Solid Modeling for Building Extraction from Aerial Images

A. FISCHER AND V. STEINHAGE

Institut für Informatik I, Universität Bonn
Römerstraße 164, D-53117 Bonn
http://miro.cs.uni-bonn.de/~steinhag/txt/luft.html
email:andre|steinhag@cs.uni-bonn.de

Abstract

This paper discusses the role of solid modeling in the application area of building extraction in aerial images. The demands of the application require an explicit representation of all visible and geometrically relevant components of the building surfaces on the one hand and a generic structure to cover all occuring shapes of buildings. These requirements are met by a hybrid modeling architecture. To represent building structures instead of general-purpose solids the solid modeling scheme has to be enhanced to a building modeler incorporating domain specific knowledge. At last modeling has to cover the gap between three-dimensional object models and two-dimensional images by integrating object modeling and sensor modeling.

1 Introduction

There is an increasing need for 3D building extraction from aerial images for various applications such as town planning, environmental- and property-related studies or transmitter placement in telecommunication.

Aerial images usually reveal on one hand a certain amount of information not relevant for the given task of building extraction, e.g. vegetation, cars, building details. On the other hand there is a loss of relevant information due to occlusions, low contrasts or disadvantageous perspectives. Therefore a promising concept for automated building reconstruction must incorporate a sufficiantly complete model of the objects of interest.

In BRAUN *et al.* 1995 we gave an overview on related work in the field of model-based 3D building extraction. Among all the systems generic approaches are from outstanding importance to meet the modeling complexity of the building domain. FUA AND HANSON 1987 employ only simple box-type primitives but propose an explicit representation of legal primitive combinations to describe complex building aggregates. The appoaches from DICKINSON *et al.* 1992 and BERGEVIN AND LEVINE 1993 combine generic object models composed from general volumetric primitives with an explicit modeling of projective object appearances in terms of an image model but neglect the definition of elaborated schemes for representing domain specific variations and combinations of the primitives.

In this paper we develop a *domain specific* and generic modeling scheme for buildings in the application field of aerial image analysis based on the principles of a solid modeler by defining appropriate volumetric primitives and combination schemes which allow to represent building *variations* as well as building *combinations*. At last our modeling approach reveals also a tight coupling between 3D object models and the corresponding 2D projective object appearences in the sense of an explicit image model.

2 Solid Modeling

A modeling scheme for buildings in the application field of building extraction from aerial images has to meet especially two requirements:

(1) The modeling scheme must allow the explicit representation and processing of all *visible* building components, i.e. surfaces, edges and corners.

(2) The modeling scheme has to capture not only different dimensions of some basic building models but also combinations and variations of building structures.

The first requirement is obviously met by ***boundary representation*** (B-Rep) schemes which encode each solid in terms of dividing its surface into a collection of faces in some convenient fashion. Employing half-edge structures for the implementation of boundary models allows the direct adressing of edges, corner vertices and edge loops.

The second requirement is met by ***constructive solid geometry*** (CSG) which encodes each solid in terms of a CSG tree where the leaf nodes correspond to instantiated volume primitives and the inner nodes correspond to regularized set operations combining the volume primitives successively to complex object shapes.

Thus both requirements suggest the combination of B-Rep and CSG for building modeling based on a solid modeling approach.

We employ the solid modeler BN-SOLID described by FISCHER AND STEINHAGE 1995 as the basis for our approach on building modeling. The primary data structures of BN-SOLID encode boundary models of solids in terms of half-edge structures. The operational facilities of constructive solid geometry are implemented as macro operations for instantiating volumetric primitives and combining given object models to new ones with regularized Boolean set operations.

