Submitted to GIS journal 4/96.

One-Eye Stereo System for the Acquisition of Complex 3D Building Descriptions

Roman Englert
University of Bonn
Institute of Computer Science III
Römerstr. 164
D-53117 Bonn
`englert@cs.bonn.edu`

Eberhard Gülch
University of Bonn
Institute of Photogrammetry
Nußallee 15
D-53115 Bonn
`ebs@ipb.uni-bonn.de`

Mai 1996

Abstract: An easy usable system for the semi-automatic acquisition of detailed 3D building descriptions from a multitude of images is provided. This approach tackles robustly and efficiently most of the problems of 3D building reconstruction, namely occlusions, inverse mapping, and noise. The 3D modeling is based on Constructive Solid Geometry (CSG) and various automated and supporting tools. Our experiences on the acquisition of an extensive scene are evaluated.

Zusammenfassung: Monokulares Stereo System für die Erfassung komplexer 3D-Gebäudebeschreibungen Für die semiautomatische Erfassung detaillierter 3D-Gebäudebeschreibungen aus einer Menge von Bildern wird ein einfach handhabbares System vorgestellt. Auf robuste und effiziente Weise werden die meisten Probleme bei der 3D-Gebäuderekonstruktion – Verdeckungen, inverse Abbildung und Rauschen – angegangen. Die 3D-Modellierung basiert auf der Constructive Solid Geometry (CSG) und zahlreichen automatisierten und unterstützenden Werkzeugen. Unsere Erfahrungen mit einer großflächigen Erfassung werden evaluiert.

1 Introduction and Motivation

An increased demand of 3D building descriptions for environmental- and city planning, air distribution- and air pollution simulation, transmitter placement for telecommunication, to choose a few out of many examples, has been observed. To satisfy this demand a variety of real scenes (cities, suburban areas, etc.) containing complex buildings have to be acquired. Performing this task virtually saves time and replaces the need for physical models. Our goal is the 3D reconstruction of complex buildings from a multitude of images. Complex buildings have a polygonal ground plan and a detailed roof shape. Thus they are in general combinations and variations of the basic building types as shown in Figure 1. Higher degrees of details contain also canopies, dormers, oriels, chimneys, and overhanging eaves (cf. Figure 2.2). In this context we regard acquisition as a structured, topological, and geometrical modeling process using background knowledge. A variety of approaches for the recognition of objects already exists [SFH92]. To reconstruct 3D objects, current methods of Analytical and Digital Photogrammetry, laser scanning, and digital surface model analysis have to tackle with several problems: occlusions, inverse mapping, and noise [BKL$^+$95]. The latter both methods are currently complete inadequate to acquire a detailed 3D building description due to their limited resolution. But they can support other approaches providing estimates to them. For very moderate requirements on the complexity of 3D building structures [BDR96], i.e. ground plan as polygon and one height for the building, the operational, but fully interactive Analytical Photogrammetry, is a good solution. With increasing demands on the degree of detail the efforts on the modeling are growing enormously for extended scenes. To make this task tractable a trained operator is needed. In contrast to these approaches Digital Photogrammetry offers the possibility to automate the modeling process in partial, and thus enables untrained operators to perform measurements. Previous work in this area has amongst others been performed by [LS93]. We adopted and substantially extended their approach using the Constructive Solid Geometry (CSG) for the modeling of complex buildings.

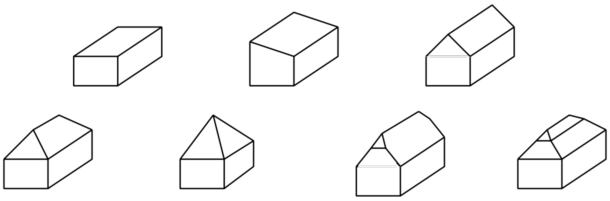

Figure 1: Some basic building types with different roof shapes: flat roof, pent roof (lean-to roof), saddleback roof (gable roof), hip roof, broach roof (pavilion roof), hipped-gable roof, mansard roof (gambrel roof).

