A STEP TOWARDS SEMANTIC-BASED BUILDING RECONSTRUCTION USING MARKOV-RANDOM-FIELDS

Ansgar Brunn
Institute for Photogrammetry, Bonn university, Germany
Brunn@ipb.uni-bonn.de

KEY WORDS: Building reconstruction, Digital surface models, Markov-Random-Fields, Topology

ABSTRACT

In this paper we describe a new concept for the reconstruction of buildings. In contrast to most of the published approaches, we link the reconstruction process with the building interpretation. With this linkage we want to enhance the reconstruction result and to yield semantic information about the buildings. We introduce building models based on their topology. We also may use data from different sensor types. The analysis is done locally using statistical building information for the interpretation in a Markov-Random-Field and using e. g. geometric or radiometric "appearance" models for the reconstruction. A real data example from laserscanner observations demonstrates the approach.

KURZFASSUNG

In diesem Artikel beschreiben wir ein neues Verfahren zur Rekonstruktion von Gebäuden. Im Gegensatz zu den meisten in der Literatur bereits veröffentlichen Verfahren, verbinden wir die Rekonstruktion mit der Interpretation. Dadurch verbessern wir das Ergebnis der Rekonstruktion und erhalten zusätzlich semantische Information über das Gebäude. Wir verwenden ein Gebäudemodell, das auf der Topologie der Gebäude definiert ist. Außerdem ist die Integration von Daten unterschiedlicher Sensortypen möglich. Die Analyse der Daten erfolgt bei der Interpretation mit lokalem statistischen Gebäudewissen und bei der Rekonstruktion mit lokalen Modellen der "Erscheinungsform" der Gebäude (z. B. geometrisch oder radiometrisch). Ein Beispiel mit realen Entfernungsdaten demonstriert den Ansatz.

1 INTRODUCTION

In this paper a new concept for the reconstruction and interpretation of building data is introduced. We present an approach for multi sensorial building data analysis which combines reconstruction and interpretation.

1.1 Motivation

Since a few years a large demand for urban and suburban 3D data can be recognized. Various applications need 3D data for planning. Others need 3D data as background information for visualization and analysis. There is an increasing number of applications (from GIS) who ask for interpreted data, which allow the application to distinguish between important and non-important parts of urban data.

Although some approaches for building reconstruction have been presented in the last years, only a few approaches are able to use different sensor types, mainly as different parts of a specific work flow, e. g. Haala and Brenner (Haala and Brenner, 1997) starting from map data adding laser-scanner data they reconstruct buildings. Others just use data of one sensor type (Brunn and Weidner, 1997, Haala and Brenner, 1997, Vosselmann, 1999, Baillard et al., 1999, Moons et al., 1998). Fischer et. al. (Fischer et al., 1999) have described on a conceptual level a tower of feasible algorithms which could be used for the reconstruction of buildings in general. Multi sensorial reconstruction and interpretation is important, because different data from different sensor types can support the reconstruction result from different aspects. Only with sensor fusion it becomes possible to use the new developed aerial sensors like laser-scanners, digital cameras or three-lines-cameras together. In this paper, an algorithm for the use of different sensor types is described. All sensor types are handled in an equal manner, no one dominates the algorithm.

In most of the published algorithms for building reconstruction interpretation is done only as a side step. The interpretation is mostly done by classifying the type of geometric structure of the reconstruction. Mostly information of classifications of the neighborhood is not used. Except Lang (Lang, 1999) achieves some semantic interpretation, considering geometric classifications from the neighborhood.