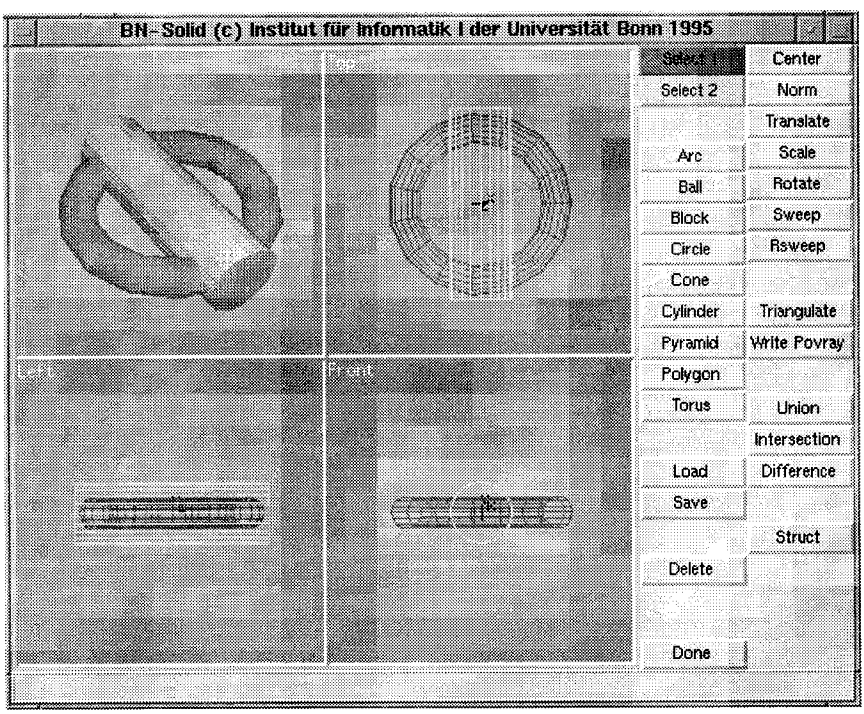


Figure 1: The hybrid modeler BN-SOLID.

The standard primitives offered by BN-SOLID are the parallelepiped (*Block*), the sphere (*Ball*), the cylinder (*Cylinder*), the cone (*Cone*) and the torus (*Torus*). Each instantiated primitive is represented by a half-edge structure. Curved surfaces are approximated by polygonal patches (s. figure 1).

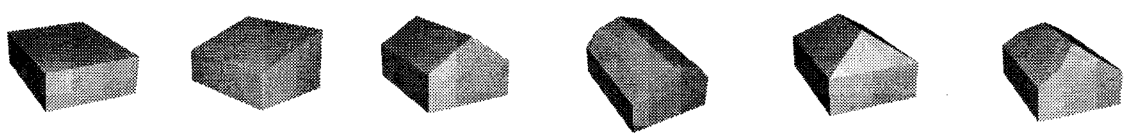
BN-SOLID also offers the definition of sweep objects in the sense of generalized cylinders. Each generalized cylinder is defined by a cross section employing the operations *Polygon*, *Circle* or *Arc* and the extrusion of the cross section along a straight or cyclic path employing the operations *Sweep* or *RSweep*, respectively.

To apply the regularized Boolean set operations *Union*, *Intersection* and *Difference* on two models in an appropriate way spatial transformations (*Translate*, *Rotate* and *Scale*) are available.

The results of all operations are visible within one window under arbitrary perspectives and three windows under orthogonal wire frame projections parallel to each axis of the coordinate system. Especially the orthogonal projections are helpful in positioning two objects for applaying a regularized Boolean set operation.

3 From Solid Modeling to Building Modeling

Towards a building modeler appropriate building primitives which represent a large number of basic building types are necessary. Figure 2 shows the shapes of some building primitives employed within our approach.



Bungalow | Penthouse | Saddle Roof | Mansard Roof | Hipped Roof | Gambrel Roof

Figure 2: Shapes of some building primitives.

Obviously building models have to describe more than pure geometry. To describe buildings instead of general solids we label all features with domain specific attributes:

- all faces F_i are labeled to be roofs or walls:
  $$L_F(F_i) \in \{\mathcal{R}, \mathcal{W}\} ;$$
- all edges E_j are labeled to be
  - vertical, horizontal or sloped,
  - placed between two walls, between two roof surfaces or between a roof and a wall:
    $L_E(E_j) \in S_E$ with
    $S_E := \{\rightarrow, \uparrow, \nearrow\} \times \{\mathcal{RR}, \mathcal{RW}, \mathcal{WW}\}$;
- all corners C_k are labeled with the number and types of joining edges (m is the largest number of edges joining in a corner):
  $$L_N(C_k) \in \bigcup_i {S_E}^i,\ 3 \leq i \leq m.$$

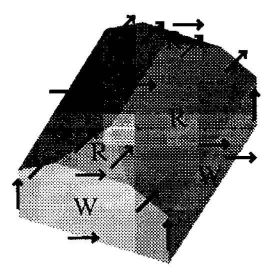


Figure 3: Labeled building primitiv.

