BUILDING RECONSTRUCTION: THE DILEMMA OF GENERIC VERSUS SPECIFIC MODELS

Stephan Heuel and Thomas H. Kolbe
Institute for Photogrammetry, Institute for Cartography & Geoinformation
University of Bonn
e-mail: heuel@ipb.uni-bonn.de, kolbe@ikg.uni-bonn.de

ABSTRACT

Automatic building extraction from images is a particularly hard object recognition problem, because both the image data and the models to be reconstructed reveal a high complexity. Whereas models have to be generic in order to fit most of the observable different building shapes they also have to be building specific to discriminate buildings from other objects in the images. This situation describes the typical dilemma that all building recognition system have to cope with.

In this article we present and discuss two approaches for automatic building reconstruction that were developed at the University of Bonn during the last eight years. It is shown how different AI methods were employed to solve the numerous problems concerning modeling, inference and uncertain reasoning, matching, and evaluation. Both approaches have been implemented and were successfully applied to real data. Due to complementary limitations, the integration of both would be desirable in order to develop a more comprehensive solution. Unfortunately, a number of open questions still have to be answered which are discussed at the end of the article.

INTRODUCTION

The extraction of buildings from digital images is a difficult object recognition problem that has been addressed in Computer Science since the 70's in the field of Artificial Intelligence. Photogrammetry as a subfield of geodesy deals with exact measurements from images since the invention of photography. Especially aerial images were and are common image data for photogrammetrists. Nowadays researchers of both fields work on finding solutions for the automatic detection of a building, its interpretation and classification, and the reconstruction of its shape.

Most applications for reconstructed buildings are related to the increasing use of 3D city models for tasks related to telecommunication, transport planning, environmental investigations, and 3D geoinformation systems. 3D city models are going to play an important role for virtual or outdoor augmented reality systems: for example, the projection of real buildings into a virtual scene clearly needs a sound and precise 3D description of those real buildings. Similarly, the 3D description of existing buildings is a prerequisite for the localization and the projection of background information into the field of vision in outdoor augmented reality systems.

3D city models describe an urban area up to a specified geometric precision and interpretational detail depending on application specific needs. Telecommunication companies are particularly interested in the height information of buildings, so they can optimize the positioning of their transmitters with respect to maximum coverage. Roofer companies are interested in the exact shape of building roofs so that they can start their planning phase without driving to the client and measure important roof informations. City planners want to have access to building classifications like one family houses or buildings with attached garages.

In the past, different solutions for building reconstruction from aerial images have been developed, though not in the ambitious AI sense: the first approach is completely manual and independently measures 3D points with a digital photogrammetric station. From these points, one can construct with some effort a 3D description. To reduce the manual work of the first solution, so-called semi-automatic systems came up which allow the operator to manually extract not only points, but complete building parts, e.g. by interacting with a wireframe building model which is continously projected in the image, see [GML99].

In contrast, truely automatic systems aim to completely remove the operator's work during the extraction process. Here, it is the task of the computer to obtain an appropriate 3D model of a building from a matrix of numbers (the aerial images). Currently these systems are only research prototypes; to our knowledge, up to now no fully automatic system is flexible and reliable enough to be used regularly within an industrial process.

So why is it tough to do automatic building extraction from a set of digital images? Compare a satellite image with 1m resolution covering a rural area in the U.S. with an aerial image with 10cm resolution covering the city of London. Because of the qualitative differences of the datasets it is hard to think of one universal algorithm which covers both cases and will result in two adequate (i.e. application dependent) site descriptions. In general, the main reasons for the high complexity are:

- the simple raster representation containing 2D information has to be converted into a complex and application-dependent representation of 3D information;
- buildings usually show a high degree of variability with respect to shape, size and colors;
- beside buildings the image contains much more information about terrain, vegetation or other man-

made objects which has to be discriminated from the buildings;

- last but not least, the observability of the buildings is not guaranteed due to poor image quality, image resolution or occlusions.

