Abstract

The automatic interpretation of 3D point clouds for building reconstruction is a challenging task. The interpretation process requires highly structured models representing semantics. Formal grammars can describe structures as well as the parameters of buildings and their parts. We propose a novel approach for the automatic learning of weighted attributed context-free grammar rules for 3D building reconstruction, supporting the laborious manual design of rules. We separate structure from parameter learning. Specific Support Vector Machines (SVMs) are used to generate a weighted context-free grammar and predict structured outputs such as parse trees. The grammar is extended by parameters and constraints, which are learned based on a statistical relational learning method using Markov Logic Networks (MLNs). MLNs enforce the topological and geometric constraints. MLNs address uncertainty explicitly and provide probabilistic inference. They are able to deal with partial observations caused by occlusions. Uncertain projective geometry is used to deal with the uncertainty of the observations. Learning is based on a large building database covering different building styles and façade structures. In particular, a treebank that has been derived from the database is employed for structure learning.

1 Introduction

The need for high-resolution three-dimensional (3D) building models has dramatically increased in the last decade. They are used in various tasks and applications such as e-planning for the efficient communication and management of sophisticated urban spatial environments (Engel and Döllner 2012). An automatic extraction of such models from images or 3D point clouds is a challenging task. This is attributed to the variety and complexity of human-made objects. In order to cope with this difficulty, the incorporation of prior knowledge represented as models is a key issue. To this end, an explicit modeling of semantics is an essential task. In this context, CityGML has been introduced (Gröger et al. 2012) which, besides semantics, defines 3D geometry, topology, and appearance of urban objects in different levels of detail.

In the context of 3D building reconstruction, formal grammars receive increasing attention (Musialski et al. 2012). Formal grammars adequately describe the aggregation of buildings, as well as the relations existing between their parts. They are used for the generation of synthetic city models (procedural modeling) (Vanegas et al. 2010), as well as for the reconstruction of buildings and their parts. So far, except for a few approaches (Ripperda and Brenner 2009;
Becker 2009; Martinović and Van Gool 2013) that tried to extract the grammar rules from given data (e.g. an image) the rules are mostly manually derived. However, this is an expensive and laborious task that needs expert knowledge.

Our goal in this article is to make a step towards the automatic learning of such rules. We propose a novel approach for the automatic learning of a weighted attributed context-free grammar (WACFG) for the identification and reconstruction of façades in 3D building models. Our method addresses not only planar façades but also façades with sophisticated 3D structures. Figure 1 shows that the proportion of buildings with displaced façades and oriels cannot be neglected. Attribute grammars – a key concept in our approach – 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.

The main contribution of this article is the automatic learning of a weighted attributed context-free grammar (WACFG) for 3D building reconstruction. In contrast to procedural methods we propose a declarative approach that separates the representation of buildings and their parts from the reconstruction task. We used our WACFG 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 Markov Logic Networks (MLNs).

Figure 2 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 façade parts in a recursive way according to rules stemming from the induced weighted context-free grammar (WCFG). An excerpt of these rules is shown in the box on the top. The $p_i$ indicates 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 to the left at the bottom of the figure. The weights $\lambda_j$ have been automatically learned and denote the importance of the associated formula. The formulas 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 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 into the distribution and importance of different structural patterns by the weights that expand the classical context-free grammar rules. WACFGs enable the modeling of objects with an a-priori unknown number of parameters such as the number of floors and windows. Our approach explicitly addresses the uncertainty of observations by uncertain projective geometry, probabilistic rules and MLNs. All in all, WACFG enables us 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. To the best of our knowledge, this is the first demonstration of the impact of adapting recent machine learning techniques – more precisely statistical relational ones – for 3D building reconstruction.
The remainder of this article is structured as follows: Related work will be discussed in Section 2. Section 3 introduces the necessary theoretical background of the used methods. After that, our method will be explained in detail in Section 4. Section 5 presents and discusses the results we achieve applying our approach. The article is summarized and concluded in Section 6.