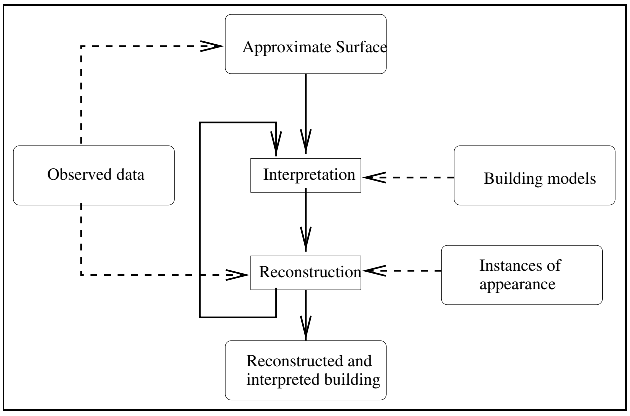


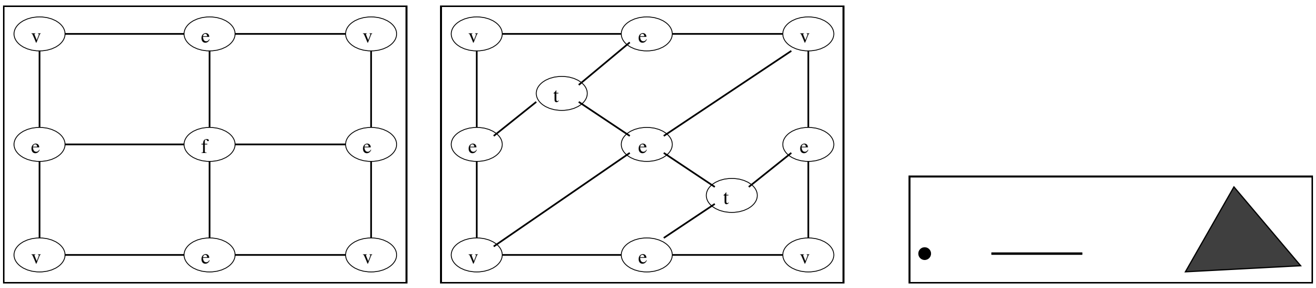
Figure 1: Work flow of the combined interpretation and analysis.

1.2 Overview

Here we assume that the problem of detecting buildings in an arbitrary 2D data set is already solved. A detection procedure based on bayesian nets can be found in (Brunn et al., 1998). For each building the detection algorithm gives a bounding polygon which encloses the complete building.

Starting from the subparts of several data sets the work flow is a follows (cf. fig. 1): From the selected part of the data of some not necessary all sensor types we generate an approximate description of the building, which is a graph representation and contains the topology of its surface. We choose the topology of the surface graph as the base representation of the building because the topology is independent of the used sensor type (cf. sec. 2). Each graph element is connected to a set of appearance models (e. g. geometric coordinates or attributes or radiometric attributes). Further steps of reconstruction and interpretation optimize the graph representation including the attributes. Ideally interpretation and reconstruction should be done in a common step. We approximate the common step by iterating separate steps of interpretation (cf. sec. 3) and reconstruction (cf. sec. 4).

The paper closes with some conclusions and an outlook on further research (cf. sec. 6).



(a) The polymorphic graph of one side of a cube (v=vertex, e=edge, f=face).

(b) Polymorphic graph of the simplex representation of one side of a cube (v=vertex, e=edge, t=triangle)

(c) Simplices are the basis elements of the simplicial complexes (v=vertex, e= edge and t=triangle).

Figure 2: Representations of topology.

2 MODELING BUILDINGS BY THEIR TOPOLOGY

2.1 CW–complexes

We introduce CW–complexes (Jänich, 1994) as a representation of the topology of buildings. The building surface can be divided in corner, bounding edge and faces. The corner are called 0-cells, the edges 1-cells and the faces 2-cells (0, 1 and 2 are the degree of the cells). They are handled as open sets, which means that the union of all points, edges and faces yields the complete surface, the intersection of all pairs of elements is empty.

We define a graph whose vertices are all cells of the three types and whose edges show neighborhood relations: Two cells are neighbored if the cell of lower degree is the border of the cell of higher degree. The neighborhood relation is symmetriced. We call the resulting graph a polymorphic graph (Fuchs and Förstner, 1995). In figure 2(a) the polymorphic graph of one side of a cube is shown. The graph of the complete cube consists of eight 0-cells, twelve 1-cells and six 2-cells.

2.2 Simplicial complexes

CW-complexes are a very general method to describe topology, esp. they are not limited to a number of cell types. Simplicial complexes can be viewed as a specialization of the CW-complexes, where 2-cells are triangles. The basis elements of this representation are called simplices. They are shown in figure 2(c) up to order two. On the one hand the simplification of the complexity of the topological boundary polygon to just three sides limits the complexity of the 2-cells. On the other hand it provides an easy access to the shape of the 2-cells. Analogously to the CW-complex we define a polymorphic graph on the simplices. Figure 2(b) shows the polymorphic graph of one side of a cube in simplex representation. Simplicial complexes are widely used in the approximation of triangulated surfaces: e. g. Halmer et al. (Halmer et al., 1996) use simplicial complexes for the approximation of surface models, but they stick to geometric triangles of a triangulation of the surface. In the following we focus on simplicial complexes.