The diverse automatic extraction systems can be categorized by different criteria, for an overview see [May99]. One category could be the type of building model, or more general the type of the scene model: a quite generic description would be a polyhedron containing straight line segments between two points, and planar faces surrounded by line segments. This description would fit to a high number of actual buildings. On the other hand a polyhedron does not contain any building-specific knowledge; a parameterized saddle-roof building for example has four parameters, namely length, width, height and roof-height. This parameter model is quite specific and can be directly related to a building.

As was explicated above, the required degree of specific building modeling varies with the type of application, for example, telecommunication companies do not need building details or classifications. Nevertheless, the question of which model to take within the reconstruction processes not only depends on the application: an approach based on a pure generic model cannot succeed, since the system would not be able to relate its results to buildings, therefore cannot distinguish between a building and a non-building. Being too specific on the other hand has the disadvantage of restricting the number of buildings which the system is capable to extract.

Questions that arises in this context are: at which time in the extraction process do we have to include what kind of building-specific knowledge? How long can we survive with a rather general concept of the scene?

In this article we want to introduce two approaches which have been developed at the University of Bonn in the last years within the DFG research project "Semantic Modelling". One is based on a generic polyhedral model, the other one is based on a more specific building part-model. We will roughly describe the two approaches and highlight important design decisions. We will also point out several AI related techniques which were used within each system. Then we will compare these aproaches according to their building models and discuss the open question of choosing an appropriate model.

POLYHEDRAL RECONSTRUCTION

We first want to present an example of a rather generic system used for building reconstruction, cf. [HLF00]. It models the building as a polyhedron, containing points, straight lines ending at two points and planar surfaces surrounded by straight lines. Therefore we assume a purely geometric description of a building.

The overall strategy contains a transition from pixel to 2D symbols, from 2D symbols to 3D aggregates, so-called corners, and from 3D aggregates to polyhedral surfaces, see figure 1.

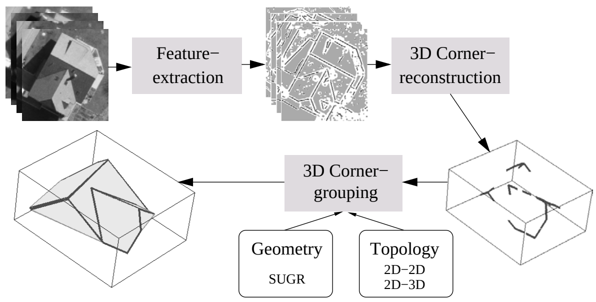

Figure 1: *Overview of polyhedral reconstruction from digital images (top left) to polyhedral surfaces (bottom left).*

The first step in the reconstruction process is the *extraction of symbolic features* from a digital image, resp. a set of digital images. We use the feature extraction FEX developed at the IPB [Fuc98], which generates points, straight lines and homogeneous areas in each image. Additionally, FEX can also extract the topological (neighborhood) relations between these three feature types. As a result we obtain a so-called feature adjacency graph where the nodes are identified with the geometric features and the arcs with the neighborhood relations. FEX assumes a rather general scene model with piecewise smooth surfaces, which includes polyhedral objects.

One important property of every feature extraction from digital images is the inherent uncertainty of the result: since the image pixels are only uncertain radiometric observations of a complex scene, the resulting features are not only uncertain in shape, but can also be plainly wrong or missing. This makes the search for the correct result in the subsequent steps much harder.

In the second stage, one has to find correspondences between the independently extracted features, i.e. points, lines and image areas, in each image. This is done by a *3D corner extraction* system developed by Felicitas Lang, cf. [Lan99]. For this task, we assume that a 3D corner in the scene is an aggregate of basic 3D features, namely of one 3D point, two or more half-lines starting at this point and one or more half planes defined by two half lines. It is important to note that in our work, this corner is not interpreted yet, i.e. at this point there is no relationship established between the extracted corners and buildings. The correspondences are established by identifying 2D corners in the images which could be projections of one unique 3D corner and are used for the estimation of the unknown 3D corner.

Having a set of 3D corners extracted, the next step is to *group the extracted 3D corners* to polyhedral surfaces. We want to introduce an approach which tries to group the corners to polyhedral surfaces which in turn have to be grouped to parts of polyhedral objects[1]

The goal of the following reasoning process is to identify and connect independently extracted 3D corners belonging to the same polyhedral surface. It is based on two steps: first we want to identify corner sets according to their topological properties, then we want to verify them on their geometric validity.