2 Related Work

Grammar-based modeling for 3D building reconstruction plays a prominent role in several works. In the field of procedural modeling, which consists of the generation of a huge number of 3D models, the focus of our work is on the modeling of historic buildings. To this end, a grammar-based approach is presented in [11], which is based on the combination of procedural modeling and Markov Logic Networks (MLNs). MLNs are a form of statistical relational learning (SRL) that can be used to model complex relationships in real-world data. In contrast to traditional machine learning approaches, MLNs provide a natural way to express complex dependencies and constraints between entities in a domain. Moreover, MLNs can be learned from training data, which allows them to be adapted to new situations and scenarios. This makes MLNs an attractive approach for modeling the complex relationships that exist between the different parts of buildings, such as windows, doors, and roofs. Figure 1 shows 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 cannot be always adequately modeled as 2D faces. The percentage of buildings with such structures is significant.

Figure 1  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 cannot be always adequately modeled as 2D faces. The percentage of buildings with such structures is significant.

Figure 2  The main façade of the Poppelsdorf castle in Bonn modeled with our weighted attributed 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).
of synthetic buildings based on a-priori designed grammar rules, Müller et al. (2007) proposed a system for the generation of consistent mass building models based on the so-called CGA shape grammar. Nonetheless the rules of this grammar are expensively manually designed based on expert knowledge.

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 imagery data. They infer split grammar rules from labeled images. However, their approach assumes that façades are planar 2D surfaces. Teboul et al. (2011) used shape grammars for 2D façade parsing using reinforcement learning 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. The approach assumes also that the considered façade is a 2D planar surface. Dehbi and Plümer (2011) proposed an inductive logic programming- (ILP) based method in order to learn attribute grammar rules for building parts from precise models and noisy observations. Toshev et al. (2010) introduced a method for the detection and parsing of buildings from 3D laserscan data. They derived parse trees that decompose buildings in roof surfaces and volumetric parts. The trees stem from a simple fitted grammar that does not take detailed façade parts into account. Xiong et al. (2013) identified and modeled the main structural components of an indoor environment using a material- or texture-driven recognition in order to consider significant occlusions for rectangular room reconstruction. For more information about the works in the field of 3D urban reconstruction we refer to a survey of urban reconstruction of Musialski et al. (2012).

Our approach is partly inspired by techniques from the field of natural language processing. In order to learn and model the structures of sentences, a treebank as tagged corpus plays a major role in this field. Charniak (1996) showed how treebanks as collections of parse trees can be efficiently exploited in order to build statistical parsers. In this context Tsouchantaridis et al. (2004) proposed a method based on Support Vector Machines for learning interdependent and structured output spaces. They demonstrated their method among others for a supervised learning of weighted context-free grammars based on a subset of the Penn treebank Wall street journal corpus (Taylor et al. 2003).

As mentioned above, ILP-based approaches have been used to learn building parts. However, since the observations are noisy and the building structures are complex, often pure first-order logic-based methods are not sufficient. Thus, it is difficult to find a precise set of logical rules that hold across all possible buildings and façades. Motivated by this observation, various approaches that combine logic and probabilities have been introduced in the field of SRL. A particular example of such an approach is the MLNs proposed by Richardson and Domingos (2006). MLNs have been used by Singla and Domingos (2006) for problems such as entity resolution, which is akin to the setting in this article. For example, entity resolution in text documents is concerned with the distinction of different mentions that refer to entities in a database. Generally, a mention is any reference in the text to an entity or concept. Here, we use MLNs to distinguish different types of objects of a building façade. Some of the rules used in entity resolution, for example transitivity, carry over to our setting and we also try to connect observations to the predicate differentiating two façade objects. We will describe MLNs more formally in Section 3.2.

© 2016 John Wiley & Sons Ltd Transactions in GIS, 2016, 00(00)
3 Theoretical Background

We are interested in the learning of attributed grammars as a particular type of formal grammars. A formal grammar $G$ (Chomsky 1956, 1959) is defined as a quadruple $(S, N, T, R)$ of a start symbol $S$, a set of non-terminals $N$, represented by capitalized initials, a set of terminals $T$, denoted by lowercase initials and a set of production rules $R$. A production rule is applied substituting the left hand-side by the right hand-side. Context-free grammars have the property that the left hand-side of the production rules contains a single non-terminal symbol only. In a weighted context-free grammar, each rule is augmented by a weight in order to express the likelihood of the application of this rule. In order to cope with a restricted form of context sensitivity, Knuth (1968, 1971) extended context-free grammars and introduced attribute grammars. 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.

