3D-CITY MODELING WITH A DIGITAL ONE-EYE STEREO SYSTEM

Felicitas Lang, Wolfgang Förstner
Institute für Photogrammetrie, Universität Bonn
Nußallee 15, D-53115 Bonn
Tel.: (228) 732901, Fax: (228) 732712
e-mail: Felicitas.Lang/Wolfgang.Foerstner@ipb.uni-bonn.de

Commission IV, Working Group 4

KEY WORDS: Building extraction, City-Models, 3D-GIS, CAD, Digital Photogrammetric Systems, Semiautomation

ABSTRACT

3D-city information is crucial for a number of applications in city planning, environmental control or for telecommunication. We describe a semiautomatic system for acquiring the 3D-shape of buildings as topographic objects. Buildings are either modeled as a freely structured union of basic shape primitives or as prisms with an arbitrary ground plan, covering a large percentage of existing buildings. Interaction takes place in only one image, requiring the operator to specify the approximate structure and shape of the buildings. 3D-reconstruction including both, height determination and form adaptation, is performed automatically using various matching tools. The paper describes the features of the system and reports on its efficiency based on an extensive test.

1 MOTIVATION

There is an increasing demand for three-dimensional (3D) building extraction as such data is needed for a large number of tasks related to measurement, planning, construction, environment, transportation, energy, property management and transmitter placement. However, systems being capable to capture entire cities in a relatively short time, at a reasonable cost and with topographic accuracy are not available. The reasons are manifold. One certainly is the lack of efficiency, or the costs for data acquisition, which are higher by a factor 2-5 than comparable conventional 2D-data. But also the diversity of users and the resulting lack of clear specification for 3D-data hinder the developments of operational systems. Even if such systems would exist, practitioners would hesitate to acquire such data without having a clear cut application, as current geoinformation systems do not support the management of 3D-data.

Though analytical film based systems to a certain degree can be used for 3D-data acquisition of buildings, efficiency can only be increased by introducing automation, i. e. image processing tools into the systems. While orientation algorithms in Digital Photogrammetric Systems (DPS) are highly developed and the production of digital orthophotos and the (semi)automatic acquisition of Digital Elevation Models (DEMs) [Krzystek, 1991] becomes a standard in Photogrammetry, nearly no support for the extraction of cartographic features is provided by DPSs. Intensive support of the operator by automatic tools for mapping will highly increase the acceptance of DPSs due to the overall higher productivity.

For complex city structures fully automatic approaches for cartographic feature extraction are of high interest. The state of the art of automatic and semiautomatic approaches has been presented at the Workshop in Ascona [Grün et al., 1995] (see also [Liedtke et al., 1991], [Burns and Riseman, 1992], [Price and Huertas, 1992], [Bignone et al., 1996]).

These procedures, however, do not seem to be operational yet. The main reason is the lack of models for linking the object classes (buildings, roads, vegetation or rivers) with their appearance in aerial images and the lack of strategies for bridging the gap between the iconic low level description, i. e. the image raster, with the semantic high level description, i. e. the concepts of the cartographic objects in concern. Therefore, the present goal of most developments is to achieve partial automation by matching and mensuration tools and leaving the application dependent decisions to the human operator (cf. [Quam and Strat, 1991], [Mundy et al., 1992], [Leberl et al., 1994], [Heuel and Nevatia, 1995]). This allows a smooth transition to more and more automated solutions.

The semi-automatic approach for model-based building extraction from digital images, was motivated by these needs. A first version of the system has been described by [Lang and Schickler, 1993] and [Lang et al., 1995]. It has been extended by the volumetric modeler (cf. [Englert and Gülch, 1996]).

The system is based on the following rationales:

- CAD-type modeling for easing link to thematic attributes.
- Attributed Boundary Representation for supporting interaction and visualization.
- Multiple Image Evaluation for increasing completeness , flexibility and reliability of data acquisition.
- One-Eye Stereo for easing interaction and increasing acceptability by non-experts.