In Section 2 the system HASE$^+$ is described and discussed. The modeling process is done in monocular mode (one-eye stereo) assisted by various supporting and automating tools for the form and pose adaptation of a large amount of CSG primitives. Our experiences on an extensive real scene are evaluated and compared with earlier experiments in Section 3. Finally, we conclude and give an outlook for future work (Section 4).

2 The System HASE$^+$

The system is designed to overcome the problems of occlusions and inverse mapping in 3D building acquisition by an one-eye stereo approach using a multitude of images (currently two) and three-dimensional primitives. We suppose that the interior and exterior orientation data of the images are known. The 3D acquisition process for buildings (cf. Figure 2) is divided into two phases. First, in the **navigation phase** the operator may zoom stepwise down into the aerial image and focus his interest on a particular region in one image. Having focussed on a building of this region the **modeling phase** is performed by a semi-automatic form and pose adaptation of 3D models. One homologous point is measured in the images in order to compute 3D world coordinates. The result of the building acquisition process is a 3D building description, more precisely an attributed CSG tree, which can also contain several buildings. For further data analysis and visualization a boundary representation (B-rep) is derived.

In the following we describe the Constructive Solid Geometry applied to 3D building acquisition and specify the supporting tools (flagpole, display and edit functions) and automated tools (matching and gluing of lines and faces of CSG primitives) for form and pose adaptation.[1]

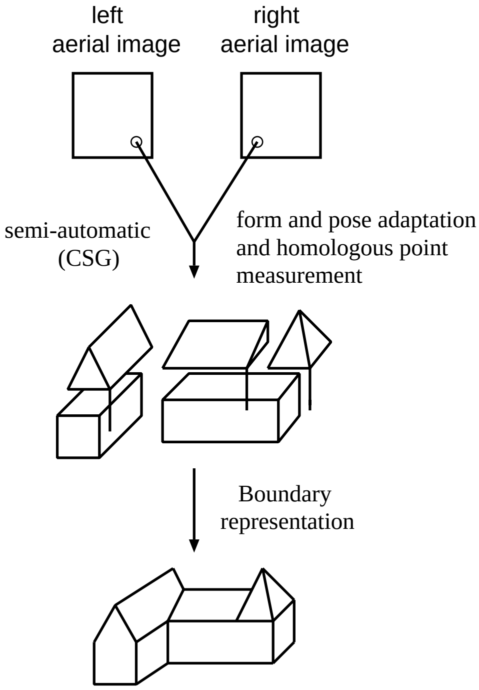

Figure 2: 3D building acquisition from stereo aerial images. CSG primitives are adapted in one image resulting in a binary tree.

---

[1] All the described modules are implemented in the system HASE$^+$ 3.0 which operates on a UNIX^{TM} workstation.

2.1 Constructive Solid Geometry in 3D Building Acquisition

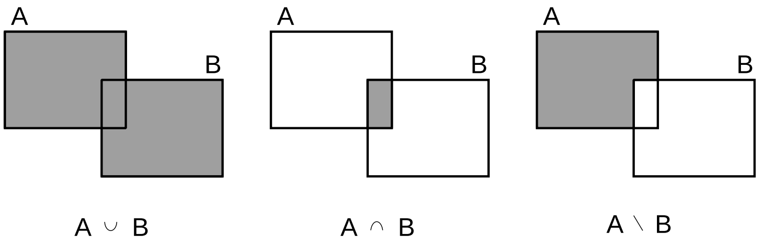

Figure 3: The three CSG operations (union, intersection, and difference) for sake of simplicity showed in 2D.

Object reconstruction, particularly building reconstruction, requires knowledge about all dimensions. Since an image has two dimensions, external knowledge about the third dimension is needed. To make this knowledge available we follow a model-driven approach which is well-known from the **Constructive Solid Geometry (CSG)** [Hof89]. Within this approach buildings are reconstructed by combining a series of 3D atoms, so-called CSG primitives (short: primitives), until the complete building has been modeled.

