further research are emphasised (Section 7.2).

### 7.1 Conclusions

As stated in the introduction, the information value from HSI, DSM, and the edge information inherent in both datasets is worth more than thousands of “words”. The aim of this thesis is to exploit these pieces of information in order to extract building polygons. Three main objectives and a minor one are set in Chapter 1 to pursuit this aim.

The first objective is fusion of HSI and DSM based on edge information. Edge probability computation results in edge probability values for every pixel in an image, unlike edge detector algorithms, which provide set of edge pixels with their strengths. For HSI and DSM, the edge probabilities are computed on the basis of material difference in abundance maps and height differences in DSM, respectively. The edge probability computation is applicable on other images than HSI and DSM. However, it is sensitive to any image edges, and captures also image edges due to missing data and uncorrected systematic errors.

It has been shown that edges in a RS image appear at different scales and that analysis of the image at a set of scales contributes to more reliable edge probability values. A scale selection using maximal edge probability over all scales leads to better results than scale selection on the basis of confidence probability proposed by Marimont and Rubner (1998). The joined edge probabilities compensate for the missing edge information in one of the datasets. The mixing coefficients for joining edge probabilities are chosen according to the prior knowledge about the datasets or by relation between confidence probabilities. The former approach requires knowledge of the quality of the datasets, whereas the latter is a fully automatic data driven fusion. The fully data driven fusion performs better, if there is no significant quality difference between the datasets. A prerequisite for edge probability fusion is accurate coregistration between the datasets. The number of datasets for fusion is theoretically unlimited.

The second objective is to define the mathematical model for adjustment of rectilinear BP. Two models are defined, first the GH model with a strict rectilinearity constraint, and the second GM model, in which the rectilinearity constraint is relaxed through weights of the groups of observations. For both adjustment models, an approximate BP is required as an input, whose position, size, and orientation is estimated during the adjustment. Moreover, both models allow for the joined edge probabilities to be included as weights in the adjustment. The adjusted BP by the model with strict rectilinearity constraint are more accurate only if all the buildings under considerations have a rectilinear shape. Otherwise, the model with relaxed constraint balances better between fitting of the BP to the data and rectilinearity constraint. It allows for more general shapes of the BP, but influences the accuracy of adjusted BP. Under an assumption that the approximate BP is correct and rectilinear, it is adjusted with sub-pixel precision, i.e. the PoLiS metric of the adjusted BP employing the first and the second model are 0.23 pixel and 0.38 pixel, respectively.

An approximate BP is required as an input for both adjustment models. It is created by the iMBR method, which adds and subtracts MBR to delineate binary building regions by a polygon. The main novelty of the iMBR method, in comparison to other MBR-based methods for BP extraction, is that it automatically selects a level of complexity of the approximate BP by analysing the cost function. The

level of complexity of the approximate BP is steered by a parameter related to the size of a minimum building part.

The third objective is to define a metric for comparison of polygons, which quantifies their dissimilarity in shape and positional accuracy. The proposed PoLiS metric is a metric in mathematical sense, it compares polygons with different numbers of vertices on polygons' edges, and is insensitive to the number of vertices on polygons' edges. It has been shown through an example that it is monotonic, and has a nearly linear response to small changes in translation, rotation, and scale. PoLiS metric and the commonly used quality rate are both symmetric and they both account for shape and geometric accuracy differences. In contrast, the units of the PoLiS are absolute, e.g. m, pixel, whereas the units of the quality rate are relative. Thus, the PoLiS metric provides a different way of quantifying the dissimilarity between polygons.

Finally, the minor objective is to join all the proposed methods in one workflow, i.e. joined edge probabilities from real RS images are used as weights in the adjustment of the approximate BP created by the iMBR method. The workflow is tested on 17 HSI–DSM dataset pairs, with four different GSD. The joined information from multi-modal images results in better quality of adjusted BP, than relying only on information from one dataset. For instance, even for datasets with  $GSD = 4.0$  m the completeness, correctness and quality rate values of extracted BP are better than 0.83, 0.68, and 0.60, respectively. The PoLiS metric value is better than 1.38 pixel for the BP from datasets with  $GSD \geq 1$  m. Inaccuracies of the images, such as holes in DSM or imperfect DSM for HSI orthorectification, are influencing the accuracy and localisation of edge probabilities and consequently also the accuracy of adjusted BP. The values of all evaluation measures are impacted by the imperfect ground truth polygons, which do not include roof overhangs.

## 7.2 Outlook

Topics for further study on urban object extraction from HSI and DSM regarding the methods presented in this thesis are listed below.

- ◇ Utilisation of the scale space for edge detection in RS images has a potential to provide more reliable and accurate results than detection of edges at a single scale. Edge probability computation to support object extraction is not limited to buildings, but can be extended to other objects, such as streets, trees, water bodies, and others.
- ◇ The relief displacement in the orthorectified images can be estimated and incorporated in the edge probability computation with the view of achieving better localisation of the edges.
- ◇ Next to the edge detection in scale space, also features like junctions and blobs can be detected (Lindeberg, 1994). E.g. blob detection in scale space can be used to extract tree crowns from RS imagery.
- ◇ Stereo matching methods, such as SGM, allow for incorporation of edge priors in the minimisation function. The main reason to use such priors is to overcome over-smoothing on the boundaries of the objects, such as buildings. Thus, the edge probabilities computed on the basis of spectral properties can be used as a prior with the goal to achieve sharper edges in the DSM.
- ◇ Both mathematical models for rectilinear buildings use fixed assignment of the boundary points to the edges of an approximate BP. This fixed assignment limits the range in which the parameters of the approximate BP, e.g. size, position, and main orientation, are adjusted. In a case when the approximate BP is not close enough to the actual building outline the fixed point assignment prevents more accurate adjustment of the BP. Two possible solutions are listed below.