As we learn from noisy observation, representation of and dealing with uncertainty plays a fundamental role. Questions such as: “Are two windows the same?” “Are they aligned?” “Is there a symmetry between the balconies on the left and the right side?” occur frequently. For this aim we used the formalism of uncertain projective geometry. For geometric reasoning under uncertainty, Heuel (2004) integrated the potentials of projective geometry (Dorst et al. 2007) and statistics. His approach draws upon the modeling of error propagation during the reasoning process, which consequently enables testing uncertain spatial relations between geometric entities based on a chi-squared statistical hypothesis test. In the following, the used methods for the learning of a weighted attributed context-free grammar are explained.

3.1 Support Vector Machine for Learning a Weighted Context-free Grammar

The learning of models that take into account functional dependencies between arbitrary input (e.g. variable length) and outputs (e.g. structured data) is a further challenge for machine learning. In contrast to multinomial classification with a finite number of atomic 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 $(x_1, y_1), \ldots, (x_n, y_n) \in X \times Y$. Hereby a fix a-priori unknown probability distribution is assumed. To this end, a discriminant function $F : X \times Y \rightarrow \mathbb{R}$ that measures the compatibility of $x$ and $y$ is defined in order to induce a mapping $f$:

$$f(x; w) = \arg\max_{y \in Y} F(x, y; w)$$

(1)

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 derivable trees are evaluated using the sum of the $w_i$ of the belonging nodes. These 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:

$$F(x, y; w) = (w, \Psi(x, y)) = \sum_{g_i \in \text{rules}(y)} w_i$$

(2)

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 Cocke-Younger-Kasami algorithm (CYK) (Manning and Schütze 1999). CYK is a bottom-up parsing algorithm for context-free grammars based on dynamic programming. The system \( \text{svm}_\text{cfg} \) (http://www.cs.cornell.edu/people/tj/svm_light/svm_cfg.html) provides an open-source implementation of the support vector machine algorithm for learning a weighted context-free grammar.

### 3.2 Markov Logic Networks

MLNs (Richardson and Domingos 2006) allow a combination of logic and probabilistic graphical models. In contrast to classical logic, a weight is assigned to each formula. A grounded MLN, i.e. a variable free set of formulas where each variable is assigned a constant, represents an instantiation of a Markov Random Field (MRF). Therefore, probabilistic inference techniques, such as Gibbs sampling or Belief Propagation, can be applied to MLNs as well. For more details on MRFs and probabilistic graphical models the reader is referred to Koller and Friedman (2009).

Essentially, MLNs are defined by weighted first-order formulas. A first-order formula is constructed using constants, variables, functions, and predicates. To restrict complexity we will restrict ourselves to function-free first-order logic. Constants represent objects in our domain, e.g. Window1, Window2, Door1, and so on. A logical variable \( x \) ranges over the domain of objects. Predicates return a truth value and define attributes of objects or relations between objects depending on their arity. For example, the unary predicate isDoor(\( A \)) is true if the object \( A \) is a door, whereas the binary predicate similar(\( A, B \)) is true if objects \( A \) and \( B \) are of the same type, possibly including topological and geometrical properties. An atom is a predicate applied to a tuple of terms and a term can be a constant or a variable. A ground term is a term not containing any variables and correspondingly, a ground atom is a predicate whose arguments are all ground terms. Hence, \( x \) and “Window1” are terms and similar(\( x, \text{Window1} \)) is an atom.

Following the work of Richardson and Domingos (2006), an MLN is defined as a set of pairs \((F_i, \lambda_i)\). Here, \( F_i \) is a formula in first-order logic and \( \lambda_i \in \mathbb{R} \) is the weight of formula \( F_i \). More details on first-order logic can be found in De Raedt (2008). In addition, a set of constants \( C=\{C_1, \ldots, C_j\} \) is needed. Now an MLN can be used in order to construct an MRF as follows. For every ground predicate in the MLN a binary variable is added to the MRF, e.g. similar(Window1, Window2) corresponds to one binary variable in the MRF. The value of the random variable models the truth state of the ground predicate. Due to the finite set of constants, this results in a finite set of random variables. Furthermore, one potential function for each possible grounding of a formula \( F_i \) is added. For example, if one instantiation of formula \( F_i \) is sameWidth(Window1, Window2) \( \rightarrow \) similar(Window1, Window2), then a potential exp(\( \lambda_i f_k \)) is added to the MRF. The feature \( f_k \) is defined over all ground predicates appearing in the formula and evaluates to 1 if the formula is satisfied and 0 otherwise. Essentially, MLNs serve as a template engine for MRFs that allow us to easily model independences based on logic rules.

