Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 818

Hai Huang

Bayesian Models for Pattern Recognition in Spatial Data

München 2018

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5229-1

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 341, Hannover 2018

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 818

Bayesian Models for Pattern Recognition in Spatial Data

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Verleihung der Lehrbefähigung
für das Fachgebiet "Photogrammetrische Computer Vision"
genehmigte Habilitationsschrift

Vorgelegt von

Dr.-Ing. Hai Huang

Geboren am 12.01.1976 in Shanghai

München 2018

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5229-1

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover
ISSN 0174-1454, Nr. 341, Hannover 2018

Adresse der DGK:

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 331 – 288 1685 ● Telefax +49 – 331 – 288 1759

E-Mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission:

Vorsitzender: Univ. Prof. Dr.-Ing. Winrich Voß

Hauptberichter: Univ. Prof. Dr.-Ing. Monika Sester

Mitberichter: Univ. Prof. Dr.-Ing. Christian Heipke
Univ. Prof. Dr.-Ing. Bodo Rosenhahn
Univ. Prof. Dr.-Ing. Konrad Schindler (ETH Zürich)

Tag der Einreichung: 20.04.2017
Tag des Habilitationskolloquiums: 02.02.2018

ISSN 0065-5325 ISBN 978-3-7696-5229-1

"The theory of probability is basically just common sense reduced to calculus."
– Pierre-Simon Laplace

Acknowledgments

First and foremost I would like to thank Prof. Monika Sester. I appreciate her contributions of advice, encouragement and funding providing me an exciting and richly diversified postdoctoral experience in Hannover. Her support makes this habilitation possible.

Prof. Claus Brenner enrolled me in his exciting project of building reconstruction via generative models, which becomes now one of my major research interests. He also shared his courses with me, from where I started to gather the experience of teaching.

I sincerely thank Prof. Winrich Voß, Prof. Christian Heipke, Prof. Konrad Schindler and Prof. Bodo Rosenhahn for their generous acceptance to become chairman and assessors of the habilitation committee, respectively. Their comments and suggestions not only improved and perfected this thesis but also inspire my current and future work.

Prof. Helmut Mayer is my "Doktorvater" (PhD supervisor) as well as my boss now at Bundeswehr University Munich. For me, his advice and support are all-round and sustained. Helmut supports my habilitation since I came back to Munich. He thoroughly reviewed and corrected this thesis just like he did for my doctoral thesis. Prof. Christian Heipke is another one who revised this thesis word by word. Christian and I have known each other at the beginning of my PhD and I learned more from him when I worked in Hannover.

I am grateful to Frau Evelin Schramm, Birgit Kieler, Sabine Hofmann, Yu Feng, Alexander Schlichting, Malte Jan Schulze, Lijuan Zhang, Jens-André Pfaffenholz and all the colleagues at the Institute of Cartography and Geoinformatics for making it such an inspiring and pleasant place to work.

I say thank you to Frau Gisela Pietzner and all the colleagues at the Chair of Visual Computing, Institute for Applied Computer Science at Bundeswehr University Munich, for sharing their intelligence and passions at work as well as their support for my habilitation.

Prof. Hao Jiang and the colleagues at the Department of Computer Science, Boston College, are appreciated. The "research summer" at Chestnut Hill was exciting and joyful.

My final but no less gratitude goes to my wife Na Liu for going with me through a lot more than she anticipated. Charlène is the most wonderful present in my life.

Hai Huang
Neubiberg, June 2018

The work described in Section 2 was supported by German Science Foundation (Deutsche Forschungsgemeinschaft – DFG).
The work described in Section 5 was supported by United States National Science Foundation (U.S. NSF).

Abstract

The improvement of measurement and particularly surveying technologies results in a large as well as rapidly increasing amount of spatial data. These data stem from various measurement techniques as well as platforms and, therefore, may compile quite different densities, qualities, and error characteristics. Effective tools are required to understand and interpret them. The challenges include efficient processing, robustness against data flows and uncertainty, rationality of modeling, and the potential of automation and learning. This thesis presents an exploration of the use of statistical models and related techniques in spatial data analysis. The foundation of the methodology employed in the scope of this thesis consists of Bayesian statistics and Markov models. Selected approaches conceived by the author, including 3D building reconstruction, semantic building classification, pattern recognition in trajectories, and segmentation of RGBD data, demonstrate their potential in spatial data modeling and interpretation.

Zusammenfassung

Die fortschreitende Entwicklung der Vermessungstechnologien führt zu einer großen sowie schnell wachsenden Menge an räumlichen Daten. Diese Daten stammen aus verschiedenen Messtechniken sowie Plattformen und können daher ganz unterschiedlichen Dichten, Qualitäten und Fehlercharakteristiken zusammenstellen. Effektive Werkzeuge sind erforderlich, um sie zu verstehen und zu interpretieren. Zu den Herausforderungen gehören effiziente Verarbeitung, Robustheit gegen Datenfehler und Ungewissheit, Rationalität der Modellierung und das Potenzial von Automatisierung und Lernen. Diese Arbeit stellt eine Erforschung der Verwendung von statistischen Modellen und verwandten Techniken in der räumlichen Datenanalyse vor. Die Grundlagen der im Rahmen dieser Arbeit eingesetzten Methodik sind die Bayes'sche Statistik und die Markow-Modelle. Ausgewählte Ansätze des Autors, darunter 3D-Gebäuderekonstruktion, semantische Gebäudeklassifikation, Mustererkennung in Trajektorien und Segmentierung von RGBD-Daten, zeigen ihr Potenzial in der räumlichen Datenmodellierung und -interpretation.

Contents

Part I

Synopsis

Chapter 1

Introduction

We live in an era of abundant data. The improvement of measurement and particularly surveying technologies results in a large as well as rapidly increasing amount of spatial data. This is especially true for the densely inhabited urban areas, which in practice attract most attention. The spatial data stem from various measurement techniques, e.g., laser scanning, photography, and radar, as well as platforms such as terrestrial, airborne, space-borne, and mobile mapping systems. They, therefore, compile quite different densities, qualities, and error characteristics. Effective tools are required to understand and interpret these data.

This thesis presents an exploration of the use of statistical models and related techniques in spatial data analysis. The foundation of the methodology employed in the scope of this thesis consists of Bayesian statistics and Markov models. Selected approaches conceived by the author demonstrate their potential in spatial data modeling and interpretation.

1.1 Spatial data and the challenges

Spatial data, mostly known as (but not limited to) geospatial data or spatial/geographic information, refers to data or information concerning the location and shape of spatial features and their relationships, or briefly "data with a spatial reference". Spatial data are stored in the form of geometry (coordinates) and attributes/features. In contrast to data analysis in other branches, the reference of data to location and time contains more crucial information and is explicitly used (Goodchild and Haining 2003).

Spatial data are acquired by means of numerous different sensor platforms and stored in various formats. There are, thus, many different way to categorize them. In the scope of this thesis, we consider spatial data as two groups: Raster data and vector data. An overview is presented in Figure 1.1.

In raster data, the space is uniformly divided into cells. Each cell has a set of evenly distributed coordinates and assigned features. The level of detail is determined by the cell size, i.e., the resolution. The space can in principle be both two dimensional (2D) and three dimensional (3D) with the corresponding terms for cells "pixels" and "voxels", respectively.

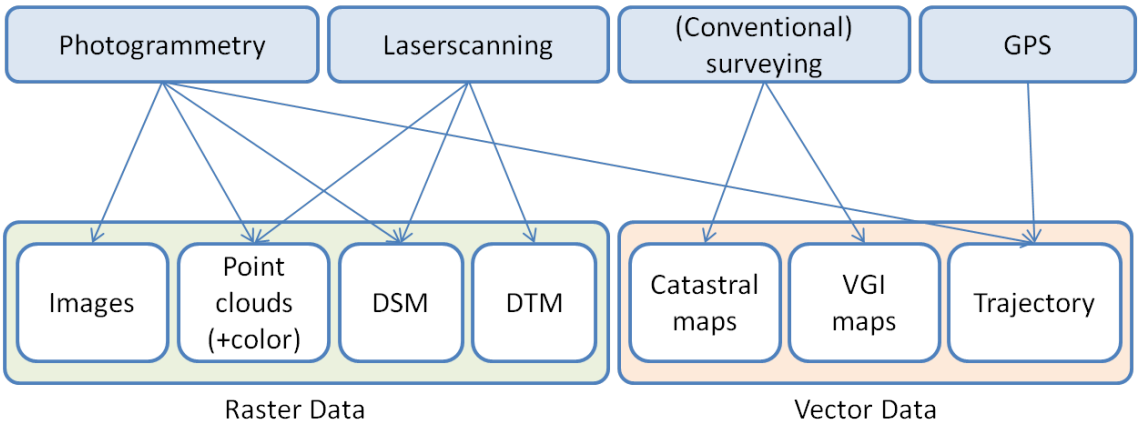


Figure 1.1: Overview of spatial data (in the scope of this thesis) and the related surveying methods

Images (digital or digitized) are inherently raster data, which, in the scope of this thesis, mainly refer to airborne and space-borne ***imagery*** from nadir view.

3D ***point clouds*** are not conventional raster data. They are generated by means of one of the following techniques: (A) Laserscanning, (B) depth estimation from stereo images or (C) depth cameras. Point clouds have no uniform cell size in 3D. Although they can still be seen as raster data with "flexible cell size" and treated with some sophisticated methods, raster is often advisable in order to reduce the data redundancy and to adapt to existing algorithms or tools. Point clouds can be turned into voxels with a 3D raster. In spatial analysis, however, they are often treated as a 2D raster:

- A raster based on the X-Y coordinate plane indicating the ground surface and Z coordinate as the height value above it, which results in ***DSM/DTM*** (Digital Surface Model/Digital Terrain Model) data, or
- a raster based on a given projection plane in 3D, which results in ***RGBD (RGB-Depth)*** data. The plane can be the projection plane of the sensor or a specific plane of interest, e.g., a building facade.

The point clouds are categorized as raster data not only because they are often rasterized in practice, but also because they share the following characteristics with raster data (in comparison with vector data):

- Only points represent coordinates
- Simple and clear data structure, but often large amount of data
- No topological relationships (besides spatial neighborhood).

In vector data, points, lines and polygons are used to represent geographical features based on coordinates. The vector data in the scope of this thesis are (digital) ***maps*** and ***trajectories***. An digital map represents the topography of the real world in a coordinate system. Trajectory data take temporal information into account.

1.1.1 Characteristics of spatial data

Spatial data are generally characterized as huge in quantity and indefinite in quality. The improvement of data acquisition methods and the introduction of new surveying technology make these characteristics more prominent:

Large amounts of data: In the last decades spatial data rapidly increased concerning both quantity and quality. Airborne laserscanning and image acquisition deliver improved accuracy and resolution. Satellite images have a high enough resolution for the detection of the small structures like buildings or even vehicles. Surveying methods such as mobile mapping systems combine multiple sensors, i.e., laserscanner, GPS, cameras, etc., on one platform. Furthermore, VGI (Volunteered Geographic Information) data, e.g., OSM and Flicker images, considerably contribute to both the variability and amount of spatial data.

Flaws and uncertainty: System or instrumental errors of sensors and measurement errors of surveying methods have been reduced due to the improvement of surveying technologies. The errors, however, cannot be totally avoided and, most important, the quality of the data are not uniformly distributed over all areas or objects. The measurement data can be very detailed (of, e.g., dense urban areas and landmark buildings with multiple sensors and high resolution) and very poor (of, e.g., rural areas with low resolution LiDAR or satellite imagery) at the same time. Redundancy and absence of data often happen in the same dataset. For instance, in the UAV (Unmanned Aerial Vehicle) photogrammetry with a camera a large number of photos might have been taken from all possible positions with the intention to gain as much overlap as possible. Yet, the large redundancy does not guarantee a complete coverage of the target object. Many problems, e.g., occlusions, concave shapes, specular reflections, homogeneous surface textures (leading to difficulties in matching), result in gaps in data as well as uncertainties. Occlusions and reflections also affect LiDAR and Kinect data.

1.1.2 Challenges

The availability of large amounts of high resolution data as well as the changing data characteristics lead to requirements not only concerning the ability to efficiently process huge data from various platforms, but also to model with more detail with respect to geometric resolution and semantic interpretation. The overall trends show demands for: 1. Transition from 2D to 3D models and 2. enhanced semantic descriptions added to the geometric attributes. The challenges can be summarized as follows:

- **Efficient processing**

  The increasing data volume renders traditional data parsing methods inefficient or even infeasible, as the computational effort shows linear or polynomial growth in relation to the number of data entries. A search for patterns needs to be performed in high-dimensional solution spaces. In the processing, top-down methods should be preferred because of their potential to deal with large data redundancy.

- **Robustness against data flaws and uncertainty**

  An increasing data volume does not guarantee more accurate or complete observations. Patterns, including geometric features (e.g., edges and planes) as well as complete models (e.g., buildings and trees), should be extracted in a stable fashion despite artefacts, gaps, and even occlusions. Robustness is also required when dealing with data from different sensor platforms and data fusion.

- **Plausibility of models**

  In practical applications, "plausible" results, i.e., complete models with reasonable parameters, are desired. The definition of plausibility is, however, in most cases tricky because, in additional to the data themselves, knowledge about the target objects is always required. This implies that prior information should be applied before the modeling and/or learned during the processing.

- **Potential of automation and learning**

  The data amount has reached a level that conventional manual analysis is no more feasible and an automatic or, at least, partially automatic processing becomes a necessity. A high degree of automation requires methods which are robust and can adapt to different scenarios and little human intervention, e.g., parameter tuning. When dealing with heterogeneous spatial data this implies a perceptual processing and the ability to learn within the proposed methods.

1.2 Statistical models

Statistics is a mathematical and conceptual discipline that focuses on the relation between data and hypotheses (Romeijn 2014). The data are recordings of observations and, in the scope of this thesis, the (geospatial) measurements.

The philosophy of statistics is part of the philosophical topic of scientific methodology – the general theory on whether and how science acquires knowledge. Statistical methods describe and justify the relationship between statistical theory (hypotheses) and evidence (empirical facts). Generally, a "statistical model" is defined by a set of statistical hypotheses and represented by a set of probability distributions on the data, also called sample space (Cox 2006, Bernardo and Smith 1994). Figure 1.2 shows the relations between the above concepts.

There are two major theories of statistical methods: Classical/frequentist and Bayesian statistics. From the viewpoint of philosophy or logic, the main difference between Bayesian and frequentist statistics is the way they treat "probability". The Bayesian calculus describes probability as "degree of belief". In the Bayesian probability equation (cf. Equation (1.1)) the beliefs are presented in the form of the so-called priors and posteriors. In Frequentist statistics, on the other hand, probability is used to model actual processes/frequencies.

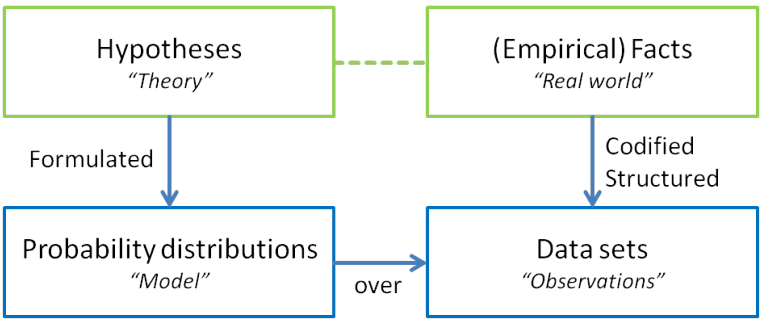


Figure 1.2: Relations between statistical models and theories in the real world

In (Bandyopadhyay and Forster 2011) the statistical inference methods are categorized into four paradigms: (1) classical statistics (frequentist inference), (2) Bayesian statistics, (3) likelihood-based statistics, and (4) Akaike-Information Criterion (AIC) (Akaike 1973) - based statistics. It needs to be pointed out that these concepts are not mutually exclusive. AIC stems from information theory and is used to compare statistical models considering the trade-off between the goodness of fit (in most cases the likelihood) and the simplicity of the models. The likelihood-based paradigm can, therefore, be seen as a subset of AIC-based methods. Furthermore, the AIC can be integrated into the Bayesian framework. More details can be found in the works described in this thesis.

Statistical models, and particularly Bayesian statistics, are considered as promising for the exploration and interpretation of spatial data. The improvement of computer technologies makes the utilization of sophisticated statistical tools nowadays possible. This Section briefly introduces Bayesian methods and Markov models as appropriate "carriers" of Bayesian statistics.

1.2.1 Bayesian statistics

Bayesian statistical models have been widely and successfully used in various areas especially in artificial intelligence/machine learning and economics. Well-known applications include (statistical) language translation, Bayesian image recognition, and (Spam) mail filtering. In the framework of Bayesian statistics we are allowed to adapt models to and learn from the given data. Bayesian probabilities are used to summarize evidences and to give statistical propositions. Prediction and learning are done in form of inference (Romeijn 2014).

In the foregoing discussion we said statistics study the relation between hypotheses and data. The Bayes theorem describes such relationships as follows:

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}) \cdot P(\mathcal{M})}{P(\mathcal{D})}\,, \tag{1.1}$$

where $\mathcal{M}$ indicates "model", i.e., hypotheses/theory, and $\mathcal{D}$ presents "data", or empirical knowledge and facts.

Bayesian theory

Classical probability theory is based on the study of independent trials. This means that knowledge of previous trials in one area has no influence on the prediction for a current trial in another, but similar area. Contrary to this, modern probability theory assumes that investigating previous observations of a similar model will very likely help us to derive better predictions for further experiments.

Since Helmholtz (1860) a basic assumption is that biologic (and also machine) vision computes the most probable interpretation(s) from input images. Let I be an image and X be a semantic representation of the world. Then, with the conditional probability $P(X|I)$ holds:

$$X^* = argmax\Big\{P(X|I)\Big\} \,. \tag{1.2}$$

In numerical statistics, usually the posterior is sampled and multiple solutions are kept:

$$(X_1, X_2, ..., X_k) \sim P(X|I) \,. \tag{1.3}$$

When studying a physical process, one often has only the observation data available rather than the underlying distributions. Statistical inference is used to estimate the adequate model along with the associated parameters based on the observed data. Bayesian inference is a means for statistical inference using Bayes' theorem. Observations, also called evidence, are used to infer the probability that a hypothesis is true.

As our knowledge about the underlying model and its parameters is "updated" along with the accumulation of more and more evidence, this can be also seen as a learning process about the statistical characteristics of the parameters from the observations.

Inference and Learning

Assume that we want to estimate the parameter θ based on observations $x = \{x_1, x_2, ..., x_n\}$. Bayes' theorem shows how to update the probability, or in other words to infer the posterior, with given evidence as follows (cf. also Equation 1.1):

$$P(\theta|x) = \frac{L(x|\theta)p(\theta)}{P(x)} \,, \tag{1.4}$$

where

- $p(\theta)$ is the prior of the parameter θ;
- $L(x|\theta)$ is the likelihood function (the conditional probability of observing evidence x given the parameter θ);
- $P(\theta|x)$ is the posterior (the conditional probability of a hypothesis given the observed evidence).

The denominator

$$P(x) = \int_{\Omega} L(x|\theta')p(\theta')d\theta' \tag{1.5}$$

is the marginal probability of x, i.e., the probability of observing x under all possible θ. This integral can also be seen as a normalization term to make sure that the posterior integrates to unity. Since $P(x) \geqq P(x \cap \theta) = L(x|\theta)p(\theta)$, the updated posterior will never become larger than 1. And more important, as it does not depend on θ, this term can be treated as constant when optimizing the posterior.

The term

$$\frac{L(x|\theta)}{P(x)}$$

describes the influence of the evidence on the belief in the hypothesis. It shows that better evidence, i.e., evidence that supports the proposed parameter, leads to a higher posterior probability for the hypothesis.

While evidence accumulates, there are two kinds of learning processes based on Bayesian inference:

- Discriminative learning is based on the posterior $P(\theta|x)$: The degree of belief in a hypothesis tends to become either very high or very low making discriminative learning suitable for decisions (accept/reject) or classification.
- Generative learning employs the joint probability $P(x, \theta)$: Knowledge, i.e., priors, about the proposed parameter(s) θ of the model is improved to the posterior by integrating verified (via the likelihood function) evidence x.

Taking the classification task as example, generative learning creates explicit models, which represent the training data of the object category. A generative classifier learns the prior $P(\gamma)$ and the likelihood $P(x|\gamma)$ of the classes γ and classifies x by maximizing $P(\gamma|x) \propto P(x|\gamma)P(\gamma)$. A discriminative approach, in contrast, learns the posterior $P(\gamma|x)$ directly. It finds the best classifier for the given data set focusing on discriminative characteristics between individual categories. Current approaches often combine these two concepts to deal with the classification problems with relatively similar categories. I.e., they learn object categories generatively and train models for each category discriminatively.

The predominant feature of Bayesian inference is that the predictions for the parameters are in the form of probability distributions (posteriors) instead of point predictions for conventional approaches. Usually, the posterior has a low entropy and the effective volume of the search space is relatively small.

Another advantage is the potential for automatic model selection (Bishop 2008): The additional uncertainty of a set of candidate models with different complexity can be addressed in

a Bayesian framework. A prior $P(M)$ is given for the different models and the posterior for the models based on the observation X can be expressed as:

$$P(M|X) \propto P(M)P(X|M) \,. \tag{1.6}$$

The model which best balances complexity and goodness of fit is determined based on optimizing the "marginal likelihood", also called "model evidence" $P(X|M)$, presenting the preference of the evidence for the candidate model M (Bishop 2008).

Optimization

Based on Bayesian inference, the objective of the optimization task is then the posterior. In comparison with the often used maximum likelihood (ML) estimation, the Maximum A Posteriori (MAP) estimation does not only observe the goodness of model (i.e., the likelihood), but also takes the plausibility of the proposed parameters (the priors) into account.

Let X be the observations and Θ the space of parameters. Then, the likelihood function can be expressed as

$$\Theta \mapsto L(\mathcal{D}) = L(X|\Theta) \,. \tag{1.7}$$

The maximum likelihood estimate of Θ is

$$\widehat{\Theta}_{ML} = arg_{\Theta}max\Big\{L(X|\Theta)\Big\} \,. \tag{1.8}$$

Based on Bayesian inference, the posterior distribution of Θ can be written as:

$$\Theta \mapsto P(\Theta|X) = \frac{L(X|\Theta)p(\Theta)}{\int_{\Omega} L(X|\Theta')p(\Theta')d\Theta'} \tag{1.9}$$

with p the prior and Ω the domain of Θ.

The denominator of the posterior does not depend on Θ. Therefore, it can be seen as a constant in the optimization. The maximum estimate for the posterior is thus equivalent to that of the numerator:

$$\widehat{\Theta}_{MAP} = arg_{\Theta}max\left\{\frac{L(X|\Theta)p(\Theta)}{\int_{\Omega} L(X|\Theta')p(\Theta')d\Theta'}\right\} = arg_{\Theta}max\Big\{L(X|\Theta)p(\Theta)\Big\} \,. \tag{1.10}$$

MAP estimation can be seen as an extension or a regularization of ML estimation: The empirical knowledge about the parameters is considered as well by integrating the prior information of them. Using the posterior also lowers the entropy of the target function to be optimized and thus the search space is narrowed down to a smaller, often more reasonable region.

Because of the augmented optimization objective, computation of the MAP estimate often has to deal with more complicated (very likely non-analytic) distributions in a much larger search space. A general numerical solution for the computation of MAP estimates is to use Markov Chain Monte Carlo (MCMC) techniques (cf. Section 1.2.2).

Please note that although MAP estimation works along with Bayesian inference, it is, strictly speaking, not a Bayes estimator, because MAP estimation provides a point estimate of the posterior distribution's mode while the prediction from Bayesian methods should be in the form of probability density functions (PDF). We prefer to regard MAP estimation as one decision rule for general purpose based on Bayesian theory. MAP estimation focuses on minimum overall error. If we pay more attention to some specific errors or costs, a minimum Bayes risk rule with loss function, which is a true Bayes estimator, should be considered.

1.2.2 Markov models

Markov models are widely used to simulate physical and natural processes. They are, more importantly here, closely connected with Bayesian statistics. The property of Markov models, i.e., that they can be represented by a directed graph with a dependency of nodes, makes them an appropriate "carrier" of Bayesian inference.

In this thesis, Markov models are used for two purposes: Modeling and calculation. Approximation methods like MCMC and its variants are employed for optimization tasks in high-dimensional solution spaces. Markov chains and Markov random fields are directly used to model trajectory data (cf. Chapter 4) and building networks (cf. Chapter 3), respectively. The computation of Bayesian models can be NP-hard (Cooper 1990). In most applications approximation is the practical or even the only way to compute the solution.