This domain specific attribute labeling is essential for image processing, as different feature types are likely to appear different in digital images (cf. KORT *et al.* 1996).

Within the task of building extraction from aerial images the dimensions of each building are a priori unknown. Thus the building modeler has to describe generic types of building shapes instead of complete instantiated building descriptions and parameter estimation is one of the crucial subtasks in building extraction.

To describe generic types of building shapes with *variable geometry* we employ a characteristic set of parameters for each building primitive. Obviously we will not find only building shapes corresponding to our set of building primitives. Instead we will find building shapes which reveal variations and combinations of building primitives. To meet this demand of *variable topology* of building shapes we introduce *building parts* in our modeling scheme.

Building parts can be classified into two types. *Terminals* are parametrized building parts which describe the shapes of the endings of building primitives. *Connectors* are parametrized building parts to describe the shapes which result by combinations of two or more building primitives. Figure 4 shows two terminals and one connector as examples of building parts. Each building part reveals closed loops of *links* defining *link faces* of the building parts (dotted loops in figure 4). Obviously terminals show one link face while connectors reveal two or more link faces. The connection of every two building parts is constrained by the shape and orientation of the corresponding link faces in topology (parallelisms, symmetries etc.) as well as in geometry (angles, distances etc.).

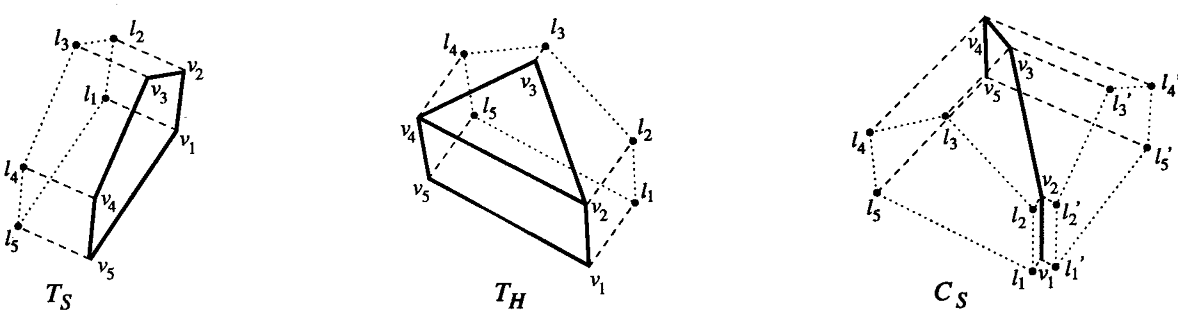
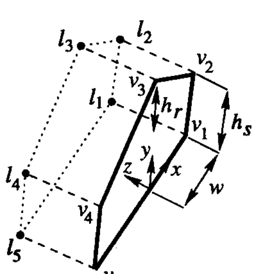


Figure 4: The saddle roof terminal T_S, the hipped roof terminal T_H and the L-shaped saddle roof connector C_S.

Each terminal and each connector is encoded by a set of equations relating the three-dimensional vertex coordinates to the shape parameters. Figure 5 shows this parametric description of the vertices $v_i = (x_i, y_i, z_i)^t$ of the saddle roof terminal T_S by it's shape parameters p_s of width w, height h_s of the storey block and roof height h_r.



$$v_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = R_i(p_s) = \begin{pmatrix} R_i(x) \\ R_i(y) \\ R_i(z) \end{pmatrix} p_s$$

v_i	$R_i(x)$	$R_i(y)$	$R_i(z)$
v_1	$+w$	0	0
v_2	$+w$	$+h_s$	0
v_3	0	$+h_s + h_r$	0
v_4	$-w$	$+h_s$	0
v_5	$-w$	0	0

Figure 5: The parametric description for the saddle roof terminal T_S.