In order to learn the weights of an MLN efficiently and in an automatic way, a pseudo-likelihood is often maximized instead of the log-likelihood \( \ln p(x; w) \) of an MRF for given training data. In order to reduce the computational complexity of the training procedure, Singla and Domingos (2005) presented a discriminative training that splits the predicates into evidence and query predicates. Then the conditional likelihood for the query predicates given the observatory predicates is maximized based on the training database. To derive the structure, i.e. the first-order formulas of an MLN, from the training dataset as well and respecting
the probabilistic nature of MLNs directly, Kok and Domingos (2005) performed a beam search guided by a pseudo log-likelihood measure. A reference implementation of MLNs, including structure and parameter learning, is provided by the open-source Alchemy-system (alchemy.cs.washington.edu/). By default, Alchemy uses MaxWalkSat (MWS) for MAP inference. Alternatively, one can use so-called message-passing algorithms such as max-product Belief Propagation.

4 Learning of Weighted Attributed Context-free Grammar for 3D Façade Reconstruction

In this section, we introduce a novel approach for the automatic learning of a weighted attributed context-free grammar for the interpretation of 3D point clouds. Figure 3 gives an overview of our approach. In order to cope with the learning task, we follow an incremental strategy consisting of learning the structure of façades first, followed by the parameters of the building parts as well as the related constraints. In the first step, a weighted context-free grammar is learned based on a relational building database (RBDB) consisting of 1,300 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. We used the open-source measureFacade (http://www.ikg.uni-bonn.de/forschung/projektarchiv/measurefacade.html) tool 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 shapes as well as the location parameters of each building part are captured. In contrast to Ripperda (2008), we used point clouds as well as images and retained the relative location of each part, e.g. floor and column location. Figure 4 shows an excerpt of the database schema of RBDB. 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 with definite columns and floors in order to describe structural information.

All data is taken manually either from undistorted, rectified and scaled images or from high resolution LiDAR 3D point clouds: 1,230 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; and 70 façades were captured by static scanning with a Leica HDS6100 laser scanner.

In order to learn façade structures we follow a supervised learning approach using Support Vector Machines for structured data as described in Section 3.1. This is performed based on input-output pairs. In the prediction stage (Figure 3, 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) in a way that is similar to the approach described in Schmittwilken and Plümer (2010). KDE enables a non-parametric estimation of a probability density function leading to the shape and location parameters of the façade parts.

The learning of a weighted context-free grammar (WCFG) enables the modeling of façade structures especially their aggregation in different parts. Furthermore, this gives insight into the distribution and importance of different structural patterns based on the rule weights. For the learning of the WCFG, we built a treebank (Manning and Schütze 1999; Charniak 1996) 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 weighted context-free grammar. We derived parse trees from observations using Treebankgenerator that enables an automatic treebank entries induction from RBDB. Each parse tree in the resulting treebank that corresponds to a given façade is automatically derived from RBDB leading to derivation trees reflecting several common architectural patterns such as floor- 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 RBDB leading to a derivation tree following filter criteria such as minimal description length. More details about the induction process are described in Burger (2012).

Once a treebank has been generated, we used a support vector machine based approach to derive and parse a WCFG as described in Section 3.1 (Figure 3, 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 of structured parse trees. Figure 5 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).
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 3.2).

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, logical ground atoms, which represent geometric and topological constraints between façade objects, are required. Thus, we generated these atoms from RBDB (Figure 3, bright brown). To this end, we performed a statistical geometric reasoning using uncertain projective geometry in order to make decisions about the similarity of geometric entities such as windows (cf. Section 3). 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 parallelism of two lines using a chi-squared statistical hypothesis test.

The logical atoms are extracted according to the predicate list in Table 1, which gives the most important predicates for our experiment but can easily be extended and modified. The full set of extracted ground atoms per façade forms the MLN training database which is used to learn the MLN, consisting of the first-order rules as well as their associated weights $w_i$.

| Predicate                  | Description                                                                 |
|---------------------------|-----------------------------------------------------------------------------|
| sameFloor(x,y)            | true iff x and y are on the same floor                                       |
| shareHorAlignment(x,y)    | true iff x and y are horizontally aligned                                    |
| sameColumn(x,y)           | true iff x and y belong to the same column                                   |
| neighborHor(x,y)          | true iff x and y are horizontal neighbors                                    |
| shareVerAlignment(x,y)    | true iff x and y are vertically aligned                                     |
| neighborVer(x,y)          | true iff x and y are vertical neighbors                                      |
| sameHeight(x,y)           | true iff x and y have the same height                                        |
| sameWidth(x,y)            | true iff x and y have the same width                                         |

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 3.2).