In the first reasoning step, we establish the possible connection of two 3D corners by their *topological* relationship. For example it is simple to state, that two corners belong to the same surface if they have this very surface as their neighbor. Unfortunately we do not have the 3D topology. But we can go back to the neighborhood relationships in the images which have been extracted by the feature extraction. From there we can relate the projection of the two corners and and "back-project" the topological 2D relation in 3D, see 2. Topological relations for polyhedra can have all combinations of neighborhood or incidence of lines, points or regions. We select the ones which are most discriminative for our task and obtain a first set of hypotheses of corners belonging to the same surface.

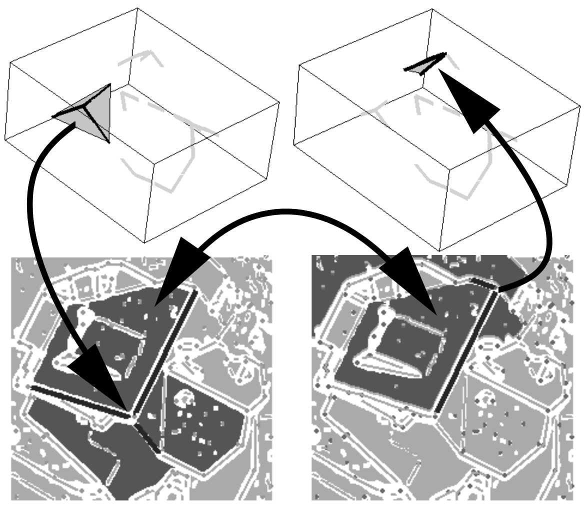

Figure 2: *Establishing a 3D topological relation (top row)by going back to the 2D neighborhood relation in the images (bottom row), see text for details.*

The geometrical reasoning is the second step to verify the initial topological guess. For example, we could verify the identity of two planes which are induced by two corners, or the incidence of a line and a plane which are supposed to belong to the same corner.

Here we are confronted with the problem of testing crisp relations like identity, coplanarity or incidence on uncertain and imprecise data, as the 3D corners are computed based imprecise data. Therefore in general geometric relations are never exactly satisfied. To overcome this problem one has to know the precision of the 3D corners, which can be computed since the errors of our data are propagated throughout the feature and 3D corner extraction. We have developed a method for statistically testing these kind of uncertain relations by formulating statistical hypotheses tests with elements from projective geometry, see [FBH00]; thus we only need a data-independent significance level to test the 13 relationships listed in 1. Using these tests on points, planes and lines in 3D we can infer the geometrical relationships of 3D corners, too.

	Point	Line	Plane
Point	$\in$	$\in$	$\in$
Line		$\cap \neq \emptyset$, $\equiv$, $\|$, $\perp$	$\in$, $\|$, $\perp$
Plane			$\equiv$, $\|$, $\perp$

Table 1: *13 relationships between points, lines and planes in 3D: incidence (denoted by $\in, \cap \neq \emptyset$), equality ($\equiv$), orthogonality ($\perp$) and parallelity relations ($\|$). Our system can perform hypotheses tests on these relations based on imprecise data, see text for details.*

We combine topological and geometrical tests to a step-wise reasoning, where we first test a grouping hypotheses by topological tests, which is only a look-up into a topological database, and then test the geometry, which gives us very discriminative results. The selection and the order of the reasoning steps are currently done manually by the system designer; automatically finding an optimal reasoning algorithm based on previous data would be very interesting.

The current output of this process is a set of polyhedral surfaces with known common line segments and points. We have tested our system on synthetic datasets and a real dataset with about 60 polyhedral surfaces and we are currently extending the approach and test them on more data. The surfaces have not been processed by an interpretation module yet thus no building-specific knowledge has been introduced other than the rather generic polyhedral model.

COMPONENT-BASED RECONSTRUCTION

The second approach that we present in this article is based on a hierarchical, component-based building model. Whereas building shapes are also modeled using polyhedra their geometry — in contrast to the previous approach — is restricted to building specific extents. Buildings are composed of volumetric, typed and parameterized building parts in a constrained but generic way.

