Institut für Geodäsie und Geoinformation
Geoinformation

Statistical relational learning of semantic models and grammar rules for 3D building reconstruction from 3D point clouds

Inaugural-Dissertation
zur
Erlangung des Grades
eines Doktors der Ingenieurwissenschaften
(Dr.-Ing.)

der
Landwirtschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Dipl.-Inform. Youness Dehbi
aus
Beni-Mellal, Marokko
Bonn 2016

Referent:	Prof. Dr. Lutz Plümer
Korreferent:	Prof. Dr. Cyrill Stachniss
Fachnahes Mitglied:	Prof. Dr. Theo Kötter
Tag der mündlichen Prüfung:	30.09.2016

Kurzfassung

Formale Grammatiken eignen sich sehr gut zur Schätzung von Modellen mit a-priori unbekannter Anzahl von Parametern und haben sich daher als guter Ansatz zur Rekonstruktion von Städten mittels 3D Stadtmodellen bewährt. Der Entwurf und die Erstellung der dazugehörigen Grammatikregeln benötigt jedoch Expertenwissen und ist mit großem Aufwand verbunden. Im Rahmen dieser Arbeit wurden Verfahren entwickelt, die diesen Aufwand unter Zuhilfenahme von leistungsfähigen Techniken des maschinellen Lernens reduzieren und automatisches Lernen von Regeln ermöglichen. Das Lernen umfangreicher Grammatiken, die die Vielfalt und Komplexität der Gebäude und ihrer Bestandteile widerspiegeln, stellt eine herausfordernde Aufgabe dar. Dies ist insbesondere der Fall, wenn zur semantischen Interpretation sowohl das Lernen der Strukturen und Aggregationshierarchien als auch von Parametern der zu lernenden Objekte gleichzeitig statt finden soll. Aus diesem Grund wird hier ein inkrementeller Ansatz verfolgt, der das Lernen der Strukturen vom Lernen der Parameterverteilungen und Constraints zielführend voneinander trennt. Existierende prozedurale Ansätze mit formalen Grammatiken sind eher zur Generierung von synthetischen Stadtmodellen geeignet, aber nur bedingt zur Rekonstruktion existierender Gebäude nutzbar. Hierfür werden in dieser Schrift Techniken der *Induktiven Logischen Programmierung (ILP)* zum ersten Mal auf den Bereich der 3D Gebäudemodellierung übertragen. Dies führt zum Lernen deklarativer logischer Programme, die hinsichtlich ihrer Ausdrucksstärke mit attributierten Grammatiken gleichzusetzen sind und die Repräsentation der Gebäude von der Rekonstruktionsaufgabe trennen. Das Lernen von zuerst disaggregierten atomaren Bestandteilen sowie der semantischen, topologischen und geometrischen Beziehungen erwies sich als Schlüssel zum Lernen der Gesamtheit eines Gebäudeteils. Das Lernen erfolgte auf Basis einiger weniger sowohl präziser als auch verrauschter Beispielmodelle. Um das Letztere zu ermöglichen, wurde auf Wahrscheinlichkeitsdichteverteilungen, Entscheidungsbäumen und unsichere projektive Geometrie zurückgegriffen. Dies erlaubte den Umgang mit und die Modellierung von unsicheren topologischen Relationen sowie unscharfer Geometrie. Um die Unsicherheit der Modelle selbst abbilden zu können, wurde ein Verfahren zum Lernen *Gewichteter Attributierter Kontextfreier Grammatiken* (*Weighted Attributed Context-Free Grammars*, WACFG) entwickelt. Zum einen erfolgte das Lernen der Struktur von Fassaden — kontextfreier Anteil der Grammatik — aus annotierten Herleitungsbäumen mittels spezifischer *Support Vektor Maschinen* (SVMs), die in der Lage sind, probabilistische Modelle aus strukturierten Daten abzuleiten und zu prädizieren. Zum anderen wurden nach meinem besten Wissen Methoden des *statistischen relationalen Lernens* (SRL), insbesondere *Markov Logic Networks* (MLNs), erstmalig zum Lernen von Parametern von Gebäuden sowie von bestehenden Relationen und Constraints zwischen ihren Bestandteilen eingesetzt. Das Nutzen von SRL erlaubt es, die eleganten relationalen Beschreibungen der Logik mit effizienten Methoden der statistischen Inferenz zu verbinden. Um latentes Vorwissen zu modellieren und architekturelle Regelmäßigkeiten auszunutzen, ist ein Verfahren zur automatischen Erkennung von Translations- und Spiegelsymmetrien und deren Repräsentation mittels kontextfreier Grammatiken entwickelt worden. Hierfür wurde mittels überwachtem Lernen ein SVM-Klassifikator entwickelt und implementiert. Basierend darauf wurden Algorithmen zur Induktion von Grammatikregeln aus Grundrissdaten entworfen.

Abstract

Formal grammars are well suited for the estimation of models with an a-priori unknown number of parameters such as buildings and have proven their worth for 3D modeling and reconstruction of cities. However, the generation and design of corresponding grammar rules is a laborious task and relies on expert knowledge. This thesis presents novel approaches for the reduction of this effort using advanced machine learning methods resulting in automatically learned sophisticated grammar rules. Indeed, the learning of a wide range of sophisticated rules, that reflect the variety and complexity, is a challenging task. This is especially the case if a simultaneous machine learning of building structures and the underlying aggregation hierarchies as well as the building parameters and the constraints among them for a semantic interpretation is expected. Thus, in this thesis, an incremental approach is followed. It separates the structure learning from the parameter distribution learning of building parts. Moreover, the so far procedural approaches with formal grammars are mostly rather convenient for the generation of virtual city models than for the reconstruction of existing buildings. To this end, *Inductive Logic Programming (ILP)* techniques are transferred and applied for the first time in the field of 3D building modeling. This enables the automatic learning of declarative logic programs, which are equivalent to attribute grammars and separate the representation of buildings and their parts from the reconstruction task. A stepwise bottom-up learning, starting from the smallest atomic features of a building part together with the semantic, topological and geometric constraints, is a key to a successful learning of a whole building part. Only few examples are sufficient to learn from precise as well as noisy observations. The learning from uncertain data is realized using probability density functions, decision trees and uncertain projective geometry. This enables the handling and modeling of uncertain topology and geometric reasoning taking noise into consideration. The uncertainty of models itself is also considered. Therefore, a novel method is developed for the learning of *Weighted Attribute Context-Free Grammar* (WACFG). On the one hand, the structure learning of façades — context-free part of the grammar — is performed based on annotated derivation trees using specific *Support Vector Machines* (SVMs). The latter are able to derive probabilistic models from structured data and to predict a most likely tree regarding to given observations. On the other hand, to the best of my knowledge, *Statistical Relational Learning* (SRL), especially *Markov Logic Networks (MLNs)*, are applied for the first time in order to learn building part (shape and location) parameters as well as the constraints among these parts. The use of SRL enables to take profit from the elegant logical relational description and to benefit from the efficiency of statistical inference methods. In order to model latent prior knowledge and exploit the architectural regularities of buildings, a novel method is developed for the automatic identification of translational as well as axial symmetries. For symmetry identification a supervised machine learning approach is followed based on an SVM classifier. Building upon the classification results, algorithms are designed for the representation of symmetries using context-free grammars from authoritative building footprints. In all steps the machine learning is performed based on real- world data such as 3D point clouds and building footprints. The handling with uncertainty and occlusions is assured. The presented methods have been successfully applied on real data. The belonging classification and reconstruction results are shown.

Contents

1 Introduction

Building models are needed for several tasks such as the calculation of escape routes or urban planning. Particularly, for the visualization in the navigation and tourism context, detailed façade models are essential. Therefore, the demand for high-resolution three-dimensional models of real-world buildings has been sharply increased over the last decade. However, the automatic derivation of such models is still an area of active and intensive research. For autonomous piloting and rescue management, automatic real time modeling and mapping of 3D buildings is of increasing importance. To this end, the research group "Mapping on Demand" (MoD) founded by the German Research Foundation ("Deutsche Forschungsgemeinschaft", DFG) has been established. This project aims to build models of inaccessible three-dimensional man-made objects during the data acquisition by a lightweight autonomously flying Unmanned Aerial Vehicle (UAV). The UAV captures 3D point clouds based on a high-level semantically specified user inquiry. The automatic interpretation of the provided observations in order to identify building parts such as windows or doors is of high interest. Despite the large number of previous work, many questions are not yet satisfactorily answered. In this context, this thesis as part of the MoD project constitutes an important step towards finding automatically semantically interpreted 3D building models.

Especially, the fully automatic extraction of high-resolution sophisticated building models from images or 3D point clouds remains a challenging task. In particular, this is the case if the claim is to generate models that cover semantic hierarchies and aggregations of building parts as well as the dependencies between these parts. The knowledge about building parts such as windows or doors and their locations in a façade provides, for example, an important information for making decisions about response actions in the context of flood planning and management. For instance, Yang and Förstner (2011) employed Conditional Random Fields (CRFs) combined with randomized Decision Forest classifiers to semantically discriminate regions in images of building façades. However, CRFs merely model dependencies in local neighborhood and do not consider constraints between not directly neighboring objects. Figure 1.1 illustrates a.o. the case of two not neighbored windows w_1 and w_3, which cannot be modeled in a wide range context via a CRF. Hence, alternative modeling frameworks are needed in order to overcome this deficiency.

Furthermore, in the context of mapping on demand and rescue management, methods which can deal with occlusions caused by vegetation (cf. Figure 1.1) or smoke are needed in order to predict hidden objects of interest. For example, a precise localization of an opening in a façade floor is decisive for the calculation of an escape route. Hence, in order to meet these expectations, highly structured models representing not only geometry but also semantics and architectural regularities are needed. Such models explicitly provide an a-priori knowledge to the interpretation process and affect to a large degree the quality and usability of the resulting interpretations. Here, the term semantics denotes the relation between terms

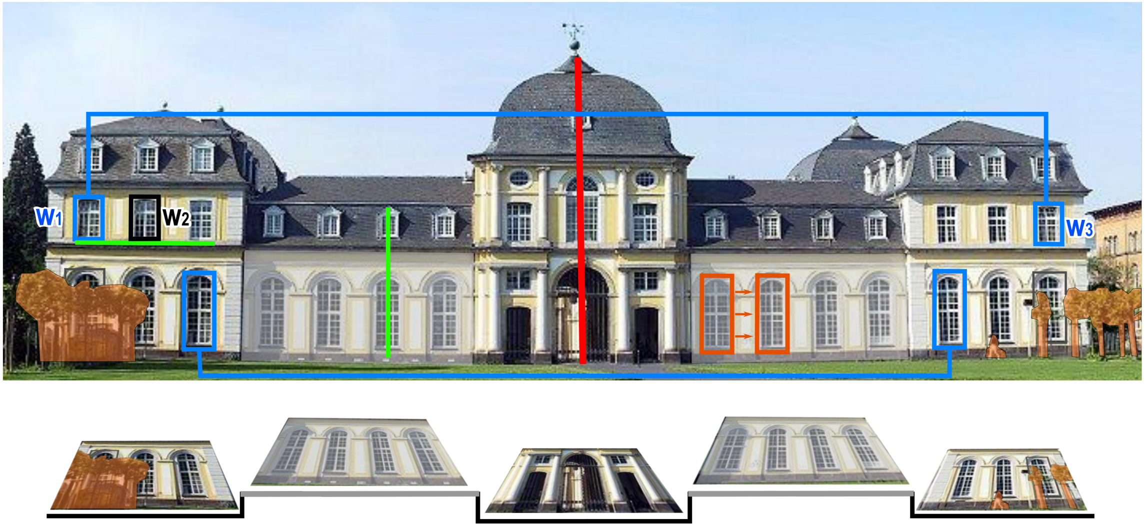

Figure 1.1: Requirements on building modeling methods exemplified by the façade of *Poppelsdorf castle* in Bonn. Models for reflecting local as well as wide range context are needed. Taking axial (red axis) and translational (orange windows) symmetries, alignments (green) and occlusions into consideration plays a key role for the façade interpretation. Protrusions in buildings and building footprints affect the resulting models substantially.

and the objects they refer to, according to the exchange data model *CityGML* (Gröger et al., 2012). Beside semantics and its hierarchies, *CityGML* defines 3D geometry, topology and appearance of urban objects in different level of details (LoD). In the context of semantics, formal grammars have proven their worth for building modeling taking the mentioned aspects into consideration.

Formal grammars represent an important pillar in the formal language theory. They are playing a key role in many application fields such as syntactical and lexical analysis for programming languages or for the processing of sentences of natural languages. In contrast to the sequential structure of texts and strings in natural languages, building models are characterized by a non-linear 3D structure and are mostly extracted from unstructured data such as 3D point clouds. In order to meet this transfer challenge, the hierarchy and taxonomy of buildings are represented using context-free grammars, whereas additional architectural patterns such as symmetries and shape distributions are modeled by attribute grammars as a special type of formal grammars. This is however up to now performed and implemented in a procedural fashion such is the case of procedural modeling approaches for the generation of consistent mass building models (Müller et al., 2007), which requires an explicit representation of regularities and architectural constraints in hard coded procedures. In this way, these approaches lack on flexibility and squander the declarative property of formal grammars for virtual building model generation at cost of reconstruction of existing buildings. Martinović and Van Gool (2013) followed the idea of inverse procedural modeling, which uses formal grammars for the generation of synthetic cities and for the reconstruction of simple 2D façades from real-world existing buildings. Their approach uses splitting heuristics based on defined lattice areas. In contrast to declarative schemes, which represent facts and assertions, procedural representations store actions and consequences. Besides, these approaches do not enable an explicit modeling of the data uncertainty. This thesis is based on a logic-based declarative approach, that separates between knowledge and inference, and thus separates between the representation of buildings and their parts from their reconstruction. Simultaneously, an explicit modeling of data uncertainty and a reasoning over uncertain observations is provided.

In comparison to classical parameter estimation, formal grammars are generic. They allow the estimation of models with a-priori unknown number of parameters like the number of floors or windows. This enables to reflect the variety and complexity of buildings. Furthermore, formal grammars do not require any assumption on the a-priori distribution of underlying data. In the Geoinformation group at the Institute of Geodesy and Geoinformation in Bonn, attribute grammars have been longstandingly used for the automatic generation and reconstruction of buildings and their parts. For instance, the works of Schmittwilken et al. (2009), Gröger and Plümer (2010) and Schmittwilken (2012) demonstrated that attribute grammars are well suited to 3D building modeling. They allow to describe structures and parameters of buildings as well as the substantial constraints between the building parts. Moreover, this type of grammars represents well the taxonomic and the partonomic structures of buildings using a set of rules and constraints. As yet, unfortunately the grammar rules are mostly manually derived.

The automatic learning of grammar rules for building modeling is a very demanding and hard task, as it requires the search over a huge hypotheses space. Particularly it is the case, if generic rules that reflect the complexity of building models are expected. Except for few approaches (Becker, 2009; Ripperda and Brenner, 2009; Martinović and Van Gool, 2013) that tried to extract rules from given data, the design of rules is an expensive and laborious process that relies on expert knowledge. However, most approaches assume façades being planar 2D surfaces and they are almost limited to grid-like designs. Due to the variety and the diversity of buildings, it is hard to manage the complexity, so a machine assisted approach is required in order to acquire a large number of fairly sophisticated grammar rules. To avoid the 2D restriction, the learning should be performed based on 3D point clouds resulting in models that support protrusions in façades or displaced façade parts such as the case in the façade of Figure 1.1. Besides, in order to generate models that reflect real-world man-made objects, the learning process has to cope with uncertain primitives and their mutual relations over unstructured data such as 3D point clouds. Moreover, the learned models have to deal with occluded measured regions of interest and should be able to incorporate prior knowledge in order to exploit the architectural building regularities such as symmetries and alignments. Figure 1.1 shows symmetries and alignments in the *Poppelsdorf castle* façade. An axial symmetry axis is depicted in red color. Two translated windows are highlighted in orange. Vertical as well as horizontal aligned windows are presented in green. Finally, the learning approach should not only deal with uncertain data but also be able to derive generic models, which reflect the likelihood of the models themselves and therefore the variety of different architectural patterns.

Since apart from syntactic differences logic programs and attribute grammars are basically the same language (Deransart and Maluszynski, 1985), logic-based machine learning methods such as Inductive Logic Programming (ILP) and Statistical Relational Learning (SRL) are applied in this thesis in order to learn the rules of a weighted attribute context-free grammar (WACFG) represented by logic programs. Therefore, an automatic approach for the learning and reconstruction of 3D building models has been developed. The proposed approach combines several sophisticated learning methods based on a variety of objects stemming from precise as well as noisy data. The data consists of terrestrial LiDAR 3D point clouds, sketches or labeled examples of façade parts as well as building footprints. In order to cope with the hard learning task, various machine learning and reasoning methods are integrated in a cutting-edge framework and used for the first time for 3D building modeling. Background knowledge such as building regularities and symmetries is modeled as well using particular grammar rules and incorporated to support façade reconstruction.

A simultaneous learning of building structures and parameters addresses their mutual interaction. However, as stated by De Raedt (2008), "Learning both the structure and the parameters[1] of a first-order probabilistic logic program is extremely hard, if not impossible." This thesis takes the approach that separates these two tasks (i.e. structure and parameter learning) from each other and makes clear that the relational learning of building models becomes feasible if the structure is learned first and the parameters afterwards. Similar to the structure learning of Bayes networks, the order affects the learning process of building structures. As stated by Koller and Friedman (2009), "a bad choice of order [of network variables] can result in poor learning results". Hence, an incremental approach with a good chosen order plays an important role in the learning process. This thesis demonstrates that a bottom-up learning starting from the smallest atomic features of a building part together with the semantic, topological and geometric constraints is a key to a successful learning of the whole building part. Structures – symmetries and aggregations – can be modeled using context-free grammars. Further parameters – location, shape and rules distribution – require probabilistic attribute grammars. This work shows that advanced machine learning methods contribute to the automatic learning and reconstruction of 3D building models. Especially, a skillful combination of heterogeneous methods contributed significantly to designing a novel learning and 3D reconstruction approach.

The main contribution of this thesis is a novel method for automatic learning of 3D weighted attribute context-free grammars of buildings and their parts. Additionally, a parsing method of the learned grammar rules is developed for the 3D reconstruction of building façades from 3D point clouds. This thesis demonstrates that it is possible to represent and reconstruct buildings with a fully automatically learned stochastic attribute grammar. For the first time, the representation and the learning of the grammar rules are performed in a pure declarative way using logical and statistical relational learning techniques. This enables not only to cope with an a-priori unknown number of parameters, but also to learn and model the likelihood of the grammar rules. Moreover, the introduced approach is able to deal and learn from both precise models and noisy observations leading to uncertain models as well. Structure hierarchies and their depths, like global and local symmetries, are also modeled and inferred if they are available. The presented approach deals successfully with complexity, uncertainty and unobservability in real-world problems. This is explicitly addressed by

[1] In this thesis, the term parameters denotes not only the weights and likelihood of the grammar rules, but also the shape and location parameters of the building parts as well as their probability distributions.

uncertain projective geometry, probability density functions, probabilistic grammar rules and Markov Logic Networks.

This thesis demonstrates how 3D building models can be represented *in a pure declarative fashion* using statistical and relational formalisms. Herewith, the thesis presents not only the learning of the WACFG but also introduces a method for the parsing of the WACFG rules in order to derive an interpreted model for given data based on different components and methods:

- Logical learning of grammar rules of building parts from precise models and noisy observations
- Statistical relational learning of weighted grammar rules for façade reconstruction
- Grammar-based learning and representation of symmetries in building footprints to support façade reconstruction

The first part deals with the investigation of declarative methods for the learning of building structures. It studies how such structures can be learned via the learning of logic programs for building parts using Inductive Logic Programming and shows that the ILP-based learning is feasible even if the structures are recursive. ILP, which has been successfully applied in many fields such as predicting protein structures (Muggleton et al., 1992), is therefore transferred for the learning of building structures. During a high-level learning, semantic and topological primitives of building parts are induced in a modular and an incremental way using positive and negative examples in addition to a set of background knowledge. In order to handle data uncertainty, this approach is extended by a low-level module, that includes the concept of uncertain projective geometry for learning noisy geometric relations on the one hand and probability density functions together with decision trees for topological constraints on the other hand.