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, logical ground atoms, which represent geometric and topological constraints between façade objects, are required. Thus, we generated these atoms from RBDB (Figure 3, bright brown). To this end, we performed a statistical geometric reasoning using uncertain projective geometry in order to make decisions about the similarity of geometric entities such as windows (cf. Section 3). 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 parallelism of two lines using a chi-squared statistical hypothesis test.

The logical atoms are extracted according to the predicate list in Table 1, which gives the most important predicates for our experiment but can easily be extended and modified. The full set of extracted ground atoms per façade forms the MLN training database which is used to learn the MLN, consisting of the first-order rules as well as their associated weights $w_i$.

The concepts described in Section 3.2 for learning MLNs and the inference based on these MLNs are now applied to building reconstruction. Our target predicate, which is always latent during the inference, is $\text{similar}(x,y)$. This binary probabilistic predicate was inferred from our database by a pairwise comparison of different objects. For two building parts, $\text{similar}$ is true if they are of the same type, have the same geometry and cannot be further distinguished by any other property in the database. Therefore, $p(\text{similar}(x,y)=\text{True})$ models the degree of similarity between two building parts. Many architectural aspects contribute to the probability $p$ of the $\text{similar}$-predicate. These aspects are among others shape parameters of the considered objects like the width and height, the vertical and horizontal alignment, neighborhood information and whether they belong to the same floor (cf. Table 1). The mentioned aspects and their influence on the similarity are represented by the MLN. To infer the similarity between pairs of building parts, we performed a structure and weight learning of an MLN (Figure 3, bright brown). In the prediction stage this MLN will be applied together with the ground atoms that will be described in the following.

So far we have derived a generic MLN model as well as a generic weighted context-free grammar. 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 are applied leading to the sequence of building part types as well as their parameter values. The latter are uncertain and may be incomplete. Therefore, uncertain projective geometry is used again in order to extract ground atoms according to the predicates in Table 1. These atoms describe geometric properties as well as relations between building parts of the façade $F$. Examples for such atoms are $\text{sameWidth}(W_1,W_2)$ and $\text{shareHorAlignment}(W_1,W_2)$ for two observed windows $W_1$ and $W_2$. 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 1 that describe structural and topological relations are derived from the predicted parse tree for the façade $F$. Examples are $\text{sameFloor}(W_1,W_2)$ or $\text{neighborHor}(W_1,W_2)$ 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, yellow background) using statistical inference as described in Section 3.2. Each leaf of the parse tree corresponds to a constant in our MLN. We are primarily interested in determining the most likely configuration of the similar-predicates, which consists of 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). Unobserved parameters, especially, can be estimated.

5 Experimental Results

We learned three different Support Vector Machine models 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 2 shows the F1-scores of the learning and test results for the three models using the svm_crg software. The prediction of derivation trees is solely based on the weak observation consisting of a string sequence of the building part types, without any information about geometric parameters and structure of floors. Nevertheless the F1-score is at least 0.8. During the learning phase 90% of the data was used, while the remaining 10% was used in the testing stage. The parsing of façade elements using different building styles can also be exploited in order to classify an a-priori unknown building type. This is a helpful tool, especially if some features such as footprints used in Henn et al. (2012) are not available.
Furthermore, we evaluated and tested our proposed MLN model. For the evaluation we used a dataset containing façades from 20 buildings in Bonn, Germany. Each building has a varying number of objects, accordingly the corresponding grounded MLNs vary in size. To make the inference task more realistic and to demonstrate that our method can cope with unobserved objects, we randomly removed some of the sameWidth and sameHeight predicates in every façade. All experiments were conducted in a 10-fold cross validation, i.e. we used 90% of the buildings for parameter and structure learning, and the remaining buildings were used for testing. We measured the performance of our results based on the well established F1-score.