The system is part of our research on automatic building extraction (cf.[Sester and Förstner, 1989], [Förstner and Pallaske, 1993], [Braun et al., 1995], [Brunn et al., 1995]). The system serves two purposes:

- It is a *test bed* for the developments in automatic image analysis, especially for integrating automatic image analysis procedures.
- It is a *practical tool* for 3D-data acquisition which allows to analyze the interface between automatic procedures and operator interaction on large data sets.

In the following the system is described in more detail (section 2). The results of empirical test are given in section 3.

2 THE SYSTEM

This section describes the most important properties of the system, namely its internal building model, its interaction capabilities and its features of automation.

2.1 Modeling Buildings

Automatic as well as interactive systems need an internal model of the objects to be acquired. Buildings show an amazingly high diversity in structure which is increasing due to new styles being developed or invented. However, a large percentage of buildings show regularities which allow to describe them using a small set of rules. Depending on the context within an acquisition system we need to distinguish volumetric, boundary and wire frame type of models.

Volumetric Models Volumetric models are most intuitive for describing solids. In our context this on one hand eases interaction. On the other hand, this type of high-level representation allows to establish a link between GIS- and CAD-systems, which seems to be important for many applications.

Many simple buildings can be described using a few parameters. Such *parametric models* can effectively be used for buildings with flat, gable or hip roof, requiring 3, 4 or 5 parameters for specifying the form. More general shapes are *prisms* with polygonal ground plan and fixed height, which are frequently found in the central part of cities. The number of parameters is $2n+1$ for a ground plan with n corners.

Obviously the ability to model buildings using only these two types of representation is limited. But combinations of these appear to be sufficient for modeling a large percentage. This suggests to immediately use the modeling tool from CAD, namely **C**onstructive **S**olid **G**eometry (CSG) for describing complex shapes. CSG uses basic parametric shape primitives and allows to combine them by boolean operations, namely union, intersection and difference. Complex objects internally are then represented as a *CSG-tree* where each node links two branches of the object with the boolean operation and a geometric transformation of the reference system of the branch into a common reference system. The leaves of the tree are instantiated primitives.

Observe that prisms do not immediately fit into that system, as they in a weak sense contain free forms, namely the polygonal ground plan.

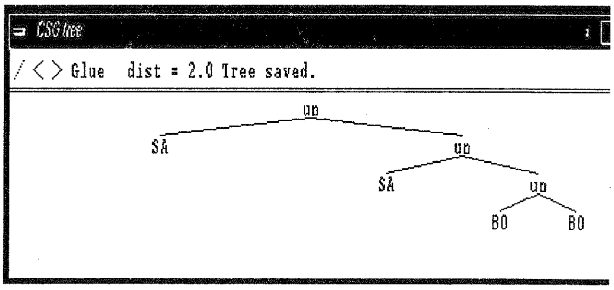
We use both types of representation. For the CSG-representation we use as primitives the box, the chock, the half chock, the cone, the cylinder and the rectangular pyramid. In order to ease acquisition we also have the above mentioned basic parametric building models as primitives.

Boundary Representation The boundary representation of parametric and prismatic models are used to support the interaction with the user and for visualization purposes.

This allows to store attributes to all faces of the buildings, especially surface texture and thus opens the door to photo-true visualization where the texture can be fused from all images [Leberl et al., 1994]. Much more, interactive measurements could be taken at a later stage, here, however, working in views which not necessarily coincide with those of the original images.

We do not provide a complete boundary representation at the moment, as Computer Graphic packages allow visualization without explicit intersection of the objects. We however will use the explicit boundary representation for two tasks: 1.) learning regularities at the object for supporting image analysis and 2.) iconic matching of the final result of the reconstruction with the content of all images for improving the accuracy and checking the consistency.

Figure 1: CSG-tree, containing 4 primitive showing the union of two saddle roofs and two boxes.



dist = 2.0 Tree saved.

Wire Frame Representation A wire frame representation is used for interaction. Its model is shown overlayed with the raster image for interactively fitting the model to the image content. A hidden line algorithms is running real time in order to ease cognition of the 3D-structure.

The wire frame model also is used within the model-image matching procedure based on extracted image edges.