For the combination of primitives different CSG operations are provided: union, intersection, and difference (cf. Figure 2.1). In our framework these operations are commutative and distributive. As primitives we are using box, chock, cone, cylinder, half-chock, pyramid, and tetrahedron (cf. Figure 2.1). Furthermore, for reasons of efficiency three **combined primitives** can be chosen for modeling: saddleback roof building, hip roof building, and lop-sided saddleback roof building. A parameterized description of these combined primitives can be found in [LS93].

During the modeling phase the operator has to perform the following simple steps, which are supported by the system: choose a primitive which will be projected as a wire frame model (with removed hidden lines) into the focussed image region, adapt the parameters of the wire frame model[2] by clicking with the mouse onto its edges and pulling them to the correct size of the modeled building part. These steps have to be repeated for each primitive until the whole building is described. During the adaptation process the system supports the operator by diverse tools: monitoring of the adaptation in all the other images, online display of parameters and 3D world coordinates, display and editing of the CSG tree, user guidance by display of system status, 3D rendering of the whole building description, flagpole principle (cf. Section 2.2), and the automated matching and gluing facilities (cf. Section 2.3). All these adaptations are performed in monocular mode.

---

[2] E.g. a box has three parameters: length, width, and depth.

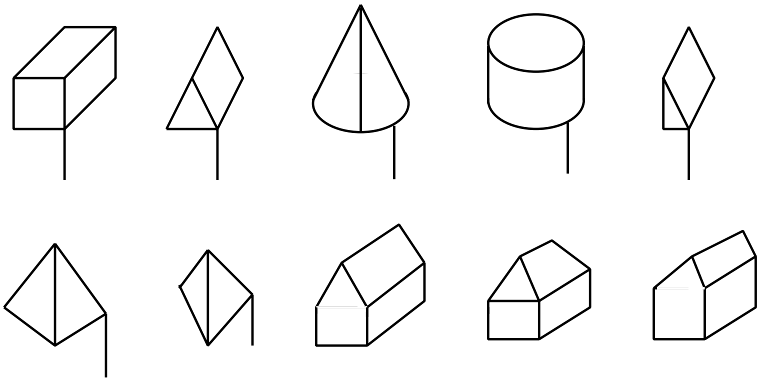

Figure 4: Primitives (with flagpole): box, chock, cone, cylinder, half-chock, pyramid, tetrahedron. Combined primitives: saddleback roof building, hip roof building, and lop-sided saddleback roof building.

The modeling process results in a CSG tree, whose interior nodes contain operations and the leaves contain instantiated primitives and attributes, e.g. form and pose parameters. Note: as shown in Figure 2.1 there are several possibilities to describe and acquire a building. The operator has the freedom of choice to determine the construction of the CSG tree. Finally, the acquired building description is visualized three-dimensionally in order to check the result (cf. Figure 3.1).

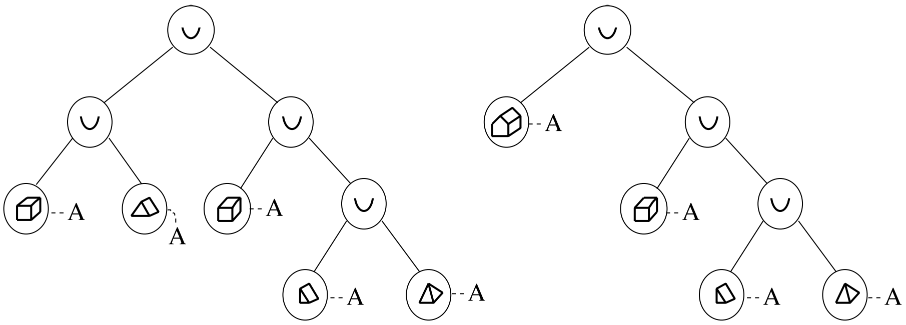

Figure 5: Alternative CSG trees of a complex 3D building description (cf. Figure 2). The interior nodes of the CSG tree contain operations and the leaves contain instantiated primitives and attributes **A**. The right CSG tree contains also a combined primitive (saddleback roof building).