We will begin our experimental evaluation based on a manually crafted MLN for which only the weights are learned using Alchemy’s discriminative learning as described in Section 3.2. This enables us to compare the results with a fully automatically learned MLN. To this end, we constructed a simple MLN that connects the similar predicate to the features by means of an implication. The goal was to predict the similar predicate based on the other predicates as features. An exemplary formula is \( \text{sameWidth}(x,y) \rightarrow \text{similar}(x,y) \) stating that the same width is a strong prior for similarity. Additionally, we added a formula expressing transitivity among the unobserved predicates. In the most simple case, all other predicates are observed and only similar is missing. In this case, our grounded MLNs contained on average 122 variables and 1,739 formulas. Since only similar predicates need to be instantiated, the number of variables exactly amounts to \(|C|^2\). C denotes the set of constants. The domain for each MLN consists of the different objects in the façade. When we start to remove observation for sameHeight and sameWidth, the grounded MLNs grow. With 25% of both predicates removed, they have on average 183 variables and 2,156 formulas. Finally, with 50% of the predicates removed, the grounded MLNs have around 242 variables and roughly 3,008 formulas. We now run MAP inference on all buildings and compare the predictions with the ground truth. The first column in Table 3 shows the averaged results for running MaxWalkSat (MWS) and the second column shows the results obtained by max-product Belief Propagation (BP). We can see that the problem is easy if none of the sameWidth or sameHeight are missing. This is not surprising because the definition of similar is strongly dependent on these predicates. If we now increasingly remove observations, we can see a drop in the F1-scores. With 25% missing, we still obtain an F1-score of 0.822 with MWS. On the other hand, one should also note the poor results (below 0.5) obtained by BP. We attribute this behavior to the construction of our MLN. In particular, we replaced the automatically learned weights for the rules expressing transitivity with near deterministic weights, i.e. very strong parameter values, in the MLN to enforce transitivity more strongly. However, BP is known to often perform poorly with near-deterministic weights. We therefore also experimented with MLNs without the transitivity rule which resulted in more balanced results for MWS and BP but the overall performance of the best MLN was lower.

We now turn our attention to MLNs automatically learned by the Alchemy system. The third and fourth column in Table 3 show the results of the cross validation based on MLNs with a maximum of two variables per formula. We can immediately see that the F1-scores are well balanced for MWS and BP. BP is even able to obtain a better F1-score in cases with more missing observations.

Overall, the results are comparable to handcrafted MLNs but do not require any expert knowledge in the rule construction. Instead, the learned MLNs reveal interesting connections expressed by the learned formulas.

For example, a rule that is commonly found is the following one:

\[ \lambda_i : \neg \text{sameHeight}(a1,a2) \lor \neg \text{sameWidth}(a1,a2) \lor \neg \text{similar}(a1,a2) \]

© 2016 John Wiley & Sons Ltd
Transactions in GIS, 2016, 00(00)
which means that the same height and width implies dissimilarity. This does not correspond to intuition. However our approach realizes this and penalizes this formula by negative weight $\lambda_i$. Hence, true groundings of this formula are also penalized. Another benefit of the automatically learned formulas is the reduced problem sizes in terms of ground formulas. Due to the fact that we have restricted the maximum number of variables in each formula, the learner is limited and cannot learn formulas such as transitivity. Therefore, even with 50% of the observations missing, the problems contain on average only 403 formulas. The number of variables is identical to the handcrafted case, because the number of unobserved predicates remains the same. We also tried permitting more variables per formula but this resulted in an overall worse performance.

In a last experiment we investigated the question of whether we could automatically combine learned MLNs with expert rules to obtain improved results over either case. To this end, we used the best performing automatically learned MLN, added additional rules specifying transitivity, symmetry and reflexivity, and then re-learned the parameters.

The results are given in columns 5 and 6 in Table 3. As one can see, however, the MLNs are not able to exploit the additional formulas to improve the predictions. Instead, the results are slightly worse. This difference however is not significant according to a paired t-test. Looking at the learned weights of the manual formulas in detail, one can see that the weights are positive. On the other hand, the learner now gives several other automatically learned formulas as weight of zero. Since learning and inference are only approximate, the inference learner cannot make use of the generally correct manual rules, but instead works better with using only automatically learned formulas.

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 6 shows the model-based reconstruction of five façades using our 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 be 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 our 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

|                  | 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      |
MLN model. Displaced façade parts such as the case in façade 2 left column and façade 3 fourth column are identified, although they are graphically not represented in the figure.