A Sample Representation The example shows the internal representation for a building with a gable roof. The coordinates of the corners are given in a local coordinate system in dependence on a set of parameters, here length l, width b height of building h_1, and height of roof h_2 The coordinates of the corner points 2 and 8 are given in the local coordinate system:

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ X_8 \\ Y_8 \\ Z_8 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} l \\ b \\ h_1 \\ h_2 \end{pmatrix}$$

Four additional *pose parameters*, three shifts and the azimuth, are needed to represent the relation of that building primitive with respect to the world coordinate system.

The model also contains all edges and faces, thus implicitly also all relations and constraints necessary to specify the polyhedral object to be a gable-roof building.

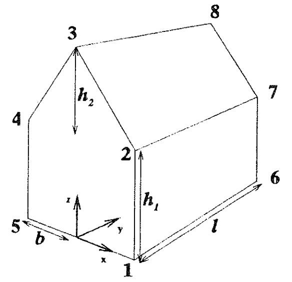
2.2 Interaction

We start with digital or digitized images with known interior and exterior orientation. A DEM is useful for providing approximate values.

Extracting CSG-primitives. The 3D-shape extraction of each single primitive of a CSG-model is performed in several steps:

1. *Choice of the type of boolean operation* to previously extracted primitives, if appropriate, and choice of type of *primitive* from a prespecified set.

Figure 2: Parameterization of the building primitive with gable roof



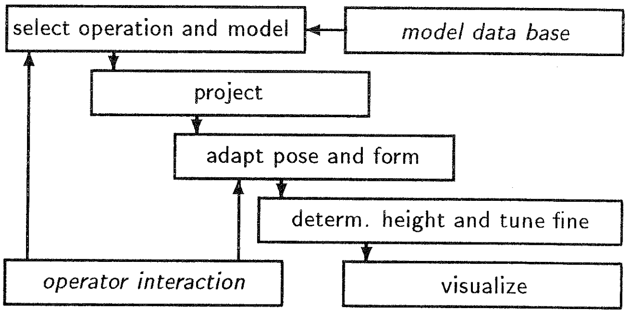
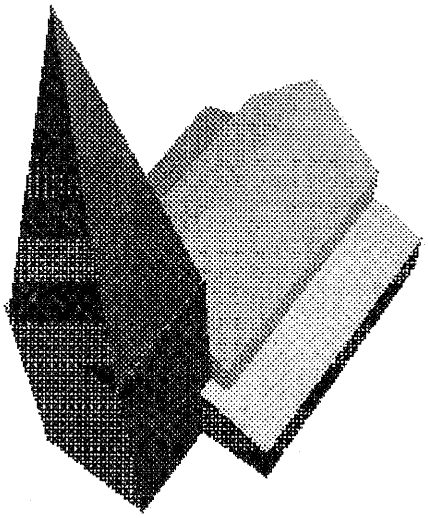


Table 1: Steps of semiautomatic building extraction

Figure 3: Extracted building block consisting of two saddle roof primitives, a box and a pyramid



2. *Projection* of the wire frame model of the primitive into the image section *after hidden line removal*. This projection is based on approximate values for the height and default values for the parameters of the primitive. If it is not the first primitive the height of the base level of the previous primitive is used. If the primitive has already been selected, the old parameters are used as default.
3. *Interactive adaptation* of the form of the wire frame model to the image content (cf. below). During this action the pose and form parameters are determined and the primitive is visualized in real time using an approximate height of the primitive. The operator may choose that resolution of the image which is suited best for this interaction.

   During this process a *snap-function* can be activated. Based on some user specified snap-radius relations to already extracted primitives are automatically established and in case they are found shown to the operator and visualized at the screen. This enables gluing of points, edges and especially faces of neighboring primitives.

   The accuracy of this adaptation can be reduced if a subsequent fine tuning takes place.
4. *Height determination* using either an automatic tool (cf. below) or interactively a second image if it is the first primitive within the CSG-model of the building.