Markov Chains

A Markov Chain is a mathematical model for stochastic systems whose states, discrete or continuous, are governed by transition probabilities. It consists of, as shown in Figure 1.3, a series of random variables, named states $X_0, X_1, ..., X_n$ and can be denoted by

$$(\Omega, P(X_0), K) \ ,$$

with

Ω: The state space;

$P(X_0)$: The initial probability – the marginal distribution for X_0, which specifies the start state;

K: The kernel – the transition probability from the previous state(s) to the current state.

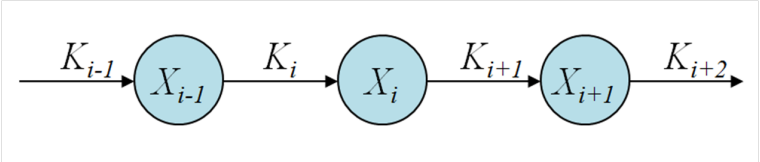


Figure 1.3: A first order Markov Chain.

A Markov process starts from the initial state X_0 and moves successively to the next state controlled by the transition probability. Each move is usually called a "step", as the Markov process can be seen as a walk in the state space Ω. Giving equal probability to every potential step, the Markov process represents a random walk. The main property of the Markov process is that the current state only depends on the most recent previous states. The typical Markov Chain, namely the first order Markov Chain (Figure 1.3), can be formulated as:

$$X_t|X_{i-1}, ..., X_0 \sim P(X_i|X_{i-1}, ..., X_0) = P(X_i|X_{i-1}) \,, \tag{1.11}$$

in which only the last state influences the current one.

The kernel K is defined as the conditional probability for subsequent variables, i.e., the probability of X_i given X_{i-1}:

$$K_i(X_i, X_{i-1}) \equiv P(X_i|X_{i-1}) \ .$$

The simplest form of a Markov Chain is the homogeneous (or stationary) one, in which the transition probability is constant for all states (independent on time or position $i \in \{0, n\}$):

$$P(X_i|X_{i-1}) = const. \tag{1.12}$$

For Markov Chains in discrete state spaces, we denote the transition probability by $T = T(X_i, X_{i-1})$. The probability of state X_i as derived from that of the previous state is given by:

$$P(X_i) = \sum_{X_{i-1}} P(X_{i-1})T(X_i, X_{i-1}) \,. \tag{1.13}$$

For continuous state spaces, the representation of a Markov Chain can be expressed using the integral kernel K:

$$P(X_i) = \int P(X_{i-1})K_i(X_i, X_{i-1})dX_{i-1} \,. \tag{1.14}$$

Stability of Markov Chains

Stability means that starting from any initial state, a Markov Chain will convergence to an invariant distribution, i.e., a stationary probability. This is a very important and attractive feature for a Markov Chain.

In practical applications, the target distribution $p(x)$ often cannot be directly observed. The goal of designing Markov Chains is then to devise a stationary probability that represents $p(x)$. I.e., samples x_i from Markov Chain states X_i simulate samples drawn from the target distribution.

The invariant distribution should be reached by

$$P(X_0)K^n \xrightarrow[n\to\infty]{} P(X)\,, \tag{1.15}$$

where n indicates the number of times that K has been multiplied. Stability means that after a limited number of transitions (steps), any initial distribution will converge to the stationary distribution. This implies that the stability of Markov Chains depends on the transition probability K. To ensure stability, K must be constructed guaranteeing the following two properties:

1. Ergodicity: A Markov Chain is called ergodic if it is possible for the chain to explore the whole state space. I.e., the probability of visiting all other states is always positive. Ergodic Markov Chains are often also called "irreducible".

2. Aperiodicity: An ergodic chain is called aperiodic (or acyclic) if there does not exist a periodic structure in the chain.

Often, Markov Chains are employed to efficiently guide the "routine" of sampling, i.e., biasing the walk towards interesting regions, and to converge as soon as possible.

Markov Chain Monte Carlo

Markov Chain Monte Carlo, or in short MCMC, is a class of statistical sampling algorithms widely applied as a general purpose computing technique for probability model simulation, integration, and optimization. It has been introduced in physics in the late 1940's and nowadays plays an important role in statistics, computer science, and econometrics. The main advantage of MCMC techniques is that they can sample efficiently in large spaces with high dimensionality.

As simulation-based approximation techniques, sampling methods like MCMC are always connected to the Bayesian applications in the fields of artificial intelligence and pattern recognition. There, practical models are usually too complex to be processed in closed form by Bayesian statistics and efficient simulation algorithms are needed for approximation. In many practical (statistical) applications it is hard to make exact inferences for the employed probabilistic models. Approximation by means of numerical sampling such as Monte Carlo technique is, thus, very important.

Monte Carlo sampling is named after the town Monte Carlo, which is world famous because of the luxurious casino. The essential idea of Monte Carlo sampling is random sampling. It can be traced to a rudimentary version invented by Fermi in the 1930s, even earlier than the first computer appeared (Robert and Casella 2011). Its first well-known practical application was for building the first nuclear reactor in 1942.

Monte Carlo simulation draws a set of samples $x_i, i = 1, ..., n$, from a target distribution $p(x)$ in a space Ω. The samples are independent and identically-distributed (iid) random variables.

The n samples can be used to approximate the target density of a function $f(x)$

$$E\{f(x)\} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \xrightarrow[n\to\infty]{} \int_{\Omega} f(x)p(x)dx\,, \tag{1.16}$$

where $E\{f(x)\}$ is an unbiased estimate of $f(x)$. According to the law of large numbers, E will converge to the integral of $f(x)$.

Theoretically, if the sampling is dense enough, the Monte Carlo samples can be directly used for optimization:

$$\widehat{x} = \arg_{\Omega'} \max p(x_i) \xrightarrow[n\to\infty]{} \arg_{\Omega} \max p(x)\,, \tag{1.17}$$

where Ω' is the sample space, $\Omega' \in \Omega$.

In practice, however, one often encounters complicated and/or combined distributions with high-dimensional sampling spaces, and then the inefficiency of Monte Carlo sampling can be fatal. More sophisticated techniques which can guide the sampling routine, e.g., MCMC, are needed.

As a general sampling method, MCMC is actually applied for both Bayesian and frequentist statistical inference. In Bayesian statistics, MCMC methods are primarily employed for numerical approximation.

MCMC generates samples x_i for the states X_i in the state space Ω using a Markov Chain guiding the sampling. The above defined stability of a Markov Chain is advantageous to ensure convergence. A typical MCMC is illustrated in Figure 1.4. The current state X_i is determined by the transition kernel K and in first order Markov Chains additionally only by the last state X_{i-1} ($X \sim P(x)$). The variable x_i is drawn stochastically ($x \sim U[a, b]$, iid) for each state X_i.

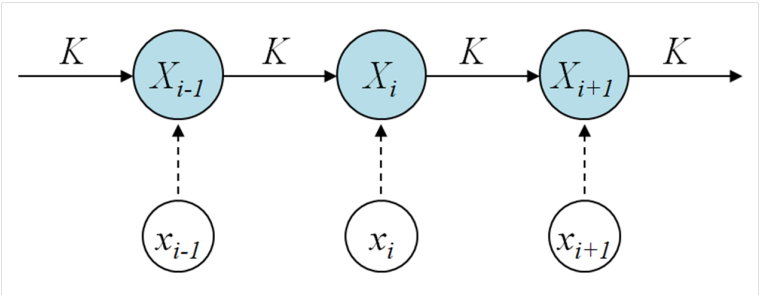


Figure 1.4: MCMC with first order Markov Chain

MCMC samplers are Monte Carlo samplers which ideally employ ergodic and aperiodic Markov Chains. The invariant distribution derived from the stationary chains can be used to

approximate the target distribution $p(x)$. One way to construct an appropriate sampler is to satisfy the reversibility, i.e., **detailed balance**, condition:

$$p(X_i)T(X_{i-1}, X_i) = p(X_{i-1})T(X_i, X_{i-1}) \,, \tag{1.18}$$

which is a sufficient, but not a necessary condition to ensure that the invariant distribution of the Markov Chain is identical to the target distribution $p(x)$.

MCMC is used as a general purpose computing technique in the following areas:

1. Simulation: MCMC produces samples of the probability model underlying the target system and represents typical states of the latter. Many natural processes are inherently stochastic, but still follow essential rules, which can be expressed in the form of probability models.

2. Integration and Computing: The computation of integrals is a typical task in scientific computing. Monte Carlo integration is used to deal with distributions in very high dimensional spaces. The expectation c is estimated as follows:

$$c = \int_{\Omega} p(x)f(x)dx \;\; ; \quad \widehat{c} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \,, \tag{1.19}$$

   where Ω is the integral space, $\widehat{c}$ the approximation with Monte Carlo, and n the number of samples drawn from $p(x)$.

3. Bayesian Inference and Learning: MCMC is an important computing tool for maximum likelihood estimation (MLE) learning of parameters $p(x;\theta)$ as well as MAP estimation (cf. Section 1.2.1). Unsupervised learning with hidden variables (simulated from the posterior) needs simulation to find suitable models. As Bayesian inference has become an important framework in many areas, e.g., image understanding and computer vision, MCMC techniques are of increasing interest.

4. Optimization: MCMC is the most often used technique for searching the global optima of complex distributions, e.g., the Bayesian posterior probability (MAP).

In summary, MCMC is a general tool for many problems with high-dimensional solution space. In some cases it is considered as the only feasible approach that provides solutions within an acceptable computation time.

An overview of techniques based on Markov models is presented in Figure 1.5, which can help to "locate" the mentioned techniques in the context of the statistical framework.

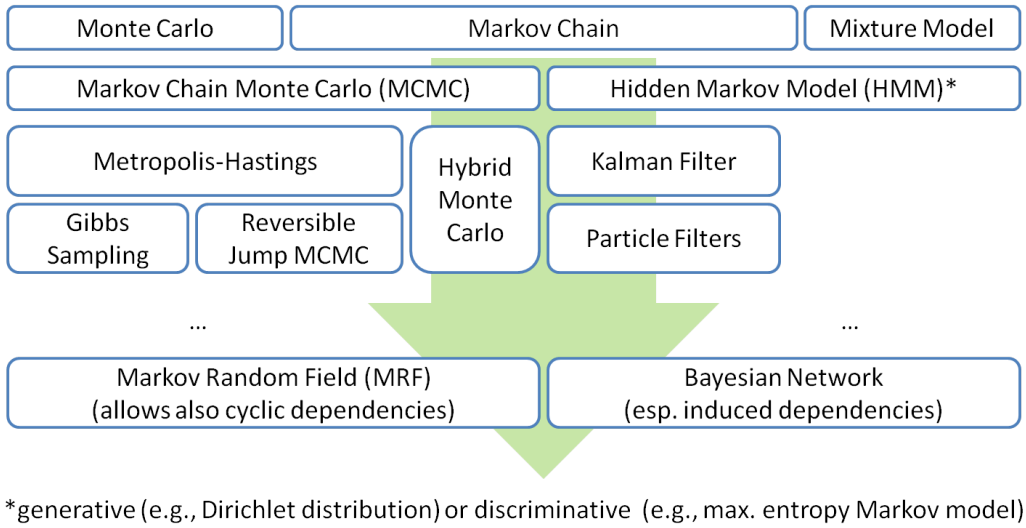


Figure 1.5: An incomplete "map" of Markov-model-based techniques: Locations and relationships

1.3 Scope and organization

This thesis is submitted in partial fulfillment of the requirements for a cumulative habilitation. It consists of two parts: (I) "Synopsis" – a summarizing synthesis of selected approaches and (II) a collection of the corresponding publications.

Selected research work that was performed and published by the author since the year of 2011 is summarized. Although the presented approaches apply various modeling and optimization/estimation techniques on a relatively wide spectrum of spatial data, they concentrate on one essential idea, namely using Bayesian statistical models for the understanding and interpretation of spatial data.

Figure 1.6 gives an overview of the publications presenting the employed data, key models and strategies as well as publication categories. For a convenient retrieval in the text, the publications are labeled by **[P#]** and marked when they appear.

From experience we have learned that the best way to expound statistical concepts and models is a demonstration with examples. The Synopsis part is organized in the form of "case-studies". The selected approaches are presented with more detail in the following chapters.

Chapter 2: Building reconstruction

Building reconstruction is one of the central topics in spatial data analysis and has been intensively studied for decades. There is a wide spectrum of data available for buildings including LiDAR and imagery from terrestrial, airborne, and space-borne platforms. Bottom-up approaches start from the data detecting basic geometric features and, thus, are susceptible to data flaws and uncertainty. Top-down methods use predefined parameterized models to fit the data, which can generally better deal with data artefacts, but often show limits because of the diversity of objects concerning both shape and appearance.

Figure 1.6: Overview over publications: Data, models, strategies, and publication categories

We employ a generative (top-down) statistical modeling scheme driven by Bayesian inference and optimization. We keep the advantages of top-down approaches: (1) The robustness of reconstruction and (2) the guarantee of complete models, while allow for (3) more flexibility to adapt the proposed method to a larger spectrum of objects. The flexibility is achieved through the inference and learning of Bayesian models. General and empirical knowledge can be integrated in the form of priors which are improved with local observations and constraints. I.e., more generic as well as flexible models are employed to cover a higher diversity of objects, which are in this case buildings with various roof types. The incremental learning integrated in MCMC leads to an efficient and powerful search in the parameter space, which renders an effective reconstruction in complex scene without additional support (such as footprints) possible. The degree of automation in building reconstruction is also improved by this means. In addition, bottom-up methods are utilized to provide reasonable priors, which improves the efficiency of the top-down inference and leads to an optimum overall performance.

Chapter 3: Building classification

VGI has become an important source of geospatial data. In comparison with conventional maps and cadastral data, they have better availability (in respect to both source and price), more spatial coverage, and higher update rate because of the entries of volunteers. For the same reasons, however, VGI such as OSM suffers form the lack of completeness and the inhomogeneous definition of the object attributes. Methods to enhance as well as complete the attribute entries are, therefore, highly desirable. This work focuses on determination of the usage (use and occupancy) information of buildings. Building usage is of great interest, e.g., for city planning, navigation and emergency management, but often not available in volunteered data.

To tackle this, the urban area is considered as a network consisting of individual buildings connected based on their neighborhood relationship. A Markov Random Field is employed to model the building network. Bayesian inference is employed to (1) derive usage attributes from existing local (geometric) features and contextual constraints of the neighborhood and (2) propagate them to other buildings in the network. By these means, the buildings are classified and enhanced with the new attribute, which is uniformly defined and covers the whole area. The proposed framework is fully automatic with the potential to be extended to other attributes.

Chapter 4: Anomaly detection in trajectories

Trajectories are space- as well as time-referenced data. Their availability has improved rapidly with the wide use of navigation and mobile communication systems equipped with GPS and motion sensors. Anomalous behavior detection is a challenge concerning both the spatial and temporal information embedded in the trajectories. It is of wide interest for navigation, driver assistance and surveillance. Current related approaches work based on a number of labeled data to train a classifier and apply it to the incoming data. Such a large training dataset, however, is neither always available nor guaranteed to cover all possible cases.

We designed an iterative Bayesian filter for GPS trajectories from cars. This operator infers anomalous behaviors (1) locally inside a single trajectory and (2) dynamically over time. There is no previous training required, though, if training data are available, they can be integrated as priors in the Bayesian framework. A Hidden Markov Model (HMM) is employed to present the trajectory as a carrier of Bayesian inference between the GPS nodes. The HMM has dynamic orders of Markov Chain to derived spatial and temporal features including “turns”, “detour factor”, and “route repetition”, which require particular long-term perspectives. Additionally to individual behaviors, a collective behavior can be derived based on the individual results which can indicate special traffic situations such as road-blocks and turn restrictions, where anomalous behaviors often occur.

Chapter 5: RGBD segmentation

RGB-Depth (RGBD) data are introduced in recent years with easy acquisition and relatively low-cost sensors. Since most RGBD sensors work in close range, indoor scene interpretation

and robot vision are of particular interest. Segmentation is the basis of scene interpretation. Although 2D image segmentation has been thoroughly studied for decades and a number of approaches have also been reported recently for RGBD data, a segmentation of complex scenes on object-level is still challenging, especially when objects with concave or irregular shapes are involved.

In comparison with most related approaches where the depth information is treated as an additional channel besides the colors on a 2D basis, we conduct full 3D parsing by introducing a novel quasi-3D model, the SVP (synthetic volume primitive). Objects are inferred not only based on color and position, but also on the underlying 3D volume intersection, which cannot be directly observed. A Markov Random Field is employed to represent the SVPs and their relationship in the scene and optimized by an MCMC sampler to find the best segmentation. By all these means free-shape including concave objects can be segmented well even in complex scenes.

These chapters are specifically structured so that the cases are clearly presented and can be easily compared with each other. Every chapter starts with a cover page including a quick overview with an illustration and a set of keywords for fast indexing of concepts and techniques. The first section of each chapter consists of an individual ***problem statement*** pointing out issues and challenges of the case. The following two sections present the ***data*** and the employed ***models***, respectively. Specific topics of interest, e.g., model selection, belief inference, and global optimization, are distributed in different chapters and discussed along with their practical use. General issues including adaptability of models and modeling strategies are given in Chapter 6 together with the final conclusions. In the chapters all challenges and questions have an item label "□", while the proposed solutions and answers are marked with "■".

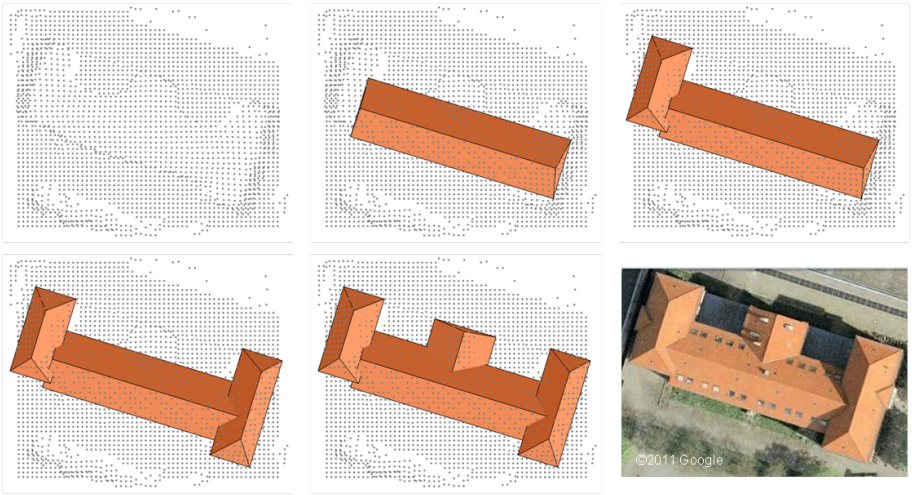
The full versions of the publications are compiled in Part II. For an easy indexing, multiple criteria including publication date, publication category as well as topic are employed.

Chapter 2

Building reconstruction

Overview

This chapter presents statistical approaches for automatic 3D building extraction and reconstruction based on 2.5D point clouds from laserscanning or stereo image matching. A hybrid framework of bottom-up and top-down methods is proposed with emphasis on the latter in the form of generative models.



Building reconstruction: Input point cloud (a), the reconstruction process (b to e), and a reference image (f)

Keywords

Object detection, 3D reconstruction, Point cloud, Generative models, RJMCMC, Model selection

2.1 Problem statement

Building is one of the most important object classes in spatial data analysis. Reconstruction of 3D building models from measurement data is of great interest for various applications including city planning, tourism, navigation, and emergency management. Detailed building models containing roof and facade geometries can be integrated into building information models (BIM) or applied to, e.g., photovoltaic planing, urban heat island (UHI) effect research, or urban turbulence and wind gust analysis (for micro wind generators and micro aerial vehicles). Building reconstruction has, therefore, been intensively studied and many approaches have been reported in the past decades. In previous work, bottom-up methods, e.g., point clustering, plane detection, and contour extraction, have been widely used. Due to the data artefacts caused by occlusion, reflection from windows or water, etc., the bottom-up reconstruction in urban areas suffers basically from incomplete or irregular roof parts. Manually given geometric constraints are usually needed to ensure plausible results.

The approaches summarized in this chapter deal with the following core challenges in 3D building reconstruction:

- (**C1**) Robust reconstruction in spite of artefacts and uncertainty of the data
- (**C2**) Complete and plausible watertight building models
- (**C3**) High degree of automation

The basic idea is to utilize the robustness of top-down strategies against the uncertainty of the data and to enhance it with powerful statistical search and plausible Bayesian inference. The complete framework includes also the approach with bottom-up methods to improve the overall performance and efficiency.

2.2 Data – LiDAR and imagery

Laserscanning data

Airborne and terrestrial laser scanner data are widely used in state-of-the-art approaches for building reconstruction due to their high accuracy and density. Airborne laser scanning data of urban areas, however, often have the following issues concerning quality (Oude Elberink and Vosselman 2011): 1. Systematic and stochastic errors in the measurement, 2. variable and relatively low point density, and 3. data gaps and artefacts. As shown in Figure 2.1, data flaws in urban areas (left) can be caused by (A) complex and detailed roof structures such as HVAC (heating, ventilating, and air conditioning) equipment, (B) the absorption of the laser pulse by water and its reflection by windows on the roof, and (C) occlusion by neighboring objects (in typical European cities mostly trees).

As a result, an accurate determination of the roof edges and the recognition of small roof structures are hard. Bottom-up reconstruction may, thus, be limited to a number of incomplete and irregular roof facets or building parts. A regularization with given constraints is

Figure 2.1: Causes of laserscanning data artefacts in urban areas (left): Complex roof structures (A), reflections (B), and occlusion (C)

typically needed during the extraction or afterwards. In many cases building reconstruction is not easy even for humans. For regularized plane detection a probability-driven edge sweeping method is proposed in (Huang and Brenner 2011, **P1**). Although it works robustly in spite of clutter and data flaws, it encounters difficulties when processing complex roofs.

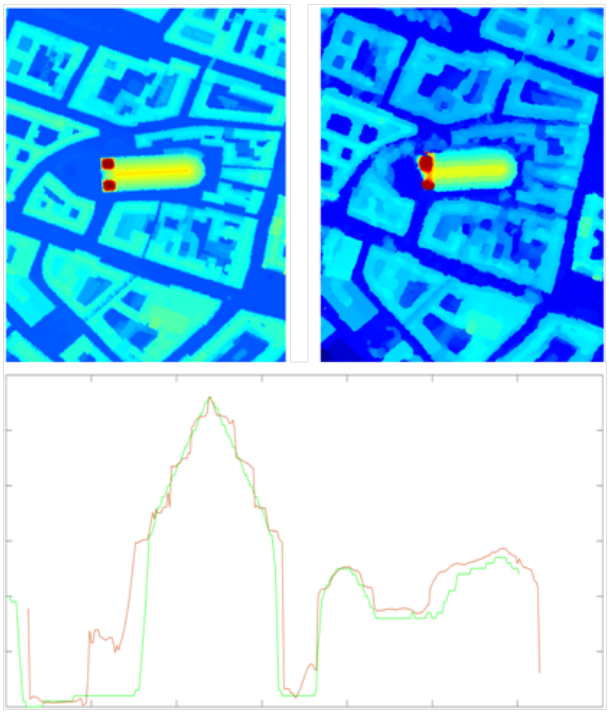


Figure 2.2: Comparison of LiDAR point cloud (left and green profile) and DSM generated from satellite images (right and red profile) (Partovi et al. 2015, **P7**)

Image data

Due to the improvement in spatial resolution of satellite imagery and the launch of numerous satellites there is a growing interest in algorithms for 3D point cloud generation by stereo

satellite image matching. Although the accuracy of the 3D point clouds from satellite images is generally lower than that of LiDAR data, we demonstrate that they can still be sufficient for building recognition and reconstruction.

The limited geometric resolution as well as the noise in the point cloud data generated from satellite imagery cause difficulties in the computation of geometrical features, e.g., normal vectors based on information of neighboring points. Thus, bottom-up methods, e.g., roof plane determination based on point cloud segmentation, is considered difficult, because of the incomplete and irregular roof planes which need geometric constraints for the reconstruction of plausible roofs. In Figure 2.2 the quality of DSM from satellite (particularly WorldView-2) imagery is compared with that from LiDAR as reference data (Partovi et al. 2015, **P7**). Elevation profiles (bottom) are given with the red line representing the DSM from satellite imagery (top right) and the green line that from the LiDAR point cloud (top left).

2.3 Model – generative models for buildings

Generative modeling can be used to (mostly randomly) generate observable data from unobservable parameter(s). Here, the observable data usually mean examples/instances that obey the underlying rules. Generative modeling is employed to estimate joint probability distributions (i.e., models defined by multiple parameters) or approximate conditional distributions via Bayesian inference. We employ generative modeling to find the best building model that fits the data. An overview of the work flow is presented in Figure 2.3.