- More flexible assignment of the boundary points to the edges of the BP by re-assignment of the boundary points in every iteration of the adjustment.
- More accurate extraction of the approximate BP. The iMBR method presented in this work uses a cost function to automatically select a level of approximate BP. The cost function is computed on the basis of RMSE between the building region and the BP and the level of complexity of the BP. The cost computation can be extended by including the sub-levels of BP, i.e. each MBR at one level gets assigned a cost separately.
- ◇ The PoLiS metric is defined with the goal of comparing building polygons, but is suitable for comparison of any polygons, where positional accuracy and shape dissimilarities have to be quantified.
- ◇ The PoLiS metric can be straightforwardly extended to a 3D PoLiS metric by computing an average distance between each 3D point of a polyhedron and a reference polyhedron and vice versa. It is also suitable to quantify the difference between a point cloud, e.g. LiDAR point cloud, and 3D polygons.
- ◇ The workflow of the joined proposed methods can be extended. A material attribute of a roof can be assigned to each extracted BP and added to the spatial database as defined in Gröger et al. (2012) and INSPIRE TWG BU (2013). Moreover, an estimation about required area of renovation of roofs can be calculated, if the spectral properties of the ageing of roofing materials are known. Another application is to provide the location and area of a target roofing material, which pose a threat to human health, e.g. asbestos.
- ◇ The accurate outlines of the objects are basis for object based coregistration between multi-modal images. Avbelj et al. (2015a) perform the coregistration of the HSI and DSM datasets on the basis of building outlines. Coregistration of the images with other modalities, e.g. thermal images, SAR, can be carried out on the basis of extracted objects with sub-pixel precision. In addition, the edge probabilities can also provide additional information in object based coregistration.