   Optional, the parameters can be *fine tuned* using a least squares adjustment (cf. below) after matching the wire frame model to the edges extracted in all available images.
5. (optional) *Visualization* of the parts extracted so far for checking the result.

Only the automatic fine tunig in step 4 requires digital images. If a sufficiently accurate adaptation of the model to the images is available the procedure can also be realized in an Analytical Plotter with Stereosuperimposition.

Extracting Prisms. The extraction of prisms relies on automatically extracted image edges. They have to be provided by some edge extraction procedure. The operator interactively *selects* a sequence of image edges representing the ground plan at eave level, thus assuming all the selected image edges to lie in a horizontal plane. Neighboring image edges are linked by intersection. In case of strong shadows where no image edges are available, the operator also can define break points of the polygonal ground plan. The geometry also may be changed interactively if necessary. Using an automatic tool or by interaction the height of the eave level is determined. Using the same procedure as for interactively adapting the parameters of CSG-primitives the height of the prism, i. e. the ground level is found.

2.3 Parameter Adaptation

The interactive adaptation of the form and pose parameters is performed in monocular mode. Stereo display is not necessary. An extension of the interactive steps for the case of a stereo display is simple.

The adaptation of the parameters is done in a sequence of steps, each specifying one or two parameters of the model. The parameters are changed in dependence of the sequence of two points specified by mouse clicks and in dependence of the type of primitive. Therefore the number of interaction steps is between 50 % and 100 % of the number of parameters. Which parameter actually is to be adapted is stored in a so called *association table* setup to support intuitive interaction.

The association table is build up according to the following rules:

- The first of the two points chosen is the reference point. It is kept fix unless the second point is identical to the first. In this case the primitive is shifted horizontally.

- The second point chosen together with the mouse key determines which parameters actually are changed.
- The second mouse key is reserved for horizontal – possibly virtual – edges of the primitive and allows a rotation of the primitive around a vertical axis through the first point.

Table 1 shows a subsection of the association table for the building shown in fig. 2. E. g. if point pair $(2,4)$ is chosen, either the form parameter b or – when pressing the second mouse key – the form parameter b together with the rotation can be changed by adapting point 4 to the image content.

point	1	2	3	4	$\cdots$
1	$v_x \quad v_y$	h_1	$h_1 \quad h_2$	$h_1 \quad b$	$\cdots$ $\cdots$
2	h_1	$v_x \quad v_y$	h_2	b $b \quad \alpha$	$\cdots$ $\cdots$
3	$h_1 \quad h_2$	h_2	$v_x \quad v_y$	h_2 $\cdots$	$\cdots$
4	$h_1 \quad b$	b $b \quad \alpha$	h_2	$v_x \quad v_y$	$\cdots$ $\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Table 2: Association table for controlling interaction

After selection of two points the parameters p_i are selected, which are to be adapted. The change of the mouse position during moving the second point determines the size of the parameter change. Using the *difference ratios* $\delta x_b/\delta p_i$ and $\delta b_b/\delta p_i$ of the image coordinates with respect of the parameters we obtain the coordinate changes Δx_b and Δy_b in the case of two parameters p_i and p_j by

$$\begin{pmatrix} \Delta p_i \\ \Delta p_j \end{pmatrix} = \begin{pmatrix} \frac{\delta x_b}{\delta p_i} & \frac{\delta y_b}{\delta p_i} \\ \frac{\delta x_b}{\delta p_j} & \frac{\delta y_b}{\delta p_j} \end{pmatrix}^{-1} \begin{pmatrix} \Delta x_b \\ \Delta y_b \end{pmatrix} \qquad (1)$$

In case only one parameter p_i is to be changed the equation system for determining the parameter is over determined. Then we use the pseudo inverse. In this case the second point moves on that image line which is determined by the specified parameter change, which appears to be meaningful.

This procedure holds for all combinations of parameters defined in the association table and is used for all types of primitives as well as for the height of the prismatic models.

Changes of the association table are simple and allow a fast adaptation to preferences of the operator.