This concept is joint work with Felicitas Lang, André Fischer, Volker Steinhage, Wolfgang Förstner, Lutz Plümer, and Armin B. Cremers.

[1] From aerial images you are not likely to see all parts of buildings. To extrapolate to invisible parts of the buildings requires more explicit building models.

Building and image model

The building model consists of a four level aggregation hierarchy representing the different semantic abstractions. The primitives on each level are further specialized into subclasses. Building specific restrictions are propagated top-down from higher to lower levels. To allow a tight coupling of 2D and 3D reconstruction processes the primitives on each level are coherently modeled in 2D and 3D.

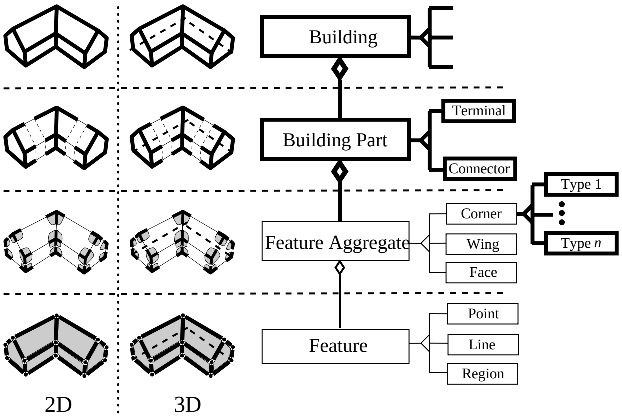

Figure 3: *Component-based building model: The concept 'building' is explicitly represented by the two highest levels and by building specific subclasses and attributes on lower levels. This model extends the general polyhedral model by the bold drawn entities.*

The highest level represents complete volumetric buildings. These are composed of building parts, which are 3D objects with one or more "open ends". Building parts with one open end are called terminators, those having more than one are called connectors. An EBNF grammar specifies aggregation rules that describe which building parts fit to others and the way building parts can be merged (c.f. [BKL+95]). Building parts are parameterized by building specific measures like width, height, roof height etc. The coordinates of all corner points are expressed in terms of these parameters. Thus, the geometric shape is completely derived from these parameters. Currently, only building parts with polyhedral shapes have been modeled. Further building knowledge is introduced by constraints, restricting single parameter values and parameter ratios to building specific values.

On the next lower level building parts are decomposed into groups of image features, namely corners, wings, and faces. These feature aggregates consist of a single image feature of one class and its incident features of the other classes. For example, corners consist of a point and the neighboured lines and regions. Analogously to the first approach we focus on corners, because they have shown to be the most robust feature configuration with respect to observability and image segmentation errors. In contrast to the first approach we further distinguish corners into different subclasses. Only corners that actually appear as parts of the chosen set of building parts are modeled (see figure 4). Currently, our model differenciates 28 corner types. Each type is characterized by its number of incident lines and regions, the topological and geometrical relations between them, and its qualitative geometric orientation (horizontal, vertical, oblique) of the incident lines (c.f. [Lan99]).

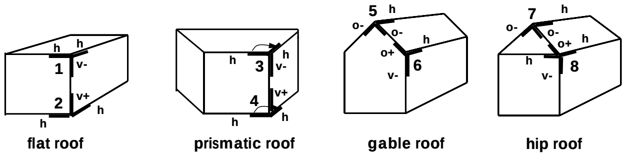

Figure 4: *Eight of the 28 different building specific corner types. Lines are labeled according to their qualitative geometric orientation.*

The lowest level consists of the three attributed feature classes points, lines, and regions. They establish the link to the symbolic image description that is derived from the original raster images by the initial image segmentation process.

Strategy

Aerial images have typical sizes of 10000×10000 pixels or more and can contain up to several hundred buildings. In a divide-and-conquer step the aerial image is segmented into regions of interest (ROI). ROI are rectangular image clips and in the ideal case contain only one building or a building block. For aerial images showing rural and suburban scenes these regions are determined by the identification of local maxima in the digital surface model of the scene.