The ILP-based relational approach focuses attention on the uncertainty of the underlying data as well as the geometric reasoning under uncertainty. In order to take the uncertainty of models into consideration as well, a statistical relational learning (SRL) approach, that combines logic and probabilities, is tailored for the automatic learning of 3D building models. In this manner, this approach serves to learn a weighted attribute context-free grammar automatically in order to reconstruct sophisticated façades. The uncertainty of observations is afresh addressed based on geometric reasoning by the use of uncertain projective geometry to provide a knowledge base for an SRL. Furthermore, the likelihood of different structural patterns is modeled using the rule weights from the context-free part of the WACFG as well as the weights of a set of logical formulas using Markov Logic Networks (MLNs). The SRL-based method is an incremental approach, that consists in learning the structure of façades, which corresponds to the context-free part of the grammar in a first step. This is performed using specific Support Vector Machines (SVMs) for structured data based on input-output pairs, that consists of a treebank of labeled façade structures. Afterwards, the context-free model is lifted leading to an attributed model, that expands the grammar rules by further constraints existing between the façade parts using an automatically learned MLN. In this manner, mapping of uncertain data into precise annotations as derivation trees for the learning of the context-free grammar played a key role in order to meet this learning challenge. Likewise, the acquisition of categorical facts from noisy observations using uncertain projective geometry for MLNs has proven to be the key to a successful learning process.

In the last part of this thesis, prior knowledge is modeled and considered in order to support the previously mentioned SRL-based façade reconstruction. It is demonstrated that formal grammars are suitable to describe a-priori hidden and analyzed regular patterns from weak observations. Especially latent symmetry information (axial as well as translational) is identified and represented with context-free grammar rules. For this task, a novel method is designed and implemented for the automatic detection and modeling of hierarchical and repetitive structures from available authoritative building footprints. This information enables the derivation of the symmetry properties of façades from the corresponding footprints without any observations of the façades. Hence, models acquired from the previous approach can be compressed becoming redundant-free. Further, due to existing occlusions in the measured point clouds a model repair or shape completion can be performed. The inherent uncertainty of the geometry of the footprint segments and their angles is considered. This is performed using a supervised classification approach for learning symmetries with SVMs. In contrast to classical statistical methods, for SVMs assumptions on the a-priori distribution of the data are not required. The developed classifier is trained based on a set of polylines labeled as axial symmetric or non axial symmetric. SVMs are simultaneously used for classifying pairs of segments in order to discriminate between translational and non translational segments. Building upon the classification results, context-free grammar rules which represent explicitly axial as well as translational symmetries are derived.

This thesis summarizes the results of my research and my publications on the automatic learning of 3D building models. The developed approaches are presented, discussed and tested on several real-world measured objects. The following publications are most relevant for this thesis and are appended to the thesis:

Dehbi, Y., Plümer, L., 2011. Learning grammar rules of building parts from precise models and noisy observations. ISPRS Journal of Photogrammetry and Remote Sensing 66, 166-176. Quality, Scale and Analysis Aspects of Urban City Models.

Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., Plümer, L., 2016b. Statistical relational learning of grammar rules for 3d building reconstruction. Transactions in GIS, doi:10.1111/tgis.12200.

Dehbi, Y., Gröger, G., Plümer, L., 2016a. Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars. Transactions in GIS, doi:10.1111/tgis.12177.

The remainder of this thesis is structured as follows: The theoretic foundations and methodological background of the used methods will be explained in Section 2. In this section it is explained why these methods are suitable to contribute to solve the addressed questions of this thesis. The developed approaches are presented in Section 3. Especially in Section 3.1, the logical learning of grammar rules from precise models and noisy observations using ILP is demonstrated. Section 3.2 shows how statistical relational learning is exploited in order to learn probabilistic attribute grammar rules. Section 3.3 gives insight into learning and modeling of regularities in building footprints using formal grammars in order to support façade reconstruction. Section 4 concludes and summarizes the thesis and gives an outlook and some perspectives.

2 Methodological background and related work

This chapter gives insight into the relevant methods for this thesis and presents related work. Section 2.1 introduces briefly geometric reasoning from noisy observations using the formalism of uncertain projective geometry, that has been used for the learning and modeling of uncertain geometric relations. Since formal grammars represent a substantial component of this thesis, Section 2.2 gives a short introduction and motivates their application in the field of building modeling and reconstruction. Section 2.3 presents a specific type of Support Vector Machines (SVM) and depicts how it can be used for the learning of building façade structures represented by weighted context-free grammar rules. This section shows as well how these grammar rules are parsed in order to find a most likely structure corresponding to given observations. Section 2.4 introduces the basic notions in logical and statistical relational learning. Beside the theoretical background, this section explains the utility and usefulness of logic and SRL for 3D building modeling and reconstruction.

2.1 Uncertain projective geometry

The process of measurement and mapping of real-world objects leads in most cases to imprecise and error-prone data. Moreover, it is not rare that a real-world object itself is noisy and uncertain. In order to deal with all these kinds of uncertainty, probabilities and statistical approaches such as SRL are suitable. For providing deterministic logical facts for the learning process with SRL or ILP, an uncertain geometric reasoning is needed. Especially if following questions for building modeling are addressed: "are two windows the same?", "are they aligned?". In addition to MLNs and probability density functions, my approach represents and addresses the quality of the given observations and the derived constraints explicitly, using concepts from *uncertain projective geometry* for learning and deriving geometric and topological relations.

Projective geometry is an alternative mathematical framework to *Euclidean geometry*, which enables a simple and consistent representation and transformation of geometric entities such as lines or planes. In order to enable geometric reasoning using projective geometry, an Euclidean entity $x \in \mathbb{R}^n$ can be in general represented as a homogeneous vector (Dorst et al., 2007) $x \in \mathbb{R}^{n+1}$. Table 2.1 gives an overview of the different entity representations. It further depicts examples of possible geometric constraints between these entities such as paralleltity and orthogonality, which are expressed as simple multi-linear relations such as cross or dot product. A list of all possible relations can be found in (Heuel, 2004).

For an efficient performance of geometric reasoning under uncertainty, Heuel (2004) integrated the potentials of projective geometry and statistics. His approach draws upon the

Geometric entity	Euclidean representation	Projective representation
point	$X = (x, y, z)$	$X = (tx, ty, tz, t)$
plane	$A = (a, b, c, d)$	$(A_h, A_0) = (a, b, c, d)$
line	$L = A \cap B$	$(L_h, L_0) = (A_h \times B_h, A_0 B_h - B_0 A_h)$
Geometric relation		**Algebraic representation**
orthogonal(planeA, planeB)		$c = A_h^T B_h \stackrel{!}{=} 0$
orthogonal(lineL, planeA)		$c = L_h \times A_h \stackrel{!}{=} 0$
parallel(planeA, planeB)		$c = A_h \times B_h \stackrel{!}{=} 0$
parallel(lineL, planeA)		$c = L_h^T A_h \stackrel{!}{=} 0$
identical(PointX, PointY)		$c = X \wedge Y$ $= (X_h Y_0 - Y_h X_0, X_0 \times Y_0) \stackrel{!}{=} 0$

Table 2.1: Projective representation of objects and constraints

modeling of error propagation during the reasoning process, which consequently enables testing uncertain spatial relations between geometric entities. This is achieved by extending the homogeneous entity x with its covariance matrix Σ_{xx} in order to get an uncertain entity (x, Σ_{xx}). A hypothesis test of spatial relations between geometric entities is usually reduced to the test whether an observed n-vector c has to be zero. For instance, the parallelity of two planes can be proven if the cross product of their homogeneous part is zero (see Table 2.1). Against the statistical test theory (Koch, 1999), this leads to the following chi-square test:

$$T = c^T \Sigma_{cc}^{-1} c \sim \chi_d^2, \tag{2.1}$$

with d being the degree of freedom of the constraint, that has to be tested. In case the covariance matrix does not have a full rank, that is if *rank* $(\sum_{cc}) = r \leq d$, the following test has to be performed:

$$T = c^T \Sigma_{cc}^{+} c \sim \chi_r^2, \tag{2.2}$$

hereby $\sum_{cc}^{+}$ denotes the pseudo-inverse of the covariance matrix $\sum_{cc}$ (Förstner, 2005). With a significance number α, T is now compared with the critical value $\chi_{r,1-\alpha}$ of the χ^2-distribution. If the following inequality

$$T > \chi_{r,1-\alpha}^2 \tag{2.3}$$

is satisfied, then the hypothesis on c is rejected.

To sum up, uncertain projective geometry provides a compact and consistent representation and transformation of geometric entities together with their uncertainties. In this thesis, statistical geometric reasoning using uncertain projective geometry is performed in order to make decisions about similarity of geometric entities such as windows and hence to answer the questions about similarity mentioned above. This enables to provide the required categorical facts for the subsequent logical and statistical relational learning. To this aim, the SUGR[1] library has been used. SUGR expects covariance matrices of 2D or 3D geometric objects like points or lines as uncertainty measures. For constructing or estimating new uncertain entities from existing ones, the error propagation is modeled enabling tests of uncertain spatial relations. In this thesis, the test for identity and similarity of building parts is performed in the feature space instead of 3D coordinate system of the façade. In this manner, for example, the test whether two windows $w1(width1, height1, depth1)$ and $w2(widtth2, height2, depth2)$ are geometrically identical (same shape parameters) is reduced to an identity test of two 3D points $p1(width1, height1, depth1)$ and $p2(widtth2, height2, depth2)$ consisting of shape parameters of the compared windows. Analogously, the verification of the alignments of windows is reduced to the verification of the parallelity of two lines using a chi-squared statistical hypothesis test. However, the verification of orthogonality for different surfaces belonging to subsequent steps in a staircase is realized in a classical way using their corresponding planes in the 3D coordinate system of the façade.

2.2 Formal grammars for building modeling

Ever since formal grammars were introduced by Chomsky (1956, 1959) for reconstructing sentences of natural language, they have also been used to generate formal languages. Likewise they have been widely applied for the generation of synthetic city models. In this context, Wonka et al. (2003) introduce *split grammars* for a rule-based generation of architectural structures inspired by the *shape grammar* works of Stiny et al. (1971) and Stiny (1982). Instead of strings from natural languages, the grammar symbols stand for geometric shapes in comparison to ordinary formal grammars.

A formal grammar G can be defined as quadruple {S, N, T, P} of a start symbol S, a set of non-terminals N represented by capitalized initials, a set of terminals T denoted by lower case initials and a set of production rules P. A special case of formal grammars are context-free grammars, which correspond to type 2 according to Chomsky's hierarchy, which distinguishes between four levels (type 0 - type 3) of formal grammars. Production rules appear in the form A $\rightarrow$ a, where A is a non-terminal, and a is a sequence of terminals and non-terminals. This rule implies that each occurrence of the symbol A can be replaced by the string a. In a weighted context-free grammar, each rule is augmented by a weight, which expresses the likelihood of this rule.

As illustrated in Figure 2.1, a possible context-free derivation, describing the structure of a gable roof building, can be modeled as follows: *Building* as a start symbol is made of the non-terminal *Corpus* and the terminal *gableRoof*. In addition to the *Facade*, *Corpus* consists of *left* side, *right* side, and *back* side. *Facade* is built of *window* and *Entrance*, which is made of *door* and *Stairs*. For the sake of simplicity of the example we assume that the façade has one single window. *Stairs* can be substituted either by one *Step* and *Stairs* as a sequence of steps or terminates with a *Step*, which in turn consists of a *riser* (vertical rectangle) and

[1] http://www.ipb.uni-bonn.de/data-software/sugr/

a *tread* (horizontal rectangle). Likewise, a building can be represented by the production rules in Listing 2.1.
The set of non-terminals is N = {*Building*, *Corpus*, *Facade*, *Entrance*, *WinArray*, *Stairs*, *Step*}; the set of terminals is T = {*gableRoof*, *left*, *right*, *back*, *window*, *door*, *riser*, *tread*}.

```
Corpus    → left Facade right back
Facade    → Entrance WinArray
WinArray  → WinArray window
Entrance  → door Stairs
Stairs    → Stairs Step
Stairs    → Step
Step      → riser tread
```

Listing 2.1: Context-free production rules of a gable roof building

In the field of procedural modeling, which consists in the generation of a huge number of synthetic buildings based on a-priori designed grammar rules, Müller et al. (2007) introduced the so-called *Computer Generated Architecture (CGA) shape grammars* and proposed a system for the stepwise generation and refinement of consistent mass building models, based on the mentioned *split grammars*. This grammar encodes the building style, structure and appearance based on procedurally implemented rules. Like split grammars, *CGA shape grammars* use a set of basic shapes together with corresponding parameters and symbols for the modeling of 3D buildings. This is performed with *split rules* describing a splitting process of a shape into several sub-shapes.

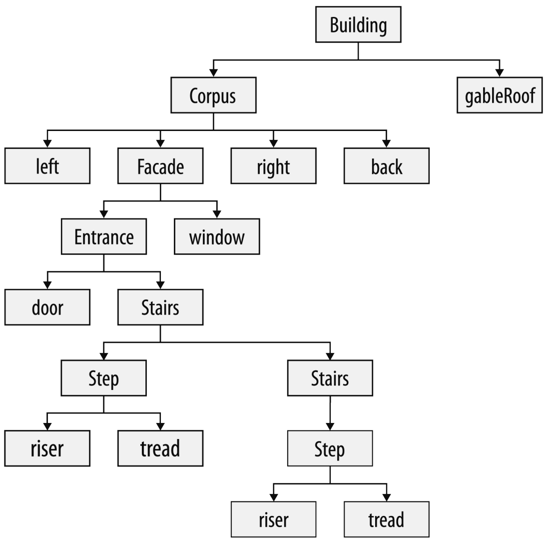

Figure 2.1: A possible derivation tree of a context-free gable roof building grammar

In order to represent geometry as well as semantics for 3D building models, Schmittwilken et al. (2009) adapted the concept of attribute grammars (Knuth, 1968, 1971), which enabled to represent shape and location parameters of the modeled building objects and express the geometric and topological constraints between these objects and their parts. Hereby, terminals and non-terminals are expanded by attributes, whereas production rules are extended by semantic rules. The latter specify the constraints among the attributes. In our example, these enable to state that all steps within the same staircase have the same dimensions, which cannot be expressed by context-free grammars. An extract of an attribute grammar of stairs is shown in Listing 2.2. The grammar symbols at the top are augmented by the attributes *height*, *depth*, and *width*, which are used in the semantic rules R2(P2) to R4(P2) at the bottom, in order to specify the identity between the shape parameters of risers and treads within the same stairs. Additionally R1(P1) ensures that the new stairs consist of exactly one more step than the one before. The superscript indices *n* and *n-1* are used to differentiate between multiple occurrences of the same symbol.

```
P1: Stairs^n → Step Stairs^(n−1)
P2: Stairs   → Step
P3: Step     → riser tread
```

```
R1(P1): Stairs^n.numberOfSteps = Stairs^(n−1).numberOfSteps + 1
R2(P2): Step.width  = Stairs.width
R3(P2): Step.height = Stairs.height
R4(P2): Step.depth  = Stairs.depth
...
```

Listing 2.2: Excerpt of an attribute grammar for stairs

In addition to (Schmittwilken, 2009) other approaches follow the idea of inverse procedural modeling, which uses formal grammars not only for the generation of synthetic cities but also for the reconstruction of real and existing buildings. Ripperda and Brenner (2009) used a-priori defined grammar rules combined with reversible jump Markov Chain Monte Carlo for supporting façade reconstruction. Martinović and Van Gool (2013) introduced an approach for learning so-called *Bayesian grammar* for two-dimensional façade generation and reconstruction from image data. They infer split grammar rules from labeled images. However, their approach assumes that façades are planar 2D surfaces, which can be reduced in 2D lattices. Teboul et al. (2011) used shape grammars for 2D façade parsing, applying reinforcement learning and taking only grid-like design patterns into consideration. Becker (2009) combined approaches for the interpretation of image as well as 3D laserscan data with split grammars in order to detect and reconstruct windows and doors in façades. During the interpretation, a so-called *façade grammar* is induced from one building façade. However, the induced rules are not generic and are only used for the reconstruction of a single 2D façade or façades of uniform architectural style buildings. The approach assumes also that the considered façade is a 2D planar surface. Gröger and Plümer (2010) elaborated an approach adapting attribute grammar rules for the generation of indoor models for route planning.

The reason why we advocate, learn and apply weighted attribute context-free grammars (WACFGs) is their declarative nature, their flexibility and their genericity. WACFGs are

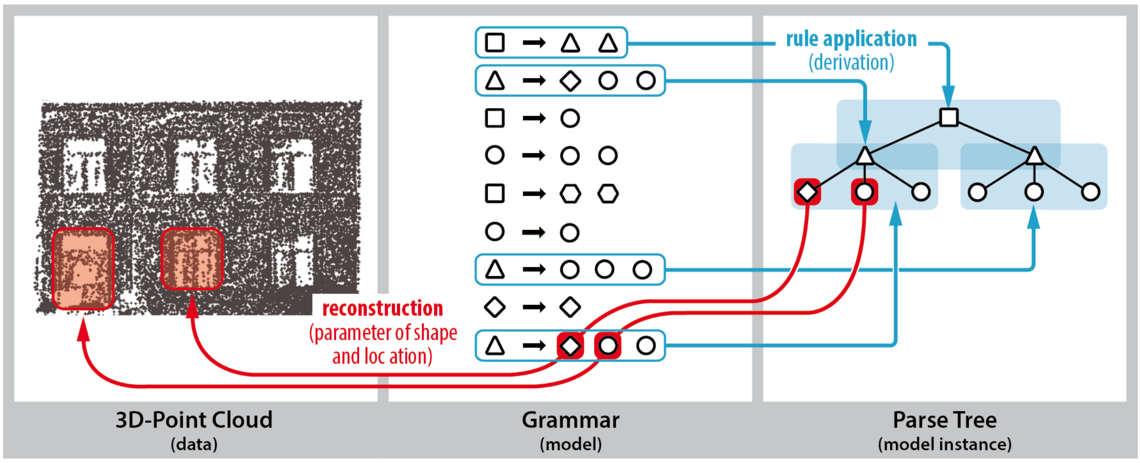

Figure 2.2: The parsing of a 3D point cloud using a formal grammar [modified figure from Schmittwilken (2012)].

able to represent a whole architectural style with an unlimited number of different buildings. WACFGs allow the estimation of building models with an a-priori unknown number of parameters like the number of floors and windows. They represent well the taxonomy (semantic hierarchy) such as different types of openings (windows and doors). Furthermore, they enable the expression of partonomies (aggregation hierarchy) like the disaggregation of floors in several windows. In contrast to CRFs, for example, wide range dependencies like symmetries can be modeled. Besides, the constraints among the building parts such as the alignments of neighboring windows can be expressed using semantic rules. The aim of this thesis is to learn stochastic attribute grammar rules automatically. Once a grammar has been learned, it is used as model or background knowledge in order to find a most likely parse tree, that matches an underlying 3D point cloud well, with successively grammar rule application (see Figure 2.2). The next sections introduce methods for the automatic learning of a weighted attributed context-free grammar and give foundations of an approach for parsing 3D point clouds using the learned grammar rules.

2.3 Support vector machines for learning a weighted context-free grammar

Towards machine learning of 3D building models the separation of structure and constraints as well as parameter learning is a key step in order to deal with the variety and complexity of buildings. This implies that the aggregation and taxonomy of buildings, especially façades, have to be a-priori learned at first. To this aim, frameworks that are able to generalize and induct models from labeled structured examples are required. The generalization from previously seen examples is the classical task of supervised machine learning. In this context, the main problem is to learn a function predicting the value of a variable y for an observed x based on a set of input and output pairs. While usually y is a simple value label, the learning scenario here demands complex outputs y for the prediction of a parse tree of a given context-free grammar.

The learning of models that take functional dependencies between arbitrary input (e.g. variable length) and outputs (e.g. structured data) into account represents a further challenge for machine learning. In contrast to multinomial classification relied to a finite number of labels, the outputs here are structured objects such as trees or sequences. In this context, Tsochantaridis et al. (2004) proposed a framework for structured interdependent output learning. They learn a model mapping inputs $x \in X$ into complex labels $y \in Y$ based on a sample of input-output pairs $(x_1, y_1) \dots (x_n, y_n) \in X \times Y$. Herewith, a fix a-priori unknown probability distribution is assumed. To this end, a *discriminant function* $F : X \times Y \to \Re$, that measures the compatibility of x and y, is defined in order to induce a mapping f:

$$f(x; w) = \operatorname*{argmax}_{y \in Y} F(x, y; w), \tag{2.4}$$

with a parameter vector w. In the case that $y \in Y$ is a labeled tree the function F is chosen such that it generates a model isomorphic to a probabilistic context-free grammar. A node in an output parse tree y for an input terminal sequence x corresponds to a grammar rule g_i with an associated score w_i. The possible trees are evaluated using the sum of the w_i of the belonging nodes. These possible trees correspond to those generating the terminal sequence x applying derivations beginning from the start symbol S. In this context, F can be introduced as a score function as follows:

$$F(x, y; w) = \langle w, \Psi(x, y) \rangle = \sum\nolimits_{g_i \in rules(y)} w_i, \tag{2.5}$$

where $\Psi(x, y)$ denotes the frequency of each grammar rule g_i in the output tree y, whereas the weight vector w consists of corresponding weights w_i. The computation of $f(x; w)$ is performed by identifying a parse tree y that maximizes $F(x, y; w)$ using the CKY algorithm (Manning and Schütze, 1999).