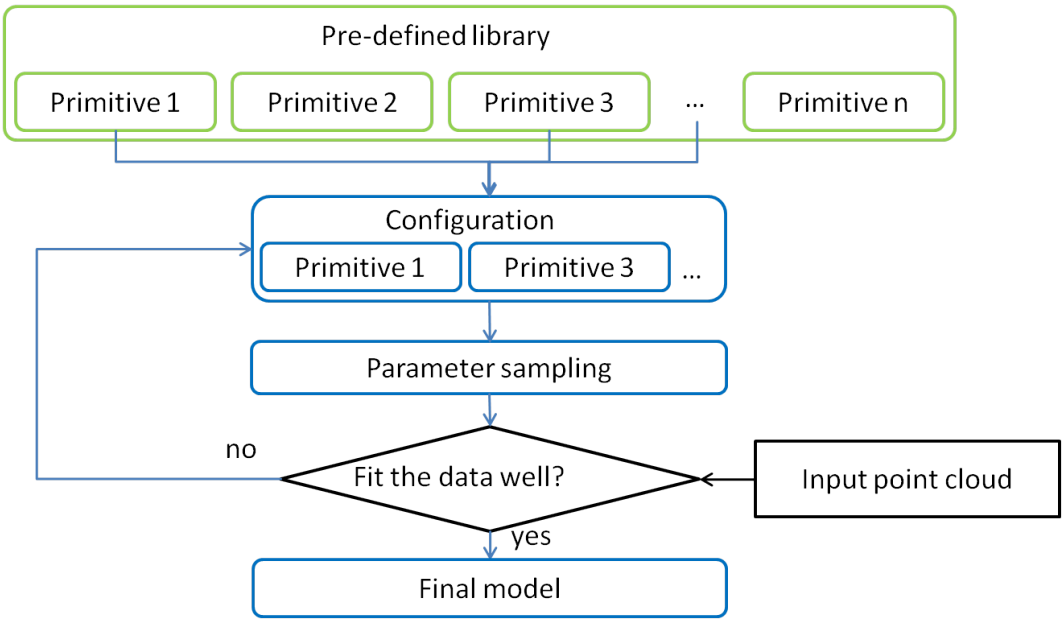


Figure 2.3: Workflow of building reconstruction via generative models

2.3.1 Primitive-based modeling

Generative modeling works based on parameterized primitives. In plane-based or primitive-based approaches the assembly of these "components" of buildings is always a crucial problem. Our strategy is to give more flexible definitions of the primitives as well as their interactions to obtain more stable reconstruction results. The basic idea is to allow overlap when combining multiple primitives. Basic geometrical rules are employed to ensure plausibility.

Figure 2.4 presents the novel primitive-based reconstruction scheme in comparison with conventional facet-based approaches by employing two example roofs. In facet-based modeling, the target roofs are seen as a group of facets (right), which are labeled with numbers. The derived Region Adjacency Graph (RAG, bottom) shows the organization of the facets and are used to guide the further reconstruction. Some roof parts, e.g., facet 10 in the first model and facet 3 in the second, however, are hard to interpret from the graph. In primitive-based modeling, on the other hand, the building roofs are considered as an assembly of regular primitives (left). The interpretation (bottom) is much simpler because not only because the related (member) facets have been predefined in the primitives, but also because the irregular fragments caused by plane intersections no longer need to be considered.

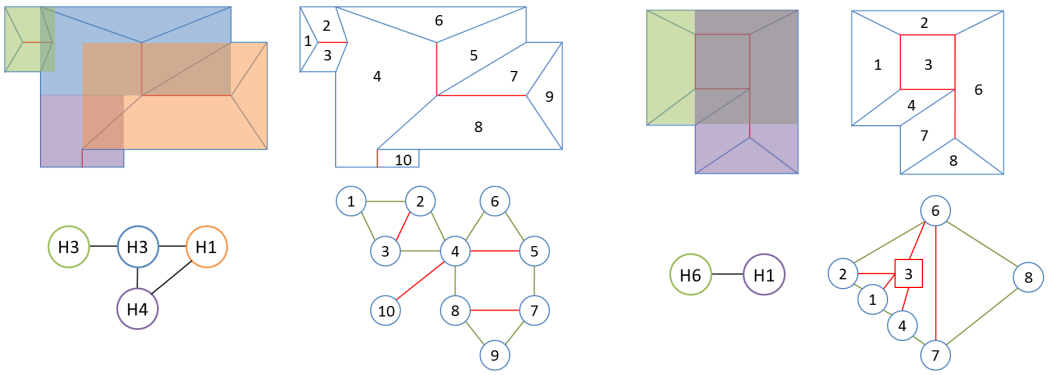


Figure 2.4: Primitive-based (left) vs. facet-based (right) approaches

The proposed scheme has two main advantages:

- Complete and plausible models: There is no irregular and incomplete roof facet or building block, which is usually caused by fitting erroneous bottom-up features like fragmented plane segments in facet-based modeling. The predefined constraints for the member facets of a primitive ensure a regularized reconstruction.
- Flexibility of modeling: A large number of different building types can be represented by a limited number of primitives and their combination. Complex roofs can be interpreted more easily.

Note that allowing overlap of primitives helps to keep the member primitives being complete during reconstruction and assembly (cf. Figure 2.4, colored blocks) instead of being cropped

to fit the ground plan or the neighbors. Besides horizontal we also allow vertical overlap (intersection). By this means, some combined roofs can be reconstructed (cf. also (Huang et al. 2013a, **P3**)) without adding new particular primitives to the library. In the final model, the redundant parts are hidden inside the assembly and can be easily removed afterwards with, e.g., CAD tools.

Primitive library and parameterization

Figure 2.5 presents a library with three groups including eleven types of roof primitives (more details cf. (Huang et al. 2013a, **P3**), Section 3.1).

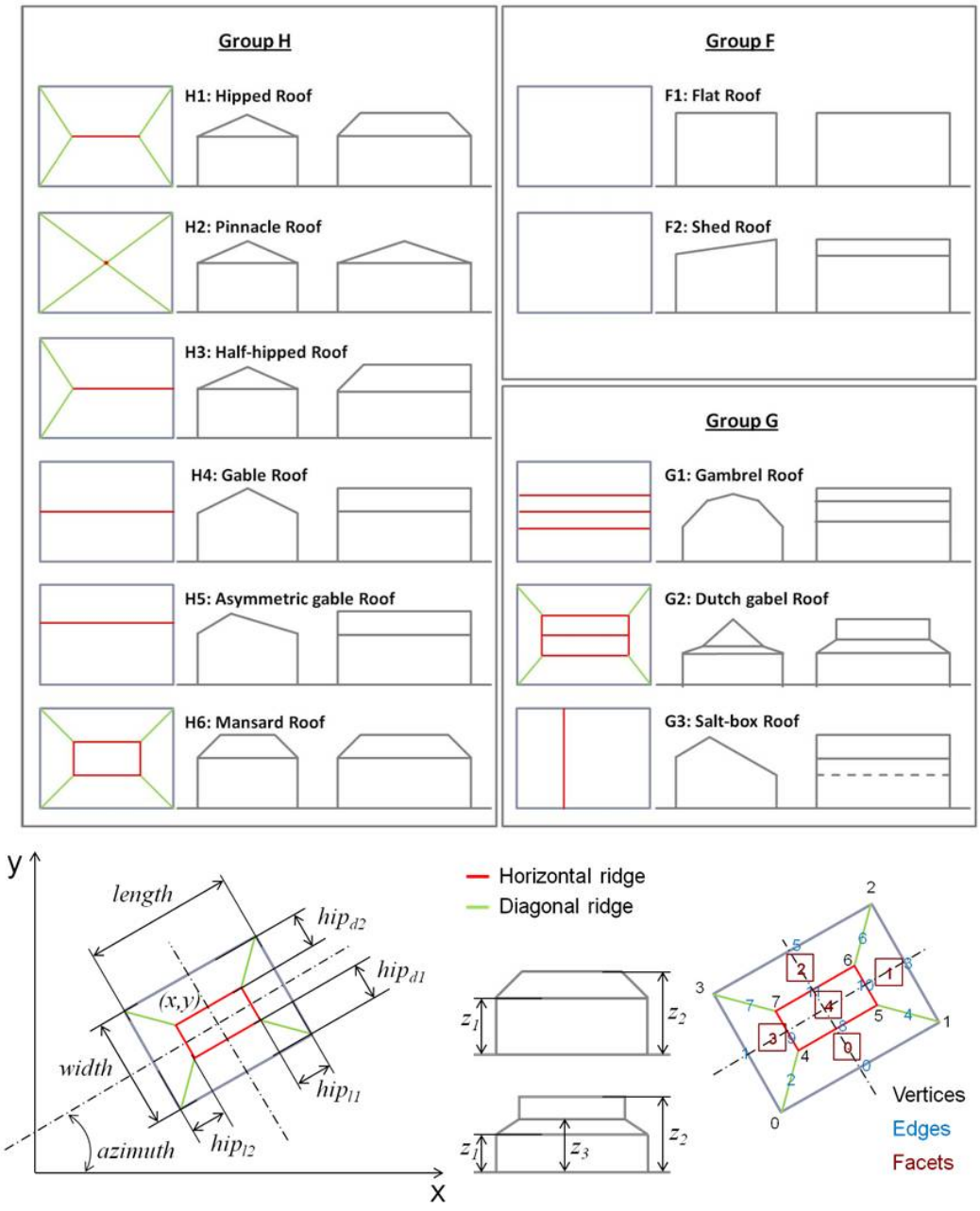


Figure 2.5: Library of roof primitives

The parameters θ of a primitive are defined as:

$$\theta \in \Theta; \Theta = \{\mathcal{P}, C, \mathcal{S}\}\,, \tag{2.1}$$

where the parameter space Θ (Figure 2.5, top right) consists of position parameters $\mathcal{P} = \{x, y, azimuth\}$ and contour parameters $C = \{length, width\}$, as all primitives are defined with a rectangular footprint. $\mathcal{P}$ and C have fixed members. $\mathcal{S}$ contains shape parameters, e.g., the ridge/eave height and the depths of hips, and depends on the primitive.

Geometrical features of different levels, i.e., vertices, edges, and facets, and their relationships are used to define the primitives (Figure 2.5, bottom right). They are the basis for primitive merging, calculation of reconstruction errors, drawing building footprints, etc.

Note that the library contains only a limited number of entries with planar shapes and rectangular footprints. For a more efficient reconstruction we prefer to keep the library simple and with a small number of primitives. The basic idea is to cover the majority of building types with a limited number of primitives. A simple building is presented with a single primitive. A complex building is modeled by a combination of multiple primitives. Planar roofs and rectangular footprints are chosen not only because they are simple (with few shape parameters), but also because they are the basic form that most buildings follow. I.e., roofs derived from the primitives or their combinations will cover most of the buildings in urban areas. Furthermore:

- The proposed primitive merging process is also able to represent buildings with non-rectangular footprints (cf. (Huang et al. 2013a, **P3**), Section 3.2).
- Elliptic roofs are not included, but can be approximated by gambrel roofs (cf. Figure 2.5, G1).

2.3.2 Extension for building generalization

3D building generalization, i.e., geometric simplification of 3D building models, is of great interest not only in cartography but also in fields such as computer graphics and GIS. It plays a central role to control the storage space as well as processing and transmission needs.

Model-driven approaches have advantages concerning completeness and plausibility. The presented building reconstruction scheme based on generative models can be easily adapted to create building models on different levels of detail (LoDs), the primitive library and the modeling rules provide a good basis. Different LoDs are usually achieved by means of geometrical modification of **existing models** with higher resolution. We, in contrary, generate **new models** with corresponding details instead. The primitives as well as the statistical search scheme are adapted to the LoDs. For the selection of primitives one can derive a multi-stage simplification routine, e.g., G2 $\rightarrow$ H1 $\rightarrow$ H4 $\rightarrow$ F1, where the jump routine is designed considering the number of parameters, which also implies their complexity and geometric inheritance (cf. Figure 2.7). For complex buildings consisting of multiple primitives, the models can be simplified by employing fewer and simpler primitives. An example is given in Figure 2.6.

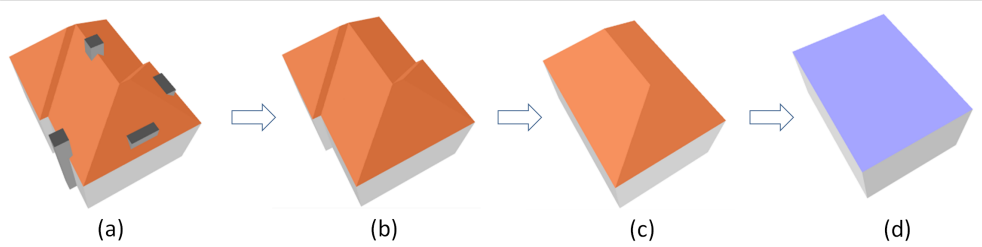


Figure 2.6: An example of building model generalization

In contrast to other approaches, the low LoD models are not derived directly from the high LoD ones but newly generated by replacing primitives with reduced number and complexity. As shown in Figure 2.6, model (b) is obtained by removing the primitives for superstructures while (c) is generated by re-sampling with less primitives. Flat roof is employed to represent the LoD1 model (d). By these means, the simplification is flexibly conducted in multiple stages and the models on every level maintain integrity and regularity. This approach corresponds to the "star"-approach in generalization (Stoter 2005), i.e., the different representations are generated from one original representation, which is in this case the raw data in terms of point clouds. It generates homogeneous representations on different LoDs, as it is based on the same set of primitives. The primitives are selected according to geometric criteria which characterize different resolutions.

2.4 Reversible Jump Markov Chain Monte Carlo

Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green 1995) is an extension of the MCMC algorithm with solution spaces of variable dimension. In RJMCMC, a sampler is allowed to switch (or called "jump") between subspaces with various dimension during the search. We utilize this concept and design a modified MCMC technique with a specified "jump" mechanism. The jumps, i.e., modeling with a varying number of parameters (dimension), are employed to represent the changes of the configuration (number as well as type of roof primitives).

In the included publications, the use of reversible jumps is first reported in (Huang et al. 2011, **P2**) for switching between different roof types, i.e., primitives. Bidirectional movements between roof types together with birth and death jumps can be found in the improved version (Huang et al. 2013a, **P3**), where all the possible jumps are defined by a mixed transition kernel:

$$\mathcal{K} = \{Sw_1, Sw_2, Bi, De\} \tag{2.2}$$

with

- Sw_1: "Switch case 1" to more complex primitive (with more parameters);
- Sw_2: "Switch case 2" to simpler primitive (with less parameters);

– *Bi*: "Birth" of a new primitive adjacent to an original one;

– *De*: "Death" of an existing primitive.

The "switches" are jumps between **different types of primitives**. Based on the characteristics of the primitives, we have narrowed down all possible switches into a specific "jump routine" shown in Figure 2.7. It ensures that each jump only changes a limited number of sensible parameters.

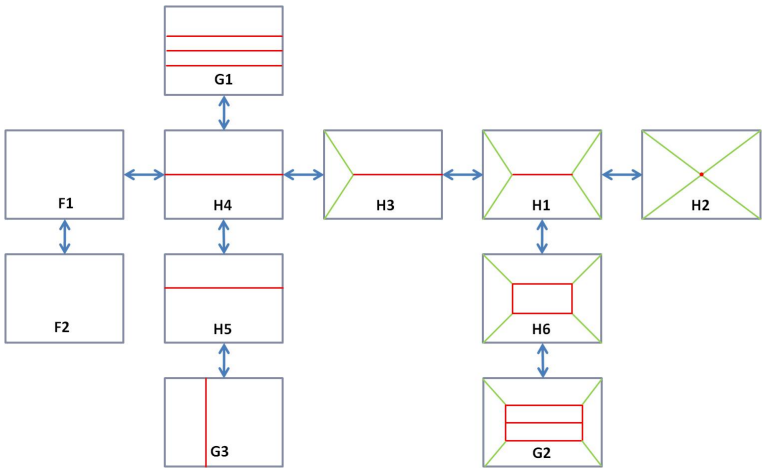


Figure 2.7: Jump routine: limited switches between primitives

Let $\mathcal{M}_i$ and $\mathcal{M}_j$ be two models in $\{\mathcal{M}_n; n = 1, ..., N\}$ with N the number of possible states, in this case the number of primitive types. The jump from i to j will be accepted according to the probability:

$$\mathcal{A}(\mathcal{M}_i, \mathcal{M}_j) = \min\left\{1, \frac{p(\mathcal{D}|\Theta_{\mathcal{M}_j})p(\Theta_{\mathcal{M}_j})}{p(\mathcal{D}|\Theta_{\mathcal{M}_i})p(\Theta_{\mathcal{M}_i})} \cdot \frac{\mathcal{J}_{i\to j}}{\mathcal{J}_{j\to i}}\right\} \tag{2.3}$$

with $\mathcal{J}_{i\to j}$ the Jacobian matrix, in which the proposal density, i.e., the probabilities for all the possible jumps from i to j, is coded. With the constraints given by the jump routine the Jacobian $\mathcal{J}$ is simplified to a fixed transition matrix for practical applications (Huang et al. 2013a, **P3**).

The "birth" and "death" are jumps between **different numbers of primitives**. Including all kinds of jumps defined in $\mathcal{K}$, the $\mathcal{M}_i$ and $\mathcal{M}_j$ can be extended to represent the states (i.e., the number as well as the types of primitives) before and after the jump. The "detailed balance" condition for the Markov Chain can be expressed as:

$$p(\mathcal{M}_i|\mathcal{D}) \cdot \mathcal{T}(\mathcal{M}_i, \mathcal{M}_j) = p(\mathcal{M}_j|\mathcal{D}) \cdot \mathcal{T}(\mathcal{M}_j, \mathcal{M}_i) \quad \forall \mathcal{M}_i \neq \mathcal{M}_j \tag{2.4}$$

with the posterior:

$$p(\mathcal{M}_i|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta_{\mathcal{M}_i}) \cdot p(\Theta_{\mathcal{M}_i})}{p(\mathcal{D})}, \tag{2.5}$$

where $P(\mathcal{D})$ is the marginal probability of the evidence and a constant. $\mathcal{T}(\mathcal{M}_i, \mathcal{M}_j)$ is the transition density from state $\mathcal{M}_i$ to state $\mathcal{M}_j$. The Markov Chain is constructed with the

Metropolis-Hastings algorithm as:

$$\mathcal{T}(\mathcal{M}_i, \mathcal{M}_j) = q(\mathcal{M}_i, \mathcal{M}_j) \cdot \mathcal{A}(\mathcal{M}_i, \mathcal{M}_j)\,, \tag{2.6}$$

where $q(\mathcal{M}_i, \mathcal{M}_j)$ indicates the proposal density for the jump between the states and $\mathcal{A}(\mathcal{M}_i, \mathcal{M}_j)$ the acceptance probability (cf. Equation 2.3).

The sampling procedure can then be summarized as follows:

1. Initialization: $(\mathcal{M}^{(i=0)}, \Theta^{(i=0)})$
2. Proposal of new state $\mathcal{M}\prime$
   - 2.1 Sampling configuration from $\mathcal{K} = \{Sw_1, Sw_2, Bi, De\}$
   - 2.2 Sampling parameters $\Theta\prime$
3. Accepting new proposal with the probability

$$\mathcal{A}(\mathcal{M}^{(i)}, \mathcal{M}\prime) = \min\left\{1, \frac{p(\mathcal{M}\prime|\mathcal{D})}{p(\mathcal{M}^{(i)}|\mathcal{D})} \cdot \frac{q(\mathcal{M}^{(i)}, \mathcal{M}\prime)}{q(\mathcal{M}\prime, \mathcal{M}^{(i)})}\right\} \tag{2.7}$$

4. $\mathcal{M}^{(i+1)} = \mathcal{M}\prime$ if accepted, otherwise $\mathcal{M}^{(i+1)} = \mathcal{M}^{(i)}$

The sampling is able to explore a wide variety of hypotheses (models) as long as the jump kernels keep the balance condition, i.e., a reversible move to the previous state is possible. Although this means that the sampling implicitly selects models, it is usually time-consuming. We, therefore, employ an explicit model selection mechanism (cf. Section 2.5) in the transition kernel. The information entropy of the model $H_{\mathcal{M}}$ is calculated as:

$$H_{\mathcal{M}} = k^{\beta} - 2ln\left(L(\mathcal{D}|\mathcal{M})\right)\,, \tag{2.8}$$

where L is the likelihood (evaluation) of the model and β represents the sensitivity concerning the number of parameters k (model complexity). β is set to a small and empirically found value of $1/12$, as we prefer a more detailed reconstructions than simpler models.

The acceptance probability in the kernel can, thus, be expressed as:

$$\mathcal{A}(\mathcal{M}_i, \mathcal{M}_j) = \min\left\{1, \frac{H^{-1}_{\mathcal{M}_j}}{H^{-1}_{\mathcal{M}_i}} \cdot q(\mathcal{M}_i, \mathcal{M}_j)\right\}. \tag{2.9}$$

In the presented sampling process, jumps are actually not proposed in every MCMC move but only when the maximum likelihood for the current configuration (with a determined primitive type as well as parameters) has been found. Note that the specified schedule and the jump routine keep the search from reaching arbitrary states in the solution space but stay in a "subspace of interest". I.e., the jumps break the stationary rules (irreducible and aperiodic) of the Markovian process by giving up the irreducibility, because of which the whole scheme is no more a conventional RJMCMC. In return, the optimization becomes more efficient and stable with the reduced search entropy.

2.5 Model selection

Model selection attracts more and more attention in spatial data analysis. The improvement of the data encourages more sophisticated representations and models. The goal of model selection is to choose the most appropriate model balancing the goodness of fit to the data and model complexity. In this section two model selection concepts based on Bayesian statistics and information theory are presented. Both of them are employed in the proposed approaches.

2.5.1 Bayesian model selection

As stated in Section 2.4, the search driven by Bayesian statistics implicitly selects models. Bayesian model selection employs probability theory to select candidates from the hypotheses. As given in Equation 2.10, the likelihood, which indicates the evaluation of the candidate model, is computed by integrating the unknown parameters.

$$P(\mathcal{D}|M) = \int_{\theta} p(\mathcal{D}|\theta)p(\theta|M)d\theta \tag{2.10}$$

The likelihood distributions of constrained models show better concentration and, therefore, reach higher probabilities because of the higher certainty of parameters. This implies that the preference of simpler models (with highly constrained parameters) over complex models (with flexible parameters) is inherently integrated. In Bayesian framework the balance of model complexity is "automatic", where over-fitting by using complicated models is always penalized (Kass and Raftery 1995).

The plausibility of two different models M_1 and M_2 are compared and assessed by the Bayes factor:

$$F_{Bayes} = \frac{P(\mathcal{D}|M_1)}{P(\mathcal{D}|M_2)} = \frac{\int_{\theta_1} p(\mathcal{D}|\theta_1)p(\theta_1|M)d\theta_1}{\int_{\theta_2} p(\mathcal{D}|\theta_2)p(\theta_2|M)d\theta_2} \,. \tag{2.11}$$

Please note that the process looks similar to, but should not be confused with maximum likelihood estimation. In the latter, the objective of estimation is the set of parameters of one model, while the Bayes factor, as given in Equation 2.11, integrates over the parameter sets θ_1 and θ_2 of two different models.

2.5.2 Information entropy and model size estimation

Information-entropy-based model selection employs information theory to find the best candidates. In the context of information theory, the entropy (or "Shannon entropy") indicates the expected value of the information contained in each message. The Akaike information criterion (AIC) (Akaike 1973) and its variants are the most-commonly-used criteria for model selection besides the Bayes factor.

An explicit model selection is performed to deal with the problem of the **model-minimizing tendency**. The latter is a typical and tricky issue for stochastic modeling, i.e., using the deviation at data points to evaluate the hypothetical models, the finally found “best" model tends to shrink to a very small size as a smaller model (corresponding to smaller comparison area and thus fewer data points for the evaluation) means less error. It follows that the model size should be reasonably estimated.

To overcome the model-minimizing tendency, a reasonable constraint for the model size is needed. We conduct an estimation employing an information criterion to balance the goodness of fit to the data and the size of the model.

Figure 2.8 (left) shows the average deviation (blue) from the proposed model to the data points depending on the model size indicated by K, which is the number of data points in the domain of the proposed model implying the model size (linear proportion for raster data). Please note that this function is neither linear nor monotonically increasing due to the influences of data flaws and clutter.