2.4 Automation of Building Acquisition

Building extraction can use two features of Digital Photogrammetric Systems for the advantage of increasing the performance:

1. Images are digital. This allows the computer to have access to the image content. Mensuration and classification are two basic tasks the operator needs to perform, which both can be heavily supported by image analysis. Mensuration is the easier task as it only determines geometric properties of prespecified parameters.
2. DPS provide a direct and fast access to all images in concern without repeated interior orientation. Using matching algorithms for more than two images all information available in a photogrammetric block can be used simultaneously, increasing efficiency and allowing a reliable self diagnosis of the automatic procedures, based on the high redundancy available.

The difficulties in automation, mentioned in the introduction, and which are the motivation for developping a semiautomatic system, will become more transparent in the following.

Using Edge Information. Image edges are lines of high contrast, which easily can be extracted from the image. They carry essential geometric information. Their precision is in the range of 1/3 to 1/20 of the pixel size, depending on the noise level and the contrast.

Unfortunately, automatic edge extraction procedures produce too many edges due to objects which are not of special interest in the chosen context, e. g. caused by windows on roofs, texture, shadow or vegetation. At the same time the procedures are likely to miss important edges, due to limited contrast or occlusions.

Therefore separating the responsibility of man and machine seems to be appropriate: the geometric information is recovered automatically, while the decisive part, namely selecting the edges of interest is left to the operator – the procedure used for defining the ground plan of prismatic models at eave level.

Model to Image Matching. In case the system is provided with a model of what to expect in the image a matching between image and model can be performed. This matching performs an instantiation of those values of the model which are not fixed. These may be geometric parameters fixing the pose or the form or class labels fixing the (sub)category the object belongs to.

In our context the interaction provides both, the internal structure of the model as well as approximate values for the parameters of the primitives. The wire frame model therefore can be projected into the image in order to find correspondencies between the edges of that projection and automatically extracted image edges. Minimizing the geometric differences in the images in a least squares adjustment yields optimal values for the parameters left free, either only the pose or both, pose and form parameters.

The accuracy of the final result can be evaluated in order to decide on further steps of the anlysis.

One-Eye Stereo and Multi Image Matching. Interaction should take place in only one image if possible. The missing 3D-information should be automatically acquired from the other images. This allows to use standard workstations, eases interaction and increases acceptability of the system by non-experts.

The before mentioned model to image matching could use the information not only of one or two images but of several images. In general it is recommendable to have multiple overlap for building acquisition just to have views of all sides of the building, being useful for either getting full information on the building, not omitting occluded details or for orthophoto production, which definitely requires to have information on all surfaces visible from zenith.

Exploiting Tools from Digital Cartography. Interactive systems for data acquisition have been developped since several decades. All tools available there should also be used here to advantage for easing interaction.

The snap function mentioned above is an example. The system automatically checks a set of relations between different primitives within a CSG-tree. The definition of that set of course is context dependent. Here we started in looking for common faces of primitives to glue them together, if the user confirms the relation.

3 EXPERIMENTAL RESULTS

The system has been tested on large data sets. The performance of the system is investigated showing promising characteristics.

3.1 Test field München

The test field *München* contains appr. 2 km^2 of the central part around the area of the university, with large but complex buildings. The data specifying the test site are given in the table. The acquisition times are given in table for the three generalization levels, adapted from [Löcherbach, 1995].

image scale	1 : 15 000
focal length	153 mm
imagery	B/W
pixel size	15 μm
area	2 km^2
# of models	578
# of prisms	230

Table 3: Information on test area München

generalization level	models	total time [min]	time/model [min]
2 (medium)	249	461	1.85
3 (high)	295	548	1.85
4 (very high)	34	67	1.97
	prisms		time/prism
2 (medium)	7	18	2.57
3 (high)	128	446	3.48
4 (very high)	95	284	2.99

Table 4: Acquisition times for the test area München, adapted from [Löcherbach, 1995]