In order to penalize parse trees that significantly deviate from the correct parse tree and tolerate those that differ in only a few nodes, loss functions (Weston et al., 2002) that measure the discrepancy between the predicted structure and the expected output are considered. To this aim, alternative loss functions instead of the *zero-one* loss function are used. In this context the following function $\Delta(y_i, y) = (1 - F_1(y_i, y))$ for two outputs y_i and y is usually considered in natural language processing. The calculation of this loss function is based on the *F1* score representing the harmonic mean of precision and recall taking the overlap of nodes between trees into consideration. Therefore, the large margin principle used in Support Vector Machines (Vapnik, 1998; Schölkopf and Smola, 2001) is generalized in order to predict structured outputs. A violation of the margin constraint with a high loss is then penalized more than a violation that generates an output with smaller loss. Iteratively, the single most violated constraint is retained together with a model for each training example untill no constraint violation occurs with respect to a given threshold. The system *svm_ cfg*[2] provides an implementation of the Support Vector Machine algorithm for learning a weighted context-free grammar.

[2]http://www.cs.cornell.edu/people/tj/svm_light/svm_cfg.html

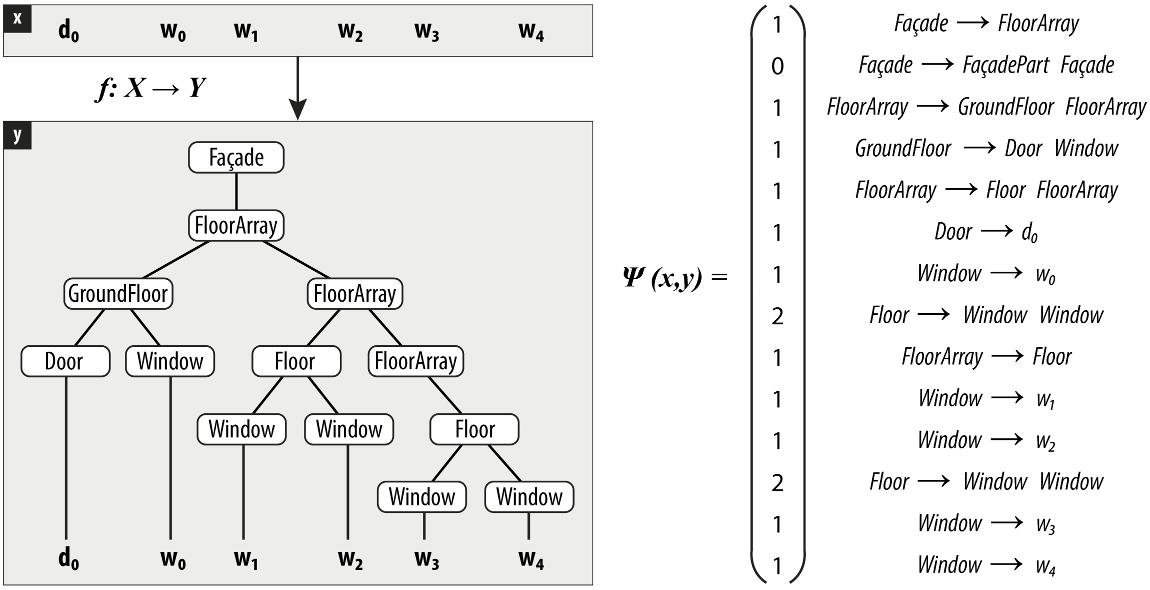

Figure 2.3: An illustrative parsing model of a façade example. The output parse tree y is derived given a building part sequence x using a.o. the rule frequency vector $\psi(x, y)$.

In this manner, a learning of a *Weighted Context-Free Grammar* (WCFG) for 3D building modeling is performed in this thesis based on a labeled treebank of annotated buildings from different building styles. The resulting WCFG consists of a set of context-free rules together with weights designating the importance of the given rule. In contrast to classical SVMs, which expect a feature vector with a fixed size and atomic labels, the feature vector here has an arbitrary size and the labels consist in structured parse trees. Figure 2.3 illustrates an instance x of the feature vector consisting of a sequence of observed façade parts (grammar terminals) as well as a parse tree y as label. Based on the terminal sequence x defining the type of some new observed building parts, a parse tree y is predicted using the learned weighted context-free grammar against equation 2.5. For example, the rule $Floor \rightarrow Window\ Window$ is applied twice in the parse tree y for generating the word x, therefore the frequency 2 is recorded in $\Psi(x, y)$ accordingly.

In summary, the method of Tsochantaridis et al. (2004) enables the learning of a weighted context-free grammar from labeled façade trees like the tree in Figure 2.3. Besides, this method provides the possibility to acquire a most likely parse tree based on a sequence of observed building parts from a 3D point cloud with respect to the induced grammar model. The acquisition of the expected input sequence it beyond the scope of this thesis. However, an approach for the identification of façade object detectors is presented in Dehbi et al. (2016c) in order to test the learned WACFG. This approach will be shortly described in Section 3.2. In this way, using the building part detectors and the learned weighted grammar, the semantic interpretation of façades from a 3D point cloud is reduced to predicting a parse tree, that fits the structure of the given parts. In the current section, the introduced method represents only the taxonomy and hierarchy of building parts using the weighted context-free grammar. This enables to represent the likelihood of different structural patterns, for example, the likelihood to divide a façade horizontally or vertically for a given building style. The next section presents methods for learning attributed rules based on logical and statistical relational learning for modeling the constraints between building parts.

2.4 Logical and statistical relational learning

Apart from formal grammars and SVMs for structured data, this thesis draws upon ideas of logical and statistical relational learning using Inductive Logic Programming (ILP) techniques and Statistical Relational Learning (SRL) methods, such as Markov Logic Networks (MLNs). ILP is a subarea of artificial intelligence, which combines machine learning with logic programming. Thus, the goal of ILP, which is inherited from inductive machine learning, consists in developing techniques to induce hypotheses from observations as well as synthesizing new knowledge from experience by using computational logic as representational scheme. ILP has been successfully applied as a learning approach in many different fields. In bio-informatics, for instance, ILP has been used to predict protein structure (Muggleton et al., 1992) and mutagenicity (King et al., 1996). Mooney (1997) has applied ILP to natural language processing. For the first time, this thesis transfers, adapts and demonstrates the potential of ILP in order to learn 3D building models inspired by these and other works.

Before describing the basics of ILP some terminology has to be introduced (De Raedt, 2008). First, a function is called *predicate* if it returns a truth value, e.g. *parallel*. Second, a *term* is a constant, a logical variable or a structured term $f(t_1, ..., t_n)$ composed of a functor f and n terms t_i. Constants represent objects in our domain, e.g.: "window1", "window2", "door1". A logical variable X ranges over the domain of objects and can take values from this domain. Third, an *atom* is a predicate symbol followed by its necessary number of terms, e.g. *line(X)* is an atom that represents a line, which is in turn represented by the term X as a variable. Fourth, a *literal* is an *atom* or its negation.

By using these definitions the key concept of a *horn clause* can be defined as an expression of the form:

$$b_1 \wedge ... \wedge b_m \rightarrow h,$$

in which h and b_i are logical atoms. The symbol '$\wedge$' symbolizes a conjunction whereas '$\rightarrow$' stands for an implication. Clauses of the form: *true* $\rightarrow h$ are called *facts*.

An example of a horn clause can be illustrated by a parallel relation in the two-dimensional space:

$$\begin{aligned} line(X) \wedge \ line(Y) \wedge \ line(Z) \wedge \ orthogonal(X, Z) \\ \wedge \ orthogonal(Y, Z) \rightarrow parallel(X, Y). \end{aligned} \tag{2.6}$$

Hereby, a new predicate *parallel* is defined as *head* of the rule (the arrow's left hand side being the rule's *body*). A line X and a line Y satisfy this predicate if there is a line Z which is orthogonal to both lines X and Y. ILP systems are able to learn such first order horn clauses (such as example 2.6) by the use of background knowledge and examples. The ability to provide declarative background knowledge to the learning system is one of the most distinct advantages of ILP. This background knowledge can be given in the form of horn clauses or as ground facts. Hereby, the prior knowledge about the predicates that appear in the learned hypothesis later is represented. For example, if instances of the predicates *orthogonal* and *line* (cf. last example clause 2.6) are provided, they can serve as background knowledge in order to learn the *parallel* relation.

Given a set E of positive and negative examples as ground facts and a background knowledge B, a hypothesis H has to be found which explains the given examples with respect to B and meets the language constraints. The hypothesis H has to be complete and consistent. H is complete if all positive examples are covered. If none of the negative examples are covered, H is considered to be consistent. The coverage of an example $e \in E$ is tested with a function *covers*(B,H,e), which returns the value *true* if H explains e given the background knowledge B, and otherwise returns *false*.

One of the most popular ILP approaches for learning first order horn clauses is the *Progol* algorithm with many different implementations. In this thesis the *Aleph* engine (Srinivasan, 2007) was used as a *Prolog* based implementation of *Progol*. The *Progol* algorithm is outlined in Listing 2.3. *Progol* first selects a positive seed example and then finds a consistent clause, which is the *most specific clause* (MSC) of this example and covers this example. This step is called the saturation step. Against the theoretical background of inverse entailment (Muggleton, 1995) the MSC can be acquired in order to guide the search. In this way, *Progol* learns by using a single example and by verifying the consistency of its generalization with the dataset. This generalization is added to the background knowledge. Afterwards, all redundant examples which have been covered by the MSC, are removed. This process is repeated until a theory is found which covers all positive examples. The coverage function is defined as follows:

$$covers(B, H, e) = true \; if \; B \cup H \models e.$$

In other words, the hypothesis H covers the example e with respect to the background knowledge B if $B \cup H$ semantically entails e.

```
1. Select a positive example
2. Saturate this example i.e. construct its Most Specific Clause (MSC)
3. Generalize MSC
4. Remove covered positive examples
5. Repeat if positive examples remain
```

Listing 2.3: Progol Algorithm

In order to examine the goodness of clauses, each clause is evaluated by a score function. In this context, the default evaluation of *Aleph* is the function *coverage*, which defines the clause utility as *P-N*, in which P is the number of positive and N the number of negative examples covered by the considered clause respectively. This can also be realized by other evaluation functions, for example, entropy.

ILP tasks are search problems in the space of possible hypotheses. *Aleph* bounds this space with the MSC as lower bound, whereas the top of the search space is bound by the empty clause. Once the MSC has been built, *Aleph* performs a top-down search in the space of possible specializations by using θ-subsumption as refinement operator, which enable a partial ordering of clauses, to decide which clause is more general than another (Lavrac and Dzeroski, 1994). In this process only the literals appearing in the MSC are used to refine clauses.

Apart from syntactic differences, logic programs and attribute grammars are closely related. Logic programs can be transformed into semantically equivalent attribute grammars (Deransart and Maluszynski, 1985). Thus, the learned logic programs can safely be used for reconstruction tasks realized by attribute grammars. This thesis exploits available ILP systems like *Aleph* in order to learn *categorical* logic programs. In contrast to procedural implementations of attribute grammars, logic programs are declarative and enable a multi-directional constraint propagation using logical inference. Declarative logic programs represent assertions based on logical facts. In this manner, they separate between knowledge and inference. Thus, for the task of building modeling, they separate between the representation of buildings and their parts from their reconstruction. Procedural approaches define building splitting schemes based on algorithms, which specify actions and consequences. This reduces the degree of flexibility of these approaches and makes the learning of grammar rules harder because they learn algorithms and we learn models.

A first-order logic programs represent a set of hard constraints on the set of ground facts. This is attributed to the categorical property of the logical propositions, which only allows the expression of true or false values. A violation of a clause with a given fact leads to a non-satisfiability of this logical formula. However, since the observations are noisy and the building structures are complex, these categorical logical methods are not sufficient. In order to overcome this deficiency, a simultaneous handling of inherent uncertainty and the exploitation of compositional structures is desirable. To this end, various approaches such as MLNs, that combine logic and probabilities, have been introduced in the field of SRL. MLNs soften the hard constraints of first-order logic in such a way that a violation of a given formula makes it less probable but not impossible.

MLNs (Richardson and Domingos, 2006) represent a simple and compact framework, that integrates ideas from logic and probabilistic graphical models. In an MLN, each formula is expanded by a weight in contrast to the deterministic logic-based approaches. This enables to handle uncertainty in a sound way tolerating imperfect and even contradictory knowledge. MLNs are considered as templates for constructing *Markov Random Fields (MRFs)*. In contrast to *CRFs*, that model conditional probabilities, *MRFs* model the joint probabilities in a generative way. An automatic grounding of an MLN specifies an instantiation of an MRF enabling the exploitation of available probabilistic inference techniques, such as Gibbs sampling or Belief Propagation. In the following, some important concepts from probabilistic graphical models are introduced.

An MRF characterizes a model for the joint distribution of a set of random variables $X = (X_1, X_2, ..., X_n) \in \mathcal{X}$ over the domain $\mathcal{X}$ and consists of an undirected graph $G = (V, E)$. V is a set of nodes where each node V_i corresponds to a random variable $X_i \in \mathbf{X}$ and E is a

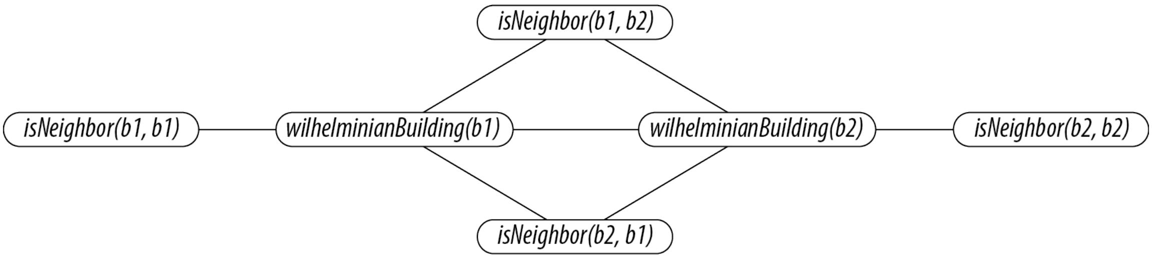

Figure 2.4: Ground Markov Random Field (MRF) obtained by the application of an MLN containing the Formula 2.8 to the constants "b1" and "b2". Equation 2.9 corresponds to the resulting grounded formula. For the sake of graphical simplicity "b1" and "b2" stand for "building1" and "building2" respectively.

set of undirected edges between the nodes. The joint distribution represented by the MRF over the variables is given as follows:

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_k f_k(\mathbf{x}_k) \ ,$$

where f_k is a non-negative, real-value potential function defined over the kth clique in G, i.e., $f_k : \mathcal{X}^{|\mathbf{x}_k|} \to \mathbb{R}^+$. $\mathbf{X}_k \subseteq \mathbf{X}$ denotes the set of nodes occurring in the kth clique. Z is a partition function, which enables a normalization ensuring a valid probability distribution.

Given an MRF statistical inference techniques are exploited for solving the MAP inference problem, i.e., finding $\mathbf{x}^*$ with

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n}{\operatorname{argmax}} \ p(\mathbf{x}) \ . \tag{2.7}$$

Here, x_i refer to an instantiation of X_i, i.e., $x_i \in \mathcal{X}_i$. More information on MRFs and probabilistic graphical models in general can be found in Koller and Friedman (2009).

An MLN is defined as a set of pairs (F_i, λ_i) (Richardson and Domingos, 2006), where F_i is a formula in first-order logic and $\lambda_i \in \mathbb{R}$ is the weight of formula F_i. Furthermore, a finite set of constants $C = \{C_1, \ldots, C_{|C|}\}$ is required in order to induce an MRF automatically. For every ground predicate in the MLN a node is added to the MRF, e.g. `isNeighbor(building1, building2)`, corresponds to a binary variable in the MRF whose value models the truth state of the ground predicate.

Equation 2.8 shows an example of a formula F_i with an attached weight 2.5.

$$\begin{aligned} 2.5: \ & wilhelminianBuilding(X) \wedge \ isNeighbor(X, Y) \to \\ & \qquad wilhelminianBuilding(Y). \end{aligned} \tag{2.8}$$

Furthermore, one potential function for each possible grounding of a formula F_i is added corresponding to a clique in the graph. For example, if one instantiation of formula 2.8 is:

$$\begin{aligned} & wilhelminianBuilding(building1) \wedge \ isNeighbor(building1, building2) \to \\ & \qquad wilhelminianBuilding(building2), \end{aligned} \tag{2.9}$$

then a potential $\exp(\lambda_i f_k)$ is added to the MRF. This means that there is an edge between two random variables if the corresponding ground predicates are appearing together in at least one grounded formula. Figure 2.4 illustrates an automatically grounded MRF from an MLN, that contains formula 2.8 with $C = \{building1, building2\}$.

The derivation of the fist-order rules of an MLN based on a given training data set is an important task, which has been a subject of intensive research. For example Kok and Domingos (2005) presented a beam search approach guided by a pseudo log-likelihood measure enabling to outperform ILP-based solutions (De Raedt and Dehaspe, 1997; Richardson and Domingos, 2006). Moreover, in order to learn the weights of an MLN in an efficient and automatic way, generative approaches perform in general a maximization of a pseudo log-likelihood instead of the log-likelihood $\ln p(\mathbf{x}; \mathbf{w})$ of an MRF with respect to a given training data set. In contrast to generative training, Singla and Domingos (2005) followed a discriminative approach splitting the predicates into evidence and query predicates in order to optimize the computational complexity. The conditional likelihood for query predicates is subsequently maximized based on the observed predicates. An implementation of MLNs is provided by *Alchemy*-system[3] allowing structure and parameter learning as well as probabilistic logic inference. Alchemy uses MaxWalkSat (MWS) for MAP inference by default. However, it is possible to make use of message-passing algorithms such as max-product Belief Propagation.

In this thesis, MLNs are used in order to generalize and extend the two-valued categorical logic by probabilities enabling a *stochastic* logic. In this way, due to measurement errors and uncertain observations, two windows that nearly have the same shape parameters are rather similar than dissimilar. This can be expressed and enforced using a first-order formula with an attached weight like Formula 2.8. Section 3.2 introduces experimental results of automatically learned rules and their weights based on logical facts extracted from annotated buildings façades. The fact extraction is performed using uncertain projective geometry as described in Section 2.1. An important benefit of MLNs is the automatic induction of an MRF from the facts without the need of a human expert to define dependencies. Based on the induced MRF, statistical inference is used in order to answer questions such as the geometrical similarity of windows. All in all, in this thesis, MLNs enable the modeling and learning of attributed weighted constraints among building parts.

[3] `alchemy.cs.washington.edu/`

3 Statistical relational learning of weighted attribute context-free grammars

This chapter is devoted to giving insights into the most relevant publications appended to this thesis. The related articles are presented and discussed in different sections. Section 3.1 deals with structure learning of building parts and demonstrates how logical and relational learning techniques are successfully applicable, despite recursive structures. The same section shows how ILP as logical relational framework can be extended by uncertain projective geometry and probability density functions in order to learn not only from precise models but also from noisy observations such as 3D point clouds. In addition to learning from uncertain data, Section 3.2 is addressed to building models considering the uncertainty of the models into account. To this aim, this section shows how probabilistic grammar rules describing façade structures are learned and how logical learning is combined with graphical models enforcing the topological and geometric constraints with uncertainties. Section 3.3 presents an automatic learning approach for the identification of inherent geometric redundancy, the modeling of latent prior knowledge and the exploiting of architectural regularities like symmetries in order to support the previously described 3D building reconstruction process.

3.1 Logical learning of grammar rules from precise models and noisy observations

This section introduces a novel approach for the automatic learning of building structures modeled by formal grammar rules described and published in the *ISPRS Journal of Photogrammetry and Remote Sensing* (Dehbi and Plümer, 2011). So far, except from few approaches that tried to extract the grammar rules from given data (Ripperda and Brenner, 2009; Becker, 2009; Martinović and Van Gool, 2013), the rules are mostly manually derived. However, this is an expensive and laborious task, that needs expert knowledge. Moreover, the mentioned approaches generated procedural models for virtual 3D scenes using split grammar rules, which are rather suitable for the generation of synthetic building models than for the reconstruction of existing complex real-world 3D buildings. Besides, most approaches assume façades being planar 2D surfaces and they are almost limited to grid-like designs. This section introduces a relational logic-based method for the learning and modeling of 3D building structures in a pure declarative scheme. This approach shows how building structures can be learned despite challenging recursive structures of some parts like stairs. Apart from the geometric and the topological constraints between the stairs and their parts, the specific challenge is to handle the recursion as a specific machine learning task. This complexity has been overcome by using a bottom-up approach, which consisted of learning the non-recursive primitives of stairs first before learning the whole stairs. This reduced the gap between low level atomic structures and high level models enabling to reduce the search space accordingly.

