Integration of 2D and 3D Reasoning for Building Reconstruction Using a Generic Hierarchical Model *

André Fischer[1] **and Thomas H. Kolbe**[2] **and Felicitas Lang**[3]

Institut für Informatik I[1]/III[2], Institut für Photogrammetrie[3], Universität Bonn
e-mail: andre@cs.uni-bonn.de, tk@cs.uni-bonn.de, Felicitas.Lang@ipb.uni-bonn.de

Keywords: building modeling, multi image correspondence analysis, mid-level feature aggregates, aspect hierarchies, constraint logic programming.

Abstract

We propose a model-based approach to automated 3D extraction of buildings from aerial images. The semantics of the concept *building* is used to control and to evaluate building extraction in all stages of the process. The semantics is encoded by means of generic 3D object modeling, which describes thematic and geometric constraints on the spatial structure of buildings, and of 2D image modeling, which integrates sensor and illumination modeling to describe the appearance of buildings specific for the given aerial imagery. 3D object and 2D image modeling are tightly coupled within a multi-layered framework which contains an *is-part-of*-hierarchy of 3D building parts and their corresponding image descriptions. The overall strategy follows the paradigm of hypotheses generation and verification and combines bottom-up and top-down processes. Due to the explicit representation of well defined processing states in terms of model-based 2D and 3D descriptions at all levels of modeling and data aggregation our approach reveals a great potential for a reliable building extraction.

1 Introduction

Due to the fact that more than about 50% of the world population live in urban or suburban environments the automation of 3D building extraction is an issue of high importance and shows an increasing need for various applications including geo-information systems, town planning or environmental related investigations.

Aerial images contain on the one hand a certain amount of information not relevant for the given task of building extraction like vegetation, cars and building details. On the other hand there is a loss of relevant information due to occlusions, low contrast or disadvantagous perspective. To compensate for these properties of image data as well as for being able to handle the overwhelming complexity of building types and building structures, a

* *Proceedings of the Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps SMATI '97, Bonn, W. Förstner Ed., Birkhäuser-Verlag, Basel, 1997*

promising concept of automated building extraction from aerial images must incorporate a sufficiently complete model of the objects of interest and their relations within the whole process of image interpretation and object reconstruction (cf. Suetens et al. 1992).

Such a strong and complete modeling approach seems not to be available up to now as the acquisition of spatial data for geo-information systems etc. today still is mainly done by human operators. Therefore, the representation of domain specific semantics is still a crucial subject of discussion and research in 3D building extraction.

We propose a model-based approach to automated 3D extraction of buildings from aerial images. The semantics of the concept *building* is used to control and to evaluate building extraction in all stages of the process. It is encoded by means of a generic 3D object model, which describes thematic and geometric constraints on the spatial appearances of buildings, and a 2D image model, which integrates sensor and illumination modeling to describe the projective appearances of buildings specific for the given aerial imagery.

1.1 Related work

Related work on 3D building extraction — or in general on 3D scene reconstruction — reveals different modeling schemes. Polyhedral models show a long tradition as approximative object descriptions (e.g. Clowes 1971, Huffman 1971, Waltz 1975, Sugihara 1986, Kanatani 1990, Heyden 1994). Obviously polyhedral descriptions are too general for the use within 3D building extraction and therefore move the burden of building modeling on additional representation schemes to represent and organize domain specific heuristics and constraints like in the MOSAIC system (cf. Herman and Kanade 1987) or in the approach of Braun 1994. Parameterized models are restricted to describe the most common building types in the sense of prototypes (cf. McKeown 1990, Quam and Strat 1991, Lang and Förstner 1996a,Lin et al. 1994a, Lin et al. 1994b, Lin et al. 1995) but show a lack to represent variations and combinations of their shapes as well as other relations. Prismatic models can describe arbitrary complex polygonal ground plans of buildings, but reveal the strong restriction to buildings with only flat roofs (cf. Weidner and Förstner 1995, Weidner 1996). CAD models are used to describe objects with fixed geometry and topology in object *recognition* tasks, especially for controlling industrial processes (cf. Hansen and Henderson 1993, Flynn and Jain 1991, Ikeuchi and Flynn 1995, Munkelt 1995). The use of CAD models in building extraction is therefore restricted to the identification of a priori known buildings (cf. Sester and Förstner 1989, Schickler 1993, Huertas et al. 1995). Generic modeling approaches promise on the one hand the greatest modeling power, but on the other hand demand effective constraints and heuristics to restrict modeling to building specific shapes. Fua and Hanson 1987 employ simple box-type primitives but propose an explicit representation of legal primitive combinations to more complex building aggregates. The approaches of Dickinson et al. 1992 and Bergevin and Levine 1993 are from outstanding importance due to the integration of 3D generic object models and an explicit modeling of 2D projective object appearances within a recognition-by-components strategy (cf. Biederman 1987). Both approaches employ volumetric primitives instead of simple box-types but neglect the description of elaborated schemes for domain dependent primitive combinations. Bignone et al. 1996 propose a generic roof model which assumes planar roof surfaces. Extracted 3D roof patches are grouped by an overall optimization according to the simplicity, compactness and completeness of the resulting roof shape. To complete the building shape vertical walls are assumed. This approach shows impressive results on some test data but obviously shows no explicit modeling of building types and building specific aggregation schemes. Groups of planar 3D patches optimized according to the criteria of simplicity, compactness and completeness are not necessarily real roof shapes.

1.2 Overview

In Braun et al. 1995 we proposed in detail the concepts and processes which have to be taken into account for a sufficient complete modeling framework for 3D building extraction. This modeling framework integrates interrelations between image data and model descriptions at different aggregation levels and in terms of corresponding 3D object and 2D projective object descriptions. Within this paper now we present a strategy for a well defined path from the unstructured image data to the model-based and highly structured 3D reconstruction of buildings.

The overall strategy follows the paradigm of hypotheses generation and verification and combines bottom-up (data-driven) and top-down (model-driven) processes. Domain knowledge constraints even the early stages of hypotheses generation due to an elaborated *is-part-of*-hierarchy of 3D building parts and their corresponding projective descriptions. The reconstruction process is carried out already for local 2D feature aggregates to allow an early domain specific classification as 3D local building feature aggregates. A stepwise and strongly model-driven aggregation process combines 3D local building feature aggregates to well defined parameterized 3D building parts and then to more complex 3D building aggregates. The resulting complex 3D building hypotheses and their components are back projected into the images to allow a component-based and robust hypothesis verification applying constraint solving techniques (cf. Kolbe et al. 1996).

Due to the explicit representation of well defined processing states in terms of model-based 2D and 3D descriptions at all levels of modeling and data aggregation our approach reveals a great potential for a reliable building extraction.

2 Concept

In this section we present the proposed building model and discuss its implications on the developed strategy.

2.1 Models

For coping with the complexity of natural scenes we propose an application specific modeling of the domain *buildings*. The general concept has been presented in Braun et al. 1995. It contains a close interaction of bottom-up and top-down strategies which is fundamental for interpretation tasks dealing with complex image data and models.