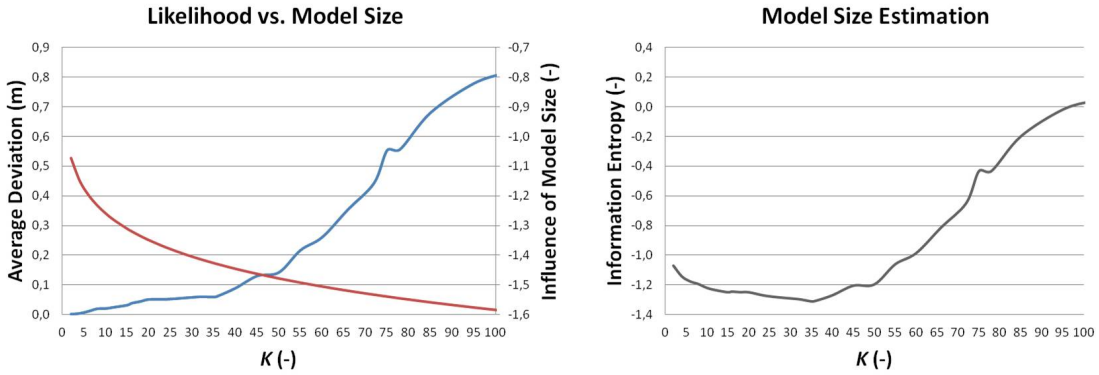


Figure 2.8: Plots of the average deviation (blue), the influence of model size $-K^{\alpha}$ (red) and the information entropy (black) over K. The minimum entropy indicates the optimal balance of goodness of fit and the model size.

We follow the basic idea of AIC,

$$AIC = 2k - 2ln(L)\,, \tag{2.12}$$

to build our goal function. In Equation (2.12), k indicates the number of parameters, which implies the complexity of the model, and L the maximum likelihood for the optimized parameters. However, in our case we do not want to prevent the model being too complicated, but the size of the model being too small. We use the evaluation of the model for L and $-K^{\alpha}$ to represent the influence of the model size. The actual number of involved points is more suitable than the area A to model the influence as the likelihood is also calculated with these points. α is the influence factor for K, which has a relatively large impact because of the usually large number of K. To balance the influences of both the model size and the goodness of fit, it is empirically determined $\alpha = 0.1$. The total information entropy of the

proposed model ($\mathcal{M}$) can then be expressed as:

$$H_{\mathcal{M}} = -K^{\alpha} - 2ln\left(L(\mathcal{D}|\mathcal{M})\right) \ . \tag{2.13}$$

By these means, a trivial improvement in fit at the cost of a decrease in model size is discouraged to prevent the model-minimizing tendency. The information criterion shown in Figure 2.8 (right) is employed to guide the parameter optimization, which is, in this case, a trade-off between reconstruction accuracy and model size (instead of model complexity, cf. Equation 2.12).

2.6 Related work

Many approaches for the reconstruction of 3D city models have been reported in the past decades. The introduction of laser scanning makes the acquisition of 3D data easier and more accurate. Overviews are given by Vosselman (2009) and Musialski et al. (2013). In this section related approaches are grouped according to the main strategy, i.e., bottom-up or top-down. A discussion of the pros and cons as well as the balance of bottom-up and top-down is given in Chapter 6.

Bottom-up approaches

Current bottom-up approaches include (Rottensteiner et al. 2008), in which roof plane delineation from LiDAR data is presented. Statistical tests and robust estimation are employed for stable edge detection in spite of clutter. Using manually given constraints, topological correctness is ensured without additional 2D data. Sampath and Shan (2010) segment and reconstruct more complicated buildings from airborne LiDAR point clouds based on polyhedral models. First, outlier points are detected by means of eigenanalysis, making the roof planar segmentation more robust. The latter is implemented through an extended fuzzy k-means clustering. An adjacency matrix is derived after segmentation. For reconstruction, the roof vertices, ridges, and edges are determined by intersecting the corresponding planes. Also starting from planar roof segments, Zhou and Neumann (2012) try to organize them as well as roof boundary segments with "global regularities" considering orientation and placement constraints. (Matei et al. 2008) and (Poullis and You 2009) present fast processes to generate simplified building models of large-scale urban areas. The input LiDAR data is segmented producing regularized buildings or building parts and simple polygon models are used for an efficient reconstruction. Meng et al. (2009) introduce a method to identify individual buildings from airborne laser data based on morphological processing. Algorithms are developed to separate ground points and then filter the other non-building parts (mostly the vegetation). Duan and Lafarge (2016) report a large-scale LoD1 urban reconstruction from stereo pair of satellite images. Instead of the conventional use of DSM (calculated from the image pair), a reconstruction directly from the imagery is presented with the consideration of semantic information for the improvement of performance.

Top-down approaches

Approaches employing top-down methods have been increasingly reported over the last several years. In (Verma et al. 2006), parametric modeling is employed for detection and reconstruction of 3D building models from airborne laser data. Relatively complex buildings can be represented by combining simple parametric roof shapes. Lafarge et al. (2010) present building reconstruction from a DSM combining generic and parametric methods. Buildings are considered as an assembly of 3D parametric blocks. 2D-support (approximate building footprints) is produced manually or automatically (Ortner et al. 2007). Based on the 2D support 3D blocks are assembled and optimized within a Bayesian framework. To deal with more sophisticated buildings, basic geometric primitives, e.g., planes, cylinders and cones are extracted and combined with mesh-patches to present irregular roof forms (Lafarge and Mallet 2012). The approach is extended with a semantic scene description to model urban environment including buildings, trees and ground surface. Based on extracted plane hypotheses, Li et al. (2016) model the geometry of buildings with a set of aligned boxes, which are represented and labeled (as parts of the buildings) via Markov Random Field. Kelly et al. (2017) introduce a framework of urban reconstruction in LoD3 based on a fusion of multiple sources of data including building footprints (GIS database), meshes and terrestrial imagery. The footprints are partitioned and the parameterized building elements are fitted in. A Convolutional Neural Network (CNN) is employed for the detection of facade elements, e.g., windows and doors.

Comparison of the proposed approach to related work

In comparison with approaches that share the “LEGO" scheme (Kada and McKinley 2009, Lafarge et al. 2010), i.e., the building is first cut down into parts and primitives are found that fit the parts, we have designed new combination rules and a novel merging process which allows overlap of primitives. By this means, a more plausible as well as stable result is obtained, because all primitives remain complete during the combination and deviations caused by random sampling can be compensated.

Unlike most of the related research, bottom-up processing, e.g., point clustering, plane detection (Sampath and Shan 2010), or 2D building footprint data (Kada and McKinley 2009, Lafarge et al. 2010), is not required in the proposed work. To compensate for the absence of the initial information (which is normally provided by the bottom-up analysis) we conduct explicit model selection for (1) the estimation of the 2D building (footprint) size and (2) the adaption of the number and type of building parts (“jump” mechanism) to guide the reconstruction.

2.7 Conclusion and Remarks

Concerning the challenges formulated in Section 2.1, the contributions of the proposed approaches can be summarized as follows:

- A generative statistical reconstruction driven by reversible jump MCMC with explicit model selection leads to robustness against uncertainty of the data (**C1**) and automatic reconstruction (**C3**).
- Primitive-based modeling with novel combination and merging rules allowing primitive overlap results in complete and plausible watertight models (**C2**).
- A hybrid scheme combining bottom-up and top-down processes improves the efficiency of reconstruction and the degree of automation (**C3**).
- A new primitive-based scheme for 3D building generalization is introduced.

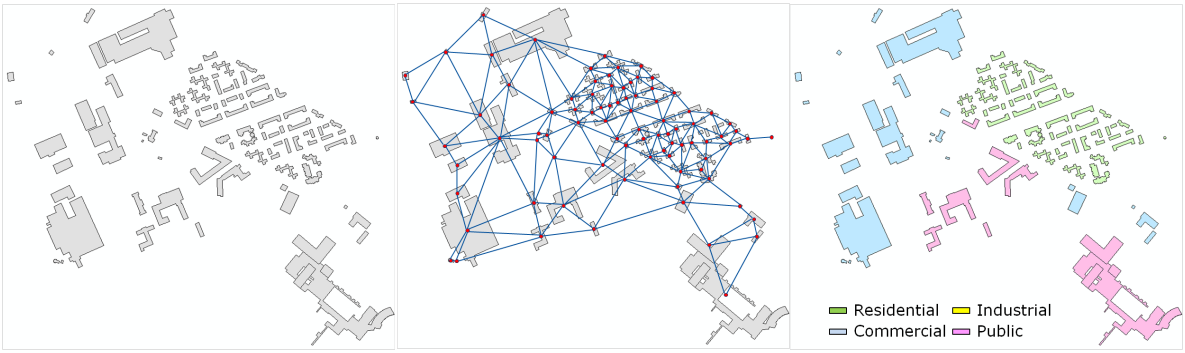
We have presented top-down methods with efficient search and flexible reconstruction. They have, however, still issues concerning uncertainty and instability. Moreover, the completeness of the reconstruction is influenced by the prior knowledge and the scene complexity. A suitable strategy, but also a real challenge, is to balance the top-down and bottom-up approaches in a hybrid framework. A specific discussion of this issue can be found in Section 6.2.2.

Chapter 3

Building classification

Overview

This chapter presents an automatic building type (concerning usage) classification based on footprint data. We propose a method to enhance maps with building usage information which exclusively uses the geometric and topological features in the footprint data. A network model of buildings is built based on a Markov Random Field (MRF) integrating both local features and contextual constraints from the neighborhood.



Building classification: Input map (left), the building network (middle), and the classification results (right)

Keywords

building network, classification/labeling, OSM, cadastral map, MRF, Gibbs sampler

3.1 Problem statement

Usage (use and occupancy) information of buildings is of great interest for many applications, e.g., navigation, city planning, and emergency management. This attribute, however, is generally not provided in volunteered data like OpenStreetMap (OSM). The volunteers cannot guarantee either complete or uniformly defined entries of attributes. Even in the official cadastral maps, the building usage information is not always available or be labeled in a consistent way either. Tools are needed to enhance this information by deriving it from existing data.

The approaches summarized in this chapter confront the following key challenges in building classification and contribute to these objectives:

- (**C1**) Derivation of semantic information solely from geometric features
- (**C2**) A model of building parsing in urban areas with consideration of both local and contextual information

3.2 Data – building footprints

The input data of the presented approach are 2D building footprints, which are represented in the form of 2D polygons defined by the given vertices as well as edges. Our experiments are performed on building footprints from OSM (Nelson et al. 2006, Taylor and Caquard 2006) and cadastral maps.

OSM is a collaborative dataset containing user-generated content in the form of free editable maps and belongs to volunteered geographic information (VGI) (Goodchild 2007). Because of the crowdsourcing concept, in comparison with conventional maps, OSM has better coverage (from a single source) and easier availability. However, at the same time the following holds:

- They can be incomplete and non-uniform entries, as the data are voluntarily provided by individuals
- Quality and reliability are not ensured, as the data are generated by people without formal training.

Cadastral maps are used in cadastral management. They are a register of the real estate or real property's metes-and-bounds comprising the ownership, the precise location, the dimensions as well as the area and the value of individual parcels. Besides the boundary of the parcels, geometric attributes may also include the location and shape of buildings. The data are generated by professional surveying and managed by official authorities. Their pros and cons can then be listed as follows:

- High precision concerning location and geometry
- High reliability for all entries
- Relatively complete and uniform attribute definition and entries
- Limited availability because of costs and data privacy.

Please note that in this work we only use the geometric properties of buildings from OSM or cadastral maps. All other building attributes including sensitive information such as ownership has been removed before processing. The footprint geometry provides the following basic features:

Direct features: Area, vertices and edges

Derived features: Dimensions of the bounding box (length, width, and orientation), centroid of the building, and medial axis (also "skeleton", cf. Section 3.3.3).

3.3 Model – Markov Random Field for building network

Network models of buildings integrate local attributes of buildings and their contextual relationships. They are of special interest in dense urban areas, where the influence as well as constraints between neighboring buildings play an important role.

3.3.1 Markov Random Field

MRF (Kindermann & Snell, 1980), also known as Markov network, is widely used in image processing and computer vision (Li, 2009) for the labeling/segmentation of image pixels or sub-regions and other applications like point cloud grouping, economics and sociology. It is defined as an undirected graph model (G, Figure 3.1):

$$G = (V, E) \tag{3.1}$$

with V indicating the set of vertices and E the set of edges. For the random variable X of the vertices

$$X = \{X_v\}, v \in V \tag{3.2}$$

hold Markov properties, i.e., it maintains global spatial consistency by only considering (relatively) local dependencies. The Markov properties of MRF can be summarized as follows:

1. **Pairwise Markov property**: Any two non-neighbor vertices are independent.
2. **Local Markov property**: A vertex is conditionally independent to all other vertices given its neighbors.

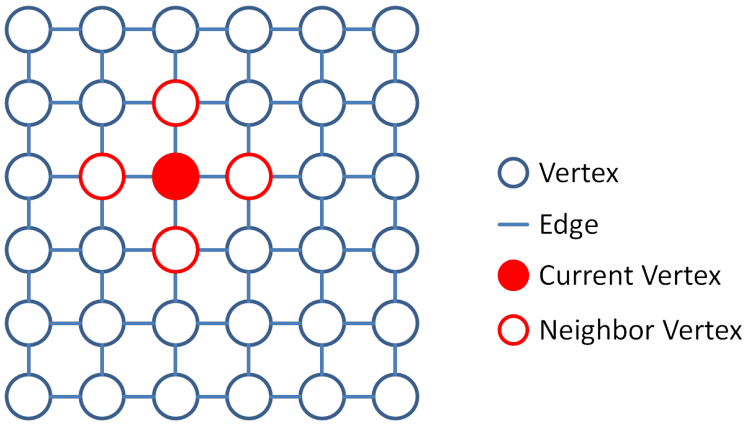


Figure 3.1: Undirected graph model of MRF: Vertices, edges and neighbors

3. **Global Markov property**: Any two non-adjacent subsets are conditionally independent given a separating subset.

Unless a MRF can be decomposed or simplified to some particular sub-classes, e.g., tree structures, for which there are polynomial-time inference algorithms, computation tasks for MRF are often NP-complete problems. In practical applications, approximation techniques, such as MCMC, are usually employed.

Please note that, in comparison with another important graph model – CRF (conditional random field), MRF is a generative model, while the latter is a discriminative model. In an MRF the posterior is inferred from likelihood and prior in a Bayesian framework. Maximum A Posteriori (MAP) is usually employed for the parameter estimation.

3.3.2 Network of buildings

In this work we use MRF to model the buildings and their neighborhood relationships in dense urban areas. The individual buildings are represented as vertices, $v = v_i, i \in V$, and edges, $e = e(i, j), \{i, j\} \in E$, connect pairs of vertices. Any pair of non-neighbor vertices is conditionally independent given all other vertices, i.e.:

$$v_i \perp\!\!\!\perp v_j, if\ \{i, j\} \notin E. \tag{3.3}$$

For the building network the Markov properties mentioned above hold as well. I.e., the characteristics of one building are only conditionally dependent on its neighboring building(s).

One key problem when establishing a building network is the definition of neighbors. In comparison with an MRF in image processing, where the pixels are uniformly distributed and have the same size, the distribution of buildings in dense urban areas is more complex. Besides this, buildings may have very different sizes and shapes. As shown in Figure 3.2, we define the neighborhood of buildings based on the determination of Voronoi cells of

the buildings' centroids (distance-based approach). That is the polygons divide the whole area into seamless cells (Figure 3.2, left). For each centroid there is a corresponding region consisting of all points closer to that centroid than to any other (Euclidean distance). All cells that share an edge are defined as neighbors.

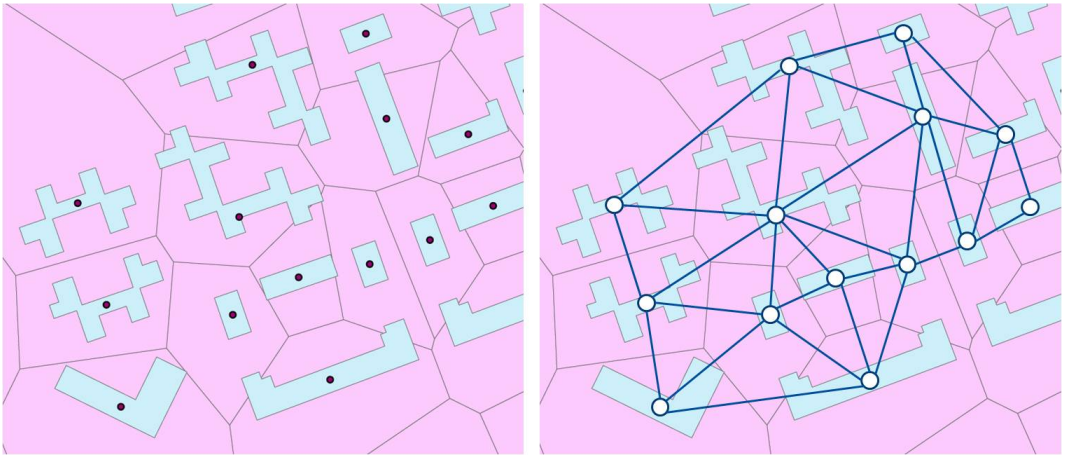


Figure 3.2: Definition of neighborhood of buildings: (left) polygons of buildings, their centroids marked with red points, and their Voronoi cells; (right) the MRF model with the edges connecting neighboring buildings

The overall energy function of the MRF consists of two components: the unary and the binary terms

$$\mathcal{H} = \sum_i u(x_i, c_x) + \sum_{i,j} b(x_i, x_j) \tag{3.4}$$

which integrate local geometric features and contextual constraints, respectively.

3.3.3 Local geometric features

The geometry of building footprints reflects to some extent the building usage, e.g., a warehouse has a much larger **size** than a single-family house, public buildings often have more complex **shapes** than industrial buildings. Simple geometric attributes like overall area and length to width ratio, however, are not discriminative enough for classification. We, therefore, propose two composite high-level features in order to provide a more sophisticated geometric description (Huang et al. 2013b, **P4**).

Effective width

The effective width (EW) can be calculated for any building with arbitrary shape. It is defined as the average width of the footprint along the centerline. As shown in Figure 3.3, the conventional building skeleton (a, black) is extended to the building boundary to be the centerline (b, red) for a more accurate calculation of EW. The effective width is of interest

in usage classification, because **it indicates the general living/movement space inside the building**. By this means, residential buildings can be distinguished very well from industrial or public ones. In many building category definitions, single-family and multi-family houses (e.g., apartment buildings) must be defined as two separate classes, because their area and complexity are significantly different. Using EW, as demonstrated in Figure 3.3 (c), the values of these two types of residential buildings show consistency, although the building areas and the complexity (calculated by "branching degree" below) are not close to each other.

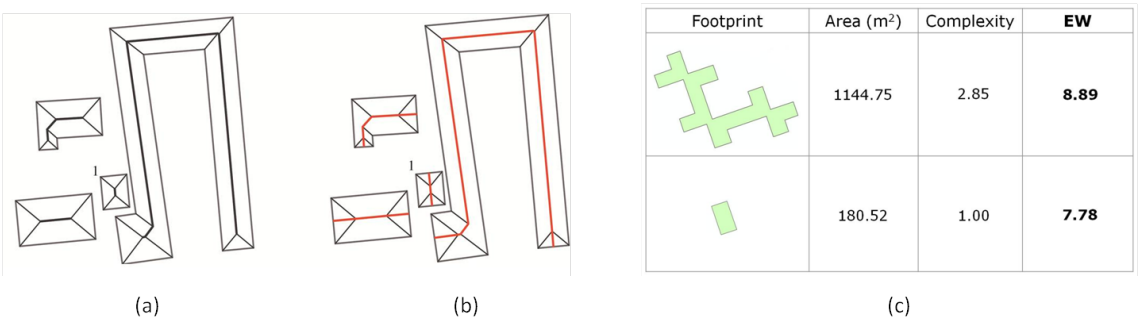


Figure 3.3: Effective width: (a) building skeleton (black bold), (b) building centerline for the calculation of EW (red bold), and (c) comparison of EWs for residential buildings with very different shapes

Branching degree

Branching degree (BD) measure the structural complexity of the building using the number and distribution of the building segments (called "branches"), which are derived from the building skeleton. Some typical examples are given in Figure 3.4. More details can be found in (Huang et al. 2013b, **P4**).

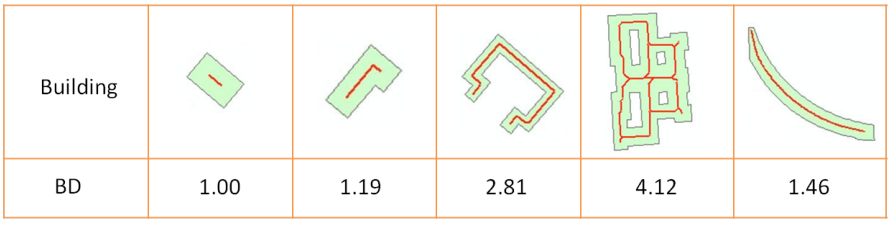
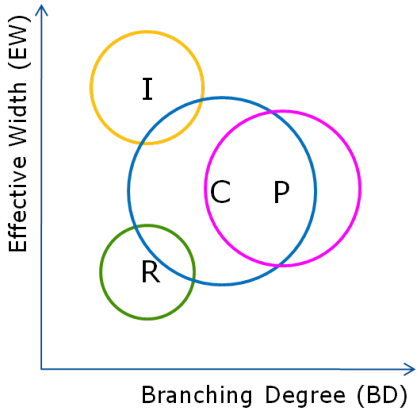


Figure 3.4: Branching degree: A measurement of the complexity of buildings

Figure 3.5 shows a rough qualitative sketch summarizing the distribution of building clusters with different usage in EW-BD space. Each building is represented in the EW-BD space. The probability that one building belongs to one of the classes is inversely proportional to its (standardized) distance to the centroid of the class, which has been empirically determined. The local potential of the building is defined by the probabilities $p_{local} = \{p_R, p_C, p_I, p_p\}$ (cf. Figure 3.5). The probabilities are standardized to a sum of 1.



Usage	BD (range/centroid)	EW (range/centroid)
Residential (R)	Low / 1.5	Low / 12
Commercial (C)	Medium / 2.5	Medium / 25
Industrial (I)	Low / 1.5	High / 40
Public (P)	High / 3.2	Medium / 25

Figure 3.5: Distribution of buildings in EW-BD space

In the following Bayesian inference (cf. Section 3.3.4), the unary energy $u(x_i, c_x)$ summarizes the local features of the individual buildings. It is calculated from the local potential

$$u(x_i, c_x) = p_{local}(x_i) \,, \tag{3.5}$$

with $x_i \in \{R, C, I, P\}$ a label assignment.

3.3.4 Contextual relationship

The contextual relationship, i.e., the plausibility of usage of the neighborhood , is encoded into the binary term. In an MRF, the binary term refers to the energy in the pairwise cliques. It is measured concerning two aspects:

1. Type consistency: Neighboring buildings tend to have the same type;
2. Logical neighborhood: It reflects reasonable city planning for adjacent areas. E.g., residential buildings are more likely be found near public buildings instead of industrial buildings.

The binary energy of each clique is defined as

$$b(x_i, x_j) = N(i, j) \,, \tag{3.6}$$

with a given matrix N. As shown in Figure 3.6, N is a symmetric matrix, in which the rewards as well as the penalties of neighborhood proposals are embedded.

The goal is to find the maximum overall energy $\mathcal{H}$ of the graph with the configuration $\mathcal{K}$, i.e., the grouping. Let $p(i, j)$ indicate the state of each pair in $\mathcal{K}$ (if connected $p(i, j) = 1$, otherwise 0). The goal function can be expressed as:

$$\begin{aligned} &\widehat{\mathcal{K}} = \underset{\mathcal{K}}{argmax}\{\mathcal{H}\} = \underset{\mathcal{K}}{argmax}\left\{\sum_i u(x_i, c_x) + \sum_{i,j} b(x_i, x_j)\right\} \\ &\textit{subject to:} \\ &\textit{i and j are guaranteed to be disconnected if p(i,j)=0.} \end{aligned} \tag{3.7}$$

N	R	C	I	P
R	1	0	-1	0.5
C	.	1	0	0.5
I	.	.	1	-1
P	.	.	.	1

Figure 3.6: Matrix summarizing rewards and penalties of neighborhood assignments (R – residential, C – commercial, I – industrial, and P – public)

For this application another MCMC technique, namely a Gibbs sampler, is employed to solve the global optimization task. Gibbs sampling is introduced in detail in Section 3.4.

Experiments are performed on data-sets of urban areas from OSM (Figures 3.7) and cadastral maps (Figure 3.8). The classification accuracies reach 97.8% and 89.7%, respectively (Huang et al. 2013b). As shown in Figure 3.7 (c and d), the building network model significantly improves the performance by globally optimizing the classification result.