In a first step, in order to evaluate the potential of ILP for the automatic learning of building parts, stairs have been used as a representative example of regular structures in man-made objects. Stairs are highly structured in a regular way. But the number of steps is arbitrary and consequently have to be modeled in a recursive way. Moreover, stairs are composed of aggregated parts characterized by geometric and topological constraints in each aggregation hierarchy. These constraints can be specified either by the identity of coordinates or by customized relations. In this context, the choice of the constraint modeling considerably affects the search space of rules. At this stage, it is assumed that both examples for steps and stairs and counterexamples, in the sense of ILP, are given by precise descriptions without noise. This means that the user provides consistent examples interactively.

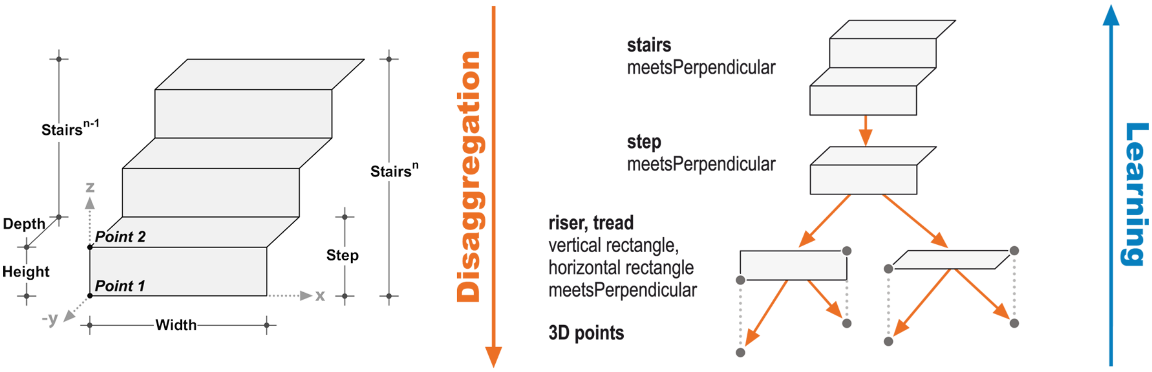

Figure 3.1: Stairs as a recursion of steps (left). Incremental learning of semantic and topological primitives of stairs (right). Semantic primitives are marked in bold, whereas topological and geometric constraints are marked in normal font.

To cope with the complexity of learning strongly interrelated concepts (aggregation and recursion, geometric and topological constraints) a modular incremental bottom-up approach has been developed in this thesis. The ability to model background knowledge enables a multi-stage learning process of stairs, which consists of two stages. First, the non-recursive parts of stairs are learned, and then, building upon these parts, the learning of the recursive clause is performed. Interestingly, only a few positive and negative examples are sufficient for the learning of this clause. This divide and conquer approach is illustrated in Figure 3.1 on the right. Stairs are disaggregated into their primitives. Disaggregation occurs in a top-down manner, whereas learning is realized by a bottom-up approach, starting from the smallest atomic feature to the whole stairs object. The disaggregation and learning directions are indicated by descending and ascending arrows, respectively. On the level of disaggregation, stairs are constructed as a recursion of steps (cf. Figure 3.1 on the right), which in turn are composed of horizontal and vertical rectangular faces. Both faces can be

Table 3.1: The whole learned logic program of stairs.

```
1) stairs([st(Point1,Point2,Height,Depth,Width)], Height,Depth,Width) ←
                                      step(Point1,Point2,Height,Depth,Width).

2) stairs([st(Point21,Point22,Height,Depth,Width),
      st(Point11,Point12,Height,Depth,Width) | Tail], Height,Depth,Width)
      ← stairs([st(Point11,Point12, Height,Depth,Width) | Tail], Height,Depth,Width),
         meetsPerpendicular(Point22,Point11), step(Point21,Point22,Height,Depth,Width).
--------------------------------------------------------------------------------
3) step(Point1,Point2,Height,Depth,Width) ← riser(Point1,X,Y,Z1,Height,Width),
                  tread(Point2,X,Y,Z2,Depth,Width), meetsPerpendicular(Point1,Point2).

4) riser(Point,X,Y,Z,Height,Width) ← point(Point,X,Y,Z),
                  vecPlus(X,Z,X1,Z1,Width,Height), point(Point1,X1,Y,Z1).

5) tread(Point,X,Y,Z,Depth,Width) ← point(Point,X,Y,Z),
                  vecPlus(X,Y,X1,Y1,Width,Depth), point(Point1,X1,Y1,Z).
--------------------------------------------------------------------------------
6) meetsPerpendicular(Point1,Point2) ← riser(Point1,X,Y,Z,Height,Width),
                  plus(Z,Height,Z2), tread(Point2,X,Y,Z2,Depth,Width).

7) meetsPerpendicular(Point2,Point1) ← tread(Point2,X,Y,Z,Depth,Width),
                  plus(Y,Depth,Y1), riser(Point1,X,Y1,Z,Height,Width).
```

defined by two user-defined 3D points. In contrast to the disaggregation, the starting point on the level of learning is a pair of two observed left and right 3D points for identifying the horizontal (tread) and the vertical rectangles (riser) respectively.

The resulting rules are shown in Table 3.1. In order to learn the end recursive rule (1-2), only four positive and two negative examples are required. This low number can be attributed to the stepwise learning strategy. It can be verified that the difference between these rules and those of the attribute grammar in Listing 2.2 (cf. Section 2.2) is merely syntactic. For instance, the attribute *Width* in the logical predicate *step* is the counterpart of *Step.width* in attribute grammar notation. It should be noted that the rules in Listing 2.2 are only a subset of the rules of Table 3.1.

As described in Section 2.4, the task of learning requires a set of positive examples, which are generalized with respect to the set of negative examples and the background knowledge. *Aleph* supports first-order horn clauses as background knowledge and is further able to learn ranges and functions with numeric data. These functions can also be used as background knowledge, and they represent a good basis to describe the geometric and topological constraints inside building parts. This is particularly important at the low level of the learning process, which starts with taking the point coordinates into consideration. Apart from the observed 3D points in the case of riser, the background knowledge includes arithmetic, namely the operations *plus* and *vecPlus*. They define scalar and vector additions, which are used to specify topological and geometric constraints between these points of the rectangle and their coplanarity. Since stairs are invariant in rotation it can be, without loss of

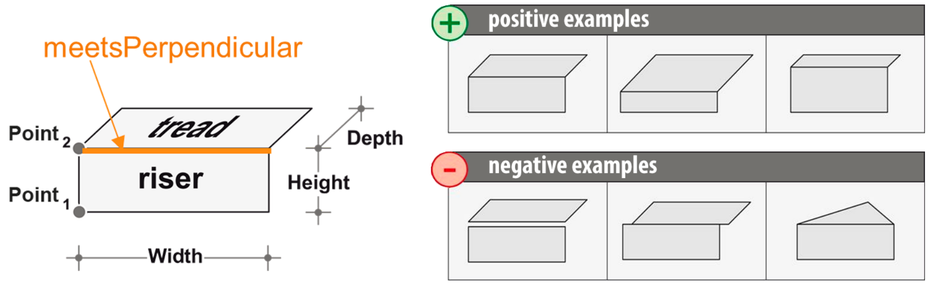

Figure 3.2: Incremental learning of stairs using ILP. The learning task requires positive and negative examples.

generality, assumed an axis parallelity according to the representation on the left in Figure 3.1.

In total, two positive examples and two negative examples are necessary to learn a riser. The learning of the concept of tread happened analogously, where the same number of examples is required. Rules 4) and 5) show the learned rules of riser and tread. Once riser and tread have been learned, they can be used as primitives in order to learn the concept of step. Therefore, they are added to the background knowledge. In the following, the incremental learning process of a step will be elaborated exemplarily. For the completion of the necessary background knowledge of a step, information about the adjacency between risers and treads is required. This is expressed by the relation *meetsPerpendicular* (see rules 6-7 and Figure 3.2), which ensures that risers and treads meet each other but only touch in their boundaries. This is ensured by the arithmetic operation *plus*, which moves the left point of riser by the distance *Height* along the z-axis. The perpendicularity is ensured by the vertical and horizontal alignment of the riser and tread, respectively.

Now it is possible to learn the concept of a step. Rule 3) shows the result clause, which defines a step. The head of this rule represents the whole step object, whereas the body defines its aggregated primitives as well as the geometric and topological relations between them. The attributes *Point1* and *Point2* serve as identifiers for riser and tread, respectively, which the whole step is composed of. The geometric description of the model is given by the location and shape parameters. The location parameter of *step* is described by the attribute *Point1*. Figure 3.2 demonstrates that this attribute represents the left point of its riser. The remaining attributes *Height*, *Depth* and *Width* constitute the shape parameters. For the consistency of the model, the location as well as the shape parameters have to be propagated, that is, they occur as attributes in the head of the rule. The relation *meetsPerpendicular* specifies the semantic rules of the attribute grammar (cf. Listing 2.2) implicitly. The acquired background knowledge together with the positive and the negative examples are the basis of the inductive learning process. In order to understand how the examples influence the learning, an excerpt of positive and negative examples, which has been used to learn a step, is illustrated in Figure 3.2.

The concept of steps is only one part of the concept of stairs. The adjacency and orthogonality between steps has to be learned as well. In contrast to the *meetsPerpendicular* in rule 6), which describes these properties within a step, the *meetsPerpendicular* in rule 7) specifies them between two neighboring steps within stairs. Now the necessary background knowledge to learn the recursive concept of *stairs* has been acquired. The rule for stairs has already been shown in Table 3.1 (cf. rules 1) and 2)). The square brackets in this rule symbolize a list in *Prolog*. In other words, stairs are represented as a list of steps, which is separated into head and tail ($stairs([head|tail])$). The new *stairs* on the left side of the rule are represented recursively and consist of the stairs from the right side concatenated with the next new neighboring *step*. As mentioned above, the neighborhood of the steps is ensured with the *meetsPerpendicular* relation. This concatenation of steps implies that new stairs consist of exactly one step more than the one before. Once again it should be noted that four positive examples and two negative examples are sufficient to learn the recursive clause *stairs*. Several ILP systems such as Golem (Muggleton and Feng, 1990) or FOIL (Quinlan, 1990) were able to learn quicksort as recursive program from categorical facts. However, man-made objects such as buildings, windows, steps and stairs are different. Especially, the learning task over imprecise observations such as 3D point clouds has to deal with non-categorical examples.

The previous step focused on the logic learning of grammar rules of stairs, where the learning examples consisted of exact descriptions of the model. The model has to be learned without taking measurement errors into consideration. This requires a precise presentation of the object parts and their mutual relations. In order to check the geometric relations between entities such as planes, it must be ensured that a certain distance measure has to equal zero. Since in reality this is not usually the case, two planes which meet with an angle around 90 degrees, for example, have also to be considered as orthogonal. Of course, the user could just provide sketches, which the system could transform into precise descriptions, using snapping functions based on thresholds and some heuristics. However, in order to model uncertainty, an appropriate representation of the geometric objects, and especially of relations between these objects, is crucial.

The previously described approach is not restricted to stairs. A summary of a generic method for the learning of building parts is depicted in Figure 3.3. For an arbitrary building part, the extended learning approach can be divided into low and high level learning as outlined in the figure. During low level learning, geometric as well as topological constraints between the elements of the considered building part are identified and learned. Uncertain projective geometry and probability distributions are used as instruments in order to accomplish the low level learning and consequently tackle uncertainty in the considered data. Once these constraints have been identified, they are converted into logical facts, which will be provided as background knowledge to the high level learner. Afterwards, the second phase can take place in an incremental and interactive way such as described in the precise learning step. The low level learning comprises the selection of the structure of interest, the forming of surfaces and learning of topological and geometric relations. During the learning from precise models these steps have been handled by ILP due to the absence of measurement errors. The high level learning deals with concepts on a higher hierarchical level and is therefore independent from imprecise measurements. Thus, in this phase the formalism of ILP can be applied similar to the learning from precise models.

The first step in the low level learning consists in estimating geometric entities which compose a building part, e.g. a staircase. From a 3D point cloud an estimation algorithm is

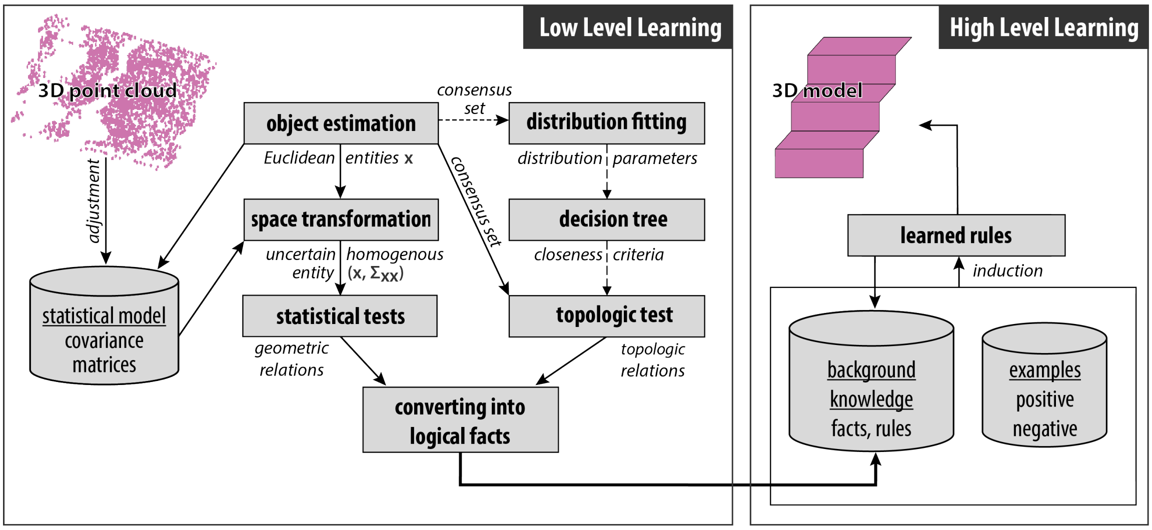

Figure 3.3: Summary of a generic method for learning buildings and their parts

performed in order to obtain the planes in the observed staircase. Before doing so, the surface normals are identified and then clustered in sets with similar normals using DBSCAN (Ester et al., 1996). This preprocessing avoids the estimation of geometric entities based on points from different surfaces. For the estimation issue the MLESAC algorithm (Torr et al., 2000) is used enabling a robust estimation of the planes from the given point cloud. In order to find all planes describing a staircase, a sequential implementation of MLESAC enables a robust estimation of these planes. After the *object estimation* has been performed a set of estimated planes in $\mathbb{R}^3$ is acquired accordingly. In order to model the error propagation, the estimation error of each entity x has to be calculated. This is achieved by determining its covariance matrix Σ_{xx}. Therefore an adjustment problem is solved based on the observations and the consensus set in order to get a *statistical model* for the estimated objects (Koch, 1999). For reasons of efficiency a transformation from the Euclidean representation of geometric entities into the homogeneous representation takes place. Afterwards, a statistical hypothesis test is performed in order to decide about the orthogonality or parallelity of two entities. In this manner knowledge about the mutual geometric relations, like parallelity and orthogonality between the different entities, is built. Together with the topological information provided in the next step, these relations complete the background knowledge to start the incremental high level learning (cf. Figure 3.3).

In order to achieve a correct description and specification of 3D geometric objects, especially stairs, a transition from reasoning with planes to surfaces is necessary. Herewith, it is necessary to determine which 3D point blobs represent horizontal and vertical surfaces that meet each other. However, this information is neither available in the planes nor in the 3D point cloud. In order to tackle this problem, a knowledge based approach, which consists of three steps is proposed. First a *distribution fitting* is performed on blobs, which represent the consensus sets of the planes, in order to estimate the data boundaries of the 3D point coordinates. For this task, bounded probability density functions such as the beta distribution (Johnson et al., 1995) are suitable. If the y-coordinates of the 3D points which

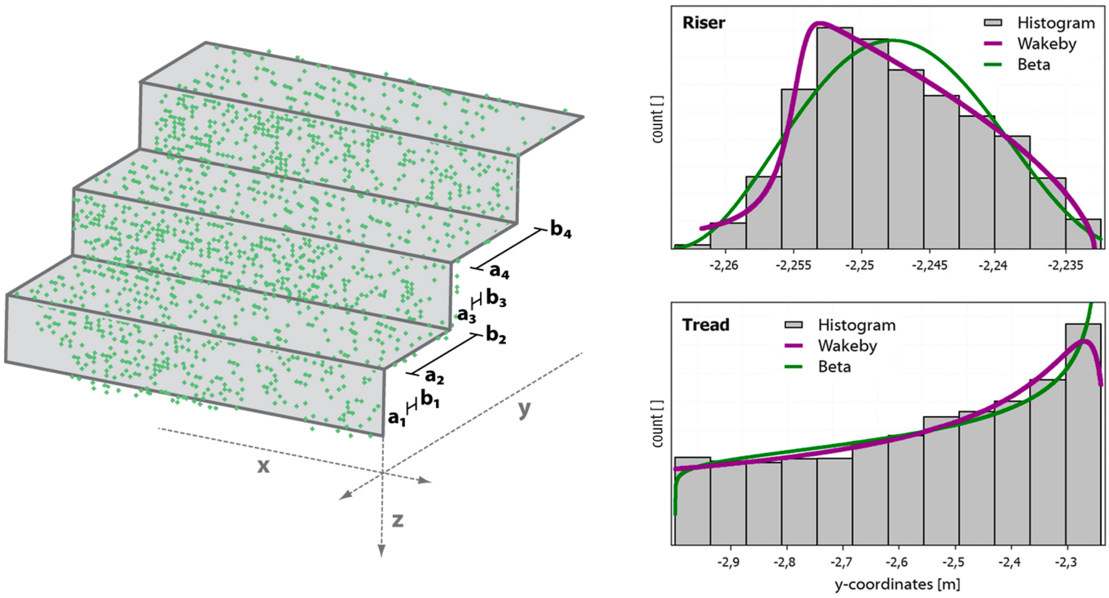

Figure 3.4: Left: 3D model of a staircase from 3D point cloud. Boundaries of the blobs in y-direction correspond to the location parameters of a probability distribution. Right: Visualization of Wakeby (purple) and beta (green) distribution fitted to the histogram of y-coordinates of a riser (top) and tread (bottom)

belong to a given blob were beta distributed, an estimation of the two last beta parameters would lead to the estimation of the y-range boundaries of the considered blob. However, for the simultaneous estimation of all four parameters of the beta distribution the first four moments have to be used. Since the usage of higher moments adds more noise than signal to the estimation process, the estimation may not be accurate. In order to tackle this problem, the so called *Wakeby distribution* was introduced by Houghton (1978). The choice of a distribution cannot be based on the theoretical arguments only, without taking the data into consideration. For this aim, about 50 distributions were examined to find the best description of the underlying 3D points representing treads and risers of different staircases. The probability distribution function which fits the data best was selected after performing three statistic tests: Kolmogorov-Smirnow, Anderson-Darling and Chi-Squared (Lehmann and Romano, 2005). For each blob at least two of these tests stated that the Wakeby distribution ranked among the two best distributions. Figure 3.4 on the right exemplarily shows the result of the data fitting for y-coordinates of 3D points which belong to a riser (top) and tread (bottom) respectively.

Furthermore, taking into account the information that Wakeby distribution is the best distribution fitting the given data, the next step consists in performing a parameter estimation to learn a closeness criterion of two different blobs. This knowledge based approach is realized on the basis of their 3D points, which build the consensus set obtained from the plane estimation. These criteria can be learned by using a decision tree, which will enable us to decide whether these blobs are neighbored. For this, the different deviations between the estimated y-boundaries from the Wakeby distributions are taken as features in order to learn a specific closeness value in a supervised way. An example of such features is the distance between the upper and the lower bound b_i and a_{i+1} of two different blobs (cf. Figure 3.4 on the left). Finally, we are interested in planar surfaces rather than in planes. Therefore, the same approach used in order to estimate the depth of the blobs turns out to be appropriate for the determination of the width and height of each blob. The boundaries obtained from the distribution fitting together with the estimated planes localize a reference point of a bounding rectangle.

To sum up, the ability to incorporate background knowledge turns out to be a crucial instrument to cope with the model complexity. A very limited number of examples are sufficient to explain to the machine a human understanding of stairs. Recursion did not turn out to be a major obstacle. This is the first step for the generation of semantic models and attribute grammar rules of man-made objects such as buildings. The learned models have been applied as strong prior in order to identify stairs in a huge 3D point cloud (Schmittwilken, 2012) using an intelligent sampling strategy. The presented method focused attention on the uncertainty of the underlying data and the geometric reasoning under uncertainty. The next section shows how the uncertainty of models can be taken into consideration as well and gives insights into a tailored statistical relational learning method, that combines logic and probabilities for the automatic learning of 3D building models.