A crucial aspect of the approach is the explicit separation of object, sensor and image model in order to reach parsimony in modeling. We start with modeling the 3D-objects

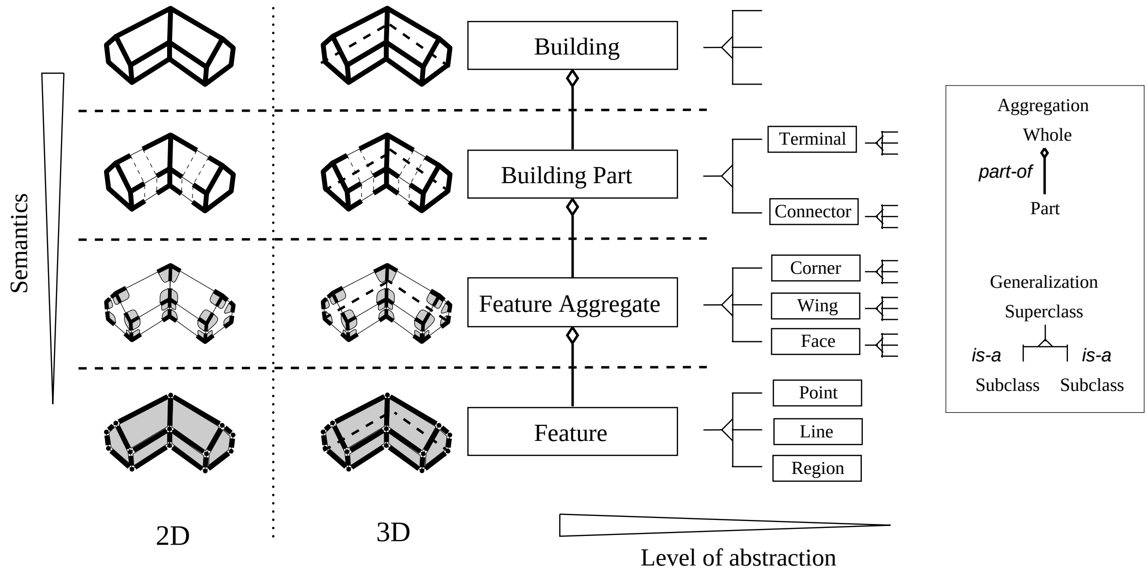

Fig. 1: Building model: The different semantic levels of the part-of hierarchy are shown in vertical direction, the different levels of abstraction of the is-a hierarchy in horizontal direction, which is only shown for the 3D-model. The 2D-image model describes the expected appearance of the building in the different levels of the part-of hierarchy, which is indicated by not showing the hidden lines.

leading to the object model. The sensor and the illumination model will transfer many of the concepts from the object to the image model describing the expected appearance of the objects. We now specify these models more detailed than in Braun et al. 1995 and explain the decisive role of the semantics.

Object Model Buildings reveal a high variability in structure which suggests to represent them as an aggregation of several simple building parts. This enables to cope with the problems caused by occlusions, low contrast, noise and disturbances.

We therefore propose a multi-level *part-of hierarchy* (cf. figure 1). It reflects different levels of the envisaged semantic abstraction. The primitives of each aggregation level are specialized by an *is-a hierarchy* into subclasses.

Each primitive is described by its semantics, its geometry and possibly its physical properties. Its class membership and its relations to other primitives formally describes its semantics. The geometry is described by pose and form parameters.

We at the moment employ four semantic levels for modeling complex buildings, which seems to be sufficient for a large class of buildings.

The first level (*feature level*) contains features F, namely attributed *points* P , *lines* L and *regions* R. Attributes for lines and regions, for instance, are the orientation classifications horizontal (h), oblique (o) and vertical (v). Regions have an additional attribute describing its role: valid values among others are wall, roof and floor. In general the set of parameters is divided into positional parameters on one hand, describing position and orientation, and form parameters on the other hand like width, height and length.

The second level (*feature aggregate level*) contains feature aggregates A which are induced by points, lines and regions, and contain all their direct neighbors. Each aggregate is defined by a feature graph, given by a set $F = \{f_1, \ldots, f_k\}$ of features and adjacency relations $R \subset F \times F$. A *Corner* C, for instance, contains one point and all its adjacent lines and regions (cf. fig. 2).

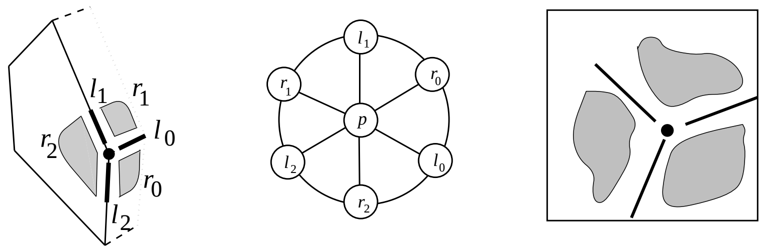

Fig. 2: A corner is a feature neighborhood of a point (left). Drawn as graph the arcs express the adjacency relation (mid). Assuming no occlusions and disturbances its expected appearance in the image reveals the same neighborhood relations between the corresponding 2D-features (right).

The third level (*building part level*) contains building parts P. Currently, they are defined as corner graphs given by a set $C = \{c_1, \ldots, c_n\}$ of corners and adjacency relations $R \subset C \times C$. They are parameterized volumetric objects. Each building part has at least one so called *plug face* which is used for connecting building primitives to each other. We discriminate *terminals* having exactly one plug face and *connectors* with two or more plug faces (cf. figure 3).

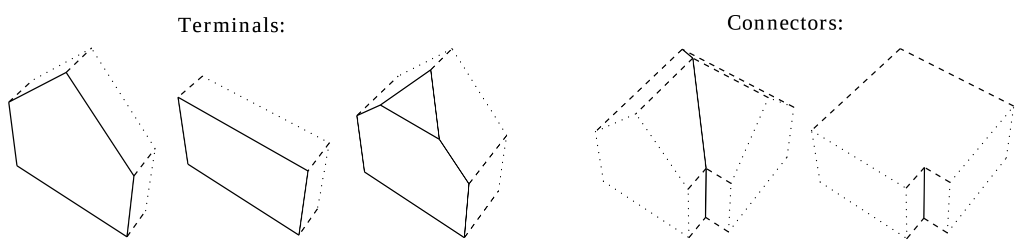

Fig. 3: Some examples of building parts. Plug faces, which are used to connect them, are drawn dashed.

The fourth level (*building level*) contains complete buildings. Buildings are defined as graphs with building parts as nodes, the arcs representing pairs of building parts connected by corresponding plug faces. Thus, the most simple building consists of two connected terminals.

Image Model The 2D image model describes the expected appearance of the building at the same levels of aggregation as the corresponding 3D-structures.[1] This guarantees coherence of the representation of 2D- and 3D-primitives being the prerequisite for all processes of 3D-reconstruction, aggregation, indexing and matching.

[1] Actually the image model contains the raster image as the lowest, say 0th level, from which the image features are extracted. This lowest level is not shown as we do not explicitely refer to it.

The image model on one hand contains all properties which are invariant under projection and can be used as constraints. This especially holds for all thematic labels, neighborhood and geometric relations. E. g. a 2D-corner, being a point induced image aggregate may be labeled a gable corner, the neighborhood relation between two regions of a roof in general can be expected to be transferred to a pair of image region, whereas – assuming weak perspective – parallel 3D-roof lines map to parallel 2D-roof lines. Constraints of the higher aggregation levels are transferred to the primitives of the lower ones, if necessary.

On the other hand all disturbing effects resulting from the illumination, the projection and the feature extraction need to be modeled. This also holds for the appearance of the model internal but not directly visible plug faces. These manifest themselves in local image aggregates, which contain open ended lines and faces, as it occurs in the feature aggregate level or in the building part level (cf. fig. 1 and 3).

The image model will be used for both, the multi view reconstruction of 3D-corners and the 2D-verification of hypotheses of building parts or of complete buildings.

Observations and templates We now are able to describe the task of linking the semantic model with the image data in detail.

As mentioned above, each primitive in the aggregation hierarchy is described by its semantics and its geometry possibly including physical properties. The semantics is manifested in the mutual relations between the different primitives or aggregates, which need to reflect their relations in reality. Semantic modeling therefore consists of describing the models of the primitives — either 3D or 2D — in terms of constraints on their attributes and relations, either on the same or on another aggregation level.