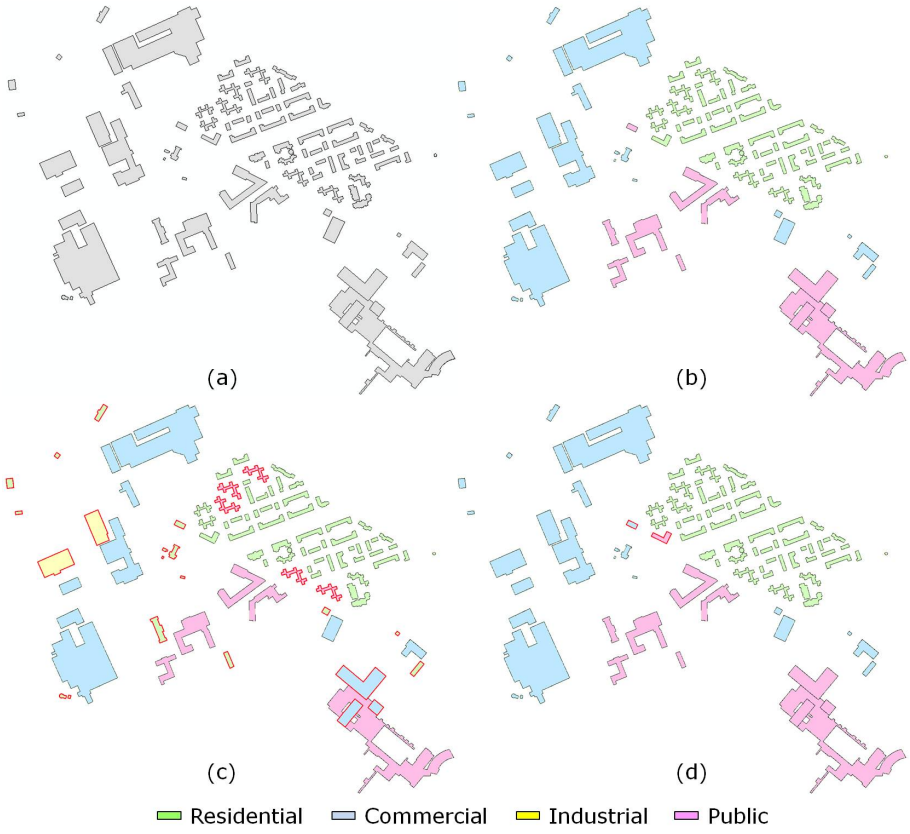


Figure 3.7: OSM data of Boston, USA: (a) input data, (b) ground truth labeled manually, (c) labeling based on local features (unary energy) only, and (d) final labeling considering both unary and binary terms. The incorrectly labeled buildings are highlighted with bold red contours in (c) and (d).

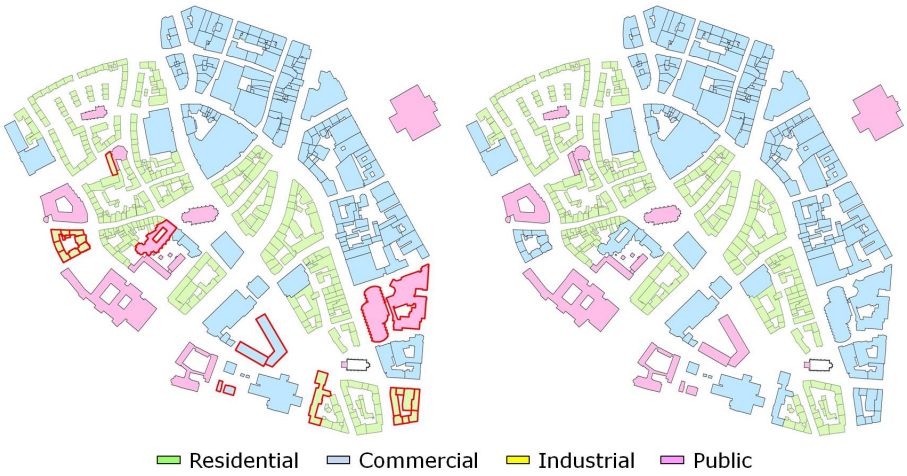


Figure 3.8: Cadastral map of Hanover, Germany: labeling result (left) and ground truth (right). The incorrectly labeled buildings are highlighted with bold red contours.

3.4 Gibbs sampler

In urban areas the buildings are densely distributed implying that the established MRF model is highly connected. In the optimization the labels of all vertices can be altered and lead to different configurations. This means that the optimization is an extremely high-dimensional problem and computationally intractable by a direct solution. We, thus, employ a statistical approximation by means of a Gibbs sampler (Geman and Geman 1984). It is particularly suitable for random sampling of potentially complicated multivariate probability distributions with a large set of variables.

3.4.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm Hastings (1970) is one of the most popular MCMC techniques. It is a generalization of the basic Metropolis algorithm (Metropolis et al. 1953) and many practical MCMC algorithms can be seen as special cases or extensions of the MH algorithm.

The basic idea of MH algorithm is to give the possibility for accepting "worse" candidates. In standard MCMC, a new step X^* proposed by the Markov Chain will only be accepted as X_{i+1} when its evaluation is higher than that for its predecessor X_i. The MH algorithm relaxes this criterion by accepting any X^* with an acceptance probability $A(X_i, X^*)$.

$$A(X_i, X^*) = \min\left\{1, \frac{p(X_i) \cdot q(X^*|X_i)}{p(X^*) \cdot q(X_i|X^*)}\right\}. \tag{3.8}$$

A basic MH algorithm is shown in Figure 3.9. If $A(X_i, X^*) >$ random number u,X^* is accepted and the chain moves forward to X_{i+1} otherwise it stays at X_i.

The acceptance probability $A(X, X^*)$ guarantees two properties:

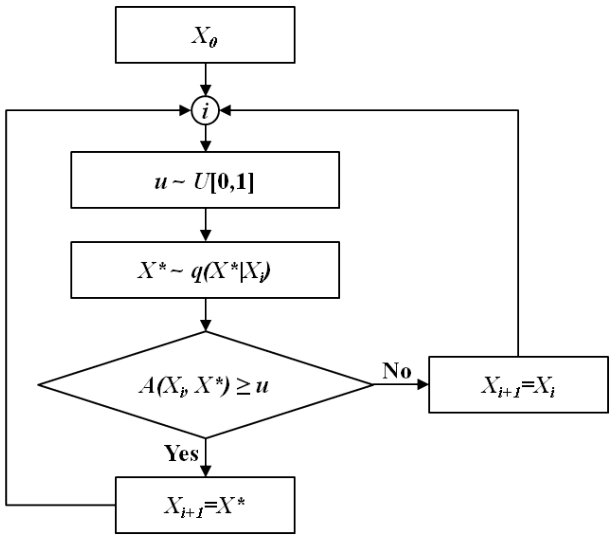


Figure 3.9: Metropolis-Hastings algorithm

1. A better candidate will always be accepted. In this case, the second term is larger than 1, $A(X, X^*) = 1$. Since u is sampled in $[0, 1)$, a better candidate will always be accepted.

2. A worse candidate will have a chance to be accepted. Instead of being rejected immediately, the probability for acceptance exists in proportion to its evaluation.

The main purpose of the MH algorithm is to overcome local minima, which often occur in complex distributions. As shown in Figure 3.10, instead of stopping at the first local minimum, the sampler has the possibility to go up (i.e., accept worse candidates) climbing the "hill" and thus to reach the global minimum. The transition kernel in MH is based on the ratio of the proposed and the current step, implying the gradient of the function. Small hills can, thus, be easily overcome, because a low gradient means a large acceptance probability.

3.4.2 Gibbs sampling

Gibbs sampling (Geman and Geman 1984, Gelfand and Smith 1990, George Casella 1992) is a practical variant of the MCMC algorithm. It was introduced in (Geman and Geman 1984) for image processing with Bayesian estimation (MAP estimation). The main goal of Gibbs sampling is to deal with multivariate probability distributions, i.e., with multiple random variables. Sampling in the joint probability distribution of a large set of variables where a simple MCMC with direct sampling is no more feasible is one of the main practical issues. Gibbs sampler tackles this problem by sampling a single variable or a limited subset of the variables in turn. I.e., complex joint distributions are approximated by (a sequences of) multiple marginal distributions.

Algorithm 1 shows the pseudo-code of the Gibbs sampling used to optimize the building network model. Gibbs sampling can be very well combined with the MH-algorithm and its

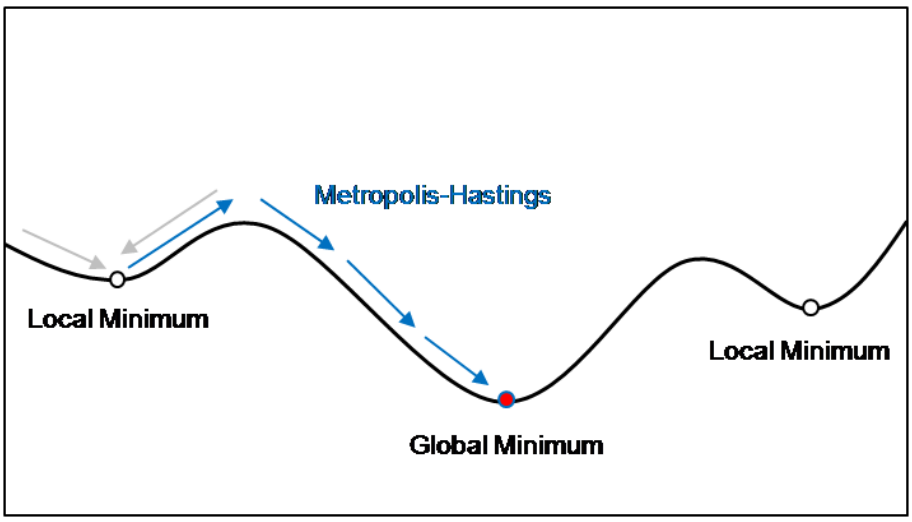


Figure 3.10: Metropolis-Hastings algorithm renders it feasible to overcome local minima by allowing "worse" hypotheses, i.e., "up-hill" moves.

variants. One concrete example sampling scheme can be found in (Huang et al. 2013b).

Algorithm 1 Gibbs sampling for building network optimization

Definitions

s: state

$\mathcal{M}$: model

$\mathcal{K}$: corresponding configuration

$\mathcal{H}$: overall energy

X, X_G, $\overline{X_G}$: universal set of vertices, the chosen vertices for Gibbs sampling, and the unchosen vertices

Initialization (using unary likelihood only)

$\mathcal{M}^s \leftarrow \mathcal{M}^{s=0}, \mathcal{K}^s \leftarrow \mathcal{K}^{s=0}$

while not converged **do**

Propose new state $\mathcal{M}', \mathcal{K}'$

1. Select $X_G \subset X$
2. Re-sample new labels for X_G
3. Calculate new labels for other vertices $\overline{X_G}$ based on new labels of X_G
4. Calculate $\mathcal{H}'$

if $\mathcal{H}' > \mathcal{H}^s$ OR $A^*(\mathcal{M}^s, \mathcal{M}') > u \sim U[0,1]$ **then**

$\mathcal{M}^{s+1} \leftarrow \mathcal{M}'$

else

$\mathcal{M}^{s+1} \leftarrow \mathcal{M}^s$

end if

end while

* Acceptance probability in MH, cf. Section 3.4.1

3.5 Related work

Haunert and Sester (2008) compare different types of skeletons that are commonly used in geographic information systems (GIS) for deriving polygon centerlines. As the basis of our work, we select a simple skeleton, the "straight skeleton", which only comprises straight lines. In contrast to the curvilinear "medial axis", straight skeletons require much less computational effort. The straight skeleton is presented by Aichholzer et al. (1995) and an example is shown in Figure 3.3 (a).

An approach to enhance OSM data semantic labels is presented in (Werder et al. 2010), proposing an unsupervised classification of spatial data solely based on the geometric and topological characteristics. Building outlines and road network information are employed. Lüscher et al. (2009) present a classification of buildings given also in the form of topographic vector data by means of an ontology-driven approach. Supervised Bayesian inference is used to deal with the vagueness in the definitions of the spatial phenomena. In (Henn et al. 2012), building types are derived using a SVM classifier from LoD1 city model data (including 2D footprints, building heights and functions) for a semantic information enrichment of the 3D models.

Current approaches include (Hecht et al. 2015), in which a building type classification of building footprints from different data source, e.g., topographic raster maps, cadastral databases or digital landscape models, is presented. A Random Forest classifier is employed and the influence of data source on the classification performance is analyzed. Kang et al. (2018) propose a classification of building type based on remote sensing as well as street view images. The latter are assigned to individual buildings with the given geographic information. Selected state-of-the-art CNN architectures are implemented for the classification and their performances are compared.

Comparison of the proposed approach to the related work

The novelty of the proposed approach consists in the use of a network model for buildings in urban areas. Besides the local geometric analysis, for which we have proposed two original features, i.e., the effective width and the branching degree, the constraints from the neighborhood have been integrated and are globally optimized. The network model contributes not only to the inference of new features, but also to the propagation of existing features making the building information more complete and uniform.

3.6 Conclusion and remarks

This work presents an approach for the automatic determination of building usage (use and occupancy) solely based on building footprint data. There are four predefined categories: residential, commercial, industrial and public.

Concerning the challenges formulated in Section 3.1, the contributions of the proposed approach can be summarized as follows:

- Two new high-level composite geometric features: effective width and branching degree. They can be used to quantify the average living space and the structural complexity of the buildings, which provide reliable information for building classification concerning usage. (**C1**)
- An MRF is employed to model the building network embedding both local features and contextual constraints. (**C2**)
- An easily extendable framework for automatic classification has been introduced. Further local building attributes and contextual constraints can be added to improve the performance.

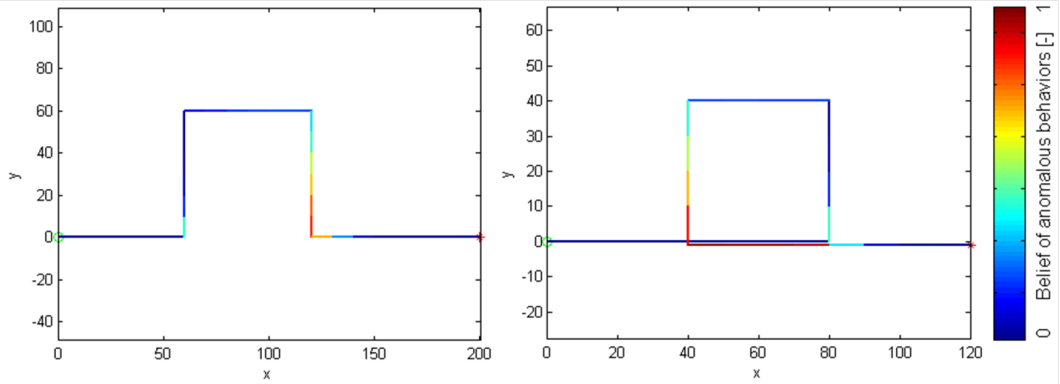
Please note that in this approach the classification works solely based on the geometric features. The result cannot be perfectly correct because some building usages, such as educational and high-hazard (used for the production or storage of flammable, radiative or toxic materials), cannot be derived from footprint data only. Additional information, e.g., materials, floor height, roof type, should be integrated for a more reliable inference.

Chapter 4

Anomaly detection in trajectories

Overview

This chapter presents an original approach to anomalous behavior detection in individual GPS trajectories of vehicles. A variant of a recursive Bayesian filter is designed for dynamic inference. The original motivation of this work war to find when and where the driver meets orientation problems, i.e., takes a wrong turn, makes a detour or gets lost, so that the navigation system can provide appropriate response in time. We propose an extended recursive Bayesian filter that detects anomalous behaviors unsupervised and dynamically during the drive.



Anomaly detection in trajectories: The belief of anomalous behaviors is inferred during the drive.

Keywords

GPS, Trajectory, HMM, Bayesian filter, Inference, Anomaly

4.1 Problem statement

Anomalous behavior detection refers to the problem of finding patterns in the data that do not conform to expected behaviors. It is of great interest for the applications of navigation, driver assistance systems, surveillance and crisis management.

Most related approaches derive normal or/and anomalous patterns from a large number of labeled data and apply the trained classifier to the new trajectories. Large amounts of training data, however, are not always available, e.g., due to privacy concerns.

The core issues and contributions of the approaches presented in this chapter are highlighted as follows:

- (**C1**) Patterns are learned locally inside a single trajectory without previous training.
- (**C2**) The detection is dynamic over time.

The system can autonomously detect anomalous behaviors, in this case orientation problems, and provide additional navigation information just in time. Besides, this detector can be utilized by transportation agencies or researchers for (1) traffic surveillance, where the anomaly (especially the collective behaviors, cf. Huang (2015, **P8**), Section 4.3) reflects traffic issues such as congestion and turn restrictions, and (2) vehicle/driver tracking to study driving habits or patterns.

4.2 Data – GPS trajectories

For the presented approach we use GPS trajectories as input data. From vehicle navigation systems to mobile devices, the rapid popularization of GPS sensors provides plenty of trajectory data. Analysis and pattern mining from the trajectories can reveal interesting information for transportation management, behavior analysis as well as vehicle development.

Trajectories are discrete data that are both space- and time-referenced. A basic GPS trajectory consists of a sequence of points with assigned IDs, coordinates, and time stamps. From a trajectory, information like position, orientation, and velocity can be derived.

Please note that GPS trajectories may contain sensitive private information such as home and work address. Clearance of privacy protection is required when using personal trajectory data, at least, personal information should be deleted or concealed from the traces. The data used in the presented approach are coming from the following sources:

- Crowdsourcing data: Driving trajectories of an anonymous commuter are gathered from a VGI dataset of Hannover, Germany. 100 trajectories are randomly selected from data over two years.

- Open dataset MapConstruction (Ahmed et al. 2014): This dataset has been published for map reconstruction from trajectories. It consists of GPS trajectories in both European (Athens) and North American (Chicago) cities. It also provides the corresponding street maps, which gives the opportunity to evaluate the detection results [1].

- Open dataset GeoLife (Zheng et al. 2008): Provided by Microsoft Research Asia, this dataset consists of a large number of GPS trajectories from 182 users over a period of years (2007 to 2012) from the city of Beijing, China. The dataset contains a broad spectrum of movements including driving, hiking, cycling, and bus/train riding. We only use the data from cars. To demonstrate the effect of collective behaviors, multiple trajectories haven been chosen from the same crossing.

- Self-acquired data: The trajectories are generated by the author himself driving a car. Please note that the routes were scheduled on purpose to pass the particular road segments and crossings with traffic problems to limit the effort and still produce an interesting dataset. Yet, while the routes are "artificial", the traffic situations and reactions of the driver were real. An advantage of the self-acquired data is that the traffic conditions and driver behaviors are available as "ground truth".

4.3 Model – Hidden Markov Model for trajectory

The input trajectory data are represented as Hidden Markov Models (HMM). The derived spatial and temporal features are considered as the measurements/observed states of the individual steps.

4.3.1 Hidden Markov Model with dynamic orders

Markov chains are commonly used in the modeling of state changes in sequences over time. The first order Markov chain is the basis of most Bayes filter variants, e.g., the Kalman filter, and is considered as an appropriate model for trajectory data. In the presented approach we use an extended higher-order Markov chain instead, in order to integrate long-term features. Let $x \in X$ be the unobserved states, i.e., the probabilities of anomalies, in the Markov process and Z the measurements, i.e., the (long-term) observations. The Bayesian network of the HMM process model is presented in Figure 4.1.

The proposed higher-order Markov chain $X = \{x_0, ..., x_n\}$ still follows the Markov assumption. I.e., the probability of the current state (x_k) given a limited number (m) of previous states is conditionally independent of the other earlier states:

$$p(x_k|x_{k-1}, x_{k-2}, ..., x_{k-m}, ..., x_0) = p(x_k|x_{k-1}, x_{k-2}, ..., x_{k-m}) \,, \tag{4.1}$$

[1] The ground truth of "anomalies" is manually generated based on the available street map. A trajectory is labeled as normal if the driver takes a reasonable route, i.e., no obvious detour or repetition, from the start point to the end point, or labeled as anomalous, otherwise.

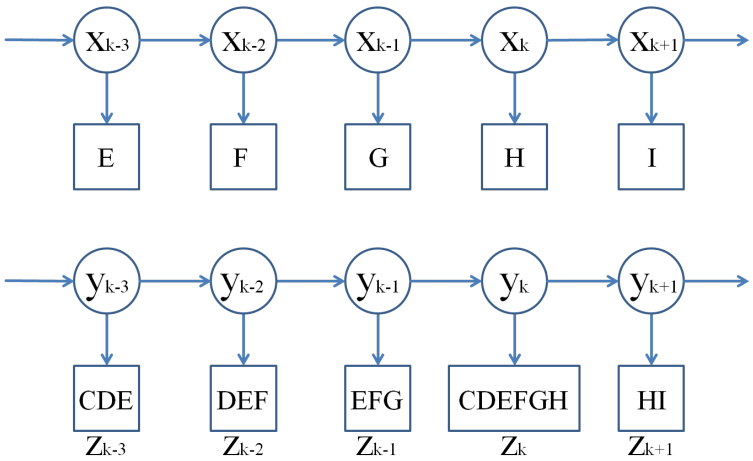


Figure 4.1: An example of a high-order Markov chain with "dynamic memory" (bottom). The last observations of each state in the first-order Markov chain X (top) are integrated into new observations Z for the new chain Y.

with $m < k$. The measurement $Z = \{z_0, ..., z_n\}$ at each state depends not only on the corresponding state, but also certain previous states:

$$p(z_k|x_k, x_{k-1}, ..., x_{k-m}, ..., x_0) = p(z_k|x_k, x_{k-1}, ..., x_{k-m+1}) \ . \tag{4.2}$$

Although the Markov chain X keeps the Markov property, the higher order also implies the following:

1. The number of parameters to be solved grows exponentially with $O(|x|^m)$ with $|x|$ the number of possible states of x and m the order of the Markov chain.
2. The reliability of the parameter estimation decreases.

Thus, we model the trajectory by constructing a new chain $Y = \{y_0, ..., y_n\}$ with an m-tuple of x states:

$$y_k = \begin{cases} (x_k, x_{k-1}, ..., x_{k-m+1}) & \forall k = (m-1, n) \\ (x_k, ..., x_0) & \forall k = (1, m-2) \\ x_0 & k = 0 \end{cases} \tag{4.3}$$

so that the new chain Y over the m-tuple is equivalent to a first order Markov chain keeping the conventional Markov property:

$$p(y_k|y_{k-1}, y_{k-2}, ..., y_0) = p(y_k|y_{k-1}) \tag{4.4}$$

with a "memory" of m, and

$$p(z_k|y_k, y_{k-1}, ..., y_0) = p(z_k|y_k) \tag{4.5}$$

as shown in Figure 4.1 (bottom).

4.3.2 Long-term spatial and temporal features

An anomaly usually cannot be observed, and therefore derived, directly at a single or few observation. It is an underlying state requiring a long-term perspective of the observations. The higher order Markov model defined above makes it possible to derive long-term features from the trajectory data. These high-level features are employed as "measurements" of anomalous behaviors. This Section gives a short summary of the features. More details can be found in (Huang 2015, **P8**).

Turns

Turning is one of the most basic movements in trajectory data. In the GPS trajectory data, the turn itself is a long-term feature as it is normally finished within several time stamps. Although a single turn dose not indicate any anomalous behavior, the combination and density of turns can represent anomalous patterns like a detour/loop. Using the dynamic property of the Markov chain model, a turn is extracted by caching the heading changes and mark the accumulated value at the position where the turn ends. For the combination of turns, we assume that sequential turns in the same direction, e.g., double or triple left turns, have more impact on the belief in anomalous behaviors than sequential turns in different directions, because they imply a potential detour (see below) or the tendency of looping. Please note that we actually do not need to count the turns and calculate their density. Instead, the influence of the density of turns is automatically reflected in the Bayesian filter (Section 4.4.1): Intensive turns in the same direction may result in a high belief (in anomaly), which is continuously increased by the newcoming turns before it decays too much.

Detour factor

A detour often happens when the driver has lost orientation or meet traffic problems, e.g., road-block and traffic congestion. The detour factor (DF) is designed to quantify the degree of detour as an anomalous feature. DF is a widely used term in various areas of transportation including road analysis (Wiedemann and Ebner 2000). The basic idea consists in the comparison of the actual route to the optimal one. We extend DF to a dynamic scheme. In comparison with conventional DFs, in the calculation of dynamic DFs the travel does not have to be finished yet and the "optimal" route (defined by given start, end points and road map) is unknown. For each individual position in the detour, the DF is defined as the ratio of the trajectory length to the direct distance from the start point to the current position.