3.2 Statistical relational learning of grammar rules

In this section a novel approach for the automatic learning of a weighted attributed context-free grammar (*WACFG*) for the identification and reconstruction of façades from 3D point clouds is proposed, as described in (Dehbi et al., 2016b). This work is published in the *Transactions in GIS* journal. In contrast to the method in the previous section, the learning approach is also able to model the uncertainty of the derived models in addition to the consideration and modeling of the data uncertainty. Furthermore, a key advantage of this approach is treating unobservability of building parts and missing values due to occlusions or noise. Besides, the proposed method addresses not only planar façades but also façades with sophisticated 3D structures. Figure 3.5 shows that the proportion of buildings with displaced façades and oriels cannot be neglected. Attribute grammars extend context-free grammars by attributes and semantic rules. In the context of 3D modeling this provides more expressive power in order to model the constraints between the primitives of the modeled 3D objects. In this way, geometric, topological and semantic constraints characterizing human-made objects can be adequately modeled. In contrast to procedural methods, a pure declarative approach, that separates the representation of buildings and their parts from the reconstruction task, is proposed. The learned *WACFG* is used for modeling as well as for reconstruction tasks. The *WACFG* describes the taxonomic and partonomic structure of buildings by a weighted context-free grammar (*WCFG*) and the substantial constraints, which are described using Statistical Relational Learning (SRL) methods, namely MLNs.

Figure 3.6 shows the main façade of the *Poppelsdorf castle* in Bonn. The parse tree reflects the taxonomic and partonomic structure of the façade. The latter is aggregated from five

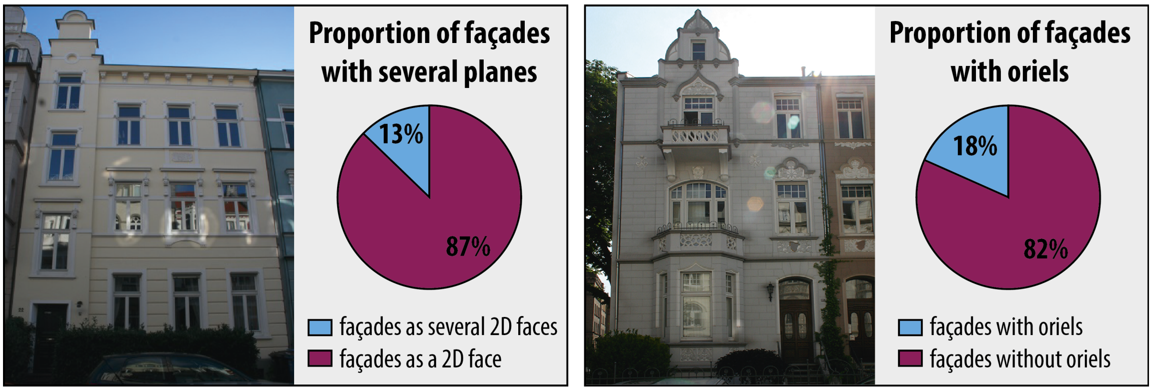

Figure 3.5: The distribution of displaced façades and oriels in selected areas of Bonn, Germany. Due to the existence of displaced façade elements such as parts (covering all floors, see left side) or oriels (not covering all floors, see right side), façades can not be always adequately modeled as 2D faces. The percentage of buildings with such structures is significant.

façade parts in a recursive way according to rules stemming from an induced *WCFG*. An excerpt of these rules is shown in the box right on the top. The p_i indicate the probability for applying a given rule according to a defined distribution over the structures generated by the grammar. The logical formulas of the MLN are depicted in the box left at the bottom of the figure. The weights λ_i have been automatically learned and denote the importance of the associated formula. The formulas have been learned as well and indicate the constraints, e.g., the vertical and horizontal alignments, and the background knowledge, e.g., neighborhood and floor information, of the underlying 3D objects.

SRL models, unlike what is traditionally done in statistical learning, seek to avoid explicit state enumeration, using a symbolic representation of states. The advantage of these models lies in the ability to succinctly represent probabilistic dependencies on an object-type level, i.e., first-order level, among the attributes of different related objects. This enables a compact representation of learned models, that allow for sharing of parameters of similar objects. Besides, SRL methods allow combining the uncertainty of the observations as well as the structural models. The learning of a *WCFG* enables the modeling of façade structures especially their aggregation in different parts. Furthermore this gives insight in the distribution and importance of different structural patterns by weights, that expand the classical context-free grammar rules. In comparison to classical parameter estimation, *WACFG*s are generic and enable to model objects with a-priori unknown number of parameters such as the number of floors and windows. The introduced approach explicitly addresses the uncertainty of observations by uncertain projective geometry, probabilistic rules and MLNs such as described in Section 2. All in all, WACFG enables to deal with the complexity and variety of real-world buildings. All components of the *WACFG* are automatically learned from examples. The grammar rules and their probabilities are learned by SVMs, the MLN is learned using statistical relational learning methods. The learned *WACFG* is applied to reconstruct buildings from observations using classification by SVMs and MLN inference.

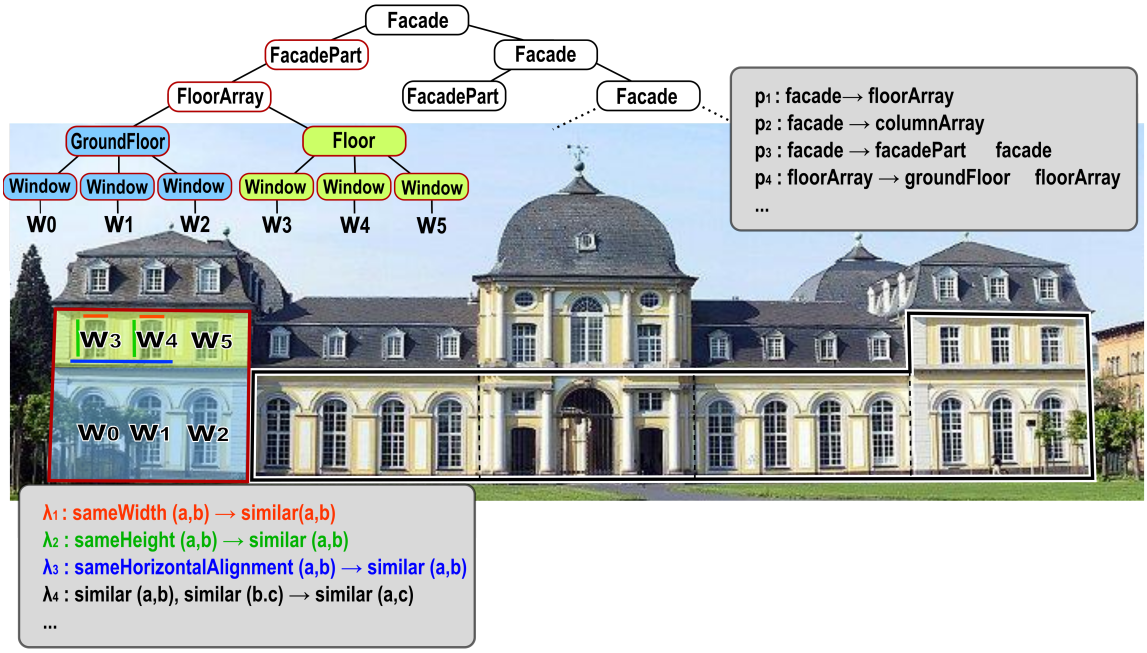

Figure 3.6: The main façade of the Poppelsdorf castle in Bonn modeled with the weighted attribute context-free grammar. The structure of the façade is described by a parse tree (top in the middle) derived from the weighted context-free part of the grammar (top on the right). The constraints and attributes of the building parts are modeled in a relational way using Markov Logic Networks (bottom).

To the best of my knowledge, this is the first demonstration of the impact of adapting statistical relational learning for 3D building reconstruction.

In order to cope with the learning task, an incremental strategy is followed consisting in learning the structure of façades first and afterwards the parameters of the building parts as well as the related constraints. Figure 3.7 gives an overview of this approach. In the first step, a weighted context-free grammar is learned based on a relational building database (*RBDB*) consisting of 1300 annotated buildings from different regions in Bonn, Germany. Their façades represent different building styles reflecting the variety as well as the complexity of building models. To this end, the open-source *measureFacade*[1] tool has been used for the interactive measurement of images as well as of 3D point clouds. The building parts of interest are façades, windows, doors, balconies and oriels. The shape as well as the location parameters of each building part are captured. In contrast to Ripperda (2008), the relative location of each part, e.g. floor and column, is retained not only from images but also from 3D point clouds. All buildings are characterized by their architectural style such as "post-war era" in Germany or "Wilhelminian" and their types e.g. single-family house or multi-family house. Furthermore, the general shape of the related footprint is stored. Each building consists of one or more façades with their relative position in the building. A façade consists of several parts such as windows or oriels. Each part is associated to definite columns and floors in order to describe structural information. All data is manually taken either from undistorted, rectified and scaled images or from high resolution LiDAR 3D point clouds. 1230 façades were taken with a Canon 350D (focal length: 18mm-55mm fixed on 18mm) or Nikon D700 (fixed focal length on 20mm) digital single-lens reflex camera with calibrated lenses. 70 façades were captured by static scanning with a Leica HDS6100 laser scanner.

In order to learn façade structures, a supervised learning approach is followed using Support Vector Machines for structured data as described in Section 2.3. This is performed based on input-output pairs. In the prediction stage (Figure 3.7, yellow background), for a given façade instance, the most likely parse tree, that represents its structural description, taxonomy and partonomy is predicted. The input is a sequence of strings, identifying the type (window, door, balcony and oriel) of the façade parts. The input sequence is acquired by applying façade object detectors, that use *kernel density estimation (KDE)* (Wand and Jones, 1994; Wang and Suter, 2004). This step, however, is beyond the scope of this thesis. An approach for the identification of façade object detectors is proposed in Dehbi et al. (2016c) for testing the learned grammar. *KDE* enables a non-parametric estimation of a probability density function without data distribution assumptions leading to the shape and location parameters of the façade parts. For example, openings such as windows or doors in the façades are interpreted as holes in the point cloud and correspond to areas with a low point density. Albeit, in order to model the uncertainty of building object parameters like the width of windows in an explicit way, Schmittwilken and Plümer (2010) showed that a normal distribution can be safely assumed. In order to avoid this assumption, parametric probability distributions can be estimated based on ground truth data from the *RBDB* such as performed for stairs in this thesis. In this context, Schmittwilken and Plümer (2010) demonstrated empirically that windows parameters are better described by a Generalized

[1] http://www.ikg.uni-bonn.de/forschung/projektarchiv/measurefacade.html

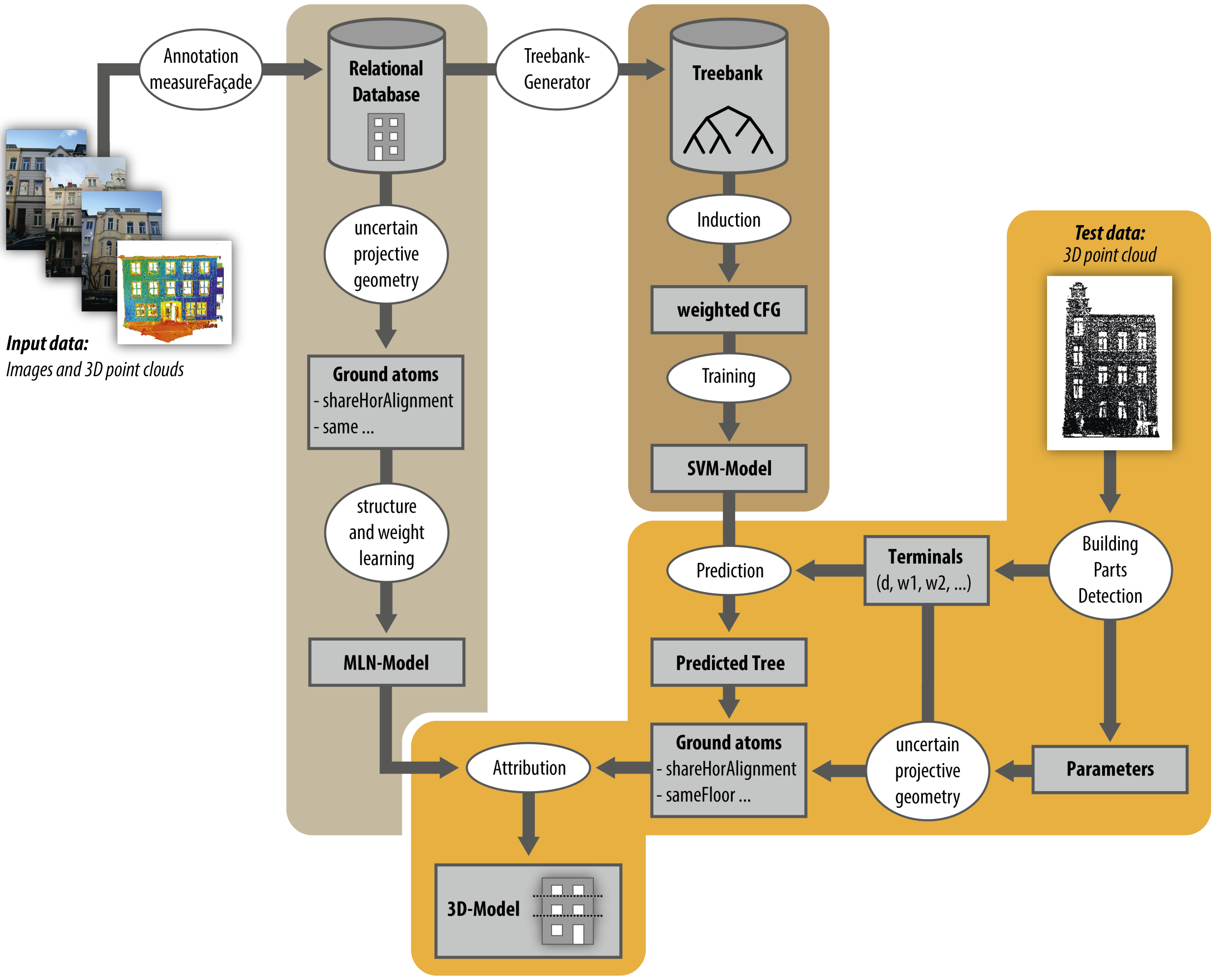

Figure 3.7: A survey of the WACFG approach. The results of the training phase are a SVM-Model (for the classification of the façade structure, dark brown) and an MLN-Model (estimation of parameter of façade parts, bright brown). These models are applied in the prediction phase (yellow) in order to derive a façade model.

Style	Train	Test
Post-war era	0.8829	0.8433
Wilhelminian	0.8402	0.8303
Without specification	0.8342	0.8072

Table 3.2: F1-scores for learning a weighted context-free grammar from a building treebank. Training and test data stem from different architectural styles.

Extreme Values distribution. This information can be integrated in the current method as input for the reasoning process using uncertain projective geometry.

For the learning of the *WCFG*, a treebank (Manning and Schütze, 1999; Charniak, 1996) has been built based on the building database *RBDB*. For this issue, several treebank types have been generated in order to prepare the basis for the following learning process of the weighted context-free grammar. Parse trees have been derived from observations using *Treebankgenerator*, that enables an automatic treebank entries induction from the *RBDB*. Each parse tree in the resulting treebank which corresponds to a given façade is automatically derived from *RBDB* leading to derivation trees reflecting several common architectural patterns such as floor-wise or column-wise splitting for grid-like structured façades. For façades that do not follow a grid structure a hybrid representation is used. The latter consists in altering columns (*columnArray*) and floor structures (*floorArray*). The split of façades is based on a reasoning process using the structural annotation from the *RBDB* leading to a derivation tree following filter criteria such as minimal description length. More details about the induction process is described in Burger (2012).

Once a treebank has been generated, an SVM-based approach was used to derive and parse a *WCFG* as described in Section 2.3 (Figure 3.7, dark brown background). The resulting *WCFG* consists of a set of context-free rules together with weights designating the importance of the given rule. In contrast to classical SVMs which expect a feature vector with a fixed size and atomic labels, the feature vector here has an arbitrary size and the labels consist in structured parse trees. An instance x of the feature vector consists of a sequence of observed façade parts (grammar terminals). The label corresponds to a parse tree y. Based on the terminal sequence x defining the type of some new observed building parts, a parse tree y is then predicted using the learned weighted context-free grammar against equation 2.5 (cf. Section 2.3, Figure 2.3).

Three different Support Vector Machine models have been learned depending on the building style. The first model is based on samples referring to buildings from the post-war era in Germany, whereas the second model represents buildings following the Wilhelminian architecture style. The last model covers the case that no information about the building style is available. The associated inferred weighted context-free grammar of the last model consists of about 450 rules. From the Wilhelminian style architecture about 160 rules have been induced. The resulting rules from the post-war era buildings amount to 135. Table 3.2 shows the F1-scores of the learning and test results for the three models using the *svm_ cfg* [2] software. The classification results of the prediction of façade structures are between 0.8

[2] http://www.cs.cornell.edu/people/tj/svm_light/svm_cfg.html

and 0.88 which is a very good classification rate since the inputs are very weak observations: list of several building types.

Up to now, the context-free grammar describes only the taxonomic and partonomic structure of façades. In this way, the topological constraints between the building parts are especially considered. In order to reflect further constraints (alignments, geometric similarity, etc.) and attributes of building parts (shape and location) the context-free weighted grammar will be augmented leading to an attribute grammar using Markov Logic Networks (see Section 2.4). However, in order to deal with the uncertainty of the observations and missing observations, MLNs and uncertain projective geometry are combined. To learn and construct an MLN model, logical ground atoms which especially represent geometric and topological constraints between façade objects are required. Thus, these atoms are generated from *RBDB* (Figure 3.7, bright brown). To this end, a statistical geometric reasoning is performed using uncertain projective geometry (cf. Section 2.1) in order to make decisions about similarity of geometric entities such as windows. The test whether two windows are geometrically identical (same shape parameters) is reduced to an identity test of two 3D-points. In order to consider the uncertainty of the data, the error propagation is modeled during the reasoning process. Analogously, the verification of the alignments of windows is reduced to the verification of the parallelity of two lines using a chi-squared statistical hypothesis test. The logical atoms are extracted according to the predicate list in Table 3.3 which gives the most important predicates for the experiment but can easily be extended and modified. The full set of extracted ground atoms per façade forms the MLN training database which was used to learn the MLN consisting of the first-order rules as well as their associated weights w_i.

The concepts described in Section 2.4 for learning MLNs and inference based on these MLNs are now applied to building reconstruction. The target predicate, which is always latent during the inference, is $\texttt{similar}(x, y)$. This binary probabilistic predicate was inferred from the database by a pairwise comparison of different objects. For two building parts, `similar` is true if and only if they are of the same type, have the same geometry and cannot be further distinguished by any other property in the database. Therefore, $p(\texttt{similar}(x, y) = \text{True})$ models the degree of similarity between two building parts. Many architectural aspects contribute to the probability p of the `similar`-predicate. These aspects are among others shape parameters of the considered objects like the width and height, the vertical and

$\texttt{sameFloor}(x, y)$	`true` iff x and y are on the same floor
$\texttt{shareHorAlignment}(x, y)$	`true` iff x and y are horizontally aligned
$\texttt{sameColumn}(x, y)$	`true` iff x and y belong to the same column
$\texttt{neighborHor}(x, y)$	`true` iff x and y are horizontal neighbors
$\texttt{shareVerAlignment}(x, y)$	`true` iff x and y are vertically aligned
$\texttt{neighborVer}(x, y)$	`true` iff x and y are vertical neighbors
$\texttt{sameHeight}(x, y)$	`true` iff x and y have the same height.
$\texttt{sameWidth}(x, y)$	`true` iff x and y have the same width.

Table 3.3: List of the important used predicates.

horizontal alignment, neighborhood information and whether they belong to the same floor (cf. Table 3.3). The mentioned aspects and their influence on the similarity are represented by the MLN. To infer the similarity between pairs of building parts, a structure and weight learning of an MLN is performed (Figure 3.7, bright brown). In the prediction stage this MLN will be applied together with the ground atoms, that will be described in the following.

As yet a generic MLN model as well as a generic weighted context-free grammar have been derived. Both models can be used for the derivation of a concrete 3D model for a specific façade F. For a given 3D point cloud, *KDE*-based object detectors were applied leading to the sequence of building part types as well as their parameter values. The latter are uncertain and maybe incomplete. Therefore, uncertain projective geometry was used again in order to extract ground atoms according to the predicates in Table 3.3. These atoms describe geometric properties as well as relations between building parts of the façade F. Examples for such atoms are `sameWidth(w1,w2)` and `shareHorAlignment(w1,w2)` for two observed windows `w1` and `w2`. These atoms are extracted for all pairs of building objects for which the relation can be derived. Ground atoms for the remaining predicates from Table 3.3, that describe structural and topological relations, are derived from the predicted parse tree for the façade F. Examples are `sameFloor(w1,w2)` or `neighborHor(w1,w2)` for two observed windows. Likewise these atoms are extracted for all pairs.