Image interpretation can be seen as a level-by-level goal driven matching process of the original sensor data with the model.

Original sensor data themselves do *not* contain any semantic information. They however provide original or derived geometric and physical measures and implicitely show certain relations. These have to be derived and organized in the same structure as the envisaged class, also in case they are aggregated. An soon as they are linked to primitives of the model by a matching process (cf. below), they result in *observations* which have a meaning and can be used in later steps of the analysis.

As at intermediate levels we obviously obtain only partially instantiated primitives we distinguish two types of instances of the primitives, templates and observations (cf. fig 4):

1. *Templates.* Templates are partially instantiated classes.

   Templates carry the semantics, which is given them during the modeling process. These are constraints on relations to templates on the same level or on the lower level. Uninstantiated attributes or relations in our context primarily refer to geometric properties, but also to labels of parts of the aggregate, thus templates on the lower level.

2. *Observations.* Observations are fully instantiated primitives.

   Observations obtain their semantics from the template through the matching process, especially by instantiation of mutual relations and relations to primitive obser-

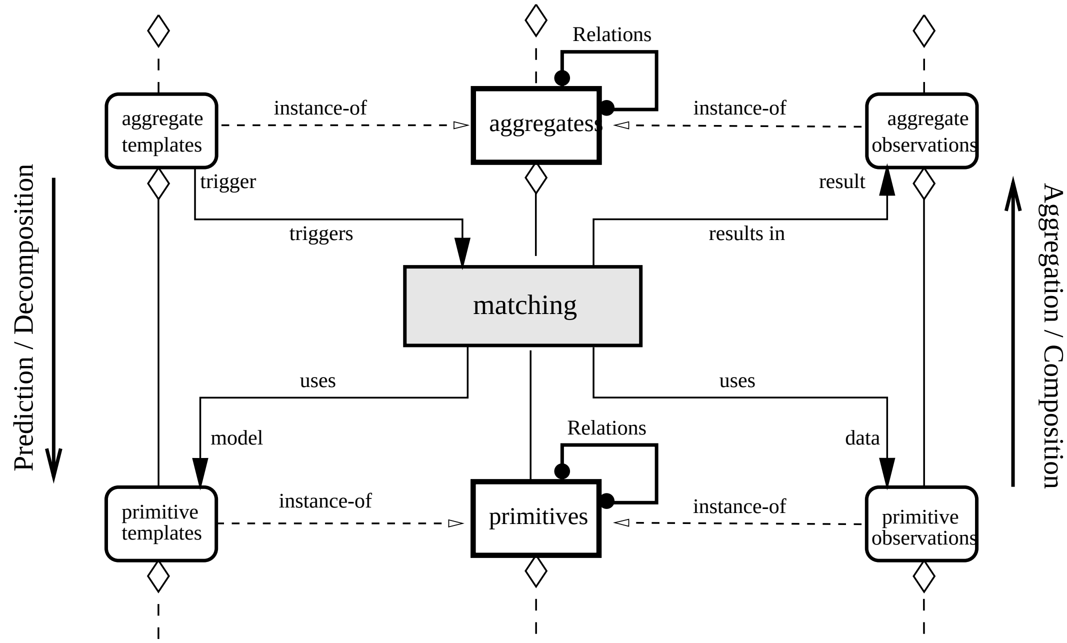

Fig. 4: Templates and observations as partially and fully instantiated primitives. Observations result from image data by a matching process.

vations. Observations can be seen as interpreted image data at a certain level of the aggregation hierarchy.

2.2 Strategy

We employ a combined, data-driven and model-driven strategy integrated in a matching process (cf. Fig. 4). It always links two levels within the *part-of*-hierarchy: Corner reconstruction links the feature and the feature aggregate level, reconstructing building parts or buildings links the feature aggregate level with the building part or the building level, whereas the verification step links the feature level with the building part or the building level.

The matching consists of four steps:

1. *Trigger:* The matching is triggered by a selected template at the higher level e. g. a corner, thus initiated by the semantic model. It can be seen as a set of goals. They have the same structure, e. g. $corner(type, geometry)$, allowing to handle them independently on the specific instance.

2. *Prediction:* The selected goals are used to predict a set of primitive templates on the lower level. They also reveal the same structure and in addition are linked by a set of compatibility relations, which are given by the aggregate template.

3. *Grouping:* Based on the predicted components the primitive observations on the lower level, which play the *role of the data* in the matching process, are grouped.

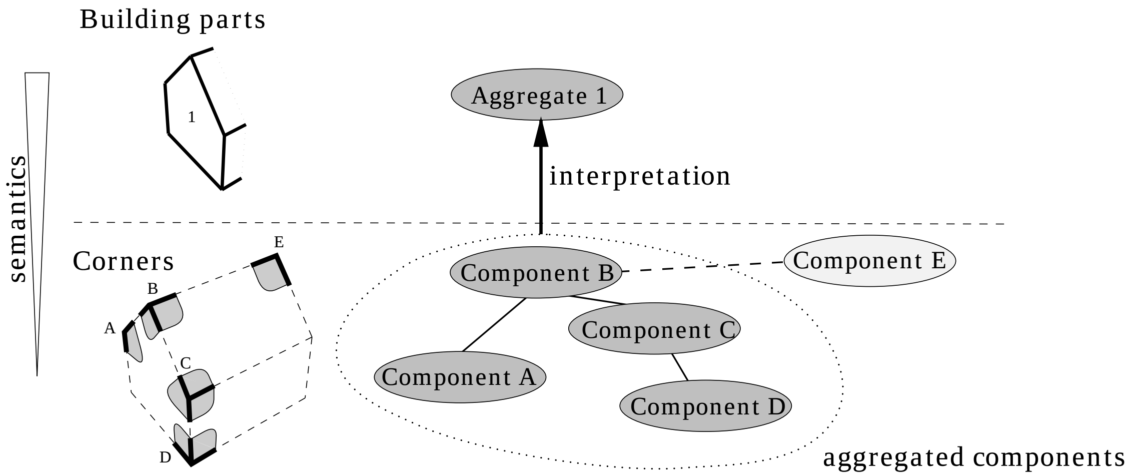

Fig. 5: Grouping and aggregation: components A to E being corners are grouped. Matching leads to a building part as aggregate (excluding E) and labeling it as a terminal gable front.

Groups thus are *candidates* for higher level aggregates. E. g. when looking for corners, point induced image structures are established. If strong positional knowledge is available, e. g. when verifying buildings or building parts, grouping consists of finding equivalent primitives, e. g. collinear straight line segments. The data structure of the groups is given by the aggregate template. However, *no final interpretation* takes place as grouping not necessarily checks the mutual relations between the primitives simultaneously.

4. *Aggregation:* The aggregate template is now matched to the found groups, establishing information at the higher level of the hierarchy. In this step all constraints on the geometry, the radiometry and possibly on the semantics of the primitives at the lower level are used. This way primitives contained in the groups may be excluded, cf. Fig. 5, where primitive E is grouped but not aggregated. The matching transfers the semantics of the templates, being part of the model, to the grouped data. This way the groups are interpreted, leading to aggregates which have a specific meaning within the model, e. g. labeling a corner as a gable corner. Using the *part-of* hierarchy, also missing semantic labels for the primitives on the lower level of the hierarchy are established, indexing them into one of the prespecified classes, e. g. labeling a horizontal 3D-line segment of a vertex as gable point line(h).

2.3 The Procedure

The complete process can be described in the following way: Our input data are given as digital raster images with multiple overlap. Further information about the aerial image flight like exterior and interior camera orientation and time stamp are used. The starting point of our analysis is the extraction of a polymorphic image description consisting of points $P^{2\mathrm{D}}$, lines $L^{2\mathrm{D}}$ and regions $R^{2\mathrm{D}}$ and their mutual relations (cf. Förstner 1994, Fuchs and Förstner 1995). It allows to derive point, line and region neighborhood aggregates $A^{2\mathrm{D}}$, where vertices $V^{2\mathrm{D}}$ are the most promising ones for starting our analysis.