In comparison with (Huang et al. 2014b, **P6**), (Huang 2015, **P8**) proposes a fully automatic and dynamic scheme for DF calculation, which can automatically detect the start and end positions of a detour.

Route repetition

In all one-way trajectories, the most prominent anomaly is route repetition, i.e., on the way to the destination the driver comes back to the same road section, from either the same or the

opposite direction. Route repetitions with opposite direction are in most cases the result of a U-turn while those with the same direction often happen after driving a loop. The repeated route is detected by finding trajectory points which fall into a buffer area around the previous trajectory. The radius of the buffer is determined considering road width, street network density and noise of the GPS signal.

4.4 Bayesian filter for belief inference

The goal of the proposed Bayes filter is to estimate one unobserved state – the belief in an anomaly, which is quantified by the probability of anomalous behaviors being performed.

4.4.1 A dynamic Bayesian filter

The proposed filter is a variant of recursive Bayesian estimators. It keeps the dynamic property and shares the prediction-updating scheme. Conventionally, the prediction and updating steps work alternately and are inputs for each other. In the proposed filter, however, either of them has a probability to be skipped in the following cases:

- During update: We employ multiple measurements, i.e., anomalous features. Sometimes more than one feature can be extracted at the same position of the trajectory. With the simplified assumption that the features are independent of each other, the update step is performed multiple times before the next prediction step.
- During prediction: Long-term features cannot be continuously observed, i.e., many positions may have no measurements. In the interval to the next position with an observation only the prediction step will be performed for multiple times at multiple positions (once at each position).

Prediction

The prediction step calculates the total probability, i.e., the integral of the products of the transition probability $p(y_k|y_{k-1})$ and the probability of the previous state $p(y_{k-1}|z_{k-1})$ over all possible y_{k-1}. We have only one variable, i.e., the belief, to be estimated and in principle it cannot be predicted based on any current measurement. We assume that anomalous behaviors are more "transitory" than the normal behaviors and, therefore, use a simple exponential decay to predict the belief of the next state:

$$y_k = y_k|y_{k-1} = F \cdot y_{k-1} + w_k \ , \tag{4.6}$$

where

$$F = e^{-s \cdot k'}; w_k \sim \mathcal{N}(0, \sigma^2) \ . \tag{4.7}$$

F simulates the decay of the belief in an anomaly along the track. Gaussian noise is added by w_k. k' is used to count the number of previous states without new anomalous features being reported, i.e., the length of normal states before the current state. The accumulation of k' makes sure that the belief decays rapidly when the driver's behavior is normal again. The tendency for decay can be tuned by the factor s. We define the urban area as the default scene with $s = 1$. The filter is adapted to other scenes by tuning s:

- When driving in suburban areas with higher speed and fewer street crossings, there will be a longer time interval between potential anomalous movements and the belief decay should, thus, be set lower so that the pattern can be found.
- For pedestrian trajectories in dense urban areas, the decay speed may need to be increased to avoid continuous accumulation of the belief.

Update

The update step employs Bayesian inference. The predicted state $p(y_k|z_{k-1})$ is used as the prior, which is refined by the observations of the current state.

$$p(y_k|z_k) = \frac{p(z_k|y_k) \cdot p(y_k|z_{k-1})}{p(z_k|z_{k-1})} \quad , \tag{4.8}$$

with the measurement likelihood

$$p(z_k|y_k) = \prod_{i \in \mathcal{V}} p(z_{i,k}|y_k) \tag{4.9}$$

from multiple measurements. z_i with $i \in \mathcal{V}$ are the observations integrated into the current update step.

4.4.2 Belief inference

In this section we use three typical anomalous behaviors, i.e., detour, wrong turn and loop, to explain how the proposed belief filter works. Simulated data is used as the input trajectories for a simplified presentation. Please note that, besides the turn feature, the detour case uses only the detour factor and no repeated route, while the wrong turn and the loop case employ only the latter. Thus, the influence of the individual features can be demonstrated independently.

Figure 4.2 presents a simulated trajectory with detour (left) and the extracted high-level features plotted over time (right). The green circle and the red asterisk are used to mark the start and end positions of the trajectory, respectively. The bold red line shows the inferred belief in anomalous behaviors over time. It is also given in the trajectory with colors.

Two typical cases of route repetition – wrong turn and loop are shown in Figures 4.3 and 4.4, respectively. All states with detected repeated route have the same feature value of 1. The

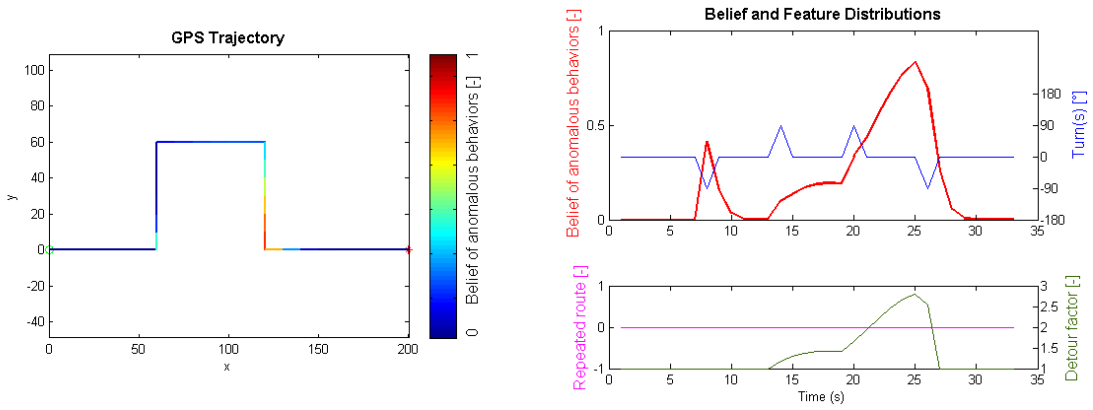


Figure 4.2: Simulated trajectory of a detour (left) with start position (green circle), end position (red asterisk) and the belief in anomaly shown in color. Three high-level features: Turns (blue), repeated route (magenta) and detour factor (green) are plotted together with the belief in an anomaly (red bold) over time (right).

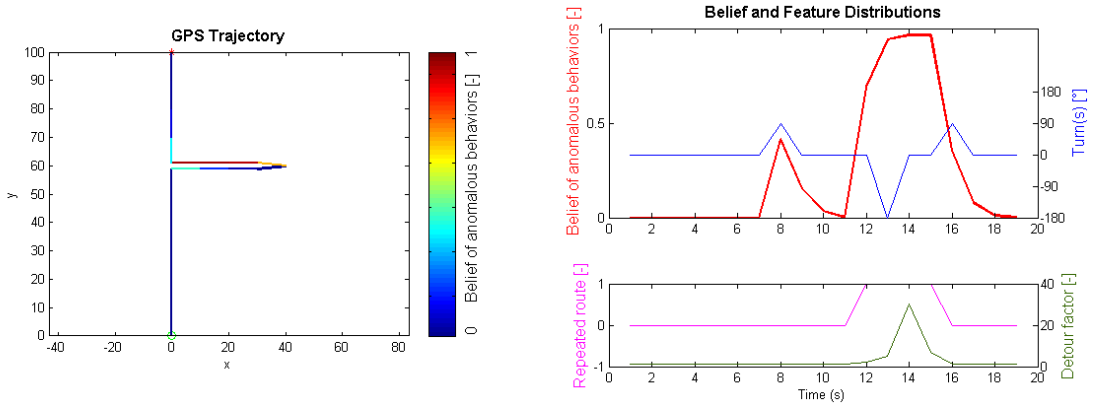


Figure 4.3: Simulated trajectory of a wrong turn: trajectory with colors indicating the belief in an anomaly (left) and the distributions of the belief and behavior features over time (right)

belief in an anomaly increases as long as the vehicle stays in the repeated route and reaches its peak at the spot where the wrong turn ends. The belief decays to the normal value after the vehicle goes back to the previous route.

These examples also demonstrate some typical situations in the inference:

- Double turns in the same direction imply a potential detour or even loop. The probability for an anomalous driving increases when the second turn happens.
- Double turns in opposite directions, in contrast, are considered normal.
- The detour factor increases and reaches its maximum value when the detour is finished. The belief in an anomaly also has its peak value at this time.
- After the detour the belief in anomaly decays fast.

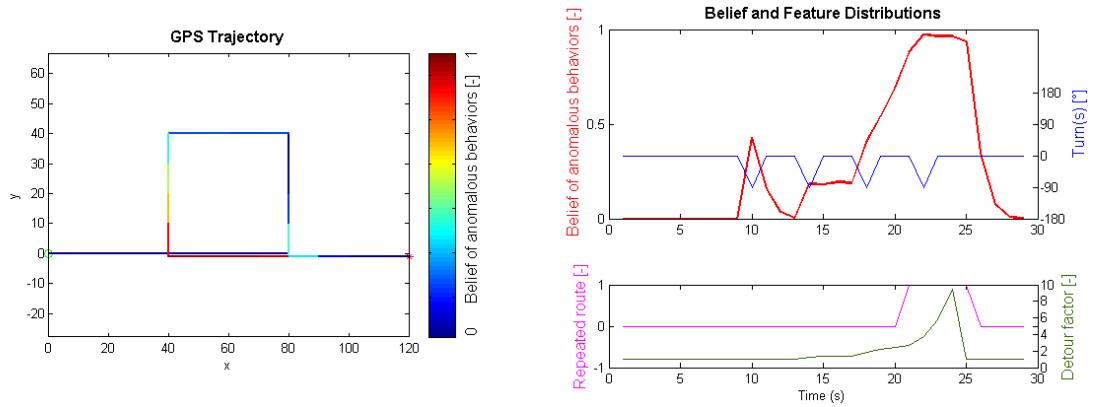


Figure 4.4: Simulated trajectory of a loop: trajectory with colors indicating the belief in an anomaly (left) and the distributions of the belief and behavior features over time (right)

4.4.3 Collective behaviors

As demonstrated above, one main characteristic of the proposed method is that it works for a single trajectory without the need to learn multiple normal behaviors to detect an anomaly. It does not mean, however, that the detection is limited to single traces. On the contrary, based on anomaly detection of individual trajectories, further information might be derived by analyzing their collective behaviors. For some urban traffic issues, e.g., road-blocks, blind alleys or turn restrictions, anomalous driving behaviors will often be found concentrated in a certain area. These collective behaviors can to a certain extent reflect the problems mentioned above.

Figure 4.5 shows self-acquired trajectories with known traffic conditions and driver behaviors in an urban area (Hannover, Germany). A temporary road-block as well as a blind alley (Figure 4.5, right) can be seen near a road crossing. Anomalous patterns are found at the end of the blind alley and on multiple sides of the road-block. As shown in the trajectories (Figure 4.5, left), driver 1 from the north saw the sign for the road-block and made a detour, driver 2 missed the warning sign for the road-block, had to make a U-turn right before the road-block and then made a detour to go on in the same direction. Driver 3 from the south turned around even earlier because of the warning sign and the traffic jam before the crossing. Although the blind alley on the west side is permanent (i.e., not a temporary setup like road-block), it sometimes also causes U-turns for drivers not familiar with this area.

Figure 4.6 presents an experiment on the open trajectory dataset “GeoLife GPS Trajectories” (Zheng et al. 2008, Zheng et al. 2009, Zheng et al. 2010) from Beijing, China. Trajectories from the bottom to the left show a collective detour to the North, while the trajectories in the opposite direction (from the left to the bottom) present no anomaly. This phenomenon indicates a possible left turn restriction, which is proven by the street map shown in the same figure, i.e., no left turn is possible here because of the cloverleaf junction and the direction restrictions in the streets.

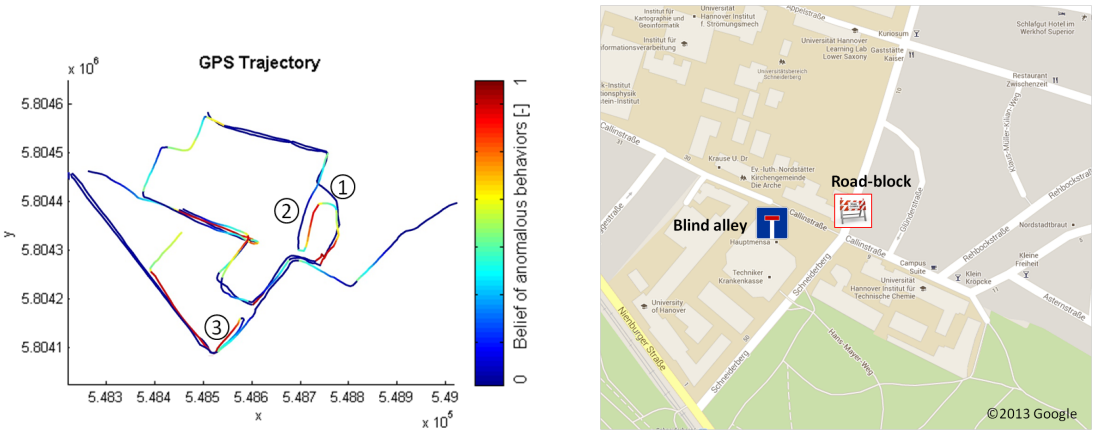


Figure 4.5: Collective behaviors for a road-block and a blind alley: a collection of trajectories in the same area and a similar time period (left) and the street map with the manually labeled locations of the road-block and the blind alley (right)

Figure 4.6: Collective behaviors in the case of a turn restriction: Trajectories passing the road intersection from the bottom to the left show a collective detour and indicate a turn restriction, which is proven by the street map.

4.5 Related work

Chandola et al. (2009) summarize the techniques developed for anomalous pattern detection up to this point in time with the following classes: classification based, parametric or non-parametric statistical, nearest neighbor based, clustering based, spectral techniques and

information theoretic. A significant amount of work related to automated anomaly detection in trajectory data involves trajectory learning, i.e., cluster models of trajectories corresponding to normal cases are learned from training trajectories. New trajectories are typically assigned an anomaly score based on the distance to the closest cluster model or the likelihood concerning the most probable cluster model (Morris and Trivedi 2008). Hu et al. (2006) propose an algorithm for automatic learning of motion patterns and use these patterns for anomaly detection and behavior prediction. Trajectories are clustered hierarchically using spatial as well as temporal information and a chain of Gaussian distributions is used to present each motion pattern. Based on the learned motion patterns, statistical methods are used to detect anomalies and predict behaviors. Besides a cluster based trajectory learning method, Piciarelli et al. (2008) propose a trajectory learning and anomaly detection algorithm based on a one-class Support Vector Machine. The algorithm can automatically detect and remove anomalies in the training data. They first evenly sample points from the raw trajectory and then model each trajectory with a fixed-dimensional feature. Bu et al. (2009) build local clusters using continuity characteristics of trajectories and monitor anomalous behavior via pruning strategies. Ma (2009) presents a method for real-time anomaly detection for users following normal routes. Trajectories are modeled as a discrete-time series of axis-parallel constraints ("boxes") and are then incrementally compared with a weighted average trajectory.

To learn motion patterns and detect anomalies in complicated human trajectories, Suzuki et al. (2007) use a HMM to model time-series features of human positions. A similarity matrix of the HMM mutual distances is formed and multi-dimensional scaling based on eigenvector decomposition provides trajectories in a low-dimensional space. K-means clustering of the projected data leads to the motion patterns. Anomalies can be detected by the use of likelihood scores for the HMM representing motion patterns. Sillito and Fisher (2008) use an incremental semi-supervised one-class learning procedure to detect anomalous behavior of pedestrians. They combine unlabeled trajectories with occasional examples of normal behavior labeled by a human operator. In (Kim et al. 2011), Gaussian process regression is used for the recognition of motions and activities as well as anomalous events given already learned normal patterns of objects in video sequences.

Current approaches include (Laxhammar and Falkman 2014), in which an online learning and sequential anomaly detection scheme is proposed with an adaptive anomaly threshold to minimize parameter tuning during the detection. Pang et al. (2013) adapt likelihood ratio test statistic to learn traffic patterns and detect anomalous behavior from taxi trajectories, to monitor the emergence of unexpected behavior in the Beijing metropolitan area. Wang et al. (2017) present a traffic anomaly detection from GPS trajectories of vehicles in certain road segments and time intervals. Besides the detection in single road segments, the anomalies implied by the inconsistency of neighboring road segments are considered as well.

Recursive Bayesian estimation (or Bayes filter) (Masreliez and Martin 1977), e.g., the Kalman filter (Kalman 1960) for linear and normally distributed variables, is widely used in signal processing, navigation and robot/vehicle control. A main characteristic of Bayes filters is the dynamic estimation (in two steps: prediction and update) of the underlying vari-

able(s) based only on the most recently acquired measurement data. The Kalman filter and its extensions have been proved appropriate for trajectory analysis. Recent work includes (Prevost et al. 2007), which presents an extended Kalman filter to predict the trajectory of a moving object based on measurement data from a moving sensor – an UAV. An Unscented Kalman filter is used in (Sun et al. 2012) for trajectory tracking based on satellite data with weak observability and an inherent large initial error.

Comparison of the proposed approach to the related work

In summary, most related approaches share the same strategy: derive either normal or anomalous patterns from given training data and apply the trained classifier to test trajectories. The strategy is strongly related to the data. One advantage is that less empirical information is needed to differentiate normal and anomalous behaviors, which is learned from the given data. It can, however, easily lead to bias, because of the incompleteness of the data, which is hard to be avoided (cf. Section 4.1). We, on the contrary, have proposed an unsupervised anomaly detector with generic prior knowledge and Bayesian inference. It can work in a single trajectory only based on local features. No training data are required. Besides, the detection is dynamic over time, i.e., the anomalous behaviors can be determined as soon as it has been finished (or sometimes even in process), while in many related approaches the anomaly detection must be performed on the whole trajectory.

4.6 Conclusion and remarks

An original approach to anomalous behavior detection in the trajectory data by means of an extended recursive Bayesian filter is presented. Concerning the challenges formulated in Section 4.1, the contributions of the proposed approach can be summarized as follows:

- A recursive Bayesian "belief" filter for dynamic anomaly detection in a single trajectory over time. (**C1**, **C2**)
- Unsupervised detection without previous training (**C1**)
- Collective behavior analysis based on detection results for multiple individual trajectories.

In a single trajectory, the result indicates where the driver is likely meeting orientation problems and assistance is needed. Furthermore, a potential for detecting traffic problems is demonstrated by analyzing the collective behaviors of multiple trajectories.

Please note that the proposed filter basically produces probabilities instead of binary labels, i.e., it does not make decision but provide proposals with quantified estimation. The driver might perform a loop on purpose to reach a certain sequence of destinations; a route weaving around an impassable area, e.g., a lake or park may look like a detour. In these cases additional information, e.g., reliable road maps with information about traffic restrictions, can be

employed to improve the results. The decision by means of a fixed threshold is rather subjective, depending very much on the individual definition of anomaly, the confidence of the decision maker/observer and the data characteristics. The proposed detector, on the contrary, tries to provide information as objectively as possible in the form of probabilities, which can be reliably used to support the decision-making, e.g., to warn concerning getting lost by the driver assistance system or concerning suspicious behaviors for surveillance.

We are aware that using geometric features only limits the performance of the proposed method. Further features along with information about the road network and traffic could be integrated to derive anomalous behaviors more reliably. One simple example is the detection of speeding, which is also a typical anomaly in driving, by using the feature ***velocity*** (directly derived from the GPS data) as long as the speed limit of the route is available. A sensitivity analysis of the decision threshold can also be considered for future work.

In detour detection, for the proposed detector that the belief in anomaly increases directly along with the rise of the detour factor. By this means, a detour is detected when it is (almost) finished and the dynamic character of the detector is maintained. In some applications, however, the begin of the detour is of interest because it may indicate the position of a traffic congestion, a road-block or where the driver starts to meet an orientation problem. Although a detour can, in principle, not be determined at the first turn, we are able to mark it "later" as soon as the detour has been detected. In the proposed work the beginning position is already buffered in the dynamic high-order Markov model for the calculation of the DF.

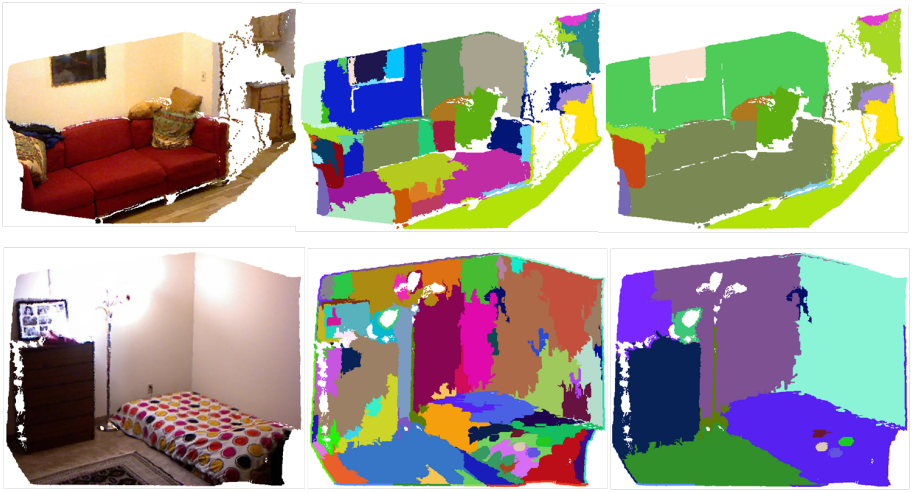
We have demonstrated one application of the proposed belief filter – the detection of specific driving behaviors in GPS trajectory data. We, though, assume that the developed scheme can be (1) extended to find other trajectory patterns given appropriate features and (2) adapted to the trajectories of pedestrians or animals, which can be derived from various sensors such as movement trackers and camera networks.

Chapter 5

RGBD Segmentation

Overview

This chapter presents a fully automatic object-level segmentation of RGBD data. A novel fully 3D parsing scheme and global optimization with an MRF are proposed for complex scenes with freeform objects. The raw RGBD data are first converted to 3D point clouds and the points are regrouped into patches with homogeneous color and geometry. A hypothetical quasi-3D model – "synthetic volume primitive" (SVP) is constructed by extending each 3D patch with a synthetic extrusion in 3D. SVPs vote for and assemble themselves into objects by considering not only color and surface geometry but also the underlying shared 3D volume. They produce more plausible results and show advantages when dealing with concave and freeform objects.



RGBD segmentation: The input data (left) are first over-segmented into patches (middle) as basis for the SVPs, which assemble themselves into objects (right).

Keywords RGBD, Segmentation, Depth image, Point cloud, SVP, MRF, Super-pixel

5.1 Problem statement

As a new category of spatial data, RGB-Depth (RGBD) data attracts increasing attention, because of easy accessibility and fast data acquisition. They are of particular interest in robotics and general indoor scene interpretation, as most of the sensors focus on close range measurement. In the community of image processing and computer vision, image segmentation is an intensively studied topic. In recent years, a number of approaches for RGBD segmentation have been reported, where the depth information is mostly treated as an additional channel to the colors. High-level, i.e., object-level, segmentation is always important and desired for robot vision and scene interpretation. One challenge remaining in both RGB and RGBD segmentation is to find objects with irregular shapes in a complex scene, e.g., a jacket lying on a sofa with other cushions (cf. overview figure, top). The core issues that the approaches included in this chapter tackle are:

- (**C1**) A fully 3D instead of 2.5D parsing of RGBD data
- (**C2**) Object-level segmentation from single RGBD image
- (**C3**) Segmentation of freeform including concave objects.

The key idea is to utilize the 2.5D (depth) information and restore the data into 3D space for analysis, i.e., to observe the data from all possible angles instead of only the original acquisition angle.

5.2 Data – RGBD

RGBD data refer to data containing (conventional) color information as well as an additional channel with depth information for each pixel. Currently, most RGBD data are acquired with platforms like the Microsoft™ Kinect, the Asus™ Xtion, and the newly introduced Google™ Tango. They work with a combination of a camera and a "depth sensor", i.e., an active sensor used to measure the depth. Figure 5.1 presents an example from Kinect. Another kind of platform is equipped with only passive sensors, i.e., stereo cameras, and the depth information is calculated by dense image matching. Passive measurement platforms have lower demand for power supply and higher concealment. It requires, however, more sophisticated algorithms and, in some cases, the ability of real-time computation. One such sensor system has been recently introduced by Roboception™. Both types of RGBD sensors are focusing on close-range measurement[1]. The Kinect-like sensors are limited by the power and accuracy of projectors while the stereo system is restricted by the limited baseline of the cameras.

[1] Although some aerial surveying and mobile mapping systems equipped with both laserscanner and cameras can also acquire RGB and corresponding depth information, such data are usually not referred to as "RGBD".