3 INTERPRETATION

In our context interpretation means classification of the simplices of the building representation by complexes. In the two step scenario of separate interpretation and reconstruction the interpretation links the observations with the reconstruction (cf. fig. 3). Knowing the likelihood functions the classification can make use of the observations.

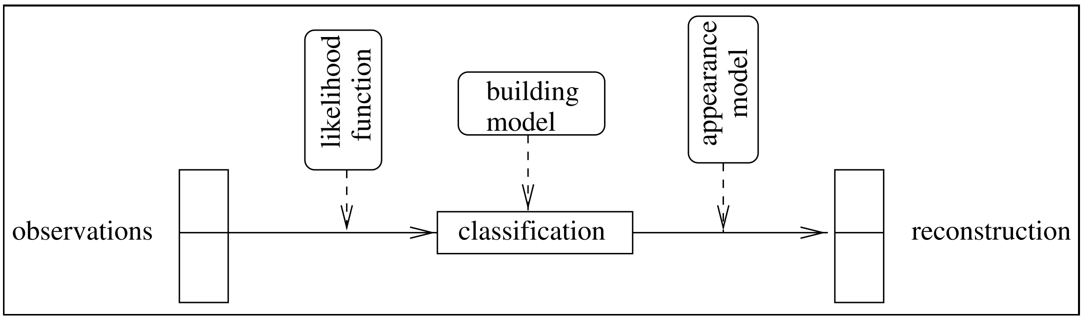


Figure 3: Principle of classification and reconstruction paradigm.

3.1 Principle

The interpretation is done by classification using statistical models, which need to be known a priori or automatically learned (cf. sec. 3.3). To find a final classification means to search for a set of classes which have a maximal probability. To reduce the complexity of this problem, which in general is exponential in the number of possible classes for each random variable, we assume only local dependencies between neighboring random variables. Therefore we only need local statistical models.

We establish a *Coupled Markov-Random-Field* (CRF) (Li, 1995), which consists of three random-fields, one for each simplex type. This is equivalent to associate a random variable s to each node of the polymorphic graph, which represent vertices, edges and triangles,

$$v, e, t \rightarrow s$$

The classes for possible classifications are explained in the next section.

Inside the CRF each random variable s_i is classified. This is achieved by finding that class which leads to the maximal probability for the random variable s_i.

$$\hat{s}_i = \arg\max_{s_i} P(s_i)$$

The probability distribution $P(s_i)$ is deduced using Bayes' Theorem

$$P(s_i|\boldsymbol{y}_i) \propto P(\boldsymbol{y}_i|s_i)P(s_i)$$

with $\boldsymbol{y}_i$ being a vector of observations. $P(\boldsymbol{y}_i|s_i)$ is called likelihood function and $P(s_i)$ prior distribution. Thus we are able to connect different probability distributions. Here we use the Bayes' Theorem by introducing background knowledge ∂s_i about the random variables in some neighborhood of s_i (markovinity) (Koch and Schmidt, 1994)

$$P(s_i|\boldsymbol{y}_i, \partial s_i) \propto P(\boldsymbol{y}_i|s_i, \partial s_i)\; P(s_i|\partial s_i).$$

Further on we assume that the likelihood function only depends on the random variable s_i itself, not on the neighboring random variables

$$P(\boldsymbol{y}_i|s_i, \partial s_i) = P(\boldsymbol{y}_i|s_i).$$

Therefore eq. 1 changes to

$$P(s_i|\boldsymbol{y}_i, \partial s_i) \propto P(\boldsymbol{y}_i|s_i)\; P(s_i|\partial s_i) \tag{1}$$

where the distribution $P(s_i|\partial s_i)$ inherits the information about the building model.