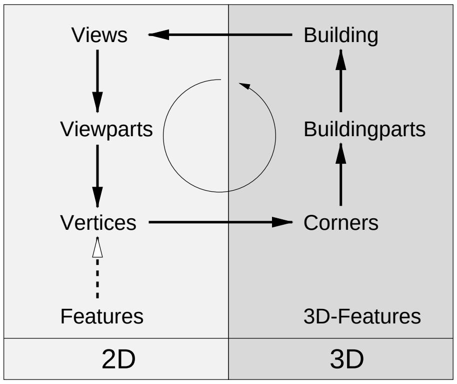

Fig. 6: Information transfer within the whole process. The close integration of 2D and 3D reasoning is performed by an iteration loop. The dashed arrow marks the initialization step. 3D-reconstruction, generation and verification of hypotheses for building parts and buildings is repeated. Verification is based on generated views using matching on the feature level.

Our procedure consists of three main tasks, which are executed subsequently (cf. Fig. 6):

1) **reconstruction of 3D corners:** Since the expressiveness of 3D data is higher than that of 2D data, we aim at an early transition from 2D to 3D in the reconstruction process, this way reducing the overall number of future hypotheses. This is done by finding feature groups, namely vertices in the images, which may correspond to building corners. Hypothesized 2D corners of different images are matched, leading to 3D corners.

2) **generation of building hypotheses:** The 3D corners now are taken to find building parts, which can explain these corners. When there is needed more than one building part to explain the 3D observations, the building parts are combined to a building. In this step some components and parameters of the building parts may stay undetermined.

3) **verification of building hypotheses:** For verification the 3D building templates are projected back into the different images, resulting in 2D views for buildings, or views of building parts, (cf. Fig. 1). The predicted feature templates and their interrelationships are then matched with the extracted image features and their relations. The matching possibly determines free parameters of the template.

The successful sequence of matching steps results in an iteratively improved gain in knowledge. The three steps are repeated until no further hypotheses can be generated: The verified building hypotheses lead to predicted unobserved 2D primitives on the lower levels, giving additional information for reconstructing previously undetected corners, initiating a second iteration of 3D-reconstruction, generation of building hypotheses and verification.

3 Realization

3.1 Reconstruction of 3D Corners

In this section we describe the procedure for 3D-reconstruction and interpretation on the level of feature aggregates (cf. fig. 1). The reconstruction is based on local image aggregates A^{2D}, which can be directly derived from our polymorphic feature description. We distinguish corners and vertices, corners being interpreted vertices. As for vertex structures V^{2D}, being the projections of 3D vertices V^{3D} into the images, the imaging geometry gives strong restrictions during the correspondence analysis, we focus on the 3D reconstruction of corners C^{3D}.

Corner Model Corners are specialized into different subclasses. Each corner is described by the corner point, several lines and planar faces (cf. fig. 2). The partitioning of corners is described by their line attribution, given by the semantic labels (h), (v) or (o) (cf. Gülch 1992). This description is further refined by distinguishing vertical and oblique lines due to their slope into (v+), (v-), (o+) and (o-). Finally, different geometric constaints between the corner components like e. g. symmetry and orthogonality are considered.

We also model corner pairs. The previous classification of single corners is used for deriving restrictions like collinearity of lines and coplanarity of planes on compatible corner pairs.

All these semantic attributed subclasses are collected in the set Ω_C, one element for instance being the quadruple { (h), (o-), (o-), symmetry (o-,o-) }, which corresponds to the vertex at the ridge of a hip roof.

Construction of Corner Hypotheses Starting point for the analysis are the 2D vertices $V^{2\mathrm{D}}$ (cf. section 2.3). In the first step we generate 3D vertices. Selected tuples of corresponding vertex structures form the basis for the transition to 3D-vertices $V^{3\mathrm{D}} = \{v_1^{3\mathrm{D}}, \ldots, v_n^{3\mathrm{D}}\}$ by a joint forward intersection of all corresponding 2D-vertices $\boldsymbol{v}^{2\mathrm{D}} = (v_1^{2\mathrm{D}}, \ldots, v_N^{2\mathrm{D}})$. It is established by heuristically selecting a suitable sequence of vertices, which are then evaluated. The correspondence analysis uses the epipolar geometry and the structural similarity of matching candidates. Relational matching of the features $F^{2\mathrm{D}}$, which describe the vertex structures $v_i^{2\mathrm{D}}$ of the correspondence tuple $\boldsymbol{v}^{2\mathrm{D}}$, leads to a 3D Vertex $v^{3\mathrm{D}}$.

Establishing corner hypotheses uses the corner model in the following way. We interpret the generated 3D vertices $V^{3\mathrm{D}}$ by a first *classification*, leading to admissible corner subclasses Ω_C. First, the line attributation is used to obtain a subset Ω'_C of Ω_C, which in case the line attributations do not correspond to a valid corner description, is the empty set. For each element of Ω'_C the geometric constraints, associated by the model, yield a further reduced $\Omega''_C \subset \Omega'_C$, each element in Ω''_C being an admissable interpretation of the vertex $v^{3\mathrm{D}}$. This way we obtain several corner hypotheses $c^{3\mathrm{D}}$.

This principle of domain reduction of possible interpretations is also applied for the generation of corner pairs $(c_i^{3\mathrm{D}}, c_j^{3\mathrm{D}})$.

Verification of Corner Hypotheses As the corner classification for each vertex $v^{3\mathrm{D}}$ generally will be ambiguous, we perform a second rigorous classification by statistical analysis. This is an optimization problem for finding the best interpretation of the data. For each hypothesis, we estimate the geometric (and probably radiometric) parameters by a maximum likelihood parameter estimation. The functional model describes the relation between observations and the unknown parameters for the unconstrained corner. Additionally hard and soft constraints $\Theta_{\omega_C}^{3\mathrm{D}}$, given by the corner class ω_C, are introduced. The result of the estimations are evaluated corner reconstructions $c^{3\mathrm{D}} = c^{3\mathrm{D}}(v^{3\mathrm{D}}, \omega_C)$. Assuming the same image feature observations, all possible corners are tested against the unconstrained corner to decide for the optimal interpretation of the vertex data $\boldsymbol{v}^{2\mathrm{D}}$.

Further details of the corner reconstruction approach can be found in (cf. Lang and Förstner 1996b, Brunn et al. 1996).