The **degree of generalization** determines which details of a building are reconstructed, e.g. if chimneys have to be modeled. The choice of the degree of generalization has to be made by the user (cf. Table 1). We are primarily interested in the most detailed acquisition of roof and building structures as possible, and thus

Generalization level	Description
1 (low)	Detailed roof structure and ground plan + combinations of primitives
2 (medium)	Combined primitives
3 (strong, for XY)	Blocks of buildings with roof structure
4 (very strong, for XYZ)	Blocks of buildings of constant height

Table 1: Description of generalization levels (adopted from [Löc95]).

referring to a **low generalization level**. In the following we are considering how to adapt three-dimensional primitives into the image.

2.2 Flagpole Principle

This section describes the modeling of those primitives which do not touch the ground level, more precise complex roof structures, canopies, dormers, smoke stacks, chimneys, oriels, etc. To enable the positioning of primitives at the intended height level we are following the **flagpole principle** which offers further advantages to the operator. All primitives (except the combined ones) are equipped with a pole (cf. Figure 2.1) along which they are moved up or down by the operator in order to adjust their height above ground. To illustrate the flagpole principle consider the following task: Acquire a 3D description of a saddleback roof building. To solve this task you just have to put a chock onto a box and to perform a few adaptations of the primitives. With other words: the chock must be “hung up” at the flagpole according to the height of the already adapted box. Note, the ground height of the chock (which equals the lower end of the flagpole) is automatically inherited by the box. No further homologous point measurement is required.

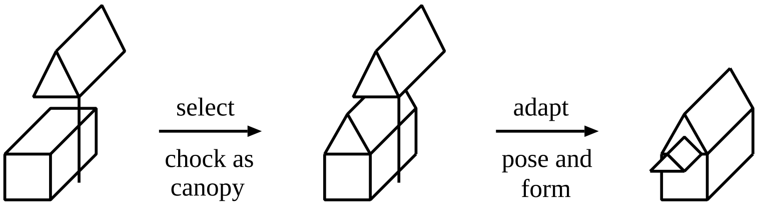

Figure 6: The flagpole principle in order to adapt primitives above ground level. This figure shows a saddleback roof building with a large canopy consisting of a box and two chocks.

This task could in principle also be solved by matching the bottom face of the chock with the top face of the box, both faces of similar size. A further merit of this principle, docking of primitives without matching (cf. Section 2.3), can be easily seen by a modification of the given task: assume the saddleback roof building has a large canopy which has to be modeled. Fortunately, the flagpole principle offers also an easy solution to this challenge (cf. Figure 2.2), where matching fails due to the different size of the docking faces.

2.3 Matching and Gluing of Building Parts

Describing a building by the combination of primitives or combined primitives requires a precise "docking" of the primitives. This docking is supported by matching and gluing facilities. The former allows to match at least two edges of different primitives and the latter matches and glues exactly two faces of different primitives together. These functions are based on a user-defined radius (cf. Figure 2.3). Note that the matching or gluing faces need not to be parallel, since these operations use a radius of a sphere and are thus extended to three dimensions. The matching and gluing are performed automatically with already instantiated primitives and even with invisible lines, respectively faces. This enables the operator to dock the current primitive easily to neighboring primitives.

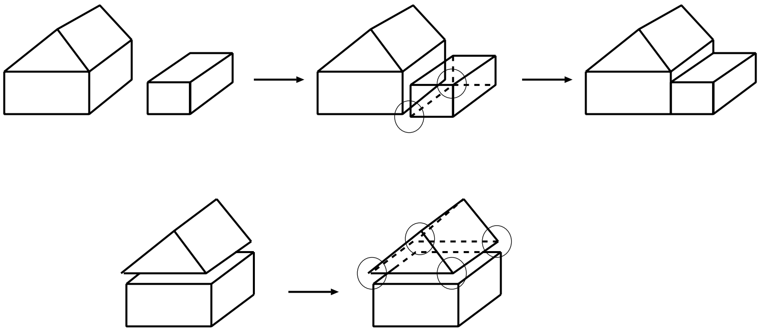

Figure 7: Automatic matching process of two edges (first row) and automatic gluing process of two faces (second row). The circles show the user-defined radius of a sphere, where matching and gluing is performed.