Figure 5.1: RGBD data: color map, depth map and registered 3D point cloud

RGBD data are usually seen and treated as an "image" rather than 3D data. Even though many approaches work directly on the registered point cloud, they process the data of a single shot like a 4-channel image and not like a 3D point cloud. This is reasonable, especially in the community of computer vision and image processing, as the depth information is inherently rasterized and aligned to the pixels. Thus, existing algorithms and methods can be directly applied. Strictly speaking, "depth" implies, as termed in geoinformation and surveying, "2.5D" instead of "3D".

5.3 Model – A novel synthetic model for spatial data parsing

A novel model – Synthetic volume primitives (SVP) – is proposed for fully 3D parsing of RGBD data. It is established based on one observation and one assumption:

> *Observation:* Point clouds (from RGBD and all other forms of surveying sensors) reveal actually only the (partial) **surface** of the target objects.
>
> *Assumption:* Every 3D patch (group of points) in the point cloud, whose member points have homogeneous color and geometry, represents (a part of) the surface of an individual **underlying solid body** in the 3D world.

SVP is a 3D solid body model with a 3D patch as basis and a given synthetic volume. By this means SVPs can be used to simulate 3D objects or their components. The hypothetic volume helps the analysis of 3D relationship and, thus, the merging of SVPs.

5.3.1 Synthetic volume primitives – SVP

An SVP consists of two components: a base patch and a hypothetic volume. The base patch is a group of 3D points and can have freeform.The patches are segmented by an extended superpixel method. More details can be found in (Huang et al. 2014a, **P5**). The hypothetic volume is generated as a solid body extension along the normal direction of the base patch.

Figure 5.2 shows how the SVPs are constructed for a 3D cube. We assume that the underlying solid body is always behind the patch. The direction of the extrusion is along the normal vector that points towards the viewing direction. The SVPs vote for a common object by intersection with each other in 3D space. We, thus, use SVP models as "bricks" to build the 3D world.

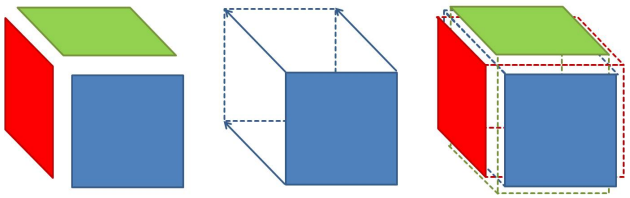


Figure 5.2: Definition of SVPs: (left) 3D patches, (middle) an SVP with hypothetical extrusion and (right) the intersection of SVPs

The construction of SVPs is shown in Figure 5.3. A surface patch may have an arbitrary shape (top left). In practice, we model the extrusion volume with multiple "sticks" (top right) to approximate the shape. The stick model is a discrete representation of the 3D extrusion. For simplification the sticks are given a uniform diameter, which is equal to the average distance of neighboring points. A tricky problem is how to define a reasonable length for the extrusion. With too short extrusions, the SVPs will fail to intersect with other object components. Too long extrusions, on the other hand, will result in the merging of multiple objects into a large one. Without any prior information, we can only assume a quasi cubic shape for a hypothetical 3D body. We empirically found, that it is reasonable to use the average edge length of the patch's bounding box as the length of the extrusion. An example scene with SVPs is shown in Figure 5.3 (bottom).

5.3.2 Freeform object voting

The hypothetic volume of SVPs can be used to analyze how different patches of an object interact with each other. Multiple SVPs may vote for a common object by intersection with each other in 3D space.

Object-level segmentation is challenging since the objects may have different sizes, colors and shapes. To decide if two 3D patches belong to the same object, a simple way would be to check if they are neighbors and their back sides face each other. This, however, will fail in many cases, e.g., two patches belong to one object, but are not directly adjacent, two patches are connected with each other, but do not belong to the same object, the patch is non-planar, or contains more than one plane, etc. To improve the results, geometric constraints and/or top-down models (with simplified and regular primitives) could be employed to ensure a more reasonable segmentation. Yet, even then, such methods are confined to limited, and mostly convex, shapes.

Instead of exploring the various possibilities of patch combination, we enforce a single condition: If two patches belong to the same object, the hypothetic volumes behind them should

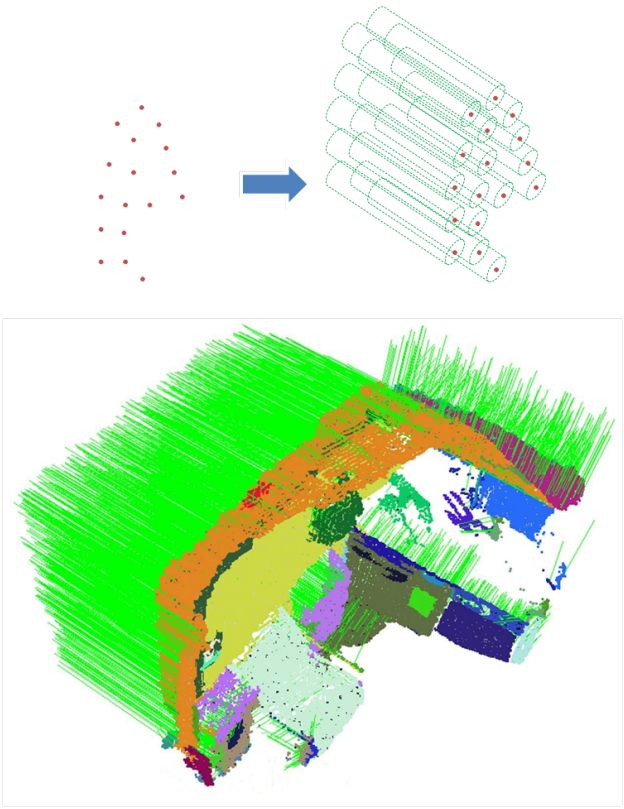


Figure 5.3: SVP models: "Sticks" (cylinders) are employed to model the extrusion of a freeform patch (top). In an example scene (bottom), the sticks are visualized by their axes (green beams) for simplification.

have a certain overlap as they represent (parts of) the same 3D object. This is a simple but reasonable condition which describes the actual 3D relationship of the patches and the underlying object. As illustrated in Figure 5.4, with the help of SVPs the hypothetical components are "assembled" from arbitrary angles to form an object. Remote object parts or patches of concave shapes, as shown in Figure 5.4 (c, d), may not link with some neighbors, but can still be included via other object members from a different direction.

An "intersection degree" is introduced to quantify the intersection of two SVPs. Besides a discrete approximation of the 3D volume, the above mentioned stick model also provides an easy way to quantify the intersection of 3D objects by just counting the intersecting sticks of different objects instead of calculating the real 3D volume overlap. We define the degree of intersection for a patch pair i and j as follows:

$$\mathcal{I}(i,j) = max\{\mathcal{J}_{i\to j}, \mathcal{J}_{j\to i}\}\,, \tag{5.1}$$

with the intersection degree of the individual patch

$$\mathcal{J}_{i\to j} = \frac{m_i}{n_i} \cdot \frac{t}{m_i \cdot n_j^{0.5}} = \frac{t}{n_i \cdot n_j^{0.5}}\,, \tag{5.2}$$

where n is the number of sticks, t the total number of intersections and m the number of sticks involved in the intersection. The intersection degree $\mathcal{J}$ is defined as the product of

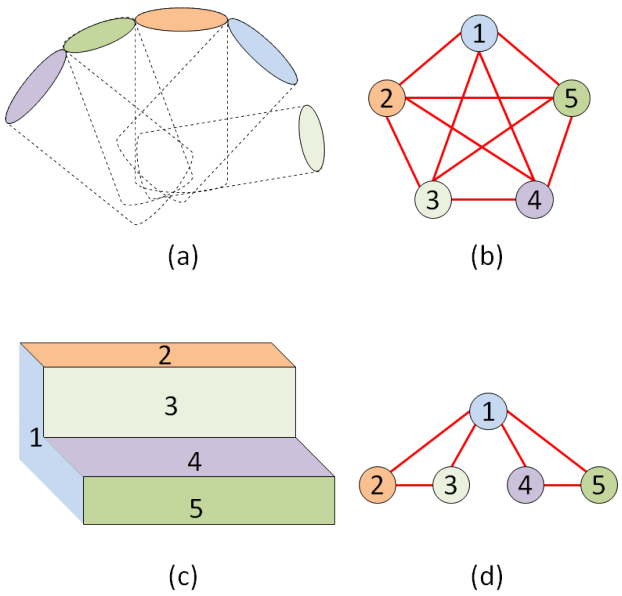


Figure 5.4: SVPs voting for freeform objects: SVPs can be assembled from arbitrary angles (a) and form concave shapes (c) when they are either fully (b) or partially (d) connected.

the percentage of the intersected sticks and the degree of intersection depth. We use $n_j^{0.5}$ to approximate the maximum possible depth. The larger value is taken for the patch pair (i, j). This score is then used as the likelihood for a valid grouping of these two patches.

SVPs provide the following advantages for 3D scene understanding:

1. Freeform objects can be represented by the intersection of SVPs from arbitrary angles.
2. There is no limit for the number or size of component SVPs.
3. Primitives (SVPs) are merged to objects in the CSG (Constructive Solid Geometry) style.

5.4 Global optimization with Markov Random Field

A global optimization is beneficial because the parameter settings of the segmentation can be significantly influenced by the various scales of different objects in the scene. An MRF is employed to model the relationship of the SVPs. The SVPs are represented as vertices and their neighborhood relationship as edges. In this undirected graphical model each SVP is only related to its first-order neighbor. The SVPs are defined as neighbors if the corresponding 3D patches have common boundaries.

The MRF model is defined as:

$$G = (V, E)\,, \tag{5.3}$$

with $v = v_i, i \in V$ the vertices and $e = e(i,j), \{i,j\} \in E$ the pairwise neighbor relationship. Any pair of non-neighboring vertices is conditionally independent given all other vertices, i.e.:

$$v_i \perp\!\!\!\perp v_j, if\ \{i,j\} \notin E. \tag{5.4}$$

Please note that there is no unary energy in the MRF. Different from most labeling tasks, the number and types of groups in this task are unknown. Also no likelihood can be derived from the local features of an individual vertex. We, thus, only observe the binary energy of the pairwise cliques defined as:

$$\mathcal{B}(i,j) = \begin{cases} 0.7 \cdot \mathcal{P}(i,j) + 0.3 \cdot C(i,j) & \forall v_i, v_j\ coplanar \\ 0.7 \cdot \mathcal{I}(i,j) + 0.3 \cdot C(i,j) & \forall v_i, v_j\ intersecting \\ -1 & otherwise \end{cases} \tag{5.5}$$

The binary energy is calculated for the intersection ($\mathcal{I} \in [0,1]$) and the coplanarity ($\mathcal{P}$=1, otherwise 0) in combination with the color coherence ($C \in [0,1]$). More details can be found in (Huang et al. 2014a, **P5**).

For the coplanar and the inherent, the binary energy represents the probability that the pair of patches belongs to the same object. We empirically found it to be advantageous to give 70% of the weight for the geometric relationship and 30% for the color. In the other cases we assume that the patches do not belong to the same object, which will be penalized with "-1" to discourage any group to include this pair.

The goal of the optimization is to find the maximum overall energy $\mathcal{H}$ of the graph model with the configuration $\mathcal{K}$, i.e., the grouping scheme. Let $p(i,j)$ indicate the state of each pair in $\mathcal{K}$ (if connected $p(i,j) = 1$, otherwise 0). The goal function can then be expressed as:

$$\begin{aligned} &\widehat{\mathcal{K}} = \underset{\mathcal{K}}{argmax}\{\mathcal{H}\} = \underset{\mathcal{K}}{argmax}\left\{\sum \mathcal{B}(i,j) \cdot p(i,j)\right\} \\ &subject\ to: \\ &i\ and\ j\ are\ guaranteed\ to\ be\ disconnected\ if\ p(i,j)=0. \end{aligned} \tag{5.6}$$

By this means the main objects in the scenes are segmented without previous knowledge or an assumption about the number of objects. A MCMC sampler is employed to tackle this high-dimensional optimization task. More details can be found in (Huang et al. 2014a, **P5**).

5.5 Related work

Image segmentation employing depth information has been intensively studied. Silberman and Fergus (2011) use RGBD data from the Kinect sensor to improve the segmentation of indoor scenes. Each segment is classified as one of seven categories, e.g., bed, wall and floor. Silberman et al. (2012) additionally extract the support relationship among the segments. In (Koppula et al. 2011), RGBD images are segmented into regions of 17 object

classes, e.g., wall, floor, monitor and bed, for office or home scenes. Li et al. (2011) propose a method to segment engineering objects that consist of regular parts from point clouds. They consider the global relations of the object parts. In (Bleyer et al. 2012), the depth is estimated from a stereo image pair and an unsupervised object extraction is conducted maintaining physical plausibility, i.e., 3D scene-consistency.

Additionally, many methods have been proposed to segment geometric primitives, e.g., cubes and cylinders, or regular objects such as buildings, for which reconstruction rules can be derived. An overview of point cloud processing is given by Vosselman (2009). Rabbani et al. (2007) present an approach for the labeling of point clouds of industrial scenes. Geometric constraints are employed in the segmentation in the form of the primitives cylinder, torus, sphere and plane. Current research for 3D building extraction is reported by Lafarge and Mallet (2012) and Huang et al. (2013a), in which the buildings are modeled as an assembly of primitive components. Current approaches focus on the use of Deep Neural Networks (DNNs) including (Qi et al. 2017), in which an extended neural network with an additional neighborhood graph given to the point cloud is proposed. The new architecture shows its capability of connecting the object parts in a long range. Wang and Neumann (2018) adapt the standard CNN for RGBD data by introducing depth-aware convolution and average pooling to utilize the additional depth information.

Comparison of the proposed approach to the related work

Compared to other approaches, the proposed method focuses on fully 3D parsing of the scene by working directly on the 3D point cloud derived from the raw data instead of dealing with 2D images with the depth values as an additional channel. A novel modeling scheme is proposed with **solid** 3D primitives – SVPs. No specific physical constraints or top-down modeling is required to ensure plausible results, because the essential 3D spatial constraints are embedded into the primitives.

5.6 Conclusion and remarks

We have proposed a novel method for object-level segmentation of RGBD data. The synthetic volume primitive – SVP is introduced to parse the 3D geometrical relationships between the pre-segmented data patches. The proposed method demonstrates its potential in finding the dominant objects in indoor scenes including the walls and the floor without using domain knowledge of specific object classes.

Concerning the challenges formulated in Section 4.1, the contributions of the proposed approach can be summarized as follows:

- A hypothetical quasi-3D model introduced by the SVP (**C1**)
- A novel segmentation scheme for object voting based on the assembly of SVPs (**C2**, **C3**)

- Global optimization of indoor scene scenes via MRF (**C2**).

In this work SVP presents an extension from 2.5D to 3D. It is actually, more strictly speaking, a link between data of the object ***surface*** and the underlying ***solid body***. We are aware the fact that almost all of the data generated by conventional surveying technologies (in the scope of this thesis) are measurements of surfaces only (even though the object has been observed from all possible directions). We, thus, assume that, besides a single Kinect acquisition, the proposed approach can be adapted to all kinds of 3D point clouds, which may come from fused Kinect data, LiDAR, dense image matching, mobile mapping systems, etc.

Chapter 6

Conclusion and Discussion

In this thesis we have demonstrated the potential of Bayesian statistical approaches for spatial data understanding and interpretation, where conventional methods may reach their limits concerning efficiency and even feasibility if the data are extremely unequally distributed concerning both quantity and quality. An overview of the data flow is given in Figure 6.1.

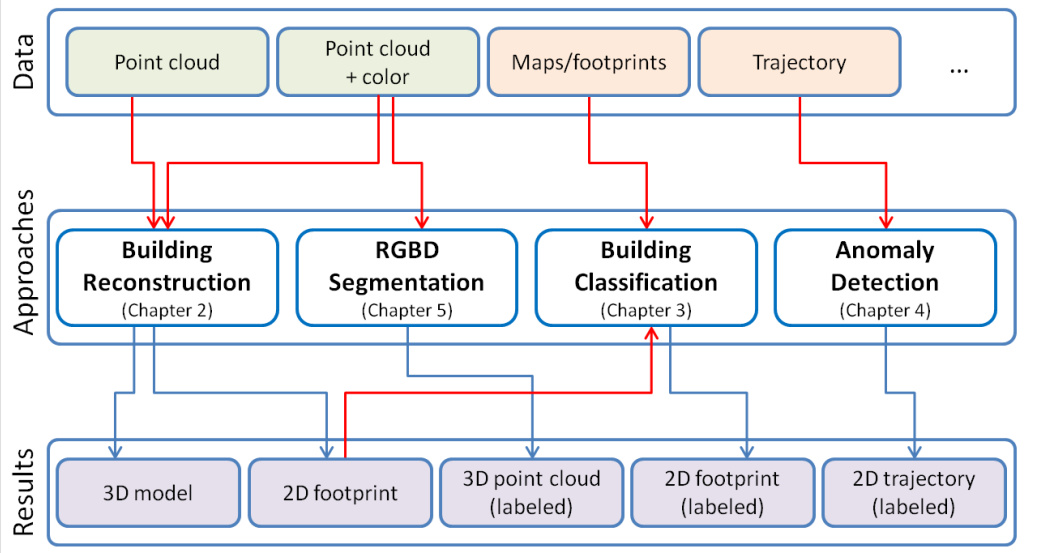


Figure 6.1: Overview of data flow (red: input, blue: output)

In this chapter we first respond to the challenges raised in Chapter 1 by summarizing the advantages of using Bayesian statistical methods for spatial data processing. Afterwards, general issues including limits and potential are discussed. The discussion is limited to the scope of the included approaches and leans toward practical applications.

6.1 Answers to Challenges

The challenges "□" raised in Section 1.1 are answered in this section by relating them to the contributions "■" of the proposed approaches:

□ **Efficient processing**

Efficiency is achieved concerning two aspects: Less dependence on data volume and adaption to new and incrementally increasing data. We can, therefore, formulate two specific contributions:

- **Independence of data quantity**: In comparison with data-driven methods, the amount of processing needed for top-down approaches based on Bayesian models does not necessarily directly depend on the amount of data. For instance, for generative modeling, the sampling of candidate models (driven by MCMC) is generally independent of the data density. Although for the evaluation (likelihood) the computation effort is still proportional to the data amount, as every data point should vote for the model, the dependence is limited to local data inside the boundary of the candidate model. The involved data population is, thus, (mostly stochastically) reduced to fit to the complexity of the chosen models.
- **Incremental processing**: Inference in the Bayesian framework is inherently suitable for the processing of incremental data entries. The possible large amount of existing data does not have to be re-considered individually for a new entry, but all entities contribute collectively to the inference in the form of priors. The new data are labeled with the posterior, for which the only variables are the added observations.

□ **Robustness against data flaws and uncertainty**

- **Independence of data quality**: As long as the data is dense and accurate enough to identify the target object in the presence of other objects and/or the background, statistical modeling is robust against outliers and producing plausible models. Prior information and the capability of learning provided by the Bayesian framework further reduce the influence of data artefacts. On the other hand, in case the data are not good enough, the model or classifier still has the possibility to derive a reasonable estimate from the prior and contextual constraints.

□ **Plausibility of models**

- The plausibility refers to the completeness of models and the rationality of their parameters, which is inherently ensured in top-down modeling. First, the final models are variants or combinations of predefined primitives. This guarantees that every proposal of a model is complete and the components, e.g., the roof planes and walls of a building model, are correctly assembled. Furthermore, in a Bayesian framework the determination of parameters starts from an empirical range with given priors and is incrementally improved with new evidence. This contributes to the rationality of the models with plausible parameters (e.g., the height and the size of regular buildings) and, thus, the final result.

□ **Potential of automation and learning**

Automation is the final goal. It is desired to deal with the rapidly increasing amounts of data. It is, however, limited by various issues, from not sufficient data quality, a lack of robustness of the algorithms, to an insufficient adaptability of the employed models. The former two issues have been discussed above. The last can in principle be solved by (1) a flexible model and (2) the capability to learn from local data.

- **Flexible model**: It is important for automation that a model is flexible enough to be applied for a large range of cases. We employ generative statistical models driven by Bayes' rule, which are fully parameterized and fully probabilistic for all parameters. I.e., in comparison with discriminative models, they can explore all possible values of any parameter and, thus, all possible models. Furthermore, automatic model selection based on an information criterion (Huang et al. 2013a, **P3**) additionally enhances the flexibility.
- **Incremental learning**: It is important that the model can be adapted for a given case. During incremental data processing, new evidence contributes to improve the priors in the Bayesian framework. This implies that the parameters are automatically adapted to the particular data and the local characteristics. The degree of automation increases when manual parameter tuning is minimized.

6.2 A Start of the Exploration: Limits and Potential

At the begin of this thesis we have emphasized that we are facing an abundance of data. This thesis has additionally shown that we are challenged at the same time by the sparseness and redundancy of the data, along with the ever-present noise and measurement errors. In the scope of this thesis we have not dealt with Big Data yet. Geospatial Big Data are, however, the opportunity as well as the challenge we face now or in the near future. We will neglect the popular topics, e.g., massive storage and distributed processing, and concentrate on another essential issue: mining information from large amounts of possibly sparsely distributed data, which is what we have started to explore in this thesis.

6.2.1 Characteristics of Big Data

Spatial Big Data share the following two characteristics with conventional spatial data:

- **Low "value density"**: While the amount of data increases, the percentage of valuable data decreases. First of all, not all the data are of equal value. In many cases only a small portion of the data contains useful information. Rapidly increasing amounts of data often come along with "noise", i.e., false, redundant and meaningless data. The latter usually increase more rapidly than the useful data. As a result, the "noise" may obscure the true information and lead to wrong conclusions.

- **Lack of authenticity**: Large amounts of data are usually obtained from multiple different sources including crowd-sourcing, in which authenticity cannot always be guaranteed. Especially for social-activity-related Geoinformation, which receive increasing interest, the data can contain additionally to unintended false also intended false or subjective information.

Additionally, spatial Big Data have the following issues:

- **Representativeness**: Are Big Data a "full sample" or at least big enough to represent data necessary to solve a problem reliably? This is often doubtable. The limit does not stem from the amount of data, but from the original data definition and the constraints of data acquisition. For example, imagery data for an urban area may have a huge amount of photos but mostly come from Crowdsourcing data. They concentrate mainly on a few local sights. The image quality is limited because of the cameras/cell phones the contributors use and the upload constraints of the social media websites.
- **Data security and privacy protection**: These two significantly limit the completeness of data. This has led to the development of sophisticated methods which allow to extract salient information, while personal data is protected.

Besides the rapid growth of the quantity, all spatial data suffer from the following two disequilibria.

- **Spatial (coverage) disequilibrium**: For many urban areas detailed surveying data including cadastral maps, LiDAR data, imagery, and even 3D building models exist, while most suburban and uninhabited areas are sparsely covered. Even for an urban building, the facades on the street side might be often scanned while for the inner side no data exist.
- **Temporal disequilibrium**: Landmark buildings may have photos daily updated while the interval between the surveys of uninhabited areas can be many years.

Big data or not, we empirically learned that a method would be ideal if it works (to a certain extent) **independently of data density and artefacts**, so that it is suitable for any quantity as well as quality of data. Additionally, it should have **the capability to process incremental data entries efficiently**. As shown at the beginning of this chapter, the statistical framework based on Bayesian models provides a promising start for the further exploration.

Another important issue that is usually paid less attention is **timeliness of data processing**. Previously, the intervals of conventional terrestrial or airborne data acquisition could be years. New satellites as well as UAVs and mobile mapping platforms reduce the interval between measurements to days or hours. Particularly for the observation of earthquakes and tsunamis, the data from sensors is much more valuable if they are available (i.e., processed) in a smaller amount of time – ideally much less than the measurement interval. Timeliness

is of great interest also for most of the social-activity-related data. The demand points towards real-time data acquisition, processing and distribution. The increased amount of data, however, may delay data transport and processing.

The challenge is, therefore, high efficiency in understanding and interpretation of data, which leads to the question: Can statistical models work fast? Usually, statistical models are not considered as fast because they are driven by random processes. However, the processing time of a method depends on many factors. Besides the limits of the hardware, also the required accuracy of the results, the amount of data and the susceptibility to noise have to be considered. An efficient model should be simple and flexible at the same time, which has been demonstrated in the case studies above. The efficiency also originates from the above two characteristics – the independence concerning data quality and quantity. We have also shown that in some cases MCMC will not only be the fastest, but also the only feasible sampler in very-high dimensional search spaces (Chapter 2) and that Bayesian inference based on a Markov model has the ability to detect online pattern deviation for real-time data (Chapter 4). The answer can, therefore, be summarized as: Statistical models have advantages when the amount and complexity of data increases.