The geometric constraints on single corners $c_i^{3\mathrm{D}}$ as well as on corner pairs $(c_j^{3\mathrm{D}}, c_k^{3\mathrm{D}})$

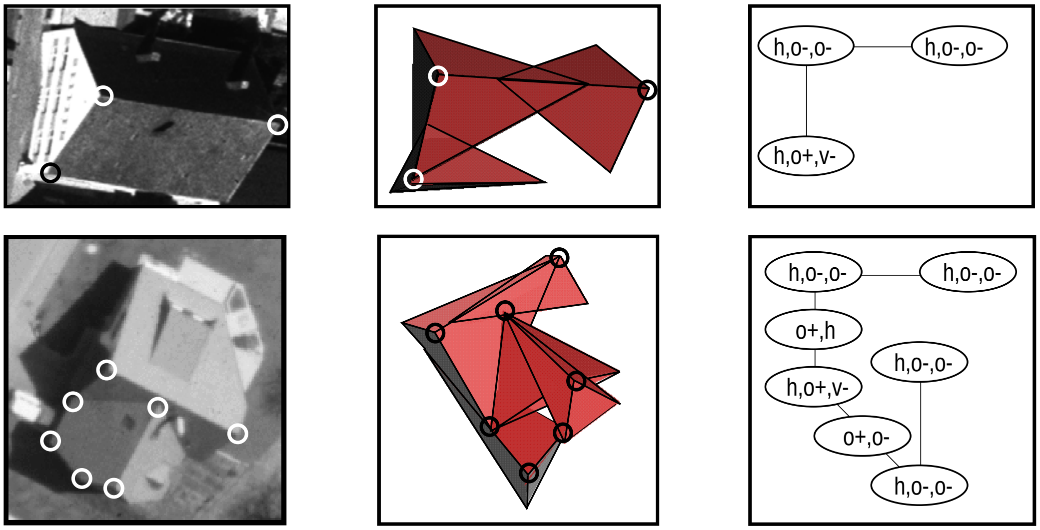

Fig. 7: Examples of 3D corner reconstruction. From left to right: feature aggregates marked in one image, reconstructed 3D corners, corner graph showing the adjacency relations.

are fundamental for a geometrically improved and verified corner reconstruction. Those corners $C^{3\mathrm{D}}$, which are accepted, are regarded as 3D-observations of the level feature aggregates and form the basis for the 3D aggregation. The grouping of corners results in a weighted graph $G = (C^{3\mathrm{D}}, R)$ where the nodes are classified corners and the arcs the adjacency relation. The weight of the arc expresses the conditional probability of the adjacency relation as derived from the classification. This forms the basis for the generation of building hypotheses as presented in the next chapter.

After the first iteration of the whole building reconstruction process, we use newly generated 2D corners $C^{2\mathrm{D}}$ (cf. section 3.3), in addition to the original vertex data.

3.2 Generation of Building Hypotheses

Building hypotheses are generated from the observed corner graph $G = (C^{3\mathrm{D}}, R)$ in a two step process. First the corners are grouped and aggregated to building part templates. In the second step these building part templates are grouped and aggrated to building templates (cf. Fischer and Steinhage 1997).

Generation of building part templates Since building parts are defined by corner graphs G_j, the generation of building part templates explaining the corner observations is done by determination of maximal subsets $C_i^{3\mathrm{D}\prime} \subseteq C^{3\mathrm{D}}$ that have a subgraph isomorphism from $(C_i^{3\mathrm{D}\prime}, R')$ onto a G_j. That is, corner observations and templates must have the same classifications and adjacency relations.

If a subgraph isomorphism of a $(C^{3\mathrm{D}\prime}, R')$ in a graph G_j of a building part p is found, the elements of $C^{3\mathrm{D}\prime}$ are aggregated to a new building part template $\dot{p}$.

Generation of building templates For the generation of building part templates structural and geometrical information are considered. In general not all building parts forming a building have corner observations. Therefore it may be necessary to hypothesize building part templates without having actual observations for them.

Buildings are constructed by grouping building parts. Our generic model allows groups of variable size as long as restrictions on pairs of building parts implied by the model are satisfied.

The grouping task is defined by iterating the following steps. Starting with the set T_0 of previously generated building part templates more complex templates are generated by merging compatible pairs of building parts. Merging two parts implies the unification of certain parameters and the introduction of a new length parameter (cf. fig. 8). Two building parts may be merged, if their plug faces are of the same type and they have compatible geometry. The latter is checked by parameter estimation. The previous step is iterated until either all elements of T_0 are part of a complete building with no plug faces or no new connections can be made.

If the first condition holds we are done, because the complete buildings in the final set T contain all elements of T_0 and therefore explain all corner observations. Otherwise, not every building part template in T_0 is part of a complete building in T. In order to assign every element in T_0 to a complete building, new building part templates have to be introduced for which there exist no observations. These parts must enable the construction of new complete buildings, consisting of as few parts as possible. Every complete building represents a building template, which is a valid aggregation of building parts.

For every building template a global parameter estimation considering all associated corner observations determines the degree of freedom for every parameter.

Generation of 2D view hierarchy For verification and determination of free parameters of the generated building hypotheses, represented by the building templates, they have to be matched with the observations in the image. In 2D, buildings are modeled by view hierarchies. With respect to the fixed image geometry all topologically different aspects of the buildings correspond to primitives on the highest level of the view hierarchy, defining the 2D building templates. In contrast to aspect graphs the viewing direction is fixed, but due to free parameters the object geometry is in general variable, which can lead to different possible aspects. After the generation of the 2D building templates, they are decomposed into primitives and their interrelationships on lower levels. A relation between primitives on a higher level is transformed to a set of relations beween their components on the lower level. Thereby semantic knowledge is propagated down through every level of our aggregation hierarchy. For each image one view hierarchy is generated.

3.3 Verification of Building Hypotheses

The generation of 3D building hypotheses leads to different possible 2D views. Since some parameters of the 3D building templates may still be free, there are free parameters

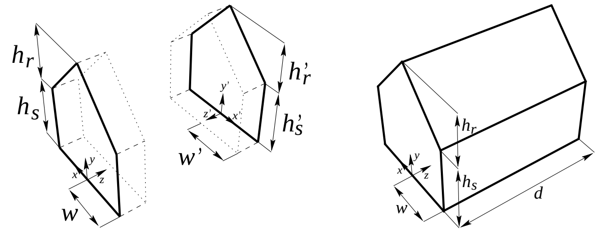

Fig. 8: The parameters w and w', h_s and h'_s, and h_r and h'_r are unified. The parameter d is introduced to describe the depth of the newly formed building.

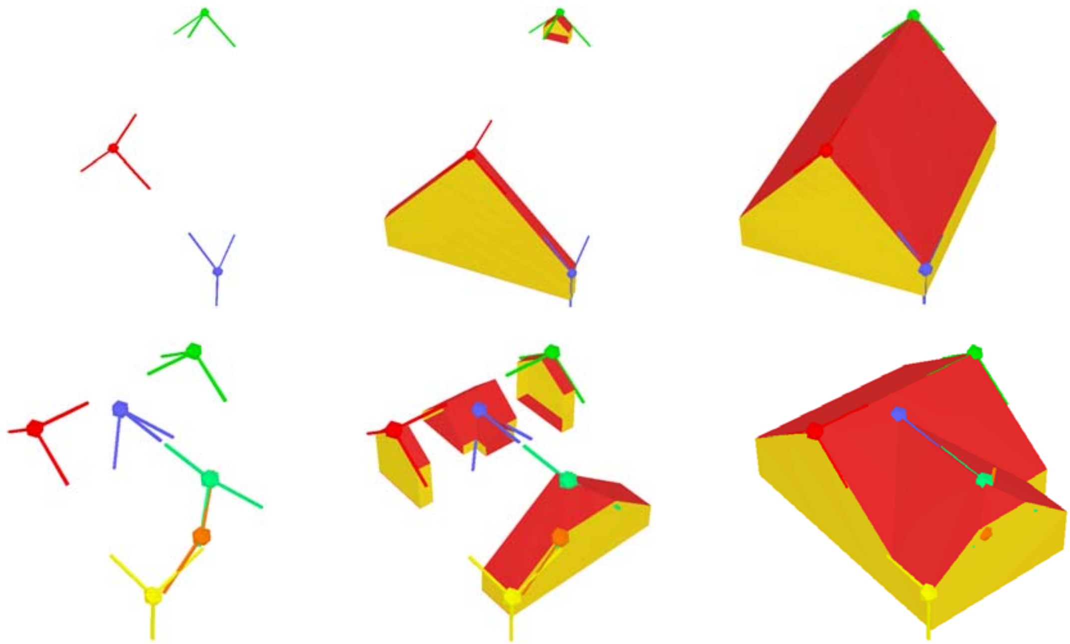

Fig. 9: Examples of building part aggregation. From left to right: reconstructed 3D corners, best matching building part hypotheses, aggregated building. Note that default values are used for free parameters.

in the projected 2D views. Therefore the first task is to identify those views which match the given observations. A match fixes free parameters of the corresponding template. Since a match between template and observations considers objects and their interrelationships this task is an instance of relational matching. We solve the resulting search problem with constraint solving methods (cf. Meseguer 1989).