# Acronyms

**ND** *N*-Dimensional

**1D** One-Dimensional

**2D** Two-Dimensional

**3D** Three-Dimensional

**ALS** Air-Borne Laser Scanner

**BP** Building Polygon

**BRDF** Bidirectional Reflectance Distribution Function

**CityGML** City Geography Markup Language

**DEM** Digital Elevation Model

**DLR** Deutsches Zentrum für Luft- und Raumfahrt

**DN** Digital Number

**DSM** Digital Surface Model

**DTM** Digital Terrain Model

**EO** Earth Observation

**FCLS** Fully-Constrained Least Squares

**FOV** Field of View

**FWHM** Full Width at Half Maximum

**GH** Gauss-Helmert

**GIS** Geographic Information Systems

**GM** Gauss-Markov

**GMM** Gaussian Mixture Model

**GNSS** Global Navigation Satellite System

**GSD** Ground Sampling Distance

**HSI** Hyperspectral Image

**IFOV** Instantaneous Field of View

**iMBR** iterative Minimum Bounding Rectangle

**INS** Inertial Navigation System

**INSPIRE** Infrastructure for Spatial Information in Europe

**IR** Infrared

**LiDAR** Light Detection and Ranging or Light Radar

**LMM** Linear Mixing Model

**LOD** Level of Detail

**LS** Least Squares

**LSF** Line Spread Function

**MBR** Minimum Bounding Rectangle

**MLE** Maximum Likelihood Estimation

**MS** Multispectral

**NCLS** Non-Constrained Least Squares

**nDSM** normalised Digital Surface Model

**NDVI** Normalised Difference Vegetation Index

**NIR** Near Infrared

**NNLS** Non-Negative Least Squares

**PAN** Panchromatic

**PoLiS** Polygons and Line Segments

**PSF** Point Spread Function

**RANSAC** Random Sampling Consensus

**RGB** Red-Green-Blue

**RMSE** Root Mean Square Error

**RS** Remote Sensing

**SAD** Spectral Angle Distance

**SAR** Synthetic Aperture Radar

**SGM** Semi-Global Matching

**SIFT** Scale-Invariant Feature Transform

**SNR** Signal to Noise Ratio

**SRF** Spectral Response Function

**SWIR** Short Wavelength Infrared

**TIN** Triangulated Irregular Network

**TIR** Thermal Infrared

**VIS** Visible

**VNIR** Visible and Near

**WV-2** WorldView-2

# List of Figures

2.1	Hypercube of a HyMAP HSI and spectral signatures of three materials collected from the image. . . . .	16
2.2	Atmospheric and other influences on the radiance measured by a passive sensor. . . . .	18
2.3	Geometric distortions of an optical RS image acquired by a push-broom sensor. . . . .	20
2.4	SGM DSM of an urban area computed from images acquired by two different sensors. . . . .	22
2.5	Image orthorectified with erroneous DSM with holes. . . . .	23
2.6	LiDAR point cloud and interpolated LiDAR DSM. . . . .	24
2.7	Spectral resolution as a function of spatial resolution of EO air- and space-borne sensors. . . . .	26
3.1	General workflow of building polygon extraction. . . . .	31
3.2	A scale space and a multi-scale representation of an image. . . . .	34
4.1	Challenge of scale selection for edge detection from DSM. . . . .	41
4.2	Comparison of the Canny edge detector and edge probability computation applied on a DSM. . . . .	43
4.3	The discrete $g_{discrete}$ , the sampled Gaussian $g_{sampled}$ kernel, and their difference. . . . .	47
5.1	Computation of a DTM from a DSM. . . . .	52
5.2	Jagged boundary, detailed polygon, and coarse polygon of two building regions. . . . .	53
5.3	Approximate building polygon creation by iMBR. . . . .	54
5.4	Two functional models for LS adjustment of a rectilinear building. . . . .	57
5.5	Gauss-Markov model for rectilinear BP. . . . .	58
5.6	Gauss-Helmert model for rectilinear BP. . . . .	59
5.7	A covariance matrix $\Sigma$ for rectilinear building for GM model and GH model. . . . .	60
5.8	The Hausdorff, the Chamfer, and the PoLiS metrics on an example of two point sets. . . . .	65
5.9	A same point set is connected into two different polygons and the influence on the Hausdorff, the Chamfer, and the PoLiS metric computation. . . . .	66
5.10	Extracted and reference polygon, their centroids, and $TP$ , $FP$ , $FN$ , and $TN$ detected areas. . . . .	68
5.11	Pseudo-code for computing PoLiS metric between two closed polygons, $A$ and $B$ . . . . .	70
5.12	The additional vertices on a line segment to test sensitivity of the Hausdorff, the Chamfer, and the PoLiS metrics. . . . .	72
5.13	Sensitivity of the Hausdorff, the Chamfer, and the PoLiS metrics to the additional vertices on a line segment and a square. . . . .	73
5.14	The Hausdorff, the Chamfer, and the PoLiS metric as a function of translation, rotation, and scale. . . . .	74
5.15	The matched rates, $RMSE_{point}$ , $RMSE_{line}$ , and distance between centroids $dc$ as function of translation, rotation, and scale. . . . .	75

5.16	Evaluation of building polygon extraction by the Hausdorff, the Chamfer, and the PoLiS metric. . . . .	76
6.1	Workflow of building polygon extraction, adjustment, and evaluation. . . . .	81
6.2	The <i>tennis court</i> dataset with the superimposed reference polygons. . . . .	84
6.3	Selected scale $t_{max}$ on the basis of maximal edge probability ( <i>tennis court</i> ). . . . .	85
6.4	Edge probability of clay material for selected scale $t_{max}$ ( <i>tennis court</i> ). . . . .	85
6.5	Edge probability of clay material for selected scale $t_{max}$ superimposed on the HySpex red channel ( <i>tennis court</i> ). . . . .	86
6.6	Confidence probability $\beta(t_{max})$ of the edge probabilities ( <i>tennis court</i> ). . . . .	86
6.7	Detail of the <i>tennis court</i> dataset with buffer zones and mean edge probability with respect to $\beta(t)$ for different buffer widths. . . . .	87
6.8	Mean edge probability for single scales and selected $\bar{p}(edge; t_{max})$ scale ( <i>tennis court</i> ). . . . .	87
6.9	“Optimal scale” $t_{opt}$ selected on the basis of confidence probability $\beta(t) \geq 0.95$ ( <i>tennis court</i> ). . . . .	88
6.10	Edge probability of clay material for selected scale $t_{opt}$ ( <i>tennis court</i> ). . . . .	88
6.11	Rasterisation of the reference building polygon by different GSD. . . . .	90
6.12	Weights for the building polygon rasterised at different GSD. . . . .	91
6.13	Rasterized non-rectilinear reference polygon and its weights. . . . .	91
6.14	Comparison of approximate BP created by iMBR method and reference BP for different GSD. . . . .	93
6.15	GSD for which a level 2 building polygon ( $BP_{k_{min}=2}$ ) is selected as a function of a minimal area of a building part. . . . .	94
6.16	PoLiS metric [m] for evaluation of the approximate and adjusted BP. . . . .	95
6.17	PoLiS metric [pixel] for evaluation of the approximate and adjusted BP. . . . .	95
6.18	The quality rate [ ] for evaluation of the approximate and adjusted BP. . . . .	97
6.19	Absolute deviation of the main orientation [°] between the reference and the approximate and adjusted BP. . . . .	97
6.20	Adjusted rectilinear and non rectilinear BP by GM $\sigma_\alpha^2 = 1^\circ$ , GM $\sigma_\alpha^2 = 10^\circ$ , and GH model for $GSD = 1.0$ m. . . . .	99
6.21	Test areas, the <i>large area</i> and the <i>small area</i> , for experiments on RS images. . . . .	102
6.22	Spectral signatures of selected materials. . . . .	103
6.23	Approximate BP creation and selection from HSI–DSM dataset pairs ( <i>small area</i> ). . . . .	106
6.24	Number of approximate BP selected from DSM with respect to the GSD ( <i>small area</i> ). . . . .	107
6.25	A number of Level 1 BP ( $BP_{k=1}$ ) selected from HSI–DSM dataset pairs with respect to the GSD ( <i>small area</i> ). . . . .	108
6.26	Comparison of the weighting of the BP boundary points by prior knowledge and by the confidence probability $\beta(t_{max})$ ( <i>small area</i> ). . . . .	109