Since every building is reconstructed independently, the following six consecutive steps can be processed in parallel for every ROI:

1. Image segmentation and feature extraction,
2. reconstruction of building specific 3D corners,
3. generation of 3D building hypotheses,
4. computation of their possible 2D projections,
5. identification of the most probable hypothesis,
6. and final parameter estimation.

In the first step the aerial images are segmented using the FEX system (see above, c.f. [Fuc98]). Based on the resulting symbolic image description the system tries to identify and match typical corner structures in the stereoscopic images. This is done in the same way that was already explained for the polyhedral reconstruction but with one difference: only those corners are 3D reconstructed which are valid instances of the 28 corner classes of the building model. The corner classification is based on the qualitative line orientations, different geometrical and topological relations and explicitly refers to their statistical characteristics (details are given in [Lan99, FKL+98]).

In the third step 3D building hypotheses are generated that are able to explain (most of) the observed 3D corners. In an initial indexing step every corner is assigned all building parts of which it is a component. The elimination of possible building parts for the corners and the aggregation to a complete, closed building hypothesis is done successively: if the 'open ends' of two neighbored building parts have the same type and opposite direction they are merged and their common parameters are unified. This step is iterated until either the aggregate has no more 'open ends' or all 3D corners are included in the aggregated hypothesis. If in the latter case the building hypothesis still contains 'open ends', they are closed by merging them with fitting terminal building parts. Figure 5

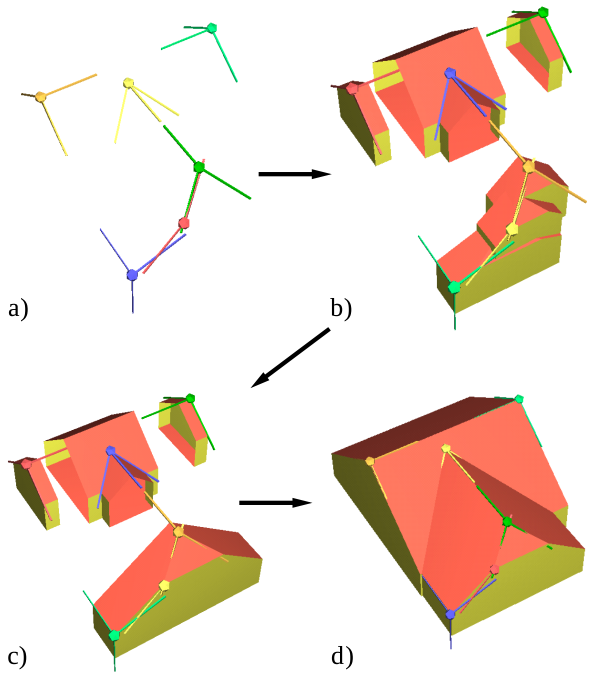

Figure 5: *A building hypothesis has to explain the set of corner observations (a). It is constructed by selection (b), merging, and parameter unification (c),(d) of the appropriate 3D building parts.*

shows an example for the aggregation of a T-shaped building hypothesis. Please note, that some building parameters may remain undetermined. In this example, the building height is unknown, because no corners were observed at ground level. In fact, the number of reconstructed 3D corners in many cases is not sufficient to completely determine the building's shape and extent. Furthermore, the set of 3D corners may include errorneous corners or corners of other image objects. Therefore, multiple hypotheses for the most probable corner explanations are generated (see [FKL+98]).

For the identification of the correct building hypothesis and the determination of its geometric extent all hypotheses have to be matched with the image data. In the fourth processing step the 3D hypotheses are projected back into the images. This is not trivial, because free building parameters not only lead to e.g. variable line lengths and angles but also can result in topologically different projections. To determine these projections we employ a stochastic approach: the free parameter space is sampled and only topologically different instances (i.e. 2D building views) are returned.