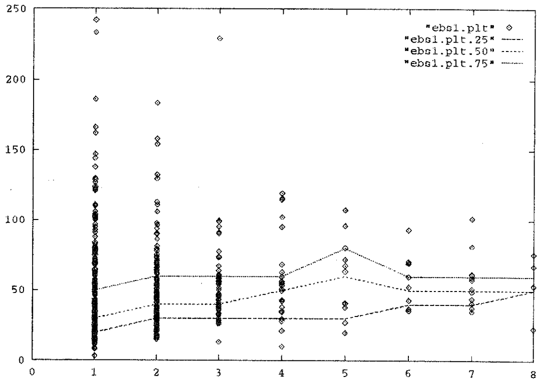
3.2 Test field Oedekoven

The testfield *Oedekoven*, near to Bonn, has been selected to acquire data for the first phase of an OEEPE-Test on *3D-City Information* (cf. [Gülch, 1996]. The area covers about 1.5 km^2, partly is sloped and contains buildings of many types. The statistics on the test data are given in the table.

image scale	1 : 12 000
focal length	150 mm
imagery	B/W
pixel size	11 μm
area	1.5 km^2
# of CSG-trees	672
# of primitives	1591
# of primitives per CSG-tree	2.5
% of boxes	66 %
% of gabled roof	18 %
% of hip-ed roof	3 %
% lop-sided gabled roof	13 %

Table 5: Information on test area Oedekoven

Figure 4: Netto time per primitive with median and 25 %- and 75 %-point,



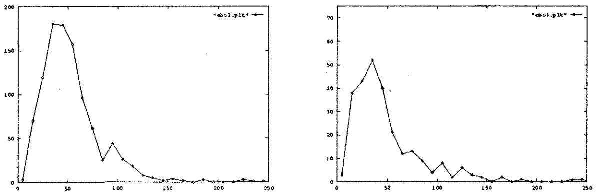
The times for acquiring the data are given in the following table. They contain all parts of the acquisition, including navigation through the image pyramid, selection of primitives, form adaptation, measurement of homologous points and 3D-visualization. We distinguished 4 levels of generalization: low generalization (1) corresponds to the finest details identifiable, medium level of generalization (2) and higher level correspond to leaving out details of 0.5 m, 1 m and 1.5 m appr.. The numbers for the different generalization levels in the table do not correspond to the same area.

generalization level	models, primitives	total time [min]	brutto time/ model [min]
1 (low)	1591	2427	1.53
a)	*1007*	*1181*	*1.17*
b)	*584*	*1246*	*2.13*

Table 6: Acquisition times for the test area Oedekoven
a) expert, mainly flat terrain
b) non-expert, partly sloped terrain

We distinguish brutto and netto time, the difference being the time for navigation, the time for deciding on modelling the structure of the building and the time for checking the result.

Figure 5: Number of primitives per netto time. Left: all primitives, right: only CSG-trees with single primitives, maximum number per time: 170, maximum time 250 s.



A closer look at the acquisition times reveils the following:

- The brutto times are shorter by appr. 25 % compared to the version used a year ago within the test *München*.

  The experienced operator, however, showed a significant higher performance.
- The netto times, covering the internal loop for fitting the primitives to the image content, however are not significantly different between the two operators (not shown)
- There is no significant difference in acquisition time between single-primitive and complex buildings (cf. fig 5). The median time per primitive lies in the range of 40 s. This time includes on-line checking using 3D-visualization. 75 % of all building primitives can be acquired in a time below a minute (cf. fig. 4)

4 OUTLOOK

This intermediate report on an semiautomatic system for 3D-data acquisition on a standard workstation with a one-eye capability has shown the efficiency of deriving 3D-city models from digitized aerial images. The next step in the development are investigations into the usefuleness of new automation procedures especially matching tools. Also an investigation into the accuracy of the acquired data is necessary. The flexibility of the setup will support further increases in efficiency for semiautomatic 3D-data acquisition.

Acknowledgements: This work was partly supported by DFG (Fo 180/2-2), BMBF (01 M 3018 F6), and DARA (50 TT 95 36) and realized in cooperation with the Institut für Informatik III, Universität Bonn.

REFERENCES

[Bignone et al., 1996] Bignone, F., Henricsson, O., Fua, P., and Stricker, M. (1996). Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery. In *Computer Vision '96*.