It is now straightforward to model building variations and combinations by combining building parts. Each building primitive (s. figure 2) is composed from two terminals of the same type. Figure 6 shows this encoding for a saddle roof primitive.

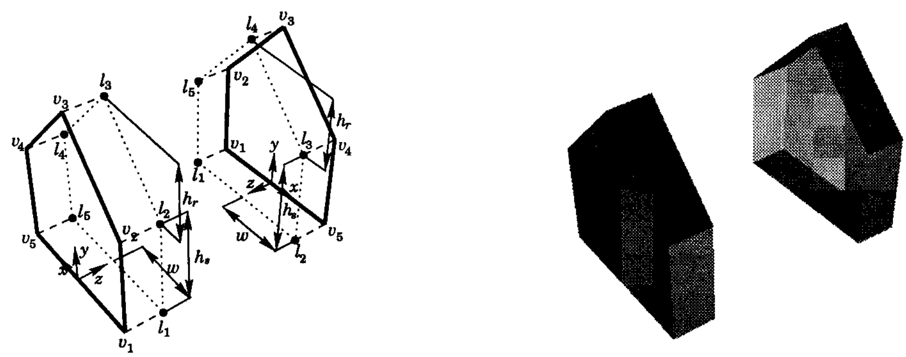


Figure 6: Constructing a saddle roof primitive by combining two saddle roof terminals.

Combining different types allows to model variants of building primitives. Figure 7 shows the example of a primitive variant with two different roof endings by combining one hipped roof terminal T_H with one saddle roof terminal T_S. Employing connectors within this aggregation process gives the opportunity to model arbitrary complex building structures. Figure 7 shows the example of the combination of a L-shaped saddle roof connector C_S with one saddle roof terminal T_S and one hipped roof terminal T_H describing a quite complex building structure showing an L-shaped ground plan and different roof structures.

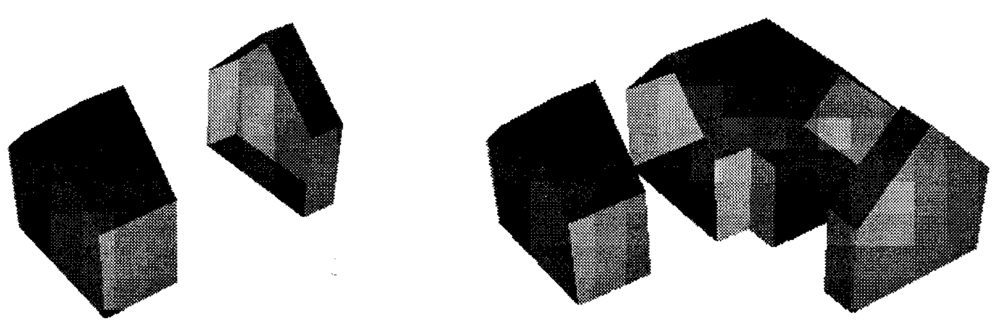


Figure 7: Left: the model of a building variant revealing a roof with a saddle roof terminal and a hipped roof terminal. Right: the model of a building combination revealing a L-shaped connector which combines a saddle roof terminal and a hipped roof terminal.

Note that all models of building parts and arbitrary complex building aggregates are represented by the building modeler as parametrized and closed b-rep descriptions where special attributes identify links, link edges and link faces. These attributes cause for example that link faces are not displayed in the shaded image.

4 Image Modeling

Model-based object recognition requires the matching of features observed in an image with object models. 2D features extracted from a image are described in a viewer-centered coordinate system and do not immediately correspond to the describing 3D features of the object models which are represented in an object-centered coordinate system. Therefore viewer-centered object models can ease the matching of image features with object models.

Furthermore viewer-centered object models can incorporate application specific sensor and illumination modeling. Note that the aerial images are oriented images, i. e. we have explicit knowledge about the camera parameters of position, orientation, focal length etc. as well as about the timestamp of the flight.