After the identification of the view, giving a semantic label to every matched feature observation, line fragments in the image are grouped according to the building template. Then previously unobserved 2D corners are generated by robust estimations on intersections of identified lines. These 2D corners are additional information that is used to reconstruct further, previously undetected 3D corners in the next iteration of the whole building reconstruction procedure.

The transformation of templates to a constraint satisfaction problem (CSP), the matching and its implementation using constraint logic programming (CLP) (cf. van Hentenryck 1989, Jaffar and Maher 1994) is explained in detail in Kolbe et al. 1996.

Transformation of view hierarchies to sets of constraints A view hierarchy enumerates the possible views for the building hypotheses with respect to one image. To evaluate the correspondence to the originally extracted image features the matching between the building templates and the observations is done on the lowest level of the hierarchy. The entities considered for matching are objects of the three feature classes points, lines, and regions and different relations, e. g. adjacency, line parallelism, region symmetry, and region contrast. The line parallelism and region symmetry relations reflect the respective 3D properties in 2D. The region contrast relation classifies the intensity ratio between adjacent regions.

A building template there is represented by a set of features and a set of relations between them. These relations may be regarded as a set of constraints which have to be satisfied simultaneously by the corresponding objects. A consistent match is an identification of a set of objects which satisfies all constraints. The decomposition of the matching problem into the simultaneous satisfaction of different constraints leads to a structure which is adequately represented by the translation of templates into conjunctions of constraints where the features are represented by variables with restricted domains. The domain of a variable depends on its associated model feature type, for instance line variables are restricted to the set of line observations that were extracted from the image. Thus the first constraint posed on a variable is the *domain constraint*. All other constraints are implied by the corresponding relations of the building template.

For each view of a hierarchy one set of constraints will be generated representing the building template.

Matching of constraint sets to the image data The matching is done seperately for every image and view by searching for a valid assignment of extracted image features to the variables of the corresponding CSP, such that all constraints are satisfied. Consistency techniques exploit the structured representation of the matching problem as a constraint network. The idea is that a constraint is used to restrict the domains of its variables and that these reductions are propagated through the network.

Standard techniques for constraint solving (Mackworth 1977, Haralick and Elliott 1980, Dechter and Pearl 1988, van Hentenryck 1989) demand that a match satisifies every constraint. Due to occlusions and disturbances in general neither every predicted model feature of a view can be found nor all their incident constraints are satisfiable. To cope with this problem we employ relaxation of constraints, allowing constraints to be neglected. Our combination of consistency techniques and constraint relaxation follows in general the scheme described in Haralick and Shapiro 1993 for inexact matching. As a language constraint logic programming provides hard constraints which are strictly enforced. A soft constraint $c(t_1, \ldots, t_{n-1})$ with $n-1$ arguments may be implemented by a hard constraint $c'(t_1, \ldots, t_{n-1}, b)$ where c' is satisfied if the boolean b is $true$ or c is satisfied. The variable b with the domain $\{true, false\}$ may act as a communication channel between the constraint and its environment. This environment may come to the conclusion that c models an unobservable relation and relaxes c by instantiating b with $true$.

The solution of the constraint satisfaction problem are the valid assignments of extracted features to the variables. On the one hand these assignments determine the free geometric parameters of the feature templates and thus determine free parameters of the whole building template. On the other hand, every assigned image feature is identified as a specific component of the building observation as defined by the building template. These matched image features are given labels, denoting their semantics with respect to the building model. Figure 10 shows on the left highlighted the image features that were successfully matched with the building template for a saddle roof house.

Since lines in general are fragmented into several line image features, corresponding line

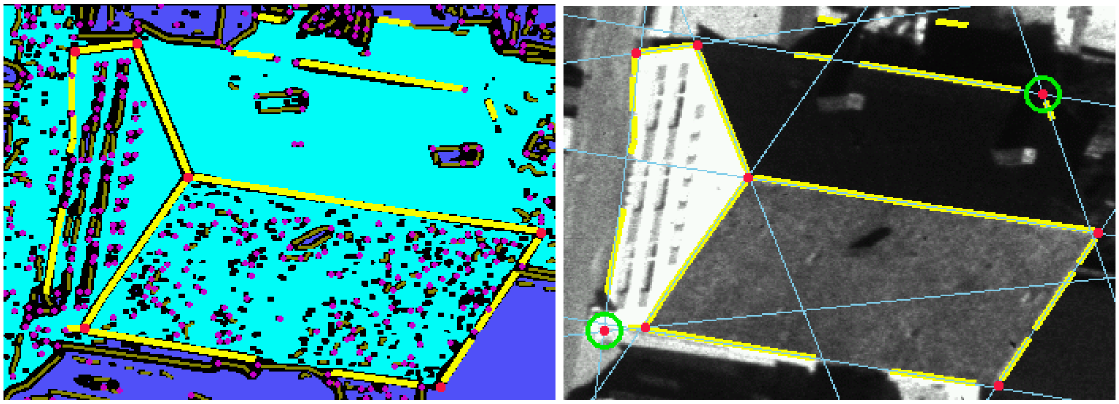

Fig. 10: Left: Matched features, right: new 2D corners

features are grouped and fitted using the labeling information to determine the lines of the building observation.

Generation of new 2D corners The solution of the CSP in general fixes further parameters of the building template. It has to be noted, that free parameters may remain, if some constraints had to be relaxed. Corners of the building template with fixed positions that have no corresponding observation of a vertex structure in the image, finally are reconstructed in 2D as illustrated by the right picture in figure 10 to provide new information for the 3D corner reconstruction in the next iteration of the reconstruction process.

As mentioned above the loop of 3D-reconstruction and generation and verification of building hypotheses will be terminated if no further hypotheses can be generated. After the final verification step the parameters of the 3D-building model will be determined simultaneously based on the observed image features and taking all geometric and possibly radiometric constraints into account.

3.4 Results

The actual result of the reconstruction procedure is presented for the international test data set which was distributed for the Ascona Workshop 1995 on *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (cf. Grün et al. 1995). The image scale is $1:5000$ and we use a resolution with $30\mu m$ pixel size in contrast to Bignone et al. 1996 who use $15\mu m$. As building `h12` is under construction, only 11 out of the 12 buildings which are contained in multiple images, are relevant for the analysis. To test the feasibility of the concept in a first instance the number of building primitives is reduced to 2 different building part terminals and their possible connectors. To resolve the footprint of the buildings, higher image resolution is necessary. Thus building heights can not be determined automatically and therefore are set to fixed values. Figure 11 shows the result after the first iteration loop (cf. Fig. 6). Results of the intermediate steps are listed in Table 1 and reflect the following aspects of the different reasoning steps:

Reconstruction of 3D corners (I): The number of reconstructed corners `RC` on average compounds 40% of the buildings corners (`C`), whereas 2 of them are incorrectely classi-