---

6.27	Edge probabilities for red roof tile material computed from HSI, for height edges computed from DSM, and joint edge probabilities ( <i>small area</i> ). . . . .	110
6.28	Adjusted BP for four HSI–DSM dataset pairs with varying GSD ( <i>small area</i> ). . . . .	111
6.29	Evaluation of the BP adjustment (GM model, $\sigma_\alpha^2 = 1^\circ$ ) by the PoLiS metric [pixel] ( <i>small area</i> ). . . . .	112
6.30	Evaluation of the BP adjustment (GM model, $\sigma_\alpha^2 = 1^\circ$ by the quality rate [ ] ( <i>small area</i> ).113	
6.31	Percentage of BP selected from HSI and DSM for VNIR20–WV-2 DSM, SWIR20–WV-2 DSM, and HyMAP–WV-2 DSM dataset pairs ( <i>large area</i> ). . . . .	115
6.32	Adjusted BP extracted from WV-2 DSM with $GSD = 2.0$ m ( <i>large area</i> ). . . . .	116
6.33	Adjusted BP from VNIR20 for four materials ( <i>large area</i> ). . . . .	117
6.34	Evaluation of the BP adjustment (GM model, $\sigma_\alpha^2 = 1^\circ$ ) by the PoLiS metric [pixel] ( <i>large area</i> ). . . . .	118
6.35	Evaluation of the BP adjustment (GM model, $\sigma_\alpha^2 = 1^\circ$ by the quality rate [ ] ( <i>large area</i> ).119	
6.36	Per-scene evaluation of the joined BP of the VNIR20–WV-2 DSM dataset pair ( <i>large area</i> ). . . . .	120

# List of Tables

1	General characteristics of the GM model and GH model for rectilinear polygon. . . . .	61
2	Selected sensor specifications of HyMAP and HySpex sensors. . . . .	80
3	Weighting of three groups of observations in GM model for sets of rectilinear and non-rectilinear BP. . . . .	92
4	Quality parameters for the test rectilinear building polygon. . . . .	96
5	Quality parameters for the test non-rectilinear building polygon. . . . .	100
6	The number of BP and per-scene evaluation of the HSI-WV-2 DSM dataset pairs ( <i>large area</i> ). . . . .	115
7	Per-object evaluation of the HSI-WV-2 DSM dataset pairs by PoLiS and 1- <i>qual</i> quality measures ( <i>large area</i> ). . . . .	117

# Bibliography

- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., Szeliski, R., Oct. 2011. Building Rome in a day. *Communication of the Association for Computing Machinery* 54 (10), 105–112.
- Ahmadi, S., Zoej, M. J. V., Ebadi, H., Moghaddam, H. A., Mohammadzadeh, A., Jun. 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation* 12 (3), 150–157.
- Arefi, H., Sep. 2009. From LIDAR point clouds to 3D building models. Dissertation, Bundeswehr University Munich, München, Germany.
- Avbelj, J., Jul. 2012. Spectral information retrieval for sub-pixel building edge detection. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-7, 61–66.
- Avbelj, J., Iwaszczuk, D., Müller, R., Reinartz, P., Stilla, U., Apr. 2013a. Line-based registration of DSM and hyperspectral images. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-1/W1, 13–18.
- Avbelj, J., Iwaszczuk, D., Müller, R., Reinartz, P., Stilla, U., Feb. 2015a. Coregistration refinement of hyperspectral images and DSM: An object-based approach using spectral information. *ISPRS Journal of Photogrammetry and Remote Sensing* 100, 23–24.
- Avbelj, J., Müller, R., Jul. 2014. Quality assessment of building extraction from remote sensing imagery. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Quebec City, Canada, pp. 3184–3187.
- Avbelj, J., Müller, R., Bamler, R., 2015b. A metric for polygon comparison and building extraction evaluation. *IEEE Geoscience and Remote Sensing Letters* 12 (1), 170–174.
- Avbelj, J., Müller, R., Reinartz, P., 2013b. Fusion of hyperspectral images and height models using edge probability. In: *Proceedings of the 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE. Gainesville, FL, USA, pp. 1–4.
- Awrangjeb, M., Ravanbakhsh, M., Fraser, C. S., Sep. 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (5), 457–467.
- Baldrige, A., Hook, S., Grove, C., Rivera, G., apr 2009. The ASTER spectral library version 2.0. *Remote Sensing of Environment* 113 (4), 711–715.
- Baltsavias, E. P., 2004. Object 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), 129–151.
- Baumgartner, A., Gege, P., Köhler, C., Lenhard, K., Schwarzmaier, T., 2012. Characterisation methods for the hyperspectral sensor HySpex at DLR's calibration home base. In: *SPIE Remote Sensing*, SPIE. International Society for Optics and Photonics, Edinburgh, UK, pp. 85331H–85331H.
- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J., 2012. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (2), 354–379.
- Bishop, C., February 2007. *Pattern Recognition and Machine Learning*. Springer.
- Bosworth, J. H., Acton, S. T., Apr. 2003. Morphological scale-space in image processing. *Digital Signal Processing* 13 (2), 338–367.
- Braun, A. C., Weidner, U., Jutzi, B., Hinz, S., 2011. Integrating model knowledge into SVM classification - Fusing hyperspectral and laserscanning data by kernel composition. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-4/W19, 57–62.
- Brédif, M., Tournaire, O., Vallet, B., Champion, N., Mar. 2013. Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework. *ISPRS Journal of Photogrammetry and Remote Sensing* 77, 57–65.