Multiview object representations describe all possible appearances of spatial objects by finite sets of view classes which were called *aspects* by (KOENDERINK AND VAN DOORN 1979). Each aspect assembles all object views which are isomorphic regarding a predefined set of outer features, eg showing the same object surfaces. The *aspect graph* assembles all aspects of an object according to their neighbourhood (s. figure 8).

The complexity of aspect graphs can be significantly reduced by focussing on domain relevant views (e.g. on top views when dealing with aerial images) and combining mirror views (e.g. due to object symmetries).

A projection of a building model becomes a line drawing obtained by perspectively projecting all visible edge segments onto the image plane. Thus the image model encodes an object view as a planar graph — the so called *junction graph* — which nodes and edges represent the junctions and the lines of the corresponding line drawing.

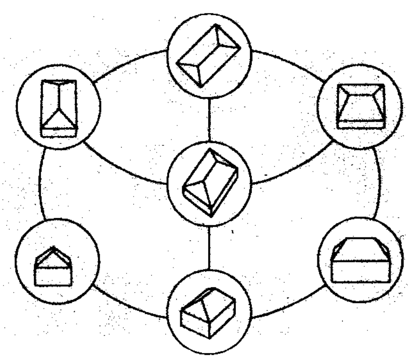


Figure 8: Aspect graph of the hipped roof building primitive.

All components of a junction graph inherit attributes from the corresponding spatial object features:

- all regions R_i defined by minimal line cycles are labeled as projections of roofs or a walls where the new label $\mathcal{B}$ denotes background segments:
$$L_R(R_i) \in \{\mathcal{R}, \mathcal{W}, \mathcal{B}\} \; ;$$
- all lines L_j are labeled according to the orientation of the corresponding edges and the labeled regions which they are separating:
$$L_L(L_j) \in S_L \quad \text{with} \quad S_L := \{\rightarrow, \uparrow, \nearrow\} \times \{\mathcal{RR}, \mathcal{RW}, \mathcal{WW}, \mathcal{WB}, \mathcal{RB}\} \; ;$$
- all junctions J_k — except the so called T-junctions which represent occlusions — are labeled according to the number, the types and the visibilty of the edges joining in the corresponding object corner:
$$L_J(J_k) \in \bigcup_i {S_J}^i, \; 3 \leq i \leq m \quad \text{with} \quad S_J := \{visible, unvisible\} \times S_L \; .$$

Two views of a building object belong to the same aspect if their visible-edge projections show isomorphic attributed junction graphs.

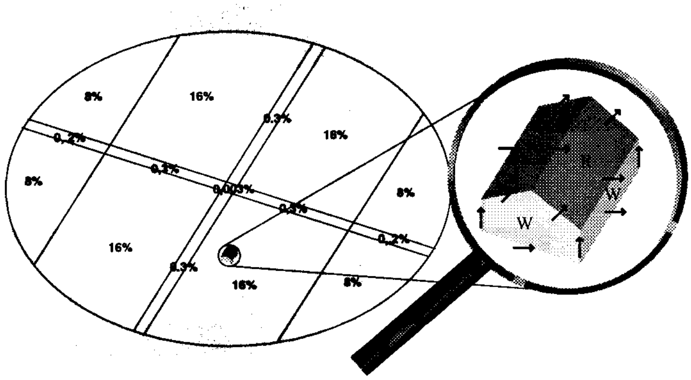


Figure 9: Viewplane and aspect regions of a building object.

The viewing space is described as a viewing plane of diameter $d = h\sin(\alpha)$ with flight height h and angle of aperture α (s. figure 9). The viewing plane is partitioned into *aspect regions* based on the approach described by FISCHER AND STEINHAGE 1996. The sizes of the aspect regions are used to define the occurrence probabilities of corresponding aspects.

To handle incomplete and uncertain results in feature extraction we employ the idea of *primitive-based aspect hierarchies* described by DICKINSON *et al.* 1992: Three hierarchy levels describe the views on building primitives, building faces, and on groups of edges and corners. To associate the representation levels we utilize likelihood methods based on the a priori probabilities derived from the partition areas.