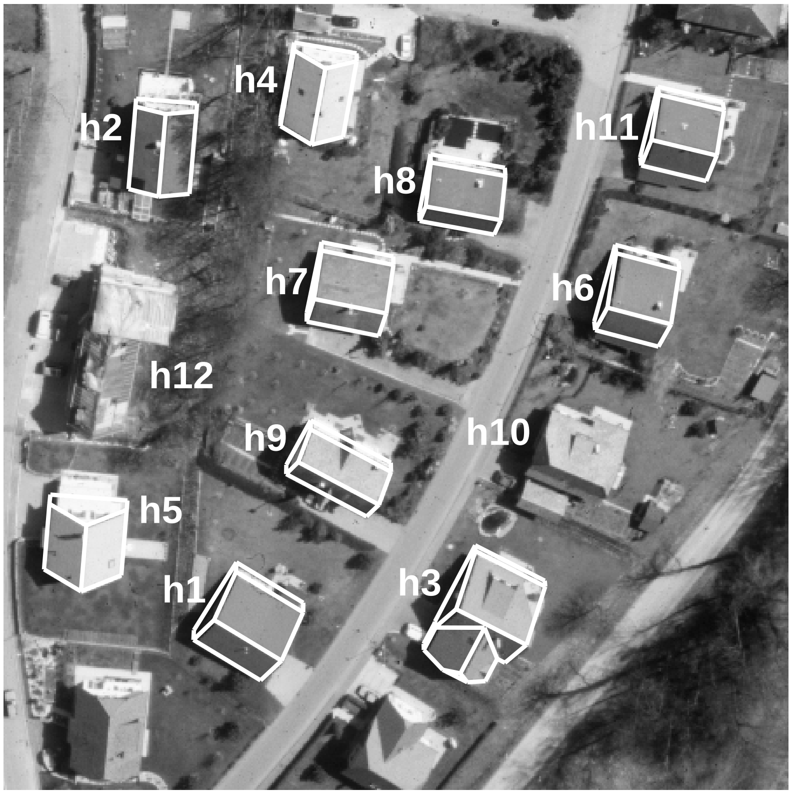

Fig. 11: Visualization of the reconstruction result for the Ascona Datset after one iteration loop (cf. Fig. 6) is performed. The images contain 12 buildings in multiple overlap. As building `h12` is under construction, we excluded it from our analysis. We are able to completely reconstruct 10 of the 11 buildings which were relevant for the analysis.

fied. We are not able to completely reconstruct every corner, because not all corners are observable in the symbolic image descriptions. An application indepent grouping as presented in Fuchs and Förstner 1995 may improve the initial symbolic image descriptions. Possibly still remaining unreconstructed corners have to be identified during the verification step of building hypotheses. The reason for incorrect classification is given by a weak intersection geometry for the line reconstruction, which is not yet considered in the modeling process. Neglecting the influence of the uncertainity of the line reconstruction, the corner point is correctely reconstructed. Actually the identification of wrong corner interpretations is performed by finding global inconsistencies during the generation of building hypotheses. For building `h10` the image resolution is not sufficient to successfully reconstruct any corner. The parameter estimation for correctely classified corners using 4 images simultaneously achieves an accuracy of the reconstructed corner point about $\sigma_x = \sigma_y = \pm 6cm$ and $\sigma_z = \pm 15cm$. The accuracy of the orientation of the corners is about $\sigma_\kappa = 0.6^o$ while the reconstruction of the slope of the roof is by $\sigma_{sl} = 2^o$.

Generation of building hypotheses (II): The column `BPH` gives the number of building part hypotheses which are generated for each of the reconstructed single corners. The combination of these building parts results in the number `BH` of hypotheses of complete buildings which are consistent with all reconstructed corners. If only one corner is given, as it is the case for `h2` and `h7`, the corresponding terminals are each completed to a closed building hypothesis with a second terminal of the same type. For each of the `BH` building hypotheses the view graph is generated. Currently the view graph of the best building hypothesis is passed to the verification of building hypotheses. The decision which is the

	I			II		III		
Building	C	RC	CE	BPH	BH	UV	GV	FP
h1	6	3	0	3/4/3	4	3	3/3/3/3	0
h2	6	1	0	3	2	5	5/5/-/-	0
h3	12	7	0	3/3/2/3/3/4/0	4	5	-/-/-/-	0
h4	6	3	0	4/3/3	4	3	3/3/3/3	0
h5	6	3	0	3/4/4	4	3	3/3/3/3	0
h6	6	3	1	3/0/4	4	4	4/4/4/4	0
h7	6	2	1	3/0	2	5	5/5/5/5	0
h8	6	3	0	3/3/4	4	3	3/3/3/3	0
h9	6	4	0	3/3/4/8	2	2	2/2/2/2	0
h10	10	0	-	-	-	-	-	-
h11	6	2	0	3/3	4	4	4/4/1/0	0

C	number of corners of the building	BH	number of building hypotheses
RC	number of reconstructed corners	UV	predicted corners
CE	number of corner classification errors	GV	predicted and verified vertices per image
BPH	number of building part hypotheses for each corner	FP	remaining free parameters

Tab. 1: Detailed results of the intermediate reconstruction processes Reconstruction of 3D corners (I), Generation of building hypotheses (II) and Verification of building hypotheses(III) for the Ascona Dataset. 10 of the 11 buildings can be reconstructed as shown in column FP.

best one is given by using the Minimum Description Length criterium following Rissanen 1987.

Verification of building hypotheses(III): For each building in each image one view graph being a building template is available and matched to the extracted image features and mutual relations. Column UV lists the number of corners which were correctly predicted during the generation of building hypotheses. The column GV shows for each of the 4 images the number of correctly matched vertices. For 80% of the images a successful matching is reached, 9 of the 10 generated building hypotheses could be verified by successful matching in at least one image. Dashes in the table denote a break-off of the analysis with complex or strongly underconstrained models, which can be resolved using appriopriate heuristics to keep the search space down. As for the buildings h3 the reconstructed corners already were sufficient to determine all parameters of the building hypothesis, in 10 of the 11 examples the building is successfully reconstructed as shown in the last column FP of the table.

4 Conclusions and Outlook

We have proposed a model-based approach to 3D building extraction from aerial images. Our modeling concept reveals a tight coupling of generic 3D object modeling and an explicit image model. Object and image model show corresponding *part-of* hierarchies describing aggregation states within a recognition-by-components strategy.

The procedure aims at a step by step increase of knowledge during the reasoning. As the semantics of a model defines the maximum achievable knowledge about an actual scene and its granularity determines the resolution for knowledge our model spans hierarchically from the smallest observable features to complex buildings and allows an adequate evaluation of hypotheses and their immanent knowledge within the different reasoning steps.

We have presented first experimental results on test data sets provided by the Landesvermessungsamt Bonn and the ETH Zürich.

Currently we are investigating and developing on

- the measurement and propagation of uncertainty within the overall reconstruction process;
- the extension of the current building modeling by more sophisticated knowledge about buildings, esp. functional aspects.
- the derivation of domain dependent heuristics to constrain search spaces;
- the handling of incomplete observations within all processes and stages of hypotheses generation and verification.

Especially we have to consider the integration of hard and soft constraints, i. e. logical and statistical knowledge into the framework of constraint logic programming. Furthermore the approach has to be compared to different automatic algorithms for segmentation, 3D reconstruction and classical photogrammetry. In order to achieve this goal an attempt is made to find a set of standardized notions about basic building models, parts of buildings and classes of buildings to ease such comparisons on a conceptual level.

Acknowledgments:

This work was largely done within the the project "Semantic Modeling and Extraction of Spatial Objects from Images and Maps" especially in the subproject "Building Extraction" which is supported by the Deutsche Forschungsgemeinschaft. We heavily profitted from discussions with Armin B. Cremers, Wolfgang Förstner, Lutz Plümer and Volker Steinhage. We thank the DFG for supporting our work.

References

Bergevin R., M. Levine (1993) *Generic object recognition: Building and matching coarse descriptions from line drawings*, IEEE T-PAMI, Vol. 15, No. 1, pp. 19–36.

Biederman I. (1987) *Recognition-by-components: A theory of human image understanding*, Psychological Review, Vol. 94, pp. 115–147.