- Brenner, C., Mar. 2005. Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation* 6 (3–4), 187–198.
- Brook, A., Ben-Dor, E., Richter, R., 2010. Fusion of hyperspectral images and LiDAR data for civil engineering structure monitoring. In: *Proceedings of the 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE. Reykjavik, Iceland, pp. 1–5.
- Brook, A., Ben-Dor, E., Richter, R., 2013. Modelling and monitoring urban built environment via multi-source integrated and fused remote sensing data. *International Journal of Image and Data Fusion* 4 (1), 2–32.
- Canny, J., November 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 679–698.
- Cerra, D., Bieniarz, J., Avbelj, J., Reinartz, P., Müller, R., 2011. Compression-based unsupervised clustering of spectral signatures. In: *Proceedings of the 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE. pp. 1–4.
- Cerra, D., Bieniarz, J., Avbelj, J., Reinartz, P., Müller, R., 2012. Spectral matching through data compression. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-4/W19*, 75–78.
- Chang, C.-I., Aug. 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory* 46 (5), 1927–1932.
- Chaudhuri, D., Samal, A., Jul. 2007. A simple method for fitting of bounding rectangle to closed regions. *Pattern Recognition* 40 (7), 1981–1989.
- Chaudhuri, S., Kotwal, K., 2013. *Hyperspectral Image Fusion*. Springer.
- Clark, J., 1989. Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1), 43–57.
- Clark, R., Swayze, G., Wise, R., Livo, E., Hoefen, T., Kokaly, R., Sutley, S., 2007. USGS digital spectral library splib06a: U.S. geological survey. Digital data series 231. URL <http://speclab.cr.usgs.gov/spectral.lib06>, [Accessed: 01 September 2015].
- Copson, E. T., 1988. *Metric Spaces*. Cambridge University Press, Cambridge; New York, USA.
- Dalla Mura, M., Benediktsson, J., Waske, B., Bruzzone, L., oct 2010. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing* 48 (10), 3747–3762.
- De Carvalho, O., Meneses, P. R., 2000. Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM). In: *Summaries of the 9th JPL Airborne Earth Science Workshop*, JPL Publication 00-18. Vol. 9. JPL Publication Pasadena, CA.
- Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pizurica, A., Gautama, S., Philips, W., Prasad, S., Du, Q., Pacifici, F., Jun. 2014. Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (6), 2405–2418.
- Doxani, G., Karantzalos, K., Strati, M. T., Apr. 2012. Monitoring urban changes based on scale-space filtering and object-oriented classification. *International Journal of Applied Earth Observation and Geoinformation* 15, 38–48.
- Eismann, M. T., Apr. 2012. *Hyperspectral Remote Sensing*. SPIE Press Bellingham.
- EnMAP, 2015. EnMAP hyperspectral imager. URL <http://www.enmap.org/>, [Accessed: 01 September 2015].
- Field, D. J., Brady, N., 1997. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research* 37 (23), 3367–3383.
- Flamanc, D., Maillet, G., Jibrini, H., 2003. 3D city models: An operational approach using aerial images and cadastral maps. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (3/W8), 53–58.

- Freire, S., Santos, T., Navarro, A., Soares, F., Silva, J. D., Afonso, N., Fonseca, A., Tenedorio, J., Apr. 2014. Introducing mapping standards in the quality assessment of buildings extracted from very high resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 90, 1–9.
- Frueh, C., Zakhor, A., Jun. 2003. Constructing 3D city models by merging ground-based and airborne views. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. 562–569.
- Fua, P., Leclerc, Y. G., 1990. Model driven edge detection. *Machine Vision and Applications* 3 (1), 45–56.
- Gallup, D., Frahm, J.-M., Pollefeys, M., Jun. 2010. Piecewise planar and non-planar stereo for urban scene reconstruction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1418–1425.
- Gamba, P., 2014. Image and data fusion in remote sensing of urban areas: Status issues and research trends. *International Journal of Image and Data Fusion* 5 (1), 2–12.
- Gatziolis, D., Andersen, H.-E., 2008. A guide to LiDAR data acquisition and processing for the forests of the pacific northwest. Tech. rep., US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Geibel, R., Stilla, U., 2000. Segmentation of laser altimeter data for building reconstruction: Different procedures and comparison. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 33 (Part B3), 326–334.
- Gerke, M., Straub, B. M., Koch, A., 2001. Automatic detection of buildings and trees from aerial imagery using different levels of abstraction. *Publikationen der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung* 10, 273–280.
- Gómez-Chova, L., Alonso, L., Guanter, L., Camps-Valls, G., Calpe, J., Moreno, J., Oct. 2008. Correction of systematic spatial noise in push-broom hyperspectral sensors: Application to CHRIS/PROBA images. *Applied Optics* 47 (28), F46–F60.
- Gonzalez, R. C., Woods, R. E., Jan. 2002. *Digital Image Processing*, 2nd Edition. Prentice Hall.
- Grigillo, D., Kosmatin Fras, M., Petrovič, D., Aug. 2012. Automated building extraction from IKONOS images in suburban areas. *International Journal of Remote Sensing* 33 (16), 5149–5170.
- Gröger, G., Kolbe, T. H., Nagel, C., Häfele, K.-H., 2012. *OGC City Geography Markup Language (CityGML) Encoding Standard, Version 2.0.0*. Open Geospatial Consortium, OGC Doc. No. 12–019.
- Gröger, G., Plümer, L., Jul. 2012. CityGML – interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing* 71 (1), 12–33.
- Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 54, 130–137.
- Haralick, R., Jan. 1984. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1), 58–68.
- Heldens, W., Jun. 2010. Use of airborne hyperspectral data and height information to support urban micro climate characterisation. Dissertation, Julius-Maximilians-Universität Würzburg, Würzburg.
- Henn, A., Gröger, G., Stroh, V., Plümer, L., Feb. 2013. Model driven reconstruction of roofs from sparse LIDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 76, 17–29.
- Heuel, S., 2004. *Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction*. Vol. 3008 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.
- Heuel, S., Kolbe, T. H., 2001. Building reconstruction: The dilemma of generic versus specific models. *Künstliche Intelligenz* 15 (3), 57–62.
- Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. 807–814.
- Hodgson, M. E., Bresnahan, P., 2004. Accuracy of airborne LIDAR-derived elevation: Empirical assessment and error budget. *Photogrammetric Engineering & Remote Sensing* 70 (3), 331–340.