Figure 10 sketches the building specific aspect hierarchy. Aspects and aspect components are encoded in terms of *relationships* among the attributed image features like adjacancies, intersections, parallelisms (thickened lines), skewed symmetries (dashed lines) etc.

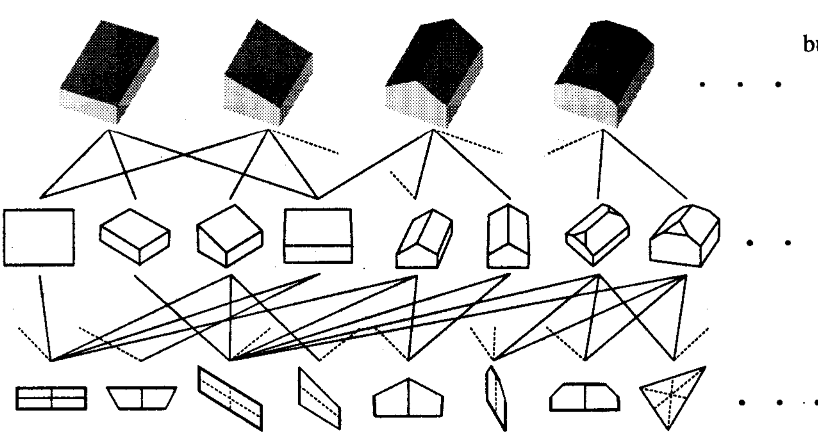


Figure 10: The building specific aspect hierarchy.

5 3D Building Extraction

The aerial images are given as digital raster images with multiple overlap. The images are oriented, i. e. their exterior and interior camera parameters as well as the time-stamp of the flight are known. Due to the overlaps and the known orientation of the images we can derive 3D information by employing multi image stereo analysis. The task of 3D building extraction from these aerial images is performed in our approach according the paradigm of hypotheses generation and verification. Figure 11 gives a sketch of the following processing steps:

(1) *Extraction of image features*: Image processing on aerial stereo images extracts relational symbolic image descriptions in terms of feature aggregates where features are lines, line junctions and regions (cf. FUCHS AND FÖRSTNER 1995).

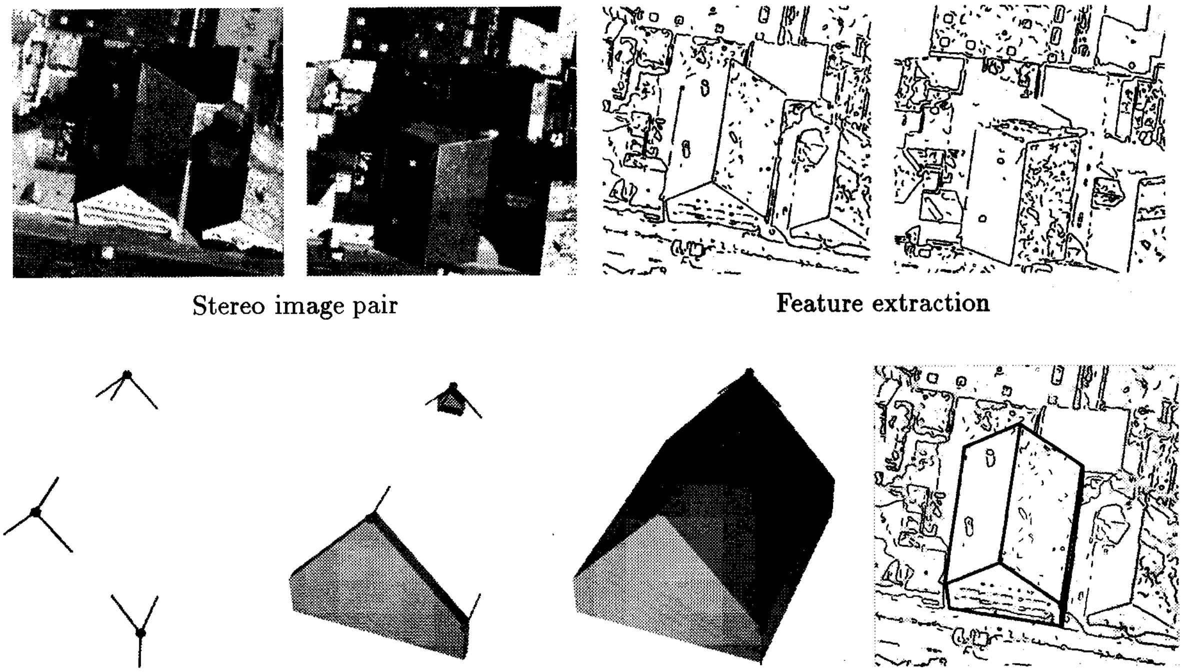
(2) *3D reconstruction of vertex aggregates*: Employing multi image stereo analysis on the extracted features aggregates results in the derivation of spatial feature aggregates, especially vertex aggregates for which the imaging geometry leads to sharp constraints within the stereo analysis (s. LANG AND FÖRSTNER 1996). Each vertex aggregate consists of the 3D coordinates of its node point and the spatial orientations of all edges and face segments joining in the vertex.