Our experience shows, however, that in practice statistical models tend to be "easy" to conceive but "hard" to adapt. I.e., despite their power in exploration and optimization driven by random processes, statistical methods may not work well without specific and sophisticated setups. A typical example is the MCMC sampler. A generic MCMC sampler with a purely random transition kernel can be implemented in only a few lines of code, but it will, in most cases, not perform an effective search. The transition kernels, e.g., have to be elaborately designed to adapt them to the individual applications.

Besides this, an appropriate integration of bottom-up and non-statistical methods plays an important role for practical applications. Top-down models driven by statistical processes are flexible, but often very inefficient. Again, the crucial part is the adaption to the particular data. Bottom-up methods can be employed to initially provide reasonable (and sometimes even key) information in the form of "priors" in the Bayesian framework. Non-statistical methods are usually straightforward and produce deterministic results. The latter can be used as improved initial values for the subsequent statistical search and, thus, often greatly contribute to the efficiency.

6.2.2 Balance between Top-down and Bottom-up

Top-down and bottom-up are two diametrically opposed strategies for the processing and interpretation/understanding of data. Although there is no strict and unified definition, they can be described as follows: A top-down approach starts with a general/overall formulation of the whole system and ends up with subsystems and components specified in detail. On the other hand, a bottom-up approach starts from individual basic elements and derives the top-level model by gathering and linking elements and subsystems.

In the context of spatial data understanding, the overall formulation consists of (abstract) geometric models of high-level objects such as buildings and streets. The models are pa-

rameterized and fitted to the measurement data, resulting in the final specified models. The measurement data such as 3D points or pixels are the input to bottom-up approaches. Higher level geometries like lines and planes are constructed to derive the top-level models.

(Huang and Brenner 2011, **P1**) presents a combined strategy with emphasis on bottom-up methods. A sophisticated joint multi-plane detection via 3D Hough transform is proposed. The Region Adjacency Graph (RAG) is used to analyze the relationship between the planes. In the top-down part, a probability-driven eave sweeping is conducted to verify the plausibility of the results. For MAP estimation, the likelihood is calculated based on the reconstruction quality and the priors are derived from given constraints: deviation of the orientation and the symmetry of roof halves.

A pure top-down approach is presented in (Huang et al. 2011, **P2**). The goal is to demonstrate and explore the potential and limits of top-down schemes. It has been demonstrated that the proposed sampler can travel over a relatively large scene and successfully locates and reconstructs the buildings with only generic prior information. To this end, an adapted search scheme have to be used (cf. (Huang et al. 2011, **P2**), Section 3.1) and more search effort is required. We are also aware that the example scene has a limited size and number of buildings, i.e., a low complexity of the joint distribution landscape. In large scenes with more buildings the complexity of the distribution as well as the number of disturbances is so high, that this search strategy will be inefficient or even fail. A complete replacement of the bottom-up process is, thus, hard to achieve and unnecessary.

In practical applications, the real challenge is to balance bottom-up and top-down processing, to achieve robustness as well as efficiency. A possible solution is given in (Huang et al. 2013a, **P3**). It is proposed to divide the data by means of pre-segmentation into individual segments with a single or at most a limited number of buildings. The advantages can be summarized as follows and a comparison of runtimes can be found in (Huang et al. 2013a, **P3**, Section 6).

1. Non-building objects, e.g., trees, can be ignored, because of their relatively small size.

2. The efficiency and stability of the further statistical reconstruction are improved. The search areas are limited to the individual segments, so that the computational complexity is significantly reduced. The search of the optimal models is conducted locally instead of in the whole scene. This avoids many local minima, which make the search unstable and time-consuming.

Please note that in the bottom-up part only simple image processing techniques such as mathematical morphology and “blob” detection are employed. This part leads to a significant performance improvement with relatively low effort. One former work (Huang and Brenner 2011, **P1**), in contrast, presents a combination of more sophisticated bottom-up methods like plane detection and a simple top-down edge sweeping, in which case the efficiency as well as the flexibility of modeling is more limited.

Handling building complexes demonstrates the "coupling" of both bottom-up and top-down processes. The simple bottom-up pre-segmentation determines the location, but cannot further decompose building complexes into individual buildings. I.e., building complexes have to be tackled as a whole in the following top-down process with sophisticated methods like RJMCMC (cf. Section 2.4). However, if the bottom-up part could reliably take over the decomposition of a building complex, e.g., by means of plane or ridge detection and analysis, the top-down processing could directly start with individual building or even primitive reconstruction and the effort for a highly demanding search of multiple buildings and their components in a larger area could be avoided [1].

In summary, the cost of both bottom-up and top-down parts should be well balanced in an approach to achieve an optimal overall performance. The goal is to build an efficient framework for specific data and a practical application.

[1] An approach for this has been recently presented in (Huang and Mayer 2017), which is not included in the main publication list of this thesis.

Bibliography

Ahmed, M., Karagiorgou, S., Pfoser, D. and Wenk, C. (2014): A comparison and evaluation of map construction algorithms, *arXiv:1402.5138v2*.

Aichholzer, O., Aurenhammer, F., Alberts, D. and Gärtner, B. (1995): A novel type of skeleton for polygons, *Journal of Universal Computer Science* **1**(12): 752–761.

Akaike, H. (1973): Information Theory and an Extension of the Maximum Likelihood Principle, *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, Hungary, 267–281.

Bandyopadhyay, P. and Forster, M. (2011): *Handbook for the Philosophy of Science: Philosophy of Statistics*, Elsevier.

Bernardo, J. M. and Smith, A. F. M. (1994): *Bayesian Theory*, John Wiley & Sons, New York.

Bishop, C. M. (2008): A New Framework for Machine Learning, *Computational Intelligence: Research Frontiers* **5050/2008**: 1–24.

Bleyer, M., Rhemann, C. and Rother, C. (2012): Extracting 3d scene-consistent object proposals and depth from stereo images, *Proceedings of the 12th European Conference on Computer Vision - Volume Part V*, ECCV'12, Springer-Verlag, Berlin, Heidelberg, 467–481.

Bu, Y., Chen, L., Fu, A. W.-C. and Liu, D. (2009): Efficient anomaly monitoring over moving object trajectory streams, *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, ACM, New York, NY, USA, 159–168.

Chandola, V., Banerjee, A. and Kumar, V. (2009): Anomaly detection: A survey, *ACM Comput. Surv.* **41**(3): 15:1–15:58.

Cooper, G. F. (1990): The computational complexity of probabilistic inference using bayesian belief networks (research note), *Artif. Intell.* **42**(2-3): 393–405.

Cox, D. R. (2006): *Principles of Statistical Inference*, Cambridge University Press: Cambridge Books Online.

Duan, L. and Lafarge, F. (2016): Towards large-scale city reconstruction from satellites, *in* B. Leibe, J. Matas, N. Sebe and M. Welling (Herausgeber), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 89–104.

Gelfand, A. E. and Smith, A. F. M. (1990): Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**(410): 398–409.

Geman, S. and Geman, D. (1984): Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **PAMI-6**(6): 721–741.

George Casella, E. I. G. (1992): Explaining the Gibbs sampler, *The American Statistician* **46**(3): 167–174.

Goodchild, M. F. (2007): Citizens as sensors: the world of volunteered geography, *GeoJournal* **69**(4): 211–221.

Goodchild, M. and Haining, R. (2003): Gis and spatial data analysis: Converging perspectives, *Papers in Regional Science* **83**(1): 363–385.

Green, P. J. (1995): Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika* **82**: 711–732.

Hastings, W. K. (1970): Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika* **57**(1): 97–109.

Haunert, J.-H. and Sester, M. (2008): Area collapse and road centerlines based on straight skeletons, *GeoInformatica* **12**(2): 169–191.

Hecht, R., Meinel, G. and Buchroithner, M. (2015): Automatic identification of building types based on topographic databases – a comparison of different data sources, *International Journal of Cartography* **1**(1): 18–31.

Henn, A., Römer, C., Gröger, G. and Plümer, L. (2012): Automatic classification of building types in 3d city models, *GeoInformatica* **16**(2): 281–306.

Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T. and Maybank, S. (2006): A system for learning statistical motion patterns, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(9): 1450–1464.

Huang, H. (2015): Anomalous behavior detection in single-trajectory data, *International Journal of Geographical Information Science* **29**(12): 2075–2094.

Huang, H. and Brenner, C. (2011): Rule-based roof plane detection and segmentation from laser point clouds, *Joint Urban Remote Sensing Event (JURSE) 2011, 11-13 April*, IEEE, Munich, Germany, 293–296.

Huang, H. and Mayer, H. (2017): Towards automatic large-scale 3d building reconstruction: Primitive decomposition and assembly, *Societal Geo-Innovation, The 20th AGILE Conference on Geographic Information Science*, Lecture Notes in Geoinformation and Cartography, Springer.

Huang, H., Brenner, C. and Sester, M. (2011): 3D building roof reconstruction from point clouds via generative models, *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS), 1-4 November*, ACM Press, Chicago, IL, USA, 16–24.

Huang, H., Brenner, C. and Sester, M. (2013a): A generative statistical approach to automatic 3D building roof reconstruction from laser scanning data, *ISPRS Journal of Photogrammetry and Remote Sensing* **79**(0): 29 – 43.

Huang, H., Jiang, H., Brenner, C. and Mayer, H. (2014a): Object-level segmentation of RGBD data, **II-3**: 73–78.

Huang, H., Kieler, B. and Sester, M. (2013b): Urban building usage labeling by geometric and context analyses of the footprint data, *26th International Cartographic Conference (ICC)*, International Cartographic Association (ICA), Dresden, Germany.

Huang, H., Zhang, L. and Sester, M. (2014b): A recursive bayesian filter for anomalous behavior detection in trajectory data, *in* J. Huerta, S. Schade and C. Granell (Herausgeber), *Connecting a Digital Europe Through Location and Place, The 17th AGILE Conference on Geographic Information Science*, Volume II of *Lecture Notes in Geoinformation and Cartography*, Springer, 91–104.

Kada, M. and McKinley, L. (2009): 3D building reconstruction from LiDAR based on a cell decomposition approach, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 38(3/W4), 47–52.

Kalman, R. E. (1960): A new approach to linear filtering and prediction problems, *Transactions of the ASME–Journal of Basic Engineering* **82**(Series D): 35–45.

Kang, J., Körner, M., Wang, Y., Taubenböck, H. and Zhu, X. X. (2018): Building instance classification using street view images, *ISPRS Journal of Photogrammetry and Remote Sensing (Available online 2 March 2018)*.

Kass, R. E. and Raftery, A. E. (1995): Bayes factors, *Journal of the American Statistical Association* **90**(430): 773–795.

Kelly, T., Femiani, J., Wonka, P. and Mitra, N. J. (2017): BigSUR: Large-scale structured urban reconstruction, *ACM Transactions on Graphics* **36**(6): 16.

Kim, K., Lee, D. and Essa, I. (2011): Gaussian process regression flow for analysis of motion trajectories, *2011 International Conference on Computer Vision*, 1164–1171.

Koppula, H. S., Anand, A., Joachims, T. and Saxena, A. (2011): Semantic labeling of 3D point clouds for indoor scenes, *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, Curran Associates Inc., USA, 244–252.

Lafarge, F. and Mallet, C. (2012): Creating large-scale city models from 3d-point clouds: A robust approach with hybrid representation, *International Journal of Computer Vision* **99**(1): 69–85.

Lafarge, F., Descombes, X., Zerubia, J. and Pierrot-Deseilligny, M. (2010): Structural approach for building reconstruction from a single DSM, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(1): 135–147.

Laxhammar, R. and Falkman, G. (2014): Online learning and sequential anomaly detection in trajectories, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **36**(6): 1158–1173.

Li, M., Wonka, P. and Nan, L. (2016): Manhattan-world urban reconstruction from point clouds, *in* B. Leibe, J. Matas, N. Sebe and M. Welling (Herausgeber), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 54–69.

Li, Y., Wu, X., Chrysathou, Y., Sharf, A., Cohen-Or, D. and Mitra, N. J. (2011): Globfit: consistently fitting primitives by discovering global relations, *ACM Transactions on Graphics* **30**(4): 52:1–52:12.

Lüscher, P., Weibel, R. and Burghardt, D. (2009): Integrating ontological modelling and bayesian inference for pattern classification in topographic vector data., *Computers, Environment and Urban Systems* **33**(5): 363–374.

Ma, T.-S. (2009): Real-time anomaly detection for traveling individuals, *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, Assets '09, ACM, New York, NY, USA, 273–274.

Masreliez, C. and Martin, R. (1977): Robust bayesian estimation for the linear model and robustifying the kalman filter, *Automatic Control, IEEE Transactions on* **22**(3): 361–371.

Matei, B., Sawhney, H., Samarasekera, S., Kim, J. and Kumar, R. (2008): Building segmentation for densely built urban regions using aerial lidar data, *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 24-26 June*, IEEE Computer Society, Anchorage, AK, USA, 1–8.

Meng, X., Wang, L. and Currit, N. (2009): Morphology-based building detection from airborne lidar data, *Photogrammetric Engineering & Remote Sensing* **75**(4): 437–442.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953): Equation of State Calculation by Fast Computing Machines, *Journal of Chemical Physics* **21(6)**: 1087–1092.

Morris, B. and Trivedi, M. (2008): A survey of vision-based trajectory learning and analysis for surveillance, *Circuits and Systems for Video Technology, IEEE Transactions on* **18**(8): 1114–1127.

Musialski, P., Wonka, P., Aliaga, D. G., Wimmer, M., Gool, L. and Purgathofer, W. (2013): A survey of urban reconstruction, *Comput. Graph. Forum* **32**(6): 146–177.

Nelson, A., de Sherbinin, A. and Pozzi, F. (2006): Towards development of a high quality public domain global roads database, *Data Science Journal* **5**: 223–265.

Ortner, M., Descombes, X. and Zerubia, J. (2007): Building outline extraction from digital elevation models using marked point processes, *International Journal of Computer Vision* **72**(2): 107–132.

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

Pang, L. X., Chawla, S., Liu, W. and Zheng, Y. (2013): On detection of emerging anomalous traffic patterns using GPS data, *Data & Knowledge Engineering* **87**: 357–373.

Partovi, T., Huang, H., Krauss, T., Mayer, H. and Reinartz, P. (2015): Statistical building roof reconstruction from worldview-2 stereo imagery, Volume XL(3/W2), 161–167.

Piciarelli, C., Micheloni, C. and Foresti, G. (2008): Trajectory-based anomalous event detection, *Circuits and Systems for Video Technology, IEEE Transactions on* **18**(11): 1544–1554.

Poullis, C. and You, S. (2009): Automatic reconstruction of cities from remote sensing data, *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 20-25 June*, IEEE Computer Society, Miami, FL, USA, 2775–2782.

Prevost, C., Desbiens, A. and Gagnon, E. (2007): Extended kalman filter for state estimation and trajectory prediction of a moving object detected by an unmanned aerial vehicle, *American Control Conference, 2007. ACC '07*, 1805–1810.

Qi, X., Liao, R., Jia, J., Fidler, S. and Urtasun, R. (2017): 3d graph neural networks for rgbd semantic segmentation, *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 22-25 July*, IEEE Computer Society, Honolulu, HI, USA, 5199–5208.

Rabbani, T., Dijkman, S., Van den Heuvel, F. and Vosselman, G. (2007): An integrated approach for modelling and global registration of point clouds, *ISPRS Journal of Photogrammetry and Remote Sensing* **61**(6): 355 – 370.

Robert, C. and Casella, G. (2011): A short history of markov chain monte carlo: Subjective recollections from incomplete data, *Statistical Science* **26**(1): 102–115.

Romeijn, J.-W. (2014): Philosophy of statistics, *in* E. N. Zalta (Herausgeber), *The Stanford Encyclopedia of Philosophy*, winter 2014 edn.

Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K. (2008): Automated delineation of roof planes in lidar data, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 36(3/W19), 221–226.

Sampath, A. and Shan, J. (2010): Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds, *IEEE Transactions on Geoscience and Remote Sensing* **48**(3): 1554–1567.

Silberman, N. and Fergus, R. (2011): Indoor scene segmentation using a structured light sensor, *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 601 –608.

Silberman, N., Hoiem, D., Kohli, P. and Fergus, R. (2012): Indoor segmentation and support inference from RGBD images, *in* A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid (Herausgeber), *Proceedings of the 12th European Conference on Computer Vision*, Springer-Verlag, Berlin, Heidelberg, 746–760.

Sillito, R. and Fisher, R. B. (2008): Semi-supervised learning for anomalous trajectory detection, *Proc. British Machine Vision Conference BMVC08*, 1035–1044.

Stoter, J. (2005): Generalisation within nma's in the 21st century, *22nd International Cartographic Conference (ICC)*, International Cartographic Association (ICA), A Coruña, Spain.

Sun, L., Li, D., Yi, D. and Liu, J. (2012): Trajectory tracking based on iterated unscented kalman filter of boost phase, *IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 232–235.

Suzuki, N., Hirasawa, K., Tanaka, K., Kobayashi, Y., Sato, Y. and Fujino, Y. (2007): Learning motion patterns and anomaly detection by human trajectory analysis, *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, 498–503.

Taylor, D. F. and Caquard, S. (2006): Cybercartography: Maps and mapping in the information era, *Cartographica: The International Journal for Geographic Information and Geovisualization* **41**(1): 1–6.

Verma, V., Kumar, R. and Hsu, S. (2006): 3D building detection and modeling from aerial LIDAR data, *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 17-22 June*, Volume 2, IEEE Computer Society, New York, NY, USA, 2213–2220.

Vosselman, G. (2009): Advanced point cloud processing, *in* D. Fritsch (Herausgeber), *Photogrammetric Week '09*, Heidelberg, Germany, 137–146.

Wang, H., Wen, H., Yi, F., Zhu, H. and Sun, L. (2017): Road traffic anomaly detection via collaborative path inference from gps snippets, *Sensors*.

Wang, W. and Neumann, U. (2018): Depth-aware CNN for RGB-D segmentation, *arXiv:1803.06791*.

Werder, S., Kieler, B. and Sester, M. (2010): Semi-automatic interpretation of buildings and settlement areas in user-generated spatial data, *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, ACM, New York, NY, USA, 330–339.

Wiedemann, C. and Ebner, H. (2000): Automatic completion and evaluation of road networks, *International Archives of Photogrammetry and Remote Sensing*, B, 979–986.

Zheng, Y., Li, Q., Chen, Y., Xie, X. and Ma, W.-Y. (2008): Understanding mobility based on GPS data, *ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea*, ACM Press, 312–321.

Zheng, Y., Xie, X. and Ma, W.-Y. (2010): Geolife: A collaborative social networking service among user, location and trajectory, *IEEE Data Engineering Bulletin* **33/2**: 32–40.

Zheng, Y., Zhang, L., Xie, X. and Ma, W.-Y. (2009): Mining interesting locations and travel sequences from gps trajectories, *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, ACM, New York, NY, USA, 791–800.

Zhou, Q.-Y. and Neumann, U. (2012): 2.5D building modeling by discovering global regularities, *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 16-21 June*, IEEE Computer Society, Providence, RI, USA, 326–333.

Part II

Publications

Publication list

In this section the publications included in this thesis are collected. For easy indexing the papers are listed concerning the following three criteria: (1) publication date, (2) type of publication, and (3) topic of research.

Index 1: Publication date

2015

[P7] Partovi et al., 2015, PIA+HRIGI
Partovi, T., Huang, H., Krauss, T., Mayer, H. and Reinartz, P. (2015): Statistical building roof reconstruction from Worldview-2 stereo imagery, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL(3/W2): 161–167.

[P8] Huang, 2015, IJGIS
Huang, H. (2015): Anomalous behavior detection in single-trajectory data, International Journal of Geographical Information Science 29(12): 2075–2094.

2014

[P5] Huang et al., 2014a, PCV
Huang, H., Jiang, H., Brenner, C. and Mayer, H. (2014): Object-level segmentation of RGBD data, Photogrammetric Computer Vision (PCV), in conjunction with the European Conference on Computer Vision (ECCV), Zurich, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II(3): 73–78.

[P6] Huang et al., 2014b, AGILE
Huang, H., Zhang, L. and Sester, M. (2014): A recursive Bayesian filter for anomalous behavior detection in trajectory data, Connecting a Digital Europe Through Location and Place: 17th AGILE Conference on Geographic Information Science, Lecture Notes in Geoinformation and Cartography, Springer, 91–104.

2013

[P3] Huang et al., 2013a, ISPRS Journal
Huang, H., Brenner, C. and Sester, M. (2013): A generative statistical approach to automatic 3D building roof reconstruction from laser scanning data, ISPRS Journal of Photogrammetry and Remote Sensing 79(0): 29–43.

[P4] Huang et al., 2013b, ICC
Huang, H., Kieler, B. and Sester, M. (2013): Urban building usage labeling by geometric and context analyses of the footprint data, 26th International Cartographic Conference (ICC).

2011

[P1] Huang & Brenner, 2011, JURSE
Huang, H. and Brenner, C. (2011): Rule-based roof plane detection and segmentation from laser point clouds, Joint Urban Remote Sensing Event (JURSE), 293–296.

[P2] Huang et al., 2011, ACM SIGSPATIAL
Huang, H., Brenner, C. and Sester, M. (2011): 3D building roof reconstruction from point clouds via generative models, 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 16–24.

Index 2: Publication category

Journal

[P3] Huang et al., 2013a, ISPRS Journal

[P8] Huang, 2015, IJGIS

Full-paper reviewed conference

[P1] Huang & Brenner, 2011, JURSE

[P2] Huang et al., 2011, ACM SIGSPATIAL

[P5] Huang et al., 2014a, PCV

[P6] Huang et al., 2014b, AGILE

Abstract-reviewed conference

[P4] Huang et al., 2013b, ICC

[P7] Partovi et al., 2015, PIA+HRIGI

Index 3: Topic

Building reconstruction

[P1] Huang & Brenner, 2011, JURSE

[P2] Huang et al., 2011, ACM SIGSPATIAL

[P3] Huang et al., 2013a, ISPRS Journal

[P7] Partovi et al., 2015, PIA+HRIGI

Building classification

[P4] Huang et al., 2013b, ICC

Anomaly detection

[P6] Huang et al., 2014b, AGILE

[P8] Huang, 2015, IJGIS

RGBD Segmentation

[P5] Huang et al., 2014a, PCV

Not included publications

The publications presented in this section are not included in the main publications of this thesis because of indirectly related topics or techniques. They may, however, provide certain extended and complementary information. These papers are listed concerning publication dates.

Zhang et al., 2017b, Remote Sensing
Zhang, W., Huang, H., Schmitz, M., Sun, X., Wang, H. and Mayer, H. (2017): Effective fusion of multi-modal remote sensing data in a fully convolutional network for semantic labeling, Remote Sensing 10(1):52.

Huang & Mayer, 2017, AGILE
Huang, H. and Mayer, H. (2017): Towards automatic large-scale 3D building reconstruction: primitive decomposition and assembly, Societal Geo-innovation, The 20th AGILE Conference on Geographic Information Science, Lecture Notes in Geoinformation and Cartography, Springer, 205–221.

Rahmani et al., 2017, Hannover Workshop 2017
Rahmani, K., Huang, H. and Mayer, H. (2017): Facade segmentation using structured random forest, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV(1/W1): 175–181.

Zhang et al., 2017a, Hannover Workshop 2017
Zhang, W., Huang, H., Schmitz, M., Sun, X., Wang, H. and Mayer, H. (2017): A multi-resolution fusion model incorporating color with elevation for semantic segmentation, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII(1/W1): 513–517.

Wang et al., 2017, Geospatial Week 2017
Wang, J., Zheng, H., Huang, H. and Ma, W. (2017): Point cloud modeling based on the tunnel axis and block estimation for monitoring the Badaling tunnel, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII(2/W7): 301–306.

Kuhn et al., 2016, ISPRS Congress
Kuhn, A., Huang, H., Drauschke, M. and Mayer, H. (2016): Fast probabilistic fusion of 3D point clouds via occupancy grids for scene classification, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III(3): 325–332.

Huang & Mayer, 2015, ACM SIGSPATIAL
Huang, H. and Mayer, H. (2015): Robust and efficient urban scene classification using relative features, 23th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 81:1–4.

Sester et al., 2015, DGPF
Sester, M., Altermatt, P.P., Holst, H., Huang, H., Schilke, H., Schöber, V., Seckmeyer, G.

and Winter, M. (2015): Vertikale Solarfassaden, Jahrestagung der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung.

Huang & Sester, 2011, GDI
Huang, H. and Sester, M. (2011): A Hybrid approach to extraction and refinement of building footprints from airborne LiDAR data, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII(4/W25): 153–158.