In the fifth step the building views are matched with the extracted image features. No simple feature based matching algorithm can be used, because it would only take feature attributes into account, but 1) lengths, sizes and angles of many lines and regions are still variable and 2) extracted image features often are fragmented. However, geometric and topologic relations have shown to be stable with respect to free parameters and segmentation errors. Therefore, relational matching is applied, where each building view is decomposed into a model graph consisting of points, lines, regions, and their interrelationships. The search for an instance of the building view in the image amounts to the identification of a subgraph isomorphism between the model graph and the huge graph of extracted image features. We apply constraint solving techniques to efficiently determine the matchings. Variables represent the model features, constraints the model relations, and the variable domains consist of the extracted image features. To cope with the uncertainty and unobservability of image features we have extended standard constraint techniques by a relaxation scheme that explicitly distinguishes between the violation and the unobservability of constraints. Constraints are weighted by an information theoretical measure expressing the similarity of a relation between model and image features in bit. These weights reflect the probabilistic properties of the different relations and are derived analytically and statistically using supervised learning.

For every building view the most probable matching with the image features is computed by maximizing the sum over all relational similarities. Due to the possible relaxation of constraints the algorithm will find a matching for every building hypothesis and finally the most probable hypothesis has to be determined. Because the matching quality is measured by the sum of the similarities of all model relations more complex building hypotheses are likely to reach higher scores than others. Thus, we cannot directly select the hypothesis with the highest matching score. The normalization of our evaluation function has been achieved by simply subtracting the model complexity, measured by its coding length in bit, from the matching score. According to the minimum description length principle this is allowed, if the model is optimally encoded and the coding length is minimal. We have developed a coding scheme for relational building hypotheses and tests with synthetic and real data have shown that is sufficient for the selection of the most probable building hypothesis. For details see [Kol00].

In the final step the parameters for the identified 3D building hypothesis are estimated in a global parameter estimation which takes into account the geometry of all assigned image features of the building hypothesis.

DISCUSSION

We have presented two different approaches to the automatic extraction of buildings from aerial images. The first one describes buildings by a polyhedral model with no other building specific knowledge. Due to its unconstrained generic shape it covers a large set of buildings, but possibly other objects in aerial images that can be represented by polyhedra. The second approach uses an explicit building model where buildings are composed of parameterized parts with a polyhedral representation. The mapping of aggregated parameterized building parts to polyhedra restrict models to building specific shapes. Because of a limited number of predefined parts not every polyhedral building can be reconstructed this way (see figure 6).

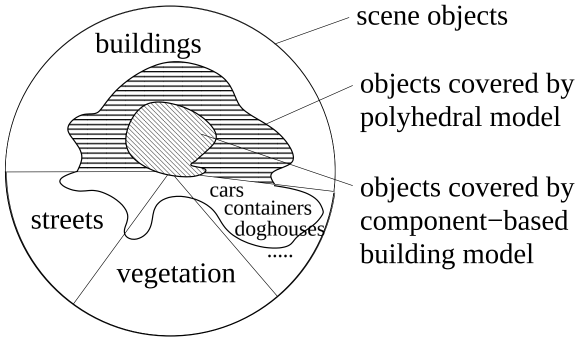

Figure 6: *Scene objects in aerial images covered by the two different modeling schemes. The polyhedral shape model covers more possible building types but also fits to non-building objects like cars, containers or even doghouses. Due to its building specific restrictions the component-based model discriminates objects better but covers a smaller subset of different possible building variants.*

Both approaches have their advantages. On the one hand, the component-based approach can cope better with incomplete observations, because it can predict unobserved parts by imposing the model. The representation of buildings in this approach is highly intensional, since most knowledge is already expressed in the model. On the other hand, the generic polyhedral approach is capable of reconstructing the complete set of buildings that can be represented by polyhedra. Here, the representation is mostly extensional.

Generalizing the above observations we have the classic dilemma of the choice between a generic, but too unspecific and a specific, but too restrictive model.