The theory of Markov-Random-Fields states that the local classification (cf. eq. 1) leads to a maximal probability of the complete random field. We calculate the probability of the classification of the complete random field by

$$\hat{\boldsymbol{s}} = \max_L P(\boldsymbol{s}|\boldsymbol{y}) = \prod_i \max_l P(s_i|\boldsymbol{y}_i, \partial s_i),$$

because in this application each subset of random variables is considered only once for local classification (Koch and Schmidt, 1994). Therefore, if the building type is unknown and from a set of building models $\mathcal{M}$ with $\mathcal{M} = \{M_i | i \in \{1, \ldots, n_M\}\}$, the maximization can be done for each building type. Then the maximum of all probabilities of the different building types gives the type of the complete building

$$\hat{\boldsymbol{M}} = \max_M P(\hat{\boldsymbol{s}}(M)|\boldsymbol{y}).$$

3.2 The building model

We use simplicial complexes, which consist of 0, 1 and 2-simplices, to represent the topology of the surface of buildings. Each simplex is associated with some appearance attributes, which could be e. g. geometric or radiometric. The simplices should be classified. Therefore we define classes for each simplex type.

We look at the classes of the simplices of the real building parts from different views (cf. fig. 4): at first we describe a semantic model which will be generalized to a geometric[1]. A semantic class scheme could consist of the following classes:

- 0-simplices: eaves corner point, ridge corner point, ground-plane corner point, eaves point (point on the eaves, not corner point), ridge point, ground-plane point, point on a wall, point on the roof and point outside the building in the ground-plane
- 1-simplices: eaves edge, ridge edge, ground-plane edge, corner edge (mostly nearly vertical), edge in the roof, edge in a wall, edge outside the building in the ground-plane
- 2-simplices: face in a wall, face in a roof, face in the ground-plane

In this paper we generalize to the following geometric model:

- 0-simplices: corner point (CP) (on the border of a least three building planes), edge point (EP) (on the border of two building planes), face point (FP) (inside one plane of the building)
- 1-simplices: breakline (BL) (intersection of two building planes), face edge (FE) (inside one building plane)
- 2-simplices: vertical (V), oblique (O) and horizontal (H) face

The definition of local neighborhoods enables us to do local classification. We define different kind of neighborhoods for the three simplex types which are shown in fig. 5. Different building types are modeled by sets of conditional probabilities which coincide with the neighborhoods. For each building type the following chosen conditional probabilities have to known are priori or have to be learned (cf. sec. 3.3)[2]:

[1] We use the notion "generalized" because the capability to distinguish between different objects in the geometric modeling is less than in the semantic modeling. Also the number of classes is reduced in the geometric model. [2] Notation: # means the "number of".

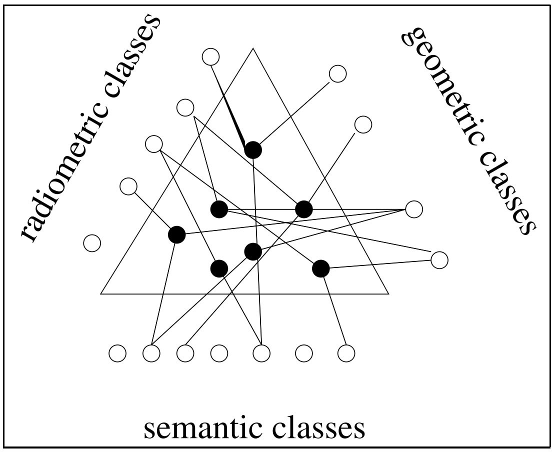


Figure 4: Three different views on the real objects result in three different sets of object groups.

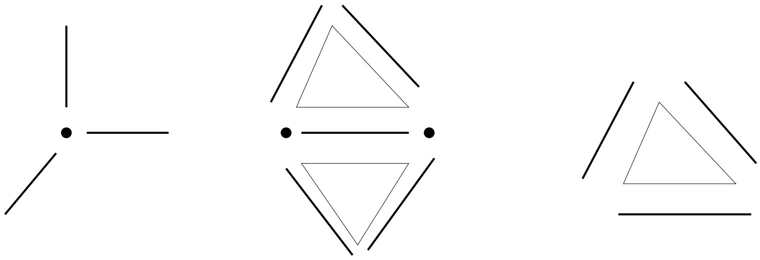


Figure 5: Neighborhood system for vertices, edges and triangles in the Markov-Random-Field.