- Hu, J., You, S., Neumann, U., Nov. 2003. Approaches to large-scale urban modeling. *IEEE Computer Graphics and Applications* 23 (6), 62–69.
- Hu, M.-K., Feb. 1962. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* 8 (2), 179–187.
- Huertas, A., Nevatia, R., Landgrebe, D., 1999. Use of hyperspectral data with intensity images for automatic building modeling. In: *IEEE Proceedings of the 2nd International Conference on Information Fusion*. Sunnyvale, CA, USA, pp. 680–687.
- HyVista Corporation, 1999. Hymap specifications. URL <http://www.hyvista.com/>, [Accessed: 01 September 2015].
- INSPIRE TWG BU, Apr. 2013. Data specification on buildings: Draft technical guidelines, version: v3.0rc3, identifier: D2.8.III.2\_v2.9. Thematic working group buildings. URL <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>, [Accessed: 01 September 2015].
- Iordache, M. D., Bioucas-Dias, J., Plaza, A., 2010. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 49 (6), 2014–2039.
- Kanani, E., 2000. Robust estimators for geodetic transformations and GIS. Dissertation, Institut für Geodäsie und Photogrammetrie, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland.
- Keshava, N., 2003. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal* 14 (1), 55–78.
- Keshava, N., Mustard, J., Jan. 2002. Spectral unmixing. *IEEE Signal Processing Magazine* 19 (1), 44–57.
- Koch, K.-R., Apr. 1999. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer Science & Business Media.
- Koenderink, J. J., 1984. The structure of images. *Biological Cybernetics* 50 (5), 363–370.
- Krauß, T., Arefi, H., Reinartz, P., 2011. Evaluation of selected methods for extracting digital terrain models from satellite born digital surface models in urban areas. In: *Proceedings of the International Conference on Sensors and Models in Photogrammetry and Remote Sensing*. Tehran, Iran.
- Kruse, F., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., Goetz, A., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* 44 (2), 145–163.
- Kurz, F., Türmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P., Leitloff, J., May 2012. Low-cost optical camera systems for real-time mapping applications. *PFG Photogrammetrie, Fernerkundung, Geoinformation* 2012 (2), 159–176.
- Kwak, E., Habib, A., Jul. 2014. Automatic representation and reconstruction of DBM from LiDAR data using recursive minimum bounding rectangle. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, 171–191.
- Lafarge, F., Descombes, X., Zerubia, J., Pierrot-deseilligny, M., May 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., Aug. 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.
- Lee, D. S., Shan, J., Bethel, J. S., 2003. Class-guided building extraction from IKONOS imagery. *Photogrammetric Engineering & Remote Sensing* 69 (2), 143–150.
- Lin, C., Nevatia, R., Nov. 1998. Building detection and description from a single intensity image. *Computer Vision and Image Understanding* 72 (2), 101–121.
- Lindeberg, T., 1994. *Scale-space theory in computer vision*. Kluwer Academic, Boston.
- Lindeberg, T., 1998. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision* 30 (2), 117–156.

- Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: Proceedings of the 7th International Conference on Computer Vision, IEEE. Vol. 2. IEEE, Kerkyra, Greece, pp. 1150–1157.
- Ma, W.-K., Bioucas-Dias, J., Chan, T.-H., Gillis, N., Gader, P., Plaza, A., Ambikapathi, A., Chi, C.-Y., 2014. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine* 31 (1), 67–81.
- Maas, H.-G., Vosselman, G., Jul. 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 153–163.
- Marimont, D., Rubner, Y., January 1998. A probabilistic framework for edge detection and scale selection. In: Proceedings of the 6th International Conference on Computer Vision, IEEE. Bombay, India, pp. 207–214.
- Mayer, H., Mar. 2008. Object extraction in photogrammetric computer vision. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (2), 213–222.
- Mayer, H., Steger, C., Apr. 1998. Scale-space events and their link to abstraction for road extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (2), 62–75.
- Mena, J. B., Dec. 2003. State of the art on automatic road extraction for GIS update: A novel classification. *Pattern Recognition Letters* 24 (16), 3037–3058.
- Mikhail, E., Ackerman, E. F., 1976. Observations and least squares. IEP, New York, NY, USA.
- Mouroulis, P., Green, R. O., Chrien, T. G., May 2000. Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Applied Optics* 39 (13), 2210–2220.
- Müller, R., Krauß, T., Schneider, M., Reinartz, P., Jun. 2010. A method for geometric processing of optical satellite images using automatically determined ground control information. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (1), 1–6.
- Müller, R., Lehner, M., Müller, R., Reinartz, P., Schroeder, M., Vollmer, B., 2002. A program for direct georeferencing of airborne and spaceborne line scanner images. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (1), 148–153.
- Neitzel, F., Petrovic, S., 2008. Total least squares (TLS) im Kontext der Ausgleichung nach kleinsten Quadraten am Beispiel der ausgleichenden Geraden. *Zeitschrift für Geodäsie, Geoinformation und Landmanagement* 133, 141–148.
- Niemeier, W., Mar. 2008. Ausgleichsrechnung: Statistische Auswertemethoden, 2nd Edition. Walter de Gruyter.
- Norsk Elektro Optikk AS, 1985. HySpex specifications. URL <http://www.hyspex.no/index.php>, [Accessed: 01 September 2015].
- Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 629–639.
- Pesaresi, M., Benediktsson, J. A., 2001. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39 (2), 309–320.
- Pfeifer, N., Rutzinger, M., Rottensteiner, F., Muecke, W., Hollaus, M., 2007. Extraction of building footprints from airborne laser scanning: Comparison and validation techniques. In: Proceedings of the IEEE Urban Remote Sensing Joint Event. Paris, France, pp. 1–9.
- Plaza, A., Martinez, P., Perez, R., Plaza, J., 2004. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 42 (3), 650–663.
- Pu, S., Vosselman, G., Nov. 2009. Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (6), 575–584.
- Ragia, L., Winter, S., Sep. 2000. Contributions to a quality description of areal objects in spatial data sets. *ISPRS Journal of Photogrammetry and Remote Sensing* 55 (3), 201–213.
- Richards, J. A., Jia, X., 2006. Remote Sensing Digital Image Analysis - An Introduction. Springer, Berlin, Germany.