Bignone F., O. Henricsson, P. Fua, M. Stricker (1996) *Automatic extraction of generic house roofs from high resolution aerial images*, Computer Vision — ECCV '96, Proc. of the 4th Europ. Conf. on Computer Vision, number 1064 in *Lecture Notes in Computer Science*, pp. 85–96.

Braun C. (1994) *Interpretation von Einzelbildern zur Gebäudeerfassung*, PhD thesis, Institut für Photogrammetrie, Universität Bonn.

Braun C., T. H. Kolbe, F. Lang, W. Schickler, V. Steinhage, A. B. Cremers, W. Förstner, L. Plümer (1995) *Models for photogrammetric building reconstruction*, Computer & Graphics.

Brunn A., F. Lang, W. Förstner (1996) *A procedure for segmenting surfaces by symbolic and iconic image fusion*, in B. Jähne, P. Geißler, H. Haußecker, F. Hering (eds), Mustererkennung 1996, DAGM, Springer, pp. 11–20.

Clowes M. (1971) *On seeing things*, Artificial Intelligence, Vol. 2, pp. 79–116.

Dechter R., J. Pearl (1988) *Network-based heuristics for constraint satisfaction problems*, Artificial Intelligence.

Dickinson S. J., A. P. Pentland, A. Rosenfeld (1992) *From volumes to views: An approach to 3-d object recognition*, CVGIP: Image Understanding, Vol. 55, No. 2, pp. 130–154.

Fischer A., V. Steinhage (1997) *Solid modeling for building extraction from aerial images*, in V. Skala (ed.), 5th Int. Conf. in Central Europe on Computer Graphics and Visualization 97, WSCG'97, pp. 114–123.

Flynn P. J., A. K. Jain (1991) *Cad-based computer vision: From cad-models to relational graphs*, IEEE T-PAMI, Vol. 13, No. 2, pp. 114–132.

Förstner W. (1994) *A framework for low level feature extraction*, in J.-O. Eklundh (ed.), Computer Vision, ECCV '94, Vol. II, number 801 in *Lecture Notes in Computer Science*, Springer-Verlag, pp. 383–394.

Fua P., A. J. Hanson (1987) *Resegmentation using generic shape: Locating general cultural objects*, Pattern Recognition Letters, Vol. 5, pp. 243–252.

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

Grün A., O. Kübler, P. Agouris (eds) (1995) *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser.

Gülch E. (1992) *A knowledge-based approach to reconstruct buildings in digital aerial imagery*, 17th ISPRS Congress, Washington D. C., Proceedings, pp. 410–417.

Hansen C., T. C. Henderson (1993) *Cad-based computer vision*, IEEE T-PAMI, Vol. 11, pp. 1187–1193.

Haralick R. M., G. L. Elliott (1980) *Increasing tree search efficiency for constraint satisfaction problems*, Artificial Intelligence, Vol. 14, pp. 263–313.

Haralick R. M., L. G. Shapiro (1993) *Computer and Robot Vision*, Vol. II, Addison-Wesley Publishing Company.

Herman M., T. Kanade (1987) *The 3d mosaic scene understanding system: Incremental recognition of 3d scenes from complex images*, in Fischler/Firschein (ed.), Readings in Computer Vision, Kaufmann, pp. 471–482.

Heyden A. (1994) *Consistency and correction of line-drawings, obtained by projections of piecewise planar objects*, Computer Vision — ECCV '94, Proc. of the 3rd Europ. Conf. on Computer Vision, number 800 in *Lecture Notes in Computer Science*, pp. 411–419.

Huertas A., M. Benjanin, R. Nevatia (1995) *Model registration and validation*, in A. Grün, O. Kübler, P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, Basel, pp. 33–43.

Huffman D. A. (1971) *Impossible objects as nonsense sentence*, Machine Intelligence 6, pp. 295–323.

Ikeuchi K., P. J. Flynn (1995) *Recent progress in cad-based vision*, CVGIP, Vol. 61, No. 3, pp. 293–294.

Jaffar J., M. J. Maher (1994) *Constraint logic programming: A survey*, Journal of Logic Programming.

Kanatani K. (1990) *Group-Theoretical Methods in Image Understanding*, Springer.

Kolbe T. H., L. Plümer, A. B. Cremers (1996) *Using constraints for the identification of buildings in aerial images*, Proceedings of the 2. Int. Conf. on Practical Applications of Constraint Technology PACT'96 in London, The Practical Application Company Ltd., pp. 143–154.

Lang F., W. Förstner (1996a) *3d-city modelling with a digital one-eye-stereo system*, ISPRS Symposium, Comm. IV, Vienna.

Lang F., W. Förstner (1996b) *Surface reconstruction of man-made objects using polymorphic mid-level features and generic scene knowledge*, Zeitschrift für Photogrammetrie und Fernerkundung, Vol. 6, pp. 193–201.

Lin C., A. Huertas, R. Nevatia (1994a) *Detection of buildings using perceptual grouping and shadowsproceedings computer vision and pattern recognition*, Proceedings Computer Vision and Pattern Recognition, pp. 62–69.

Lin C., A. Huertas, R. Nevatia (1995) *Detection of buildings from monocular images*, Automatic Extraction of Man-Made Objects from Aerial and Space Images, pp. 125–134.

Lin C. L., Q. Zheng, R. Chelappa, L. S. Davis, X. Zhang (1994b) *Site model supported monitoring of arial images*, Proceedings Computer Vision and Pattern Recognition, pp. 694–699.

Mackworth A. K. (1977) *Consistency in networks of relations*, Artificial Intelligence, Vol. 8, pp. 99–118.

McKeown D. M. (1990) *Towards automatic cartographic feature extraction from arial imagery*, in H. Ebner, D. Fritsch, C. Heipcke (eds), Digital Photogrammetric Systems, Wichmann, Karlsruhe.

Meseguer P. (1989) *Constraint satisfaction problems: An overview*, AICOM, Vol. 2, No. 1, pp. 3–17.

Munkelt O. (1995) *Aspect-trees: Generation and interpretation*, CVGIP, Vol. 61, No. 3, pp. 365–386.

Quam L., T. Strat (1991) *Sri image understanding research in cartographic feature extraction*, in C. H. H. Ebner, D. Fritsch (ed.), Digital Photogrammetric Systems, Wichmann, Karlsruhe, pp. 111–121.

Rissanen I. (1987) *Minimum description length principle*, Encyclopedia of Statistical Sciences, Vol. 5, pp. 523–527.

Schickler W. (1993) *Towards automation in photogrammetry*, Geodetical Info Magazine.

Sester M., W. Förstner (1989) *Object location based on uncertain models*, in H. Burkhardt, K.-H. Höhne, B. Neumann (eds), Mustererkennung 1989, 11. DAGM-Symposium, Hamburg, Vol. 219 of Informatik Fachbericht, Springer, pp. 457–464.

Suetens P., P. Fua, A. J. Hanson (1992) *Computational strategies for object recognition*, ACM Computing Surveys, Vol. 24, No. 1, pp. 5–61.

Sugihara K. (1986) *Machine Interpretation of Line Drawings*, MIT-Press.

van Hentenryck P. (1989) *Constraint Satisfaction in Logic Programming*, Logic Programming Series, MIT Press, Cambridge, MA.

Waltz D. L. (1975) *Understanding line drawings of scenes with shadows*, in P. H. Winston (ed.), Psychology of Computer Vision, McGraw-Hill, New York, pp. 19–91.

Weidner U. (1996) *An approach to building extraction from digital surface models*, ISPRS.

Weidner U., W. Förstner (1995) *Towards automatic building extraction from high resolution digital elevation models*, ISPRS Journal, Vol. 50, No. 4, pp. 38–49.