- Conditional probability for 0-simplex classification: the classification of a 0-simplex depends on the number of breaklines and the number of corner points in the neighborhood

$$P_M(l_v | \#BL(v), \#CP(v)).$$

With the assumption that the conditional probabilities can be separated,

$$P_M(l_v | \#BL(v), \#CP(v)) \quad = \quad P_M(l_v | \#BL(v)) \cdot P_M(l_v | \#CP(v)) \tag{2}$$

follows.

- Conditional probability for edge classification: the classification of a 1-simplex depends on the classifications of the neighboring faces, neighboring vertices and the number of other breaklines bounding the neighboring faces

$$P_M(l_e \quad | \quad \{l_{f_1}(e), l_{f_2}(e)\}, \{l_{p_1}(e), l_{p_2}(e)\}, \#BL(e))$$

We assume separability of the conditional probability:

$$\begin{aligned} P_M(l_e \quad | \quad & \{l_{f_1}(e), l_{f_2}(e)\}, \{l_{p_1}(e), l_{p_2}(e)\}, \#BL(e)) \\ = \quad & P_M(l_e | \{l_{f_1}(e), l_{f_2}(e)\}, \{l_{p_1}(e), l_{p_2}(e)\}) \cdot P_M(l_e | \#BL(e)) \end{aligned} \tag{3}$$

- Conditional probability for triangle (face) classification: The classification of 2-simplices depends on the number of breaklines bounding the 2-simplex.

$$P_M(l_f | \#BL(f)) \tag{4}$$

We assume the likelihood function as independent of the building type. Therefore, we just have to define one set of likelihood function for each combination of each simplex with each data type or, of course, a joint likelihood function for all data types, which is mostly not practicable. These functions $P(\boldsymbol{y}|\boldsymbol{s})$ transfer the observations $\boldsymbol{y}$ into probabilities for all classes of each random variable. Assuming a normal distribution for a observation vector $\boldsymbol{y}$, a likelihood may have the following form

$$P(\boldsymbol{y}|\boldsymbol{s}) = (2\pi)^{-n/2} (\det \boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_s)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_s)\right)$$

The vector $\boldsymbol{\beta}$ is a class specific parameter vector, $\boldsymbol{X}$ a matrix of linear coefficients and $\boldsymbol{\Sigma}$ the matrix of covariances of the observations.

3.3 Learning

The models of the different building types M are learned automatically. For this purpose a set of interactively classified simplicial complexes could be evaluated, yielding relative frequencies or empirical probabilities. Here another approach is used: we learn building models from just one representative classified example for each building type. Of course the deduced probabilities are less precise, however it takes less effort.

3.4 Classification

There is large variety of algorithms for optimizing MRFs available. We choose the Iterated-Conditional-Modes (ICM) (Li, 1995), because firstly we get from a initial likelihood estimate - just using the likelihood function - a good initial classification. Secondly the influence of a local classification on not neighboring random variables is quite low. During some iterations of classification we find a optimal solution, iterating until only minor changes occur in the classification of random variables. The number of interactions depends on the amount of noise in the data.

4 RECONSTRUCTION

After the interpretation step the appearance of the building is corrected (cf. fig. 3). The correction takes the classification result into account and uses local appearance models for each simplex. In case of geometric attributes, the coordinates of the corner of building description are corrected according to the geometric classification.

We use *robust Least-Squares-Estimators* which enable us to handle outliers in the data and wrong classifications of the previous step.

5 EXAMPLE

We use a laserscanner dataset to demonstrate the feasibility of the approach. Figure 6 shows a dense grid-based laserscanner data set. We want to reconstruct and interpret a flat roof building. Therefore we learn the conditional probabilities of an chosen example of a flat roof building (FR). The learning yielded the following conditional probabilities:

- Vertices (cf. eq. 2):

$$P_{FR}\left(l_v|\#BL(v)\right) = \left(\left(P_{FR}\left(l_v = j|\#BL(v) = k\right)\right)_{jk}\right) = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_{FR}\left(l_v|\#CP(v)\right) = \left(\left(P_{FR}(l_v = j|\#CP(v) = l)\right)_{jl}\right) = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

with

$$j \in \{FP, EP, CP\} \qquad \text{and} \qquad k, l \in \{0, 1, 2, \geq 3\}$$

- Edges (cf. eq. 3):

$$\begin{aligned} P_{FR}\left(l_e\right. & \left| \quad \{l_{p_1}(e), l_{p_2}(e)\}, \{l_{f_1}(e), l_{f_2}(e)\}\right) \\ &= \left(\left(P_{FR}((l_e = BL, l_v = FE)|\{l_{p_1}(e), l_{p_2}(e)\} = l, \{l_{f_1}(e), l_{f_2}(e)\} = k\right)\right)_{lk}) \\ &= \begin{pmatrix} (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (1, 0) & (0.5, 0.5) & (0.5, 0.5) \\ (0.5, 0.5) & (0.5, 0.5) & (1, 0) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) \\ (0, 1) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) \\ (0, 1) & (0.5, 0.5) & (1, 0) & (0, 1) & (0.5, 0.5) & (0.5, 0.5) \\ (0, 1) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) \\ (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) & (0.5, 0.5) \end{pmatrix} \end{aligned}$$

$$P_{FR}\left(l_e|\#BL(e)\right) = \left(\left(P_{FR}((l_e = BL, l_e = FE)|\#BL(e) = l)\right)_l\right) = \begin{pmatrix} (1, 0) \\ (0.44, 0.56) \\ (0.17, 0.83) \\ (0, 1) \\ (0.5, 0.5) \end{pmatrix}$$

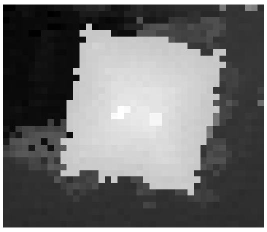


Figure 6: Part of a digital surface model interpolated from laserscanner data with grid width $0.6 \times 0.6m^2$ (Copyright by TopoSys, Ravensburg). The heights are coded in greyvalues.

with

$$\begin{aligned} j &\in \{\{CP,CP\},\{CP,EP\},\{CP,FP\},\{EP,EP\},\{EP,FP\},\{FP,FP\}\}, \\ k &\in \{\{H,H\},\{H,O\},\{H,V\},\{V,V\},\{V,O\},\{O,O\}\} \qquad \text{and} \qquad l \in \{0,1,2,3,4\} \end{aligned}$$

- Triangles (faces) (cf. eq. 4):

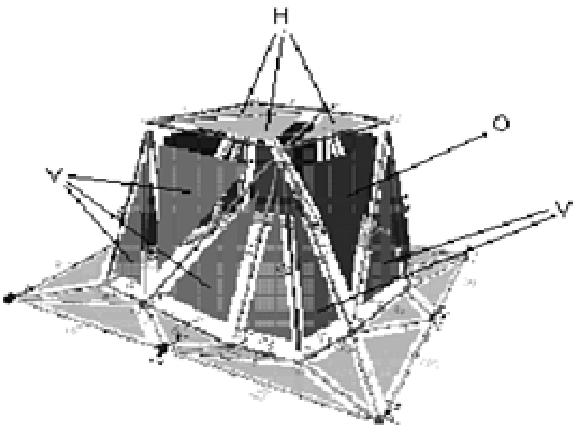
$$P_{FR}\left(l_f|\#BL(f)\right) = \left((P_{FR}(l_f = j|\#BL(f) = k))_{jk}\right) = \begin{pmatrix} 0.33 & 0.53 & 0.13 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.66 & 0.33 & 0 \end{pmatrix}$$

with

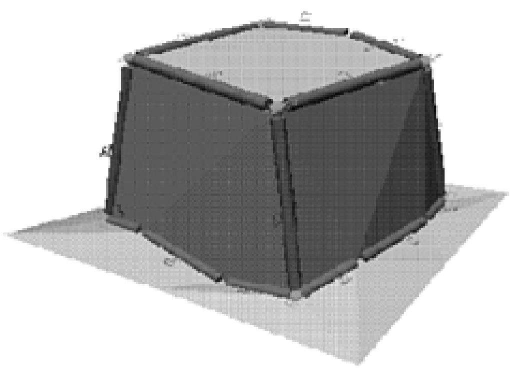
$$j \in \{H,O,V\}, k \in \{0,1,2,3,4\}$$

The likelihood function is defined a priori. We classify introducing the approximate surface and not on the original data as observations, to reduce the complexity of handling observations in the likelihood function. We choose the following deduced observations:

- for 0-simplices (vertices): the number of normals of touching triangles. For each number of possible amounts of normals a description-length (Rissanen, 1987) is calculated and transformed into a probability.
- for 1-simplices (edges): the number of normals of touching triangles (cf. 0-simplices).
- for 2-simplices (triangle): the difference angle to the vertical axis.