(3) *Indexing for building part hypotheses*: Due to their specific topology and geometry each vertex aggregate votes for certain building corners of our terminal or connector elements. Domain specific constraints are used to inspect if multiple vertices vote for the same terminal or connector element. All vertices voting for a building part determine — at least partial — its parameters of shape and location.

(4) *Aggregation of building parts to building hypotheses*: Hypotheses for complete buildings are derived by aggregating those building part hypotheses where the instantiated parameters of shape, location and orientation of their link faces match. Note that due to the 3D reconstruction of only local vertex aggregates and their ambiguous matching to different corners of building parts the indexing and aggregation steps in general will derive multiple competing hypotheses for buildings.

(5) *Generation of aspect hierarchies for building hypotheses*: Each derived 3D building hypothesis is projected back into the given aerial images. These projections are represented as aspect hierarchies (s. figure 10). In contrast to the general approach of DICKINSON *et al.* 1992 the viewing directions within our approach are well known but free parameters of the building geometry may cause even now different aspects of a building hypotheses. The aspect hierarchies are computed for all competing building hypotheses and for all aerial images.

(6) *Verification of the building hypotheses*: The aspect hierachies are now used to match the projections of the building hypotheses against the extraced edge segments of the aerial images according a *recognition-by-components*-strategy (cf. BIEDERMAN 1987). For this matching process we employ constraint logic programming techniques suggested by KOLBE *et al.* 1996. A control strategy for a backtracking approach is proposed by KORT *et al.* 1996. The matching results in a scored quality measurement of the building hypotheses on the basis of the matched image data.



Stereo image pair — Feature extraction

3D vertex aggregates — Building parts — Building hypothesis — Hypothesis verification

Figure 11: An example for the process of 3D building extraction from a stereo image pair: the results of feature extraction, three reconstructed 3D vertices indexing for two saddle roof terminals, the resulting building hypothesis of a saddle roof house with rectangular groundplan and the verification of this hypothesis by projecting it back into the images (in this figure only shown for one stereo image).

Figure 11 shows an example for the sketched process of 3D building extraction from a stereo pair image. By stereo analysis on the extracted image features among others three 3D vertex aggregates are derived which reveal shapes of building corners. These vertex aggregates are used for indexing into the data base of building parts. Among others two vertex aggregates vote for one saddle roof terminal thereby determining the shape parameters of width, height of the storey block and roof height for this terminal. Obviously we need for determining the height of the storey block a terrain model of the observed territory, i. e. the height of the ground. The third vertex aggregate votes also — among others — for a saddle roof terminal thereby only determining the height of the whole building, i. e. the sum of the storey block height and the roof height. Both saddle roof terminals meet the contraints of relative orientation, location, distance and shape to vote for a common building hypothesis, i. e. a saddle roof building with rectangular groundplan thereby unifying the common shape parameters. At last this building hypothesis is projected back into the stereo images. Another building hypothesis generated from the three vertex aggregates is for example a building with mansard roof and rectangular groundplan. But among all matches between the extracted image features of the stereo images on the one hand and the backprojections of the building hypotheses into these images the saddle roof building hypotheses of figure 11 show the best scored match and therefore verifies this building hypothesis.