These ground atoms together with the generic MLN model are the input for the attribution step (cf. Figure 3.7, yellow background) using statistical inference as described in Section 2.4. Each leaf of the parse tree corresponds to a constant in the MLN. Special attention is given to determining the most likely configuration of the `similar`-predicates, which consists in `similar` ground atoms with an associated probability of being similar. The result enables us to derive the most likely geometry of the façade and its parts (3D model). Especially unobserved parameters can be estimated.

In a further step the proposed MLN model was evaluated and tested. For the evaluation a dataset was used containing façades from twenty buildings in Bonn, Germany. Each building has a varying number of objects, accordingly the corresponding grounded MLNs vary in size as well. To make the inference task more realistic and to demonstrate that the proposed method can cope with unobserved objects, some of the `sameWidth` and `sameHeight` predicates in every façade were randomly removed. All experiments were conducted in a 10-fold cross validation, i.e., 90% of the buildings have been used for parameter and structure learning, and the remaining buildings were used for testing. Herewith, the experiments were performed on handcrafted, automatic learned and a semi-automatic learned MLN. The first experimental evaluation is based on a manually crafted MLN. For this MLN, merely the weights are learned using *Alchemy* and discriminative learning as described in Section 2.4. This enabled to compare the results with a fully automatically learned MLN. The second experiment turned attention to automatically learned MLNs by the *Alchemy* system. The last experiment addresses the question if automatically learned MLNs can be combined with expert rules to obtain improved results over the previous two experiments. To this end, the best performing automatically learned MLN was taken, in addition to further rules specifying transitivity, symmetry and reflexivity, and then a parameter re-learning was performed. In all experiments an MAP inference on all buildings were performed and the predictions

	Handcrafted		Auto. Learned		Semi-Auto.	
	MWS	BP	MWS	BP	MWS	BP
0% missing	0.971	0.483	0.998	0.998	0.998	0.973
25% missing	0.822	0.461	0.837	0.886	0.798	0.874
50% missing	0.724	0.434	0.626	0.726	0.626	0.714

Table 3.4: F1-scores for predicted `similar`-predicates based on MWS and BP. A 10-fold cross validation was used. `sameHeight`- and `sameWidth`-predicates were randomly removed.

were compared with the ground truth. Table 3.4 shows the results for running MaxWalk-Sat (MWS) and max-product Belief Propagation (BP) for the each experiment in different columns. The performance of the results is evaluated and tested based on F1-scores. The overall classification rate for the automatically learned MLN in the case of fully observed objects is 0.99, 0.83 if 25% of observations are missing and 0.62 if 50% are missing. It can be shown that handcrafted rules do not outperform the automatically learned ones, that do not require any expert knowledge. The learned MLN reveal interesting connections expressed by the learned formulas. For example, a rule that is commonly found is the following one:

$$\lambda_i : \neg\texttt{sameHeight}(a1, a2) \vee \neg\texttt{sameWidth}(a1, a2) \vee \neg\texttt{similar}(a1, a2) \ ,$$

which means that the same height and width implies dissimilarity. This does not correspond to the intuition. However the presented approach realizes this and penalizes this formula by negative weight λ_i. Hence, true groundings of this formula are also penalized. The experiments have shown that automatically learned MLNs can well capture regularities in building façades without the need of a human expert to define background knowledge and relationships. Automatically learned MLNs can additionally be smaller in size and hence allow for a faster inference.

Figure 3.8 shows the model-based reconstruction of five façades using the *WACFG*. Due to occlusions, noise or sparse point clouds the kernel density estimation (third row) does not guarantee a complete reconstruction. In façade 1 an MLN-based inference (see fourth row) enables to adapt and regularize the size and the alignment of the windows in the ground floor. In façade 2 the height of two windows as well as the door in the ground floor could not be identified by KDE due to the vegetation in the front of the façade. With the predicted parse tree for the façade, however the missing objects can semantically interpreted leading to a door on the left and further two windows. The shape parameters and the alignment constraint of these façade parts are ensured using the MLN model and a-priori learned probability distributions of model parameters from the *RBDB*. Likewise in façade 3 the false estimated shape parameters of the window in the middle due to the existence of a traffic sign are corrected by the MLN model. Displaced façade parts such as the case in façade 2 left column and façade 3 fourth column are identified, albeit they are graphically not represented in the figure.

Figure 3.8: Façades reconstructed with the WACFG. Windows are colored in green and doors in blue. The first row shows reference façade images. The second row depicts the corresponding input 3D point clouds. Row three represents a kernel density-based reconstruction of each façade. Deficiencies of the kernel density estimation are overcome using the WACFG model.

Representation of uncertainty

In this thesis, the uncertainty of data as well as of models is explicitly represented in different levels. During the annotation stage for training the MLN, we pessimistically assumed that the annotation errors[3] are normal distributed with a standard deviation of 10 cm for shape parameters. This information is incorporated in the reasoning process based on covariance matrices of the considered geometries using uncertain projective geometry in order to derive constraints and relations, i.e. alignments of windows, as logical facts for MLN. During the interpretation stage, parametric probability density functions fitted on the underlying observations are considered and integrated in the uncertain reasoning yielding logical facts analogously. Once the façade parts have been acquired using the building part detectors, the learned weighted context-free grammar is used to find the most likely parse tree explaining the detected parts. At this stage, the learned weights of the grammar rules, expressing their

[3]The annotation errors occurs due to geometric inaccuracies, that stem from the user during the annotation process and deviations resulting from imprecise semantic definitions of building parts.

likelihood, are used together with the a-priori learned SVM model. The acquired parse tree incorporates further relations such as neighborhood in floors. In the next step, statistical inference, in the sense of graphical models, is performed on an automatically induced MRF to propagate the constraints among the building parts based on the extracted facts. The learned first-order rules and their learned corresponding weights play an important role in this step. The acquisition of categorical facts from noisy observations using uncertain projective geometry for MLNs is a key to a successful learning and an interpretation process. However, this avoid the propagation of the parameter uncertainty of the detected building parts. This can be modeled using hybrid MLNs (Wang and Domingos, 2008) that allow the integration of probability distributions and continuous variables. This issue can be addressed in future research.

In order to avoid geometric redundancy and to support the building reconstruction using the presented SRL-method, the next section introduces an automatic learning approach for the identification and modeling of architectural regularities and hidden redundancies such as symmetries.

3.3 Grammar-based learning and representation of symmetries in building footprints

In several prominent architectural styles, symmetries play a key role. It is important to identify and represent them appropriately in building models. This section introduces a novel approach for the automatic identification and modeling of symmetries and their hierarchical structures in building footprints, providing an important prior for the SRL-based façade reconstruction described in the previous section. This work is described and published in the *Transactions in GIS* journal (Dehbi et al., 2016a). Buildings and man-made objects are for many reasons, such as economical or aesthetic, often characterized by symmetry properties. The latter is dominating in the design of building footprints as well as of building parts such as façades. Thus, the identification and modeling of this valuable information facilitates the reconstruction of these buildings and their parts. Figure 3.9 shows exemplarily some façades with symmetrical properties and their associated footprints. The structure of the symmetry property in each footprint is reflected in the structure of the façade. Thus, the latter can be already mostly derived from the hierarchical description of the footprint.

Symmetry models enable an accurate prediction of occluded building parts for shape completion based on sparse observations. Further, the identification and explicit modeling of axial and translational symmetries serves to discover and represent the occurrence of protrusions in buildings. These information can be used in order to improve the footprint based roof classification and reconstruction (Henn et al., 2013). Such methods decompose footprints in order to determine the appropriate roof type. But up to now they suffer from suboptimal decompositions.

The presented method provides automatically models based on weak observations towards reconstructing buildings and especially façades. This approach detects axial as well as translational symmetries in building footprints and models them, taking their hierarchical structures into account. The main problem is the inherent uncertainty of the geometry of the

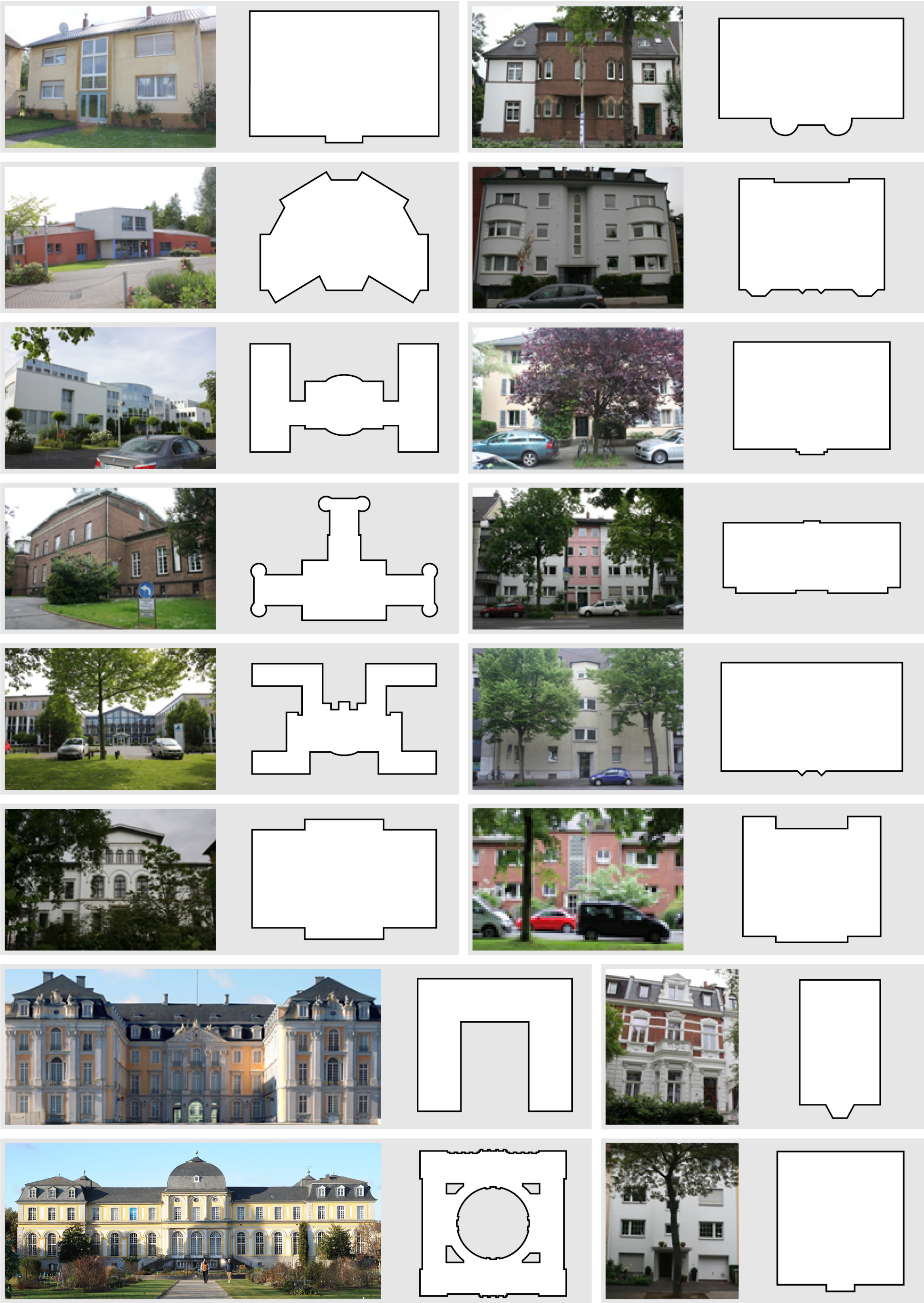

Figure 3.9: Examples of façades and their corresponding footprints. The symmetries of the façade can already be derived from the footprint symmetries.

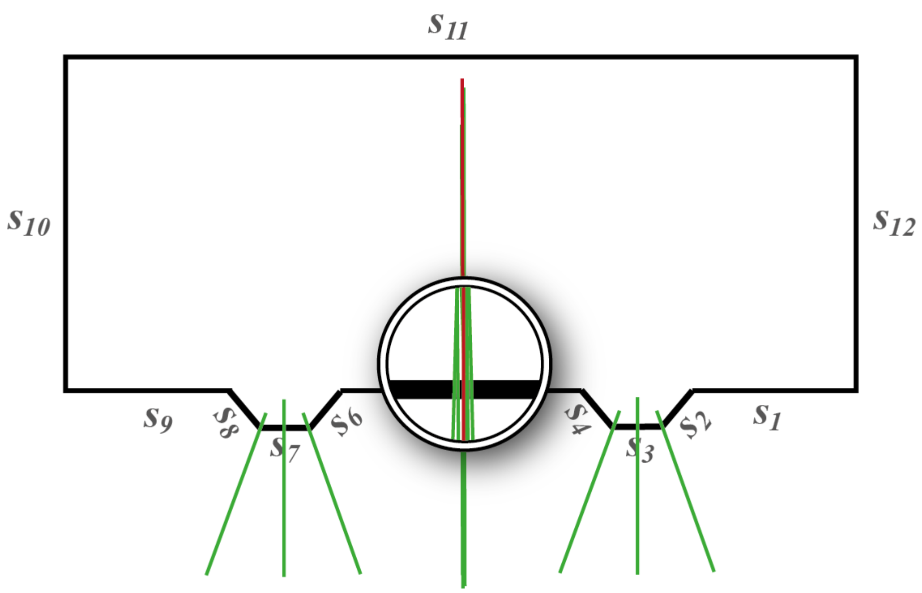

Figure 3.10: Predicted axial symmetry axes (green) of a given polyline (footprint) and its sub-polylines. Adjusted axis using least squares is highlighted (red).

footprint segments and their angles. A simple threshold based approach is not successful, since the threshold values are not known a-priori. Instead, methods of supervised learning are used in order to classify symmetric footprints and footprint parts. This is reduced to the classification of polylines and their sub-polylines using Support Vector Machines (Vapnik, 1998; Schölkopf and Smola, 2001) as a robust classifier, which enables to tackle the inherent uncertainty of the observed footprints. In this way several axial as well as translational symmetries are detected. In contrast to classical statistical methods, no assumptions on the a-priori distribution of the data are required for SVMs. These assumptions can in general not be guaranteed for footprint data due to its variability. Regression methods (least squares) are used in order to assess the quality of the identified major and minor symmetry axes. Based on the classification results a novel algorithm is developed to induce formal grammar rules for particularly representing the hierarchy of symmetry axes and the repetitive structures and regularities in the footprint.

As a first step, the approach is based on supervised machine learning and classification of axial and translational symmetries in footprints. A footprint is modeled as a polyline consisting of a set of sub-polylines, that can be discriminated in axial symmetric or non-axial symmetric (sub-)polylines. The SVM classifier is trained based on a set of polylines labeled as axial symmetric or non-axial symmetric. The acquired model is applied for the classification and determination of axial symmetric polylines. However, machine learning methods require feature vectors of fix lengths, whereas building footprints have an arbitrary number of segments. This problem is addressed by mapping polylines onto feature vectors by using aggregate functions. The axial symmetry is a property of polylines, whereas translational symmetry is a property of pairs of segments. Hence, there are two different classification tasks. SVMs are also used for classifying pairs of footprint segments into translational and

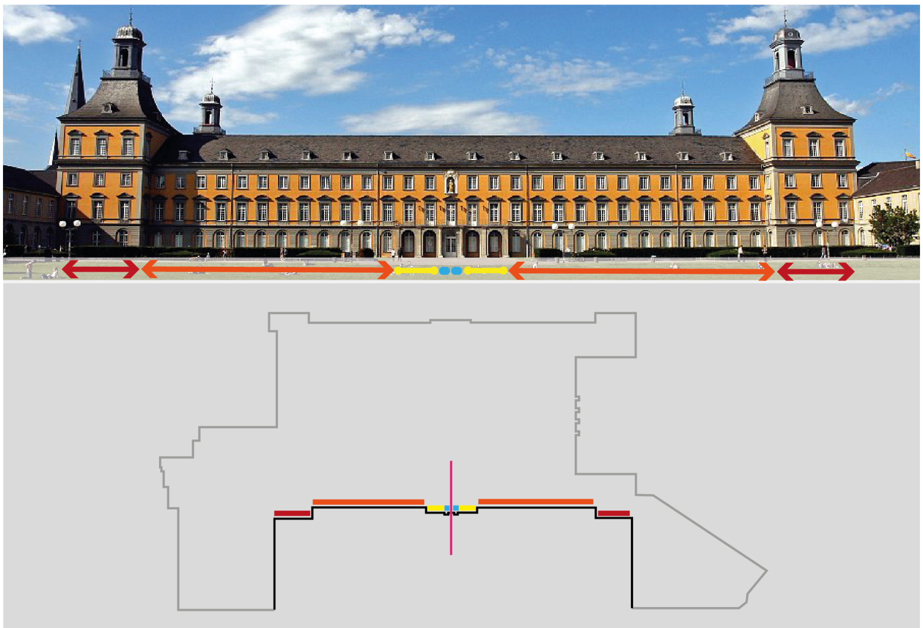

Figure 3.11: Identification of broken symmetries in footprints. Symmetric parts in the front are highlighted in same color.

non translational symmetric segments. As a result, a set of symmetry axes is obtained including major as well as minor symmetry axes. Axes that belong to sub-polylines of a given polyline deviate slightly from each other and hence are adjusted using least squares in order to fit an optimal shared axis. Figure 3.10 shows the predicted axes in green for a given footprint. One adjusted symmetry axis of the footprint from figure 3.10 is exemplary illustrated in red color. Analogously translational symmetric segments from the footprint are classified.

In the next step, the classification is exploited in order to induce and learn context-free grammar rules, which reflect the hidden hierarchical and repetitive structures in footprints. For this aim, an algorithm is developed, which uses a lexical analysis, in order to generate a derivation tree representing symmetry hierarchies and repetitive structures. The inferred rules can be used as background knowledge in order to predict building parts such as occluded parts, which are unknown a-priori. Furthermore, the grammar rules enable a compact representation of buildings and their parts especially of symmetric ones. The fact that symmetric building parts are often constructed in a similar way helps to restrict the space of hypotheses during a stochastic reasoning process for the reconstruction of 3D building models as described in (Loch-Dehbi et al., 2013). Besides, the presented approach detects symmetries and infers adequate grammar rules even though symmetry is broken in the given footprint (see Figure 3.11).

As mentioned above, a footprint is modeled as a polyline consisting of a set of sub-polylines. A *polyline* $\mathcal{P}$: $(p_1, ..., p_n)$ is specified by 2D points p_i, $i = 1 \dots n$. A footprint is represented by a special polyline, which is closed. In a closed polyline p_1 and p_n are connected by a

	true axial symm. pol.	true non-axial symm. pol.	precision (%)
pred. axial symm. pol.	1294	14	98.92
pred. non-axial symm. pol.	0	24712	100
recall (%)	100	99.94	

Table 3.5: Classification results of axial symmetric polylines. Two classes are considered: axial symmetric polylines (axial symm. pol.) and non-axial symmetric polylines (non-axial symm. pol.)

segment. A sub-polyline[4] of a given closed polyline $\boldsymbol{\mathcal{P}}$: $(p_1, ..., p_n)$ is a sub-sequence of the points of $(p_1, ..., p_n, p_1 ..., p_n)$ with maximal length n. In some cases it is more appropriate to represent a polyline as a sequence of segments $\boldsymbol{\mathcal{Q}} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n-1})$ with $\boldsymbol{s}_i = \boldsymbol{p}_i\boldsymbol{p}_{i+1}$ for $i = 1 \ldots n-1$. Axial symmetric parts of a given footprint $\boldsymbol{\mathcal{P}}$ correspond to a symmetric polyline $\boldsymbol{\mathcal{SP}}$, which can be defined as follows: A polyline $\boldsymbol{\mathcal{SP}}$: $(p_1, ..., p_n)$ is said to be axial symmetric if there is a symmetry axis A such that for each point p_i in $\boldsymbol{\mathcal{P}}$, either A passes through p_i or there is a corresponding point p'_i in $\boldsymbol{\mathcal{P}}$ such that A is the bisector of the segment $p_i p'_i$. The previous definition holds for footprints without uncertainty, which do rarely occur in real-world scenarios. The developed approach copes with uncertainty based on a classification using a trained model. The geometric properties of a polyline are represented by features. A polyline consists of an arbitrary number of segments whereas the number of features is fixed. For example the difference of number of segments left and right of the bisector Δ of the first and the last point or the difference of the sum of the azimuths left and right were used. A feature weighting is performed in the training phase using the *Relieff* algorithm (Robnik-Šikonja and Kononenko, 2003), which provides the importance of each feature. *Relieff* ranks individual features according to their relevance in the context of others. Features with a low weight are neglected. These features were used to classify polylines into symmetric polylines and non-symmetric ones. An SVM was employed, which has been trained on $\sim$ 26000 labeled polylines, which are based on about 100 authoritative cadastral footprints from the area of Bonn, Dortmund, Münster and Düsseldorf in North Rhine-Westphalia, which differ in complexity. To this end, a software, that supports a human user in order to annotate polylines and segment pairs, was implemented. The average number of the vertices of these footprints amounts to 16.1.