- Robila, S., Gershman, A., 2005. Spectral matching accuracy in processing hyperspectral data. In: Proceedings of the IEEE International Symposium on Signals, Circuits and Systems. Vol. 1. pp. 163–166.
- Roessner, S., Segl, K., Heiden, U., Kaufmann, H., Jul. 2001. Automated differentiation of urban surfaces based on airborne hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39 (7), 1525–1532.
- Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, 256–71.
- Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., Jul. 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-3, 293–298.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., Dec. 2005. Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion* 6 (4), 283–300.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., Jun. 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (2), 135–149.
- 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.
- Sampath, A., Shan, J., 2007. Building boundary tracing and regularization from airborne LiDAR point clouds. *Photogrammetric Engineering & Remote Sensing* 73 (7), 805.
- Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47 (1), 7–42.
- Schuster, H.-F., Weidner, U., 2003. A new approach towards quantitative quality evaluation of 3D building models. In: *ISPRS Commission IV Joint Workshop on Challenges in Geospatial Analysis*. Stuttgart, Germany, pp. 8–9.
- Segl, K., Heiden, U., Müller, M., Kaufmann, H., 2003a. Endmember detection in urban environments using hyperspectral HyMap data. In: *Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy*. pp. 13–16.
- Segl, K., Roessner, S., Heiden, U., Kaufmann, H., Jun. 2003b. Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1-2), 99–112.
- Shahzad, M., Zhu, X., Feb. 2015. Robust reconstruction of building facades for large areas using spaceborne TomoSAR point clouds. *IEEE Transactions on Geoscience and Remote Sensing* 53 (2), 752–769.
- Shan, J., Lee, S. D., 2005. Quality of building extraction from IKONOS imagery. *Journal of Surveying Engineering* 131 (1), 27–32.
- Shufelt, J., 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (4), 311–326.
- Sithole, G., Vosselman, G., Aug. 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (1–2), 85–101.
- Snavely, N., Seitz, S. M., Szeliski, R., Nov. 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision* 80 (2), 189–210.
- Sohn, G., Dowman, I., May 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (1), 43–63.
- Song, W., Haithcoat, T., 2005. Development of comprehensive accuracy assessment indexes for building footprint extraction. *IEEE Transactions on Geoscience and Remote Sensing* 43 (2), 402–404.
- Steadman, P., 2006. Why are most buildings rectangular? *Architectural Research Quarterly* 10 (2), 119–130.