In contrast to existing tools such as CityEngine (http://www.esri.com/software/cityengine/) which is based on a-priori manually specified grammar rules, our approach enables an automatic learning of a wide range of grammar rules. Besides, our approach expects an input object such as 3D point cloud in order to reconstruct the façade parts based on the learned grammar rules. Furthermore, with the use of MLNs we provide a flexible tool that avoids the implementation of the procedural description of the attributes and constraints such as the case in shape grammars.

6 Conclusions and Outlook

This article introduced a novel, machine learning-based approach for learning weighted attribute context-free grammar rules for 3D building reconstruction. The rules serve as strong prior knowledge and semantic models in order to support the reconstruction process of buildings and their parts. The learning of such models reduces the expense of manually designing grammar rules. In order to cope with the complexity of this task, a two-staged incremental strategy is followed. At first, the context-free part of the grammar 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 as structured data. The latter consists of parse trees representing façades and a treebank. The trees are automatically induced from a relational database of buildings from Bonn, Germany, reflecting different building styles. In addition to the grammar rules a classification model is obtained. This enables parsing 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 existing between the façade elements. Hereby, a statistical relational learning method using Markov Logic Networks (MLNs) is used for the first time in 3D building reconstruction, to the best of our knowledge. MLNs allow a probabilistic inference that enables decisions about geometry and topology of building parts. In order to learn the structure, i.e. formulas, as well as the parameters, i.e. weights of the formulas, of an MLN logical atoms are automatically generated from our relational building database. To this issue uncertain projective geometry is used in order to take the uncertainty of the underlying observations into account.

Our approach deals successfully with complexity – a varying number of objects – uncertainty and unobservability in real-world problems. This is explicitly addressed by uncertain projective geometry, probabilistic grammar rules and MLNs.

The classification results of the prediction of façade structures are between 0.8 and 0.88 which is a very good classification rate since the inputs are very weak observations: lists of several building types. The overall classification rate 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. We have shown that handcrafted rules do not outperform the automatically learned ones.

The rules, their weights, the classification model for the prediction of façade structures, the structure and the weights of MLNs are learned in a supervised fashion. This is based on a large building data set covering the variety of buildings.

In the presented approach, the variables and features of MLNs are discrete. So far we have used uncertain projective geometry 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. They allow the integration of all probability distributions from the exponential family such as multivariate Gaussians in order to model continuous properties. Furthermore, lifted inference approaches can be beneficial for the task at hand. 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). In many cases, symmetry is already reflected in the footprints and can be exploited in the reconstruction process (Dehbi et al. 2015). In particular, unobservability is addressed and hence, more efficient and reduced models are sufficient. Since the represented approach cannot deal with non-rectangular shapes, an extension of the supported shapes should be investigated in future works.