Table 3.5 presents the result of the classification with an overall accuracy of 99.94% and a precision of 98.92% for symmetric polylines derived by 10 fold cross validation. Further during the learning of the model the problem of unbalanced class frequencies has been taken into consideration using adequate penalty parameters as suggested in Vapnik (1998).

The finding and modeling of repetitive structures and regularities in footprints represents an interesting task as well. Two segments s_1 and s_2 are *translational symmetric* if they have roughly the same length and the same azimuthal angle. The identification of repetitive structures in a footprint is considered as a classification problem, which addresses uncertainty of the segments in analogy to the case of axial symmetry. Features are the

[4]This definition copes with the case that the symmetric polylines may not start at p_1 or p_n but include one of both.

	true transl. segm.	true non-transl. segm.	precision (%)
pred. transl. segm.	435	10	97.75
pred. non-transl. segm.	7	14039	99.95
recall (%)	98.41	99.92	

Table 3.6: Classification results of translational symmetric segment pairs. Two classes are considered: translational symmetric segment pairs (transl. segm.) and non-translational symmetric segment pairs (non-transl. segm.)

difference of azimuthal angles, quotient of the lengths and the minimum of the differences in x and y directions. The training and the classification phase as well as the results are comparable to the classification of axial symmetry. From 96 footprints from cities in North Rhine-Westphalia 445 segment pairs are labeled as translational symmetric and 14046 pairs as non-translational. The confusion matrix in Table 3.6 shows that an overall accuracy of 99.92% and a precision of 97.75% for translational symmetric segments have been achieved derived by 10 fold cross validation. In both classification tasks the few misclassified cases result from situations where even human observers have problems in deciding whether a polyline or a pair of segments is symmetric or not. These cases represent decisions lying in a twighlight zone due to the inherent uncertainty of the data where both decisions can be justified. Since the classification is performed using a SVM Platt scaling approach, we get a probability of the predicted class in addition to a prediction. In the misclassified cases the probabilities are not peaked (0.3869 against 0.6131).

Building upon the axial symmetric polylines and the translational symmetric segments from the previous classification steps, grammar rules for hierarchical and repetitive structures of a footprint can be induced. In a preprocessing step the footprint is permuted such that it starts with the longest symmetric sub-polyline $\boldsymbol{sp}$, yielding $\boldsymbol{s}_1\ \boldsymbol{s}_2 \ldots \boldsymbol{s}_n = \boldsymbol{sp}\ s_{k+1} \ldots s_n$. The following rule $\boldsymbol{S} \rightarrow \boldsymbol{s}_1\ \boldsymbol{s}_2 \ldots \boldsymbol{s}_n$ is generated. Starting with the longest axial symmetric polyline, the footprint is successively decomposed into parts. Rules of a context free grammar are generated, which reflect the axial symmetry. In this rule the segments of $\boldsymbol{sp}$ are replaced by nonterminal symbols N_1 and $\overline{N_1}$ reflecting the symmetry. This generates three rules: $\boldsymbol{S} \rightarrow N_1\ \overline{N_1}\ \boldsymbol{s}_k \ldots \boldsymbol{s}_n$, $\boldsymbol{N_1} \rightarrow \boldsymbol{s}_1 \ldots \boldsymbol{s}_m$ and $\overline{\boldsymbol{N_1}} \rightarrow \boldsymbol{s}_{m+1} \ldots \boldsymbol{s}_k$. In the case that the length of $\boldsymbol{sp}$ is odd the first rule is $\boldsymbol{S} \rightarrow N_1\ \boldsymbol{s_t}\ \overline{N_1}\ \boldsymbol{s}_k \ldots \boldsymbol{s}_n$ where $\boldsymbol{s_t}$ is the middle segment of $\boldsymbol{sp}$. The other rules are adapted accordingly. Symmetric polylines are represented by two rules expressing explicitly that one sub-polyline represents the mirrored sequence of the other. This procedure is recursively applied to both parts in order to reflect the inherent hierarchy between symmetry axes. For example in Figure 3.10, a major symmetry axis is located at the segment s_5, while a minor axis lies on segment s_7. Figure 3.12 depicts graphically the parse tree on the façade based on the identified axial symmetries from a given footprint. The derived context-free grammar rules are given in Table 3.7.

In a second step, the a-priori classified repetitive parts are represented by the same symbol of the grammar, modifying the grammar derived in the last step. This enables to use the *Sequitur* algorithm (Nevill-Manning and Witten, 1997), which infers a hierarchical structure from a sequence of translational symmetric segments. These hierarchies are represented

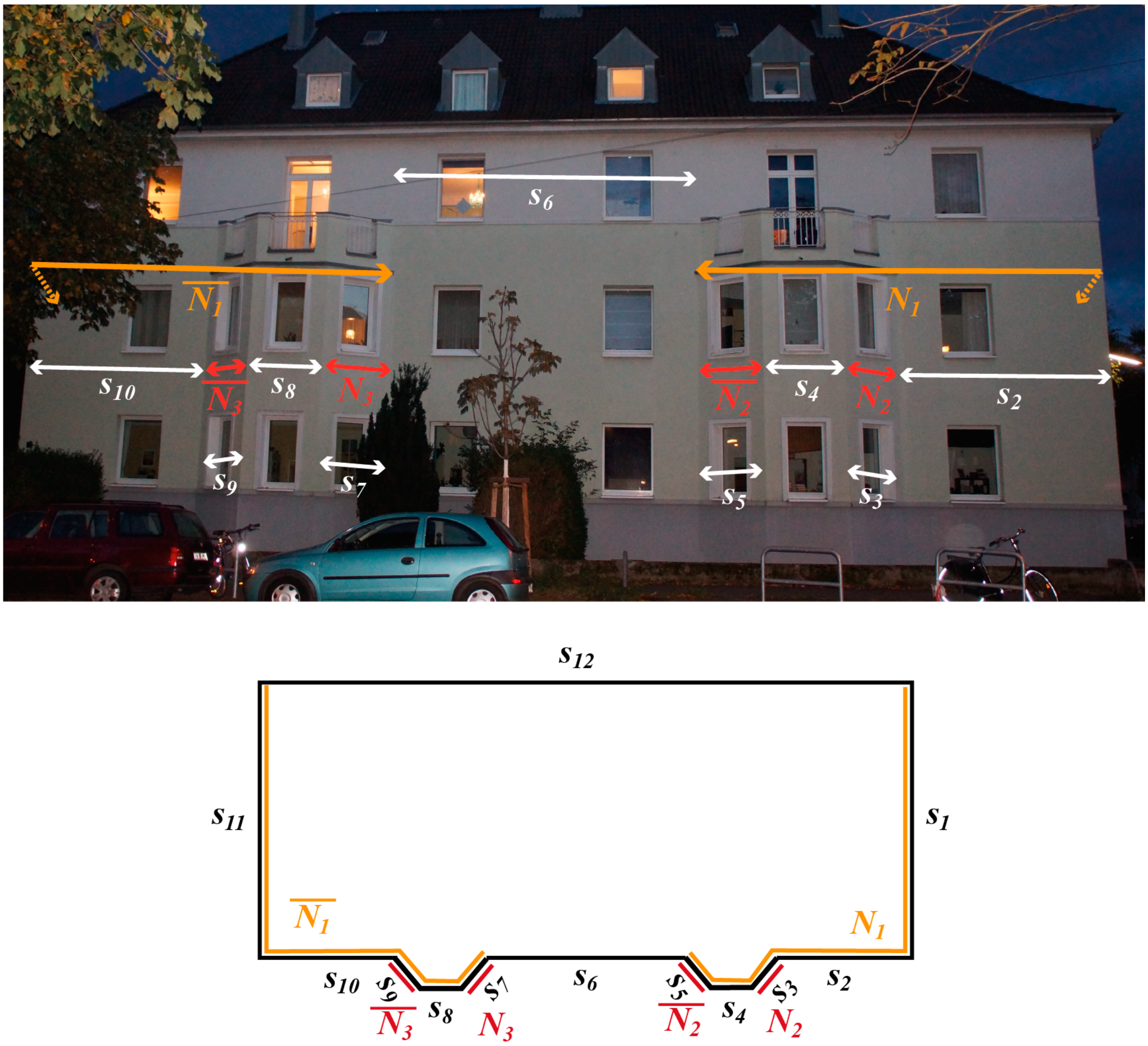

Figure 3.12: The induced grammar rules from a given footprint. The hierarchical structure of the symbols is illustrated in the façade as well as in the footprint.

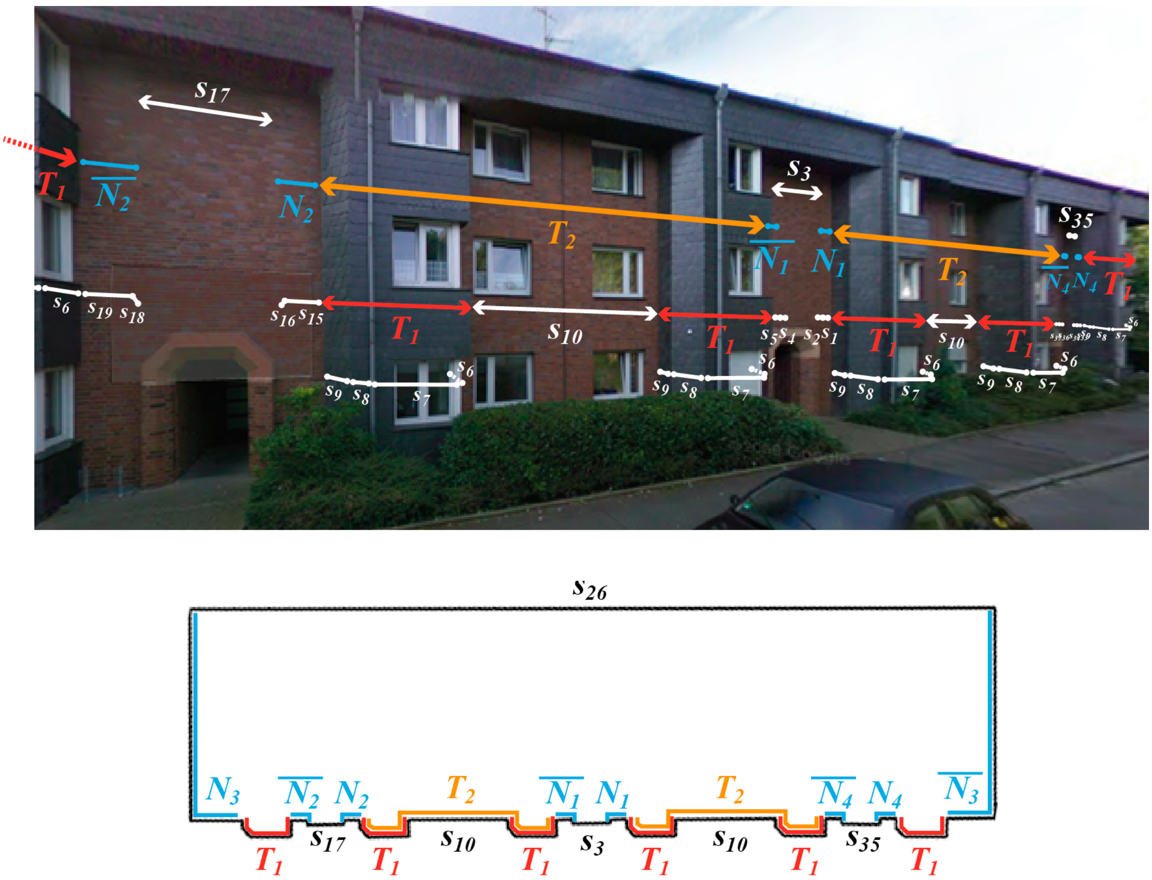

Figure 3.13: The induced grammar rules for translational footprint parts reflect the repetitive structure of the corresponding façade

by context free grammar rules. In the contrast to the standard Sequitur algorithm, which generates rules from scratch, an adaptation is developed, that extends and modifies existing rules, which already represent axial symmetries. Figure 3.13 shows a footprint that contains several translational symmetries. The application of the modified Sequitur yields the highlighted structures at the bottom. The red colored parts (protrusions) and the orange parts (two translated protrusions and a connecting wall) of the footprint as shown are translated regions in the corresponding façade, which match the visual perception. Further grammar rules for the non-terminals $N_1, N_2, \ldots$ have been previously derived in order to represent axial symmetry yielding the rules shown in Table 3.8.

To sum up, the main contribution in this section is an approach that automatically detects and models axial as well as translational symmetries in footprints. The uncertainty of the underlying observations is tackled with a Support Vector Machine based classification and a regression of symmetry axis using least squares. The classification enables to model repetitive as well as translational symmetries in an elegant way by formal grammars. This method has been successfully applied for generating façade hypotheses for CityGML LoD3 buidlings (Loch-Dehbi et al., 2013). The introduced method delivers good prior knowledge

Terminals $= \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}\}$	
Rules:	
$\mathcal{S} \rightarrow N_1\ s_6\ \overline{N_1}\ s_{12}$	
$N_1 \rightarrow s_1\ s_2\ N_2\ s_4\ \overline{N_2}$	$\overline{N_1} \rightarrow N_3\ s_8\ \overline{N_3}\ s_{10}\ s_{11}$
$N_2 \rightarrow s_3$	$\overline{N_2} \rightarrow s_5$
$N_3 \rightarrow s_7$	$\overline{N_3} \rightarrow s_9$

Table 3.7: The induced grammar from the footprint in Figure 3.12. $\overline{N}$ represents an axial symmetric part of a given part N.

accelerating a reasoning process yielding a small number of accurate hypotheses and hence reducing the search space significantly. Once one part has been detected and reconstructed, consequently the (axial or translational) symmetric part can be likewise reconstructed. This approach can be combined with the introduced SRL-based reconstruction method in order to reduce the search space and to improve the reconstruction quality. A key advantage of this approach is treating the problem of unobservability due to occlusions. More details and algorithms can be found in Dehbi et al. (2016a).

Terminals $= \{s_1, s_2, s_3, \ldots, s_{45}, s_{46}\}$	
Rules:	
$\mathcal{S} \rightarrow N_1\ s_3\ \overline{N_1}\ T_2\ N_2\ s_{17}\ \overline{N_2}\ T_1\ N_3\ s_{26}\ \overline{N_3}\ T_1\ N_4\ s_{35}\ \overline{N_4}\ T_2$	
$T_1 \rightarrow s_6\ s_7\ s_8\ s_9$	$T_2 \rightarrow T_1\ s_{10}\ T_1$
$N_1 \rightarrow s_1\ s_2$	$\overline{N_1} \rightarrow s_4\ s_5$
$N_2 \rightarrow s_{15}\ s_{16}$	$\overline{N_2} \rightarrow s_{18}\ s_{19}$
$N_3 \rightarrow s_{24}\ s_{25}$	$\overline{N_3} \rightarrow s_{27}\ s_{28}$
$N_4 \rightarrow s_{33}\ s_{34}$	$\overline{N_4} \rightarrow s_{36}\ s_{37}$

Table 3.8: The induced grammar from the footprint in Figure 3.13 taking translational symmetries into consideration. Previously extracted rules for axial symmetries have been extended. A repeated T stands for two or more translated parts.

4 Conclusion and perspectives

3D building models have been used for several tasks such as the calculation of energy balances or escape routes. This requires automatically derived 3D models which represent semantics explicitly. In this thesis, new methods and approaches have been developed for the automatic learning and parsing of grammar rules for 3D building reconstruction from 3D point clouds. A novel machine learning-based approach is introduced for the learning of weighted attribute context-free grammar rules (WACFG). In contrast to procedural methods, a declarative approach separates the representation of buildings and their parts from their reconstruction. The learned rules serve both as strong prior knowledge and as semantic models for the interpretation and reconstruction of buildings and their parts from 3D point clouds. The proposed machine aided method reduces the need for expert knowledge and the expense of designing grammar rules manually. This addresses not only the learning of the grammar rules but also the parsing of 3D point clouds leading to a semantically interpreted model. Moreover, the introduced approach is able to learn from precise models as well as noisy observations and deals with the uncertainty of models itself. The presented approach handles successfully complexity —a varying number and type of objects—, uncertainty and unobservability in real-world problems. This is explicitly addressed by uncertain projective geometry, probability density functions, probabilistic grammar rules and Markov Logic Networks (MLNs). These formalisms enable to handle and evaluate data quality in an explicit way. The developed approach consists of three components.

The first component of the learning method consists in the stepwise learning of building parts and supports both learning from precise descriptions as well as from real observations. Herewith, two learning levels, high and low level, have been introduced. In the first case, the user as teacher provides some examples of the target concept in addition to related background knowledge in order to learn the logical rules which describe this concept using Inductive Logic Programming (ILP). The background knowledge contains basic spatial relations like parallelity or orthogonality. The ability to model background knowledge enables a multi-stage learning process of a building part based on its previously learned parts. In the second case, the background knowledge is automatically extracted from a terrestrial 3D point cloud. In addition to uncertain projective geometry, which uses hypothesis tests for learning imprecise geometric relations (e.g. orthogonality), probability distributions are used in order to learn the uncertain topological relations between 3D blobs within the building parts. Only few examples were sufficient to learn from precise as well as noisy observations.

The second component consists in a two-staged incremental strategy in order to cope with the complexity of the learning task and to take the uncertainty of models into consideration as well. At first, the context-free part of the WACFG is learned. Afterwards, the rules are extended by attributes and constraints between the building parts. A Support Vector Machine-based approach was used to infer a weighted context-free grammar from input-output pairs given as structured data from a façade treebank. In addition to the grammar rules a classification model is obtained. This enables to parse a sequence of observed façade elements in order to predict the most likely tree structure. The weighted context-free grammar rules are extended by attributes and constraints, which describe the geometric as well as the topological dependencies between the façade elements. To this end, a Statistical Relational Learning (SRL) method using MLNs is applied for the first time in 3D building reconstruction. In order to learn the structure, i.e. formulas, as well as parameters, i.e. weights, of an MLN model, logical atoms are automatically generated from a relational building database consisting of 1300 annotated buildings from different regions in Bonn, Germany. It can be stated that the learned MLNs can well capture regularities in building façades without the need of a human expert to define background knowledge and relationships. Handcrafted rules do not outperform the automatically learned ones.

In order to support and provide an important prior for the SRL-based façade reconstruction, a third component has been described. It enables the automatic identification and induction of models for the representation of axial and translational symmetries in footprints. This information can be derived from the corresponding footprints without any observations of the façades. Uncertainty is explicitly addressed by using classification methods. Translational as well as axial symmetries are detected using supervised classification methods. Support Vector Machines in combination with Platt's posterior probabilities (Platt, 1999) have been used as robust classifier, that derives not only predicted classes but also a degree of certainty of the prediction. Hierarchical and repetitive structures in building footprints are induced based on the previous accurate classification results. Context-free grammar rules are then derived using a lexical analysis enabling the description of repetitive as well as axial symmetric parts in a relational and compact way.

The learned models are consistent but in general they are not necessarily redundance-free. In order to check these models for redundancy, the geometric reasoning approach presented in (Loch-Dehbi and Plümer, 2011) can be applied. It combines algebraic methods with logical inference rules which reduce the search space for valid rules. Conversely, the learned logical concepts can serve as input in order to enrich the predictions made during the reasoning process. Further, the assumption that the measurements error is greater than the uncertainty of the theoretical concept was made. Consequently, only errors from noisy observations are considered. However, an extension of the approach is possible by incorporating covariance matrices, that represent the uncertainty of the object model.

In the presented method, the variables and features of MLNs are discrete. As yet uncertain projective geometry has been used in order to extract ground atoms from imprecise observations. In order to enable the modeling of continuous variables such as the shape parameters of building parts, hybrid MLNs (Wang and Domingos, 2008) can be investigated using probability distributions from the exponential family. Furthermore, lifted inference approaches can be beneficial for the underlying task. Lifted inference (Kersting, 2012) exploits symmetries in the underlying problem structure and clusters indistinguishable objects together to increase efficiency. 3D building reconstruction is expected to benefit from such approaches due to the fact that building façades contain a lot of symmetries. Hence, additional ground and lifted inference algorithms should be evaluated, such as the lifted likelihood maximization approach presented by Hadiji and Kersting (2013).