- Taillandier, F., 2005. Automatic building reconstruction from cadastral maps and aerial images. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (3), 105–110.
- Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., Koehl, M., 2007. Model-driven and data-driven approaches using LIDAR data: Analysis and comparison. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 87–92.
- Thiele, A., Cadario, E., Schulz, K., Thonnessen, U., Soergel, U., Nov. 2007. Building recognition from multi-aspect high-resolution InSAR data in urban areas. *IEEE Transactions on Geoscience and Remote Sensing* 45 (11), 3583–3593.
- Tsai, A., Westin, C. F., Hero, A., Willsky, A., 2007. Fiber tract clustering on manifolds with dual rooted-graphs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN, USA, pp. 1–6.
- van der Meer, F. D., 2001. *Imaging spectrometry: Basic principles and prospective applications*. Kluwer Academic Publishers, Dordrecht, Boston.
- Vanegas, C. A., Aliaga, D. G., Benes, B., Jun. 2010. Building reconstruction using manhattan-world grammars. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA, pp. 358–365.
- Veksler, O., 2001. Semi-dense stereo correspondence with dense features. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. II-490–II-497 vol.2.
- Veltkamp, R., 2001. Shape matching: Similarity measures and algorithms. In: *IEEE International Conference on Shape Modeling and Applications*. Genova, Italy, pp. 188–197.
- Veltkamp, R. C., Hagedoorn, M., 1999. *State-of-the-art in shape matching*. Tech. rep., Utrecht University, Departement of Computer Science, The Netherlands.
- Verma, V., Kumar, R., Hsu, S., 2006. 3D building detection and modeling from aerial LIDAR data. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. New York, NY, USA, pp. 2213–2220.
- Vosselman, G., Maas, H.-G., 2010. *Airborne and terrestrial laser scanning*. Whittles Publishing, Dunbeath.
- Vu, T. T., Yamazaki, F., Matsuoka, M., Aug. 2009. Multi-scale solution for building extraction from LiDAR and image data. *International Journal of Applied Earth Observation and Geoinformation* 11 (4), 281–289.
- Wack, R., Wimmer, A., 2002. Digital terrain models from airborne laserscanner data—a grid based approach. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (3), 293–296.
- Wang, R., 2013. 3D building modeling using images and LiDAR: A review. *International Journal of Image and Data Fusion* 4 (4), 273–292.
- Weidner, U., Förstner, W., Aug. 1995. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50 (4), 38–49.
- Witkin, A. P., 1984. Scale-space filtering: A new approach to multi-scale description. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 9. San Diego, CA, USA, pp. 150–153.
- Xiao, J., Gerke, M., Vosselman, G., Mar. 2012. Building extraction from oblique airborne imagery based on robust façade detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 68, 56–68.
- Yokoya, N., Miyamura, N., Iwasaki, A., 2010. Preprocessing of hyperspectral imagery with consideration of smile and key-stone properties. In: *SPIE Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications*. Vol. 7857. p. 78570B.
- Zebedin, L., Bauer, J., Karner, K., Bischof, H., 2008. Fusion of feature- and area-based information for urban buildings modeling from aerial imagery. In: *Proceedings of the 10th European Conference on Computer Vision*. Vol. IV. Springer Verlag, Marseille, France, pp. 873–886.
- Zeng, C., Wang, J., Lehrbass, B., 2013. An evaluation system for building footprint extraction from remotely sensed data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (3), 1640–1652.

- Zhan, Q., Molenaar, M., Tempfli, K., Shi, W., 2005. Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing* 26 (14), 2953–2974.
- Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. *Pattern Recognition* 37 (1), 1–19.
- Zhang, J., Collins, R., Liu, Y., 2004. Representation and matching of articulated shapes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. 342–349.
- Zhang, W., Wong, A., Mishra, A., Fieguth, P., Clausi, D., Sep. 2011. Efficient globally optimal registration of remote sensing imagery via quasi-random scale-space structural correlation energy functional. *IEEE Geoscience and Remote Sensing Letters* 8 (5), 997–1001.
- Zoubir, A., Koivunen, V., Chakhchoukh, Y., Muma, M., Jul. 2012. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Processing Magazine* 29 (4), 61 –80.

# Acknowledgements

I consider my PhD period a privilege to grow professionally and personally. It would not be possible without support of my supervisors, colleagues, friends, and family.

*Hvala! Danke schön! Thank you!*

First and foremost, I would like to thank Rupert Müller (DLR) for his invaluable scientific advice, guidance, and encouragement throughout my PhD research period. He was always there to discuss my research and progress, even while travelling by over-packed and badly delayed trains. He has been a mentor to me far beyond the scientific guidance. I look up to him professionally as well as personally.

Second, I am grateful to Prof. Dr.-Ing. Richard Bamler (TU Munich and DLR) and Prof. Dr.-Ing. Peter Reinartz (DLR) for giving me the opportunity to carry out the thesis at the TU Munich and at the German Aerospace Center (DLR) and for their advice on my research. I would also like to thank Prof. Dr.-Ing. Markus Gerke (University of Twente) for showing me a different view on my research and Prof. Dr. techn. Friedrich Fraundorfer (TU Graz) for valuable discussions.

Moreover, I would like to thank the City of Munich, Department of Environment and Health (Referat für Gesundheit und Umwelt der Stadt München) for providing the reference building outlines used as ground truth in this thesis.

I gratefully acknowledge Prof. Dr. Ben-Dor (Tel Aviv University) who hosted me at his Department as a guest scientist. He and his colleagues, Gila, Ido, Yaron, Prof. Anna Brook (University of Haifa), introduced me to thermal hyperspectral imaging and made me feel a part of their research team from day one.

My first steps into research were encouraged by Prof. Dr. Krištof Oštir (Slovenian Academy of Sciences and Arts) and Prof. Dr.-Ing. Uwe Stilla (TU Munich). I appreciate honest advice and critical opinion by Prof. Dr. Dušan Petrovič (University of Ljubljana) and support by Prof. Dr. Mojca Kosmatin Fras (University of Ljubljana), Dr. Klemen Zakšek (Universität Hamburg), and Konrad Eder (TU Munich). It is great to know, I can always count on you!

I greatly enjoyed the research environment at TU Munich and DLR. I would like to thank all my TU Munich and DLR colleagues, especially Daniele, Jakub, Vroni, Dorota, Daniela, Jagmal, Olli, Chara, Nina, Sebastian, Stefan A., Stefan G., Xiaoxiang, and Ludwig, of whom many became my friends. I appreciate the effort Dr. Tobias Storch put into proofreading this thesis and thank Rolf Stätter for comments on its formatting and layout.

My deepest gratitude goes to my family. My parents, who have always been encouraging me on the ways I have chosen. Ulrike and Rolf for showing support by sending me all those personalised “relax boxes” with chocolate, tea, and books during my PhD time. At last, Jörn, thank you for making me laugh, loving, and supporting me “als sei selbstverständlicher nichts auf der Welt”.