References

Bakir G H, Hofmann T, Schölkopf B, Smola A J, Taskar B, and Vishwanathan S V N 2007 Predicting Structured Data (Neural Information Processing). Cambridge, MA, MIT Press
Becker S 2009 Generation and application of rules for quality dependent façade reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 64: 640–53
Burger J 2012 Generating Tree Bank Entries from Relational Descriptions of Facades and Their Components. Unpublished MS Thesis, Fern Universität Hagen.
Charniak E 1996 Tree-bank grammars. In Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference, Portland, Oregon: 1031-36
Chomsky N 1956 Three models for the description of language. IRE Transactions on Information Theory 2: 113–24
Chomsky N 1959 On certain formal properties of grammars. Information and Control 2: 137–67
De Raedt L 2008 Logical and Relational Learning. Berlin, Springer
Dehbi Y, Gröger G, and Plümmer L 2016 Identification and modelling of translational and axial symmetries and their hierarchical structures in building footprints by formal grammars. Transactions in GIS 20: in press
Dehbi Y and Plümmer L 2011 Learning grammar rules of building parts from precise models and noisy observations. ISPRS Journal of Photogrammetry and Remote Sensing 66: 166–76
Dorst L, Fontijne D, and Mann S 2007 Geometric Algebra for Computer Science: An Object-oriented Approach to Geometry. Amsterdam, the Netherlands, Elsevier/Morgan Kaufmann
Engel J and Döllner J 2012 Immersive visualization of virtual 3D city models and its applications in e-planning. International Journal of E-Planning Research 1: 17–34
Gröger G, Kolbe T H, Nagel C, and Häfele K 2012 OpenGIS City Geography Markup Language (CityGML) Encoding Standard. Version 2.0.0. Wayland, MA, Open Geospatial Consortium Document No. 12-019
Hadiji F and Kersting K 2013 Reduce and re-lift: Bootstrapped lifted likelihood maximization for MAP. In Proceedings of the Statistical Relational Artificial Intelligence Workshop held in conjunction with the Annual Meeting of the Twenty-seventh International AAAI Meeting for the Advancement of Artificial Intelligence, Bellevue, Washington
Henn A, Römer C, Gröger G, and Plümmer L 2012 Automatic classification of building types in 3D city models. Geoinformatica 16: 281–306
Heuel S 2004 Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction. Berlin, Springer
Kersting K 2012 Lifted probabilistic inference. In Proceedings of the Twentieth European Conference on Artificial Intelligence, Montpellier, France
Knuth E D 1968 Semantics of context-free languages. Theory of Computing Systems 2: 127–45
Knuth E D 1971 Top-down syntax analysis. Acta Informatica 1: 79–110
Kok S and Domingos P 2005 Learning the structure of Markov logic networks. In Proceedings of the Twenty-second International Conference on Machine Learning, Bonn, Germany: 441-48
Koller D and Friedman N 2009 Probabilistic Graphical Models. Cambridge, MA, MIT Press
Manning C D and Schütte H 1999 Foundations of Statistical Natural Language Processing. Cambridge, MA, MIT Press
Martinovic A and van Gool L 2013 Bayesian grammar learning for inverse procedural modeling. In Proceedings of the Twenty-sixth IEEE Conference on Computer Vision and Pattern Recognition, Portland, Oregon: 201-08
Müller P, Zeng G, Wonka P, and van Gool L V 2007 Image-based procedural modeling of facades. ACM Transactions on Graphics 26: 85
Musialski P, Wonka P, Aliaga D G, Wimmer M, van Gool L, and Purgathofer W 2012 A survey of urban reconstruction. In Proceedings of the Thirty-third Annual Conference of the European Association of Computer Graphics, Cagliari, Italy
Richardson M and Domingos P 2006 Markov logic networks. Machine Learning 62: 107–36
Ripperda N 2008 Determination of facade attributes for facade reconstruction. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37: 285-90
Ripperda N and Brenner C 2009 Application of a formal grammar to facade reconstruction in semi-automatic and automatic environments. In Proceedings of the Twelfth AGILE Conference on Geographic Information Science, Hannover, Germany
Schmittwilken J and Plümmer L 2010 Model-based reconstruction and classification of facade parts in 3D point clouds. In Proceedings of PCV 2010-ISP RS Technical Commission III Symposium on Photogrammetric Computer Vision and Image Analysis, Paris, France
Singla P and Domingos P 2003 Discriminative training of Markov logic networks. In Proceedings of the Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, Pittsburgh, Pennsylvania: 868-73
Singla P and Domingos P 2006 Entity resolution with Markov logic. In Proceedings of the Sixth International Conference on Data Mining, Hong Kong: 572-82
Taylor A, Marcus M, and Santorini B 2003 The Penn Treebank: An overview. In Abeillé (ed) Treebanks The state of the art in syntactically annotated corpora. Berlin, Kluwer: 5–22
Teboul O, Kokkinos I, Simon L, Koutsourakis P, and Paragios N 2011 Shape grammar parsing via reinforcement learning. In Proceedings of the Twenty-third IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, Colorado.

Toshev A, Mordohai P, and Taskar B 2010 Detecting and parsing architecture at city scale from range data. In Proceedings of the Twenty-third IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California: 398-405.

Tschantaridis I, Hofmann T, Joachims T, and Altun Y 2004 Support vector machine learning for interdependent and structured output spaces. In Proceedings of the Twenty-first International Conference on Machine Learning, Banff, Alberta: 104-112.

Vanegas C A, Aliaga D G, Wonka P, Müller P, Waddell P, and Watson B 2010 Modelling the appearance and behaviour of urban spaces. Computer Graphics Forum 29: 25–42.

Wand M P and Jones M C 1994 Kernel Smoothing. Boca Rotan, FL, CRC Press.

Wang H and Suter D 2004 Robust adaptive-scale parametric model estimation for computer vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1459–74.

Wang J and Domingos P 2008 Hybrid Markov logic networks. In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, Illinois: 1106-11.

Xiong X, Adan A, Akinci B, and Huber D 2013 Automatic creation of semantically rich 3D building models from laser scanner data. Automation in Construction 31: 325–37.