In our opinion, the success and the relevance of automatic building reconstruction systems depends on both concepts: On the one hand one wants to reach a high degree of semantic abstraction within a building model. For example, a GIS operator wants to be sure about the number of saddle roof houses in a certain area. On the other hand, one needs to be flexible with respect to the infinite number of possible building shapes: the GIS operator should rely on the fact that no buildings are missing because some uncommon architectural parts were not modeled by the system.

model	component-based	polyhedral
abstraction	**high**	low
representation	mostly intensional	mostly extensional
flexibility	low	**high**

Table 2: *Properties of the employed building models. The dilemma is that is desirable to have one system with a high degree of abstraction and a high flexibility at the same time.*

Looking at table 2 we need to bridge the gap between the highly abstracted component-based model and the highly flexible polyhedral model. This possibly can be achieved by different extensions to the systems that were presented above:

- Polyhedral model: after the reconstruction of a polyhedron it has to be interpreted in the context of a more specific building model. This means the identification and labelling of building parts in a 3D model.
- Component-based model: in cases where the verification of building hypotheses was nearly successful, one can make the assumption that there is a building that only can be extracted by a pure polyhedral reconstruction.
- One possible integration of both approaches could be the combination of parameterized building parts and generic polyhedral building parts. This can be done either by the identification of parameterized parts in polyhedral models or extending parameterized models by unconstrained polyhedral parts.

Although our two systems do not satisfy the practical requirements yet, progress has been achieved at different levels:

- The implementation of both systems has shown that it is already possible to automatically reconstruct detached houses in rural areas from aerial images.
- The problem of reasoning on uncertain 3D data has been solved by using categorical topological and statistical geometrical tests with no data-dependent thresholds.
- Constraint solving techniques have been extended to explicitly represent and cope with uncertainty and unobservability of image observations.
- Relational matching has been cast in the context of the Minimum Description Length Principle allowing the determination of the most probable model.
- A general feature extraction system was developed which not only simultaneously extracts points, line segments, and homogeneous regions, but also their neighborhood relations.

AUTHORS

Dipl. Inf. Stephan Heuel is a PhD student and scientific assistant at the Institute for Photogrammetry at the University of Bonn. He studied computer science with minor Photogrammetry. His research interests include computer vision with photogrammetric applications, especially 3D and 2D grouping, object reconstruction and projective geometry.

Dr. Thomas H. Kolbe is senior lecturer and researcher at the Institute for Cartography and Geoinformation at the University of Bonn. He studied computer science and graduated with a Ph.D. thesis on automatic building extraction. His research interests cover geoinformation systems, computer vision, constraint solving, and information theory.

REFERENCES

C. Braun, T. H. Kolbe, F. Lang, W. Schickler, V. Steinhage, A. B. Cremers, W. Förstner, and L. Plümer. Models for Photogrammetric Building Reconstruction. *Computer & Graphics*, 19(1):109–118, 1995.

W. Förstner, A. Brunn, and S. Heuel. Statistically testing uncertain geometric relations. In G. Sommer, N. Krüger, and Ch. Perwass, editors, *Mustererkennung 2000*, pages 17–26. DAGM, Springer, September 2000.

A. Fischer, T. H. Kolbe, F. Lang, A. B. Cremers, W. Förstner, L. Plümer, and V. Steinhage. Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D. *Computer Vision & Image Understanding*, 72(2), 1998.

C. Fuchs. *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung*. DGK, Bayer. Akademie der Wissenschaften, Reihe C, Heft 502, 1998.

A. Grün, E. P. Baltsavias, and O. Henricsson, editors. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*. Birkhäuser, Basel, 1997.

E. Gülch, H. Müller, and T. Läbe. Integration of automatic processes into semi-automatic building extraction. In *Proceedings of ISPRS Conference: Automatic Extraction Of GIS Objects From Digital Imagery*. ISPRS, September 8-10 1999.

S. Heuel, F. Lang, and W. Förstner. Topological and geometrical reasoning in 3d grouping for reconstructing polyhedral surfaces. In *Proceedings of the XIXth ISPRS Congress*, volume XXXIII, Amsterdam, 2000. ISPRS.

T. H. Kolbe. *Identifikation und Rekonstruktion von Gebäuden in Luftbildern mittels unscharfer Constraints*. Shaker Verlag Aachen, 2000. Dissertation am Institut für Umweltwissenschaften, Hochschule Vechta, 1999.

F. Lang. *Geometrische und Semantische Rekonstruktion von Gebäuden durch Ableitung von 3D Gebäudeecken*. PhD thesis, Institut für Photogrammetrie, Universität Bonn, 1999.

H. Mayer. Automatic object extraction from aerial imagery - a survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999.