[Braun et al., 1995] Braun, C., Kolbe, T., Lang, F., Schickler, W., Steinhage, V., Cremers, A., Förstner, W., and Plümer, L. (1995). Models for Photogrammetric Building Reconstruction. *Computer & Graphics*, 19(1).

[Brunn et al., 1995] Brunn, A., Weidner, U., and Förstner, W. (1995). Model-based 2D-Shape Recovery. In Sagerer, G., P. S. and Kummert, F., editors, *Mustererkennung 1995*, pages 260–268. DAGM, Springer.

[Burns and Riseman, 1992] Burns, J. B. and Riseman, E. M. (1992). Matching Complex Images to Multiple 3D Objects using View Description Networks. In *Image Understanding Workshop, San Diego, California*, pages 675–682.

[Englert and Gülch, 1996] Englert, R. and Gülch, E. (1996). A One-Eye Stereo Sytem for the Aquisition of Complex 3D-Building Structures. *GIS*.

[Förstner and Pallaske, 1993] Förstner, W. and Pallaske, R. (1993). Mustererkennung und 3D-Geoinformationssysteme. In *Proc. of 3. Int. Anwenderforum für Geoinformationssysteme, Duisburg*.

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

[Gülch, 1996] Gülch, E. (1996). Zwischenbericht "Interaktives 3D-Erfassungssystem". Interner Bericht, Institut für Photogrammetrie, Universität Bonn.

[Heuel and Nevatia, 1995] Heuel, S. and Nevatia, R. (1995). Including Interaction in an Automated Modelling System. In *Proceedings International Symposim on Computer Vision*. IEEE.

[Krzystek, 1991] Krzystek, P. (1991). Fully Automatic Measurement of Digital Elevation Models. In *Photogrammetric Week, Stuttgart, Proceedings*, pages 203–214.

[Lang et al., 1995] Lang, F., Löcherbach, T., and Schickler, W. (1995). A One-Eye Stereo System for Semi-Automatic 3D-Building Extraction. *Geomatics Info Magazine*.

[Lang and Schickler, 1993] Lang, F. and Schickler, W. (1993). Semiautomatische 2D-Gebäudeerfassung aus digitalen Bildern. *ZPF*, 5:193–200.

[Leberl et al., 1994] Leberl, F., Gruber, M., Uray, P., and Madritsch, F. (1994). Trade-Offs in the Reconstruction and Rendering of 3D Objects. In Kropatsch, W. and Bischof, H., editors, *Mustererkennung, Proceedings*, pages 58–73. DAGM.

[Liedtke et al., 1991] Liedtke, C.-E., Busch, H., and Koch, R. (1991). Shape Adaptation for Modelling of 3D Objects in Natural Scenes. In *Proc. of CVPR, Hawaii*.

[Löcherbach, 1995] Löcherbach, T. (1995). System Performance of Semiautomatic Building Reconstruction. In *2nd Course on Digital Photogrammetry*. Inst. f. Photogrammetry, Bonn.

[Mundy et al., 1992] Mundy, J., Binford, T., Boult, T., Hanson, A., Beveridge, R., Haralick, R., Ramesh, V., Kohl, C., Lawton, D., Morgan, D., Price, K., and Strat, T. (1992). The Image Understanding Environment Program. In *Proceedings Computer Vision and Pattern Recognition*, pages 406–416.

[Price and Huertas, 1992] Price, K. and Huertas, A. (1992). Using Perceptual Grouping to detect Objects in Aerial Scenes. In *Internat. Archives for Photogrammerty, Comm. III, Washington*, volume 29, pages 842–855.

[Quam and Strat, 1991] Quam, L. and Strat, T. (1991). SRI Image Understanding Research in Cartographic Feature Extraction. In H. Ebner, D. Fritsch, C. H., editor, *Digital Photogrammetric Systems*, pages 111–121. Wichmann, Karlsruhe.

[Sester and Förstner, 1989] Sester, M. and Förstner, W. (1989). Object Location Based on Uncertain Models. In *Proc. DAGM Symposium, Hamburg*.