(a) Approximated building surface with the result of the initial classification coded in intensities.



(b) Result of the interpretation and reconstruction. Only those simplices are shown which are necessary for the geometric description.

Figure 7: 3D views: Each greyvalue of vertices, edges and triangles represents a classification.

From a preprocessing, which is based on the feature extraction (Fuchs and Förstner, 1995) of the DSM, we get the geometry of the approximate building description (cf. fig. 7(a)). The a priori likelihood estimate leads to the shown simplex classification. The misclassification of a least one face is obvious. The iterative process of classification and reconstruction results in the building shown in fig. 7(b). Only essential simplices are shown in the visualization. Misclassifications and the geometry of the surface have been corrected automatically in this second step.

We found that the error tolerance of the approach is quite high, which is also known from MRFs in general. Correction of single misclassification could be done, but regions of false classifications cannot be corrected due to the defined low sizes of the neighborhoods.

6 CONCLUSIONS AND OUTLOOK

This article presents a new approach for automatic analysis of multi sensorial data. The integration of interpretation and reconstruction in a common algorithm was a major goal. A new structure for the representation of buildings based on their topology was introduced. The example shows the correction capability of the local classification and local reconstruction.

In future research, the integration of changes in topology during the classification has to done. The implementation and modeling should be extended to CW-complexes. For an empirical evaluation of the approach, additional models, large datasets and additional sensor types have to be covered.

Acknowledgments: This work has been founded by the European Community as part of the ESPRIT project "Image Processing for Automatic Cartographic Tools (IMPACT)", No. 20.243.

REFERENCES

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic line matching and 3d reconstruction of buildings from multiple views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5.

Brunn, A. and Weidner, U., 1997. Extracting buildings from digital surface models. In: 3D Reconstruction and Modeling of Topographic Objects, Vol. 32, IAPRS.

Brunn, A., Gülch, E., Lang, F. and Förstner, W., 1998. A hybrid concept for 3D building acquisition. JPRS 53(1), pp. 119–128.

Fischer, A., Kolbe, T. H. and Lang, F., 1999. On the use of geometric and semantic models for component-based building reconstruction. In: SMATI 99, pp. 101–120.

Fuchs, C. and Förstner, W., 1995. Polymorphic grouping for image segmentation. In: 5th International Conference on Computer Vision, Proceedings, pp. 175–182.

Haala, N. and Brenner, C., 1997. Interpretation of urban surface models using 2D building information. In: A. Gruen, E. Baltsavias and O. Henricsson (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag, Basel, Switzerland, pp. 213–223.

Halmer, A., Heitzinger, D. and Kager, H., 1996. 3d-surface modelling with basic topologic elements. International Archives of Photogrammetry and Remote Sensing XXXI(B4), pp. 407–412.

Jänich, K., 1994. Topologie. 4 edn, Springer, Berlin.

Koch, K. and Schmidt, M., 1994. Deterministische und stochastische Signale. Dümmler Bonn.

Lang, F., 1999. Geometrische und semantische Rekonstruktion von Gebäuden durch Ableitung von 3D-Gebäudeecken. Berichte aus der Informatik, Shaker Verlag. 184 Seiten, 60 Abbildungen, ISBN 3-8265-6619-X.

Li, S., 1995. Markov Random Field Modeling in Computer Vision. Springer-Verlag Tokyo.

Moons, T., Frere, D., Vandekerckhove, J. and Gool, L. V., 1998. Automatic modelling and 3d reconstruction of urban house roofs from high resolution aerial imagery. Computer Vision – ECCV'98, Freiburg pp. 410–425.

Rissanen, I., 1987. Minimum description length principle. Encyclopedia of Statistical Sciences.

Vosselmann, G., 1999. Building reconstruction using planar faces in very high density height data. In: Automatic Extraction of GIS objects from Digital Imagery, ISPRS.