A further open question is building footprint related symmetry model repair. In this context the posterior probabilities of the classified symmetric polylines can be exploited in order to address this issue. These probabilities, that represent a quality assessment of the predictions, give an evidence of the perfection of the symmetry. In this manner, errors stemming from data acquisition can be detected and corrected. Moreover the presented approach, that currently is limited to the induction of relational models from building footprints, can be extended in order to derive similar models for building façades. This addresses cases of symmetry that are not reflected in the building footprints.

The main contribution of this thesis is a novel method for automatic learning and parsing of 3D weighted attribute context-free grammars for 3D reconstruction of buildings and their parts from 3D point clouds. For the first time it is possible to represent and reconstruct buildings with a fully automatically learned weighted grammar in a pure declarative way using logical and statistical relational learning techniques. This enables the separation between the representation of buildings and their parts from their reconstruction from 3D point clouds. The uncertainty of data as well as of models is explicitly represented in different levels by uncertain projective geometry, probability density functions, probabilistic grammar rules, Markov Logic Networks and Support Vector Machines.

Bibliography

Becker, S., 2009. Generation and application of rules for quality dependent façade reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 64, 640–653.

Burger, J., 2012. Generating tree bank entries from relational descriptions of façades and their components. Master's thesis. FernUniversität Hagen, Rheinische Friedrich Wilhelms Universität Bonn.

Charniak, E., 1996. Tree-bank grammars, in: In Proceedings of the thirteenth National Conference on Artificial Intelligence, 1031–1036.

Chomsky, N., 1956. Three models for the description of language. Information Theory, IEEE Transactions 2, 113–124.

Chomsky, N., 1959. On certain formal properties of grammars. Information and Control 2, 137–167.

De Raedt, L., 2008. Logical and relational learning. Springer, Berlin.

De Raedt, L., Dehaspe, L., 1997. Clausal discovery. Machine Learning 26, 99–146.

Dehbi, Y., Gröger, G., Plümer, L., 2016a. Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars. Transactions in GIS 20, 645–663.

Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., Plümer, L., 2016b. Statistical relational learning of grammar rules for 3d building reconstruction. Transactions in GIS. doi:10.1111/tgis.12200.

Dehbi, Y., Plümer, L., 2011. Learning grammar rules of building parts from precise models and noisy observations. ISPRS Journal of Photogrammetry and Remote Sensing 66, 166 – 176. Quality, Scale and Analysis Aspects of Urban City Models.

Dehbi, Y., Staat, C., Mandtler, L., Plümer, L., 2016c. Incremental refinement of façade models with attribute grammar from 3d point clouds, 311–316. doi:10.5194/isprs-annals-III-3-311-2016.

Deransart, P., Maluszynski, J., 1985. Relating logic programs and attribute grammars. The Journal of Logic Programming 2, 119–155.

Dorst, L., Fontijne, D., Mann, S., 2007. Geometric algebra for computer science: An object-oriented approach to geometry. The Morgan Kaufmann series in computer graphics, Elsevier/ Morgan Kaufmann, Amsterdam.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Knowledge Discovery in Databases 96, 226–231.

Förstner, W., 2005. Uncertainty and projective geometry, in: E. Bayro-Corrochano (Ed.), Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics. Springer.

Gröger, G., Kolbe, T.H., Nagel, C., Häfele, K., 2012. OpenGIS City Geography Markup Language (CityGML) encoding standard. Version 2.0.0, Open Geospatial Consortium (OGC), Doc. no. 12-019.

Gröger, G., Plümer, L., 2010. Derivation of 3d indoor models by grammars for route planning. Photogrammetrie, Fernerkundung, Geoinformation PFG 3, 193–210.

Hadiji, F., Kersting, K., 2013. Reduce and re-lift: Bootstrapped lifted likelihood maximization for MAP, in: Statistical Relational Artificial Intelligence, Papers from the 2013 Association for the Advancement of Artificial Intelligence (AAAI) Workshop, Bellevue, Washington, USA, July 15, 2013.

Henn, A., Gröger, G., Stroh, V., Plümer, L., 2013. Model driven reconstruction of roofs from sparse lidar point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 76, 17–29.

Heuel, S., 2004. Uncertain projective geometry: Statistical reasoning for polyhedral object reconstruction. Springer, Berlin.

Houghton, C.J., 1978. Birth of a parent: The Wakeby distribution for modeling flood flows. Water Resources Research 14, 1105–1110.

Johnson, L.N., Kotz, S., Balakrishnan, N., 1995. Continuous univariate distributions. A Wiley Interscience publication. 2. ed. ed., Wiley, New York.

Kersting, K., 2012. Lifted probabilistic inference, in: Raedt, L.D., Bessiere, C., Dubois, D., Doherty, P., Frasconi, P., F. Heintz, P.L. (Eds.), Proceedings of 20th European Conference on Artificial Intelligence (ECAI-12), ECAI. IOS Press.

King, R.D., Muggleton, S.H., Srinivasan, A., Sternberg, M.J., 1996. Structure-activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. Proceedings of the National Academy of Sciences USA 93, 438–442.

Knuth, E.D., 1968. Semantics of context-free languages. Theory of Computing Systems 2, 127–145.

Knuth, E.D., 1971. Top-down syntax analysis. Acta Informatica 1, 79–110.

Koch, K.R., 1999. Parameter estimation and hypothesis testing in linear models. 2., updated and enlarged ed., Springer, Berlin.

Kok, S., Domingos, P., 2005. Learning the structure of Markov Logic Networks, in: Proceedings of the 22nd International Conference on Machine Learning (ICML), Bonn, Germany, ACM Press. 441–448.

Koller, D., Friedman, N., 2009. Probabilistic graphical models: Principles and techniques. Adaptive computation and machine learning, MIT Press, Cambridge, Mass.

Lavrac, N., Dzeroski, S., 1994. Inductive logic programming: Techniques and applications. Ellis Horwood series in artificial intelligence, Horwood, New York.

Lehmann, E.L., Romano, J.P., 2005. Testing statistical hypotheses. Springer texts in statistics. 3. ed., Springer, New York, NY.

Loch-Dehbi, S., Dehbi, Y., Plümer, L., 2013. Stochastic reasoning for uav supported reconstruction of 3d building models. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2, 257–261.

Loch-Dehbi, S., Plümer, L., 2011. Automatic reasoning for geometric constraints in 3d city models with uncertain observations. ISPRS Journal of Photogrammetry and Remote Sensing 66, 177 – 187. Quality, Scale and Analysis Aspects of Urban City Models.

Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA, USA.

Martinović, A., Van Gool, L., 2013. Bayesian grammar learning for inverse procedural modeling, in: IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013, 201–208.

Mooney, J.R., 1997. Inductive logic programming for natural language processing, in: Muggleton, S. (Ed.), Inductive Logic Programming: Selected papers from the 6th International Workshop, Springer. 3–24.

Muggleton, S., 1995. Inverse entailment and progol. New Generation Computing, Special issue on Inductive Logic Programming 13, 245–286.

Muggleton, S., Feng, C., 1990. Efficient induction of logic programs. Citeseer.

Muggleton, S., King, D.R., Sternberg, J.E.M., 1992. Protein secondary structure prediction using logic, in: Muggleton, S. (Ed.), ILP92, 228–259.

Müller, P., Zeng, G., Wonka, P., Gool, L.V., 2007. Image-based procedural modeling of façades. ACM. Transactions on Graphics 26, 85.

Nevill-Manning, C., Witten, I., 1997. Identifying hierarchical structure in sequences: A linear-time algorithm. Journal of Artificial Intelligence Research 7, 67–82.

Platt, J.C., 1999. Advances in kernel methods, MIT Press, Cambridge, MA, USA. chapter Fast Training of Support Vector Machines Using Sequential Minimal Optimization, 185–208.

Quinlan, J.R., 1990. Learning logical definitions from relations. Machine learning 5, 239–266.

Richardson, M., Domingos, P., 2006. Markov logic networks. Machine Learning 62, 107–136.

Ripperda, N., 2008. Determination of façade attributes for façade reconstruction. ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37, 285–290.

Ripperda, N., Brenner, C., 2009. Application of a formal grammar to façade reconstruction in semiautomatic and automatic environments, in: Proc. of 12th AGILE Conference on GISience.

Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of relieff and rrelieff. Mach. Learn. 53, 23–69.

Schmittwilken, J., 2009. Model selection for composite objects with attribute grammars, in: Proceedings of 12th AGILE Conference on GIScience.

Schmittwilken, J., 2012. Attributierte Grammatiken zur Rekonstruktion und Interpretation von Fassaden. Ph.D. thesis. University of Bonn.

Schmittwilken, J., Dörschlag, D., Plümer, L., 2009. Attribute grammar for 3D city models, in: Proceedings of Urban Data Management Symposium (UDMS), June 2009 Ljubljana, Slovenia, Taylor & Francis.

Schmittwilken, J., Plümer, L., 2010. Model-based reconstruction and classification of façade parts in 3D point clouds, in: Proceedings of PCV 2010 - ISPRS Technical Commission III Symposium on Photogrammetric Computer Vision and Image Analysis, Paris, France, September 1-3, 2010.

Schölkopf, B., Smola, A.J., 2001. Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning). 1st ed., MIT Press.

Singla, P., Domingos, P., 2005. Discriminative training of Markov Logic Networks, in: Proceedings, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, July 9-13, 2005, Pittsburgh, Pennsylvania, USA, 868–873.

Srinivasan, A., 2007. The Aleph manual: Version 4 and above. URL: http://www.comlab.ox.ac.uk/activities/machinelearning/Aleph/. (accessed 05.07.2016).

Stiny, G., 1982. Spatial relations and grammars. Environment and Planning B: Planning and Design 9, 313–314.

Stiny, G., Gips, J., Stiny, G., Gips, J., 1971. Shape grammars and the generative specification of painting and sculpture, in: Segmentation of Buildings for 3D Generalisation. Proceedings of the Workshop on generalisation and multiple representation, Leicester.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakisand, P., Paragios, N., 2011. Shape grammar parsing via reinforcement learning. Computer Vision and Pattern Recognition (CVPR) in Colorado Spring, USA .

Torr, H.S.P., Zisserman, A., Torr, H.S.P., 2000. MLESAC: A new robust estimator with application to estimating image geometry. Computer vision and image understanding 78, 138–156.

Tsochantaridis, I., Hofmann, T., Joachims, T., Altun, Y., 2004. Support vector machine learning for interdependent and structured output spaces, in: International Conference on Machine Learning (ICML), 104–112.

Vapnik, V.N., 1998. Statistical learning theory. 1 ed., A Wiley-Interscience publication, New York, NY, USA.

Wand, M.P., Jones, M.C., 1994. Kernel smoothing. CRC Press.

Wang, H., Suter, D., 2004. Robust adaptive-scale parametric model estimation for computer vision. Pattern Analysis and Machine Intelligence, IEEE Transactions 26, 1459–1474.

Wang, J., Domingos, P., 2008. Hybrid Markov Logic Networks, in: Association for the Advancement of Artificial Intelligence (AAAI), AAAI Press. 1106–1111.

Weston, J., Chapelle, O., Elisseeff, A., Schölkopf, B., Vapnik, V., 2002. Kernel Dependency Estimation, in: Becker, S., Thrun, S., Obermayer, K. (Eds.), NIPS, MIT Press. 873–880.

Wonka, P., Wimmer, M., Sillion, F., Ribarsky, W., 2003. Instant architecture. ACM Transactions on Graphics 22, 669–677.

Yang, M.Y., Förstner, W., 2011. Regionwise classification of building façade images, in: Photogrammetric Image Analysis (PIA2011), Springer. 209 – 220.

5 List of own publications

5.1 List of publications appended to this thesis

The following list of publications is most relevant for this thesis and appended below.

- Dehbi, Y., Plümer, L., 2011. Learning grammar rules of building parts from precise models and noisy observations. ISPRS Journal of Photogrammetry and Remote Sensing 66, 166- 176. Quality, Scale and Analysis Aspects of Urban City Models.

  In this paper I described an approach, which I designed and implemented as a first step for my thesis. I also performed the writing of the paper. The ideas and the content have been discussed with Prof. Lutz Plümer.

- Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., Plümer, L., 2016b. Statistical relational learning of grammar rules for 3d building reconstruction. Transactions in GIS, doi:10.1111/tgis.12200.

  In this paper I described how Statistical Relational Learning can be used for the learning of formal grammars for building modeling. I developed the ideas and performed the writing of the paper. The ideas and the content have been discussed amongst all authors. Fabian Hadiji supported me in designing the MLN and performing the experiments. Gerhard Gröger, Kristian Kersting and Lutz Plümer acted as supervisor for the final shape of the paper.

- Dehbi, Y., Gröger, G., Plümer, L., 2016a. Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars. Transactions in GIS, doi:10.1111/tgis.12177.

  In this paper I described a method for the derivation and modeling of symmetries from building footprints using formal grammars. I developed the ideas and performed the writing of the paper. The ideas and the content have been discussed amongst all authors. Gerhard Gröger and Lutz Plümer acted as supervisor for the final shape of the paper.

5.2 List of publications not appended to this thesis

Dehbi, Y., Staat, C., Mandtler, L., Plümer, L., 2016c. Incremental refinement of façade models with attribute grammar from 3d point clouds, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, III-3, 311-316, doi:10.5194/isprs-annals-III-3-311-2016, 2016.

Loch-Dehbi, S., Dehbi, Y., Plümer, L., 2013. Stochastic reasoning for uav supported reconstruction of 3d building models. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W2, 257-261. doi:10.5194/isprsarchives-XL-1-W2-257-2013.

Dehbi, Y., Schmittwilken, J., Plümer, L., 2010: Learning semantic models and grammar rules of building parts In: 24th Workshop on (Constraint) Logic Programming WLP' 2010, 45-56.

Dehbi, Y., Schmittwilken, J., Plümer, L., 2009: Learning semantic models and grammar rules of building parts In: Proceedings of the ISPRS WG II/2+3+4 and Cost Workshop. Lund, Sweden.

A Appended papers

A.1 Learning grammar rules of building parts from precise models and noisy observations

Dehbi, Y., Plümer, L., 2011. Learning grammar rules of building parts from precise models and noisy observations. ISPRS Journal of Photogrammetry and Remote Sensing 66, 166-176. Quality, Scale and Analysis Aspects of Urban City Models.

Abstract

The automatic interpretation of dense three-dimensional (3D) point clouds is still an open research problem. The quality and usability of the derived models depend to a large degree on the availability of highly structured models which represent semantics explicitly and provide a priori knowledge to the interpretation process. The usage of formal grammars for modelling man-made objects has gained increasing interest in the last few years. In order to cope with the variety and complexity of buildings, a large number of fairly sophisticated grammar rules are needed. As yet, such rules mostly have to be designed by human experts. This article describes a novel approach to machine learning of attribute grammar rules based on the Inductive Logic Programming paradigm. Apart from syntactic differences, logic programs and attribute grammars are basically the same language. Attribute grammars extend context-free grammars by attributes and semantic rules and provide a much larger expressive power. Our approach to derive attribute grammars is able to deal with two kinds of input data. On the one hand, we show how attribute grammars can be derived from precise descriptions in the form of examples provided by a human user as the teacher. On the other hand, we present the acquisition of models from noisy observations such as 3D point clouds. This includes the learning of geometric and topological constraints by taking measurement errors into account. The feasibility of our approach is proven exemplarily by stairs, and a generic framework for learning other building parts is discussed. Stairs aggregate an arbitrary number of steps in a manner which is specified by topological and geometric constraints and can be modelled in a recursive way. Due to this recursion, they pose a special challenge to machine learning. In order to learn the concept of stairs, only a small number of examples were required. Our approach represents and addresses the quality of the given observations and the derived constraints explicitly, using concepts from uncertain projective geometry for learning geometric relations and the Wakeby distribution together with decision trees for topological relations.

A.2 Statistical relational learning of grammar rules for 3D building reconstruction

Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., Plümer, L., 2016b. Statistical relational learning of grammar rules for 3d building reconstruction. Transactions in GIS, doi:10.1111/tgis.12200.

Abstract

The automatic interpretation of 3D point clouds is a challenging task. The interpretation process requires highly structured models representing semantics. These models serve as prior knowledge and enhance the quality of the reconstruction. In this context formal grammars play a prominent role in building modelling in the last decade. They allow to describe structures as well as parameters of buildings and their parts. As yet, the grammar rules are mostly manually derived in an expensive and laborious process that relies on expert knowledge. We propose a novel approach for the automatic learning of attribute grammar rules for 3D building reconstruction. In order to deal with the variety and complexity of buildings, we annotated buildings especially façades from different building styles. In order to cope with the learning task, we follow an incremental machine learning based strategy. The approach consists in learning the structure of façades, context free part of the grammar, in a first step. Afterwards the context free model is lifted leading to an attributed model that extends the grammar rules by further constraints reigning between the façade parts. In the first step, inspired by works from natural language processing we automatically built a treebank, collection of parse trees, in order to generate the structure of building models corresponding to the context free part of the attribute grammar. Based on the treebank entries a weighted context free grammar is learned using a support vector machine (SVM) based method. The latter infers rules and predict parse trees as structured data from a given building parts sequence as well. The second step consists in a statistical relational learning method using markov logic networks (MLNs). The latter models and enforces the topological and geometric constraints on the pre-defined terminals of the context free grammar. To this end structure as well as parameters of a MLN model are learned. Uncertain projective geometry is used in order to generate logical atoms for this learning task. This enables to deal with the uncertainty of the observations. Furthermore, MLNs addressed explicitly uncertainty allowing probabilistic inference in order to make decisions about the geometry and topology of building parts. Further MLNs are able to deal with partially unobserved values such as the case in the context of buildings e.g. due to occlusions. The learn and classification results of both parts of the learning approach are presented.

A.3 Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars

Dehbi, Y., Gröger, G., Plümer, L., 2016a. Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars. Transactions in GIS 20, 645-663, doi:10.1111/tgis.12177.

Abstract

Building and man-made objects are for many reasons, such as economical or aesthetic, often characterized by symmetry properties. The latter is reigning in the design of building footprints as well as of building parts such as façades. Thus the identification and modelling of this valuable information facilitates the reconstruction of these buildings and their parts. This paper presents a novel approach for the automatic identification and modelling of symmetries and their hierarchical structures in building footprints, providing an important prior for the façade and roof reconstruction. The uncertainty of symmetries is explicitly addressed using classification and regression methods. Symmetries are identified using supervised machine learning (Support Vector Machines). Axial as well as translational symmetries are detected. The quality of the identified major and minor symmetry axes is assessed by a least squares based adjustment. Context free formal grammar rules are used in order to model the hierarchical and repetitive structure of the underlying footprints. We present an algorithm which derives grammar rules based on the previously acquired symmetry information and using lexical analysis describing regular patterns and palindrome-like structures. This offers insights into latent structures of building footprints and describe therefore the associated façade in a relational and compact way.

Danksagung

Ich möchte an dieser Stelle all denjenigen danken, die in irgendeiner Form zu der Erstellung und Verfeinerung dieser Dissertation beigetragen haben.

An erster Stelle möchte ich mich bei Prof. Dr. Lutz Plümer für die Ermöglichung dieser Arbeit und die Unterstützung in allen Phasen bedanken. Mit seiner herzlichen und konstruktiven Betreuung hat er einen wesentlichen Beitrag zum Gelingen dieser Arbeit geleistet.

Ebenfalls möchte ich mich bei Herrn Prof. Dr. Cyrill Stachniss für die Bereitschaft, das Korreferat für diese Dissertation zu übernehmen, bedanken.

Ich möchte mich weiterhin bei meinen jetzigen und ehemaligen Kollegen vom Institut für Geodäsie und Geoinformation für die Unterstützung und die kollegiale gute Zusammenarbeit bedanken. Besonderer Dank gilt Prof. Dr. Kristian Kersting, Priv.-Doz. Dr. Gerhard Gröger, Dr. Christoph Römer, Dr. Jörg Steinrücken, Dipl. Inform. Sandra Loch-Dehbi, M.Sc. Jan Behmann und Dipl. Ing. André Henn, die mich fachlich in meiner Arbeit bestärkt haben, sowie Dipl. Ing. Michael Kneuper und Dipl. Ing. Stefan Teutsch für die stets große Hilfsbereitschaft bei der Erstellung und Überarbeitung der Grafiken in dieser Thesis.

Ich möchte mich weiterhin bei den Mitautoren meiner Veröffentlichungen bedanken. Besonderer Dank für die produktive Zusammenarbeit gilt Fabian Hadiji.

Insbesondere bedanke ich mich auch herzlich bei meiner Familie, insbesondere meiner Frau Sandra, für die Unterstützung und den Verzicht auf gemeinsame Zeit.

Abschließend möchte ich mich bei der Deutschen Forschungsgemeinschaft für die Förderung des Verbundprojektes “Mapping on Demand”, in dessen Rahmen wesentliche Teile dieser Arbeit entstanden sind.