6 Conclusion

We have presented an approach to 3D building extraction from aerial images based on a tight coupling of a generic 3D building model and an explicit 2D image model. The generic building model includes the explicit representation of variations and combinations of building shapes. Object and image model reveal a hierarchical organization which allows a robust building extraction according the *recognition-by-components*-strategy. We currently work on several test data sets which reveal complex building structures like that of the reconstruction result shown in figure 12. Future work will include an explict illumination model for predicting shadows.

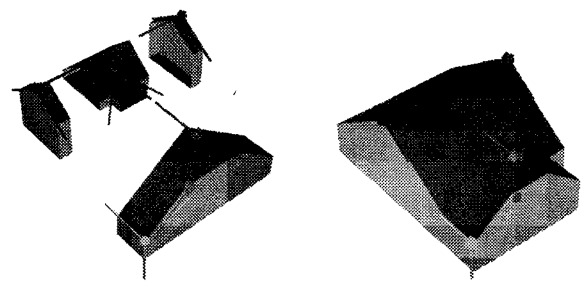


Figure 12: Derived building part hypotheses and the aggregated building hypothesis of a complex structured building.

Acknowledgements — This work was largely done in cooperation with the Institut für Photogrammetrie and the Institut für Informatik III, Universität Bonn within the project "Semantic Modeling and Extraction of Spatial Objects from Images and Maps" especially in the subproject "Building Extraction" which is supported by the German National Research Council (Deutsche Forschungsgemeinschaft, DFG).

References

BERGEVIN, R.; LEVINE, M. D. (1993): Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings. *IEEE T-PAMI*, 15(1):19–36, 1993.

BIEDERMAN, I. (1987): Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, 94:115–147, 1987.

BRAUN, C.; KOLBE, T.H.; LANG, F.; SCHICKLER, W.; STEINHAGE, V.; CREMERS, A.B.; FÖRSTNER, W.; PLÜMER, L. (1995): Models for Photogrammetric Building Reconstruction. *Computer & Graphics*, 19(1):109–118, 1995.

DICKINSON, S.; PENTLAND, A.; ROSENFELD, A. (1992): 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE T-PAMI*, 14(2), 1992.

FISCHER, A.; STEINHAGE, V. (1996): *On the Computation of Visual Events in Aspect Graph Generation.* 1996. Submitted Paper.

FISCHER, A.; STEINHAGE, V. (1995): *A Hybrid Approach on Solid Modeling Integrating B-Rep and CSG.* Int. Rep. I/95/3. Inst. f. Informatik I, Univ. Bonn, 1995.

FUA, P.; HANSON, A. J. (1987): Resegmentation Using Generic Shape: Locating General Cultural Objects. *Pattern Recognition Letters*, 5:243–252, 1987.

FUCHS, C.; FÖRSTNER, W. (1995): Polymorphic Grouping for Image Segmentation. In: *5th Int. Conf. on Computer Vision, ICCV 1995, Boston*, 175–182, 1995.

KOENDERINK, J. J.; VAN DOORN, A. J. (1979): The Singularities of the Visual Mapping. *Biological Cybernetics*, 24:51–59, 1979.

KOLBE, TH. H.; PLÜMER, L.; CREMERS, A. B. (1996): Using Constraints for the Identification of Buildings in Aerial Images. In: *Practical Applications of Constraint Technology, PACT'96*, 143–154, 1996.

KORT, A.; POGODA, A.; STEINHAGE, V. (1996): Employing Aspect Hierarchies for Building Detection in Digital Images. In: *Mustererkennung 1996, 18. DAGM-Symposium, Heidelberg*, Informatik aktuell, 217–224. Springer, 1996.

LANG, F.; FÖRSTNER, W. (1996): Surface Reconstruction of Man-Made Objects using Polymorphic Mid-Level Features and Generic Scene Knowledge. *Zeitschrift für Photogrammetrie und Fernerkundung*, 6:193–201, 1996.