Before storing the corresponding CSG tree of a 3D building description, the operator may add useful knowledge to the building description, e.g. the gutter height by clicking with the mouse onto a specific primitive in the CSG tree. The system manages 3D building descriptions with additional geometrical attributes (Gauß-Krüger coordinates, gutter height, etc.) in a flexible usable manner. This extended CSG structure can be further analyzed in standard CAD and GIS systems and the conversion into B-rep is visualized by various rendering tools.

3 Experience on an Extensive Scene

The system is a practical tool for the acquisition of complex 3D building descriptions. It has been tested on aerial images of an extended scene of about 3 km^2. We give the present status of the acquisition and compare the performance to results obtained with an earlier version of this system.

3.1 Test Field OEDEKOVEN

The test field OEDEKOVEN has been chosen to acquire data for an OEEPE[3] test on 3D city models. The image scale of the B/W aerial imagery is 1:12000, the focal length is 153 mm. A stereo pair had been digitized with a pixel size of 12.5 μm in the image or 15 cm on the ground, with a total amount of 1 Gigabyte of data for images and image pyramids. The scene had been divided into two parts for two operators. The task is to acquire detailed roof and building structures. This refers to generalization level 1 (cf. Table 1) or a high degree of detail. Earlier tests were performed only with generalization levels 2 to 4.

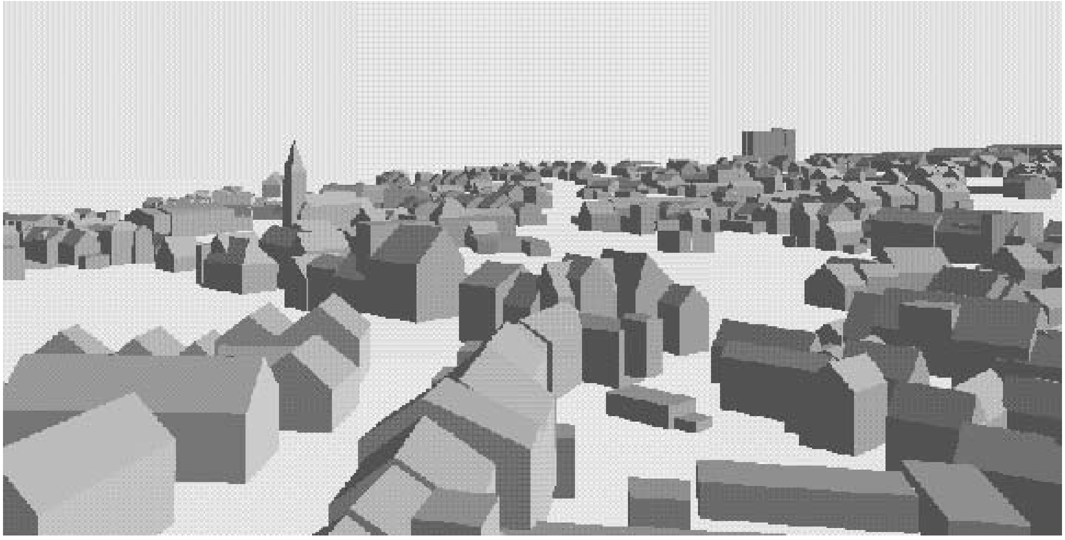

Figure 8: A visualized part of the acquired extensive scene OEDEKOVEN.

3.2 Used Primitives

Until mid April 1996 about 85% of the buildings in the test field had been measured. As primitives occurred 59% boxes, 1% chocks and half-chocks, 28% saddleback roof

---

[3] Organisation Européenne d'Etudes Photogrammétriques Expérimentales

buildings, 2% hip roof buildings, 9% lop-sided saddleback roof buildings, and 1% others. There are single buildings, building groups, garages, churches, farms, and plants. Figure 3.1 visualizes a small subset of the acquired buildings.

	Total	Op1	Op2
CSG trees	1372	503	869
Primitives	3853	1007	2846
Primitives/CSG tree	2.8	2.0	3.3
CSG with 1 Primitive	481	261	220
CSG with $\geq$ 2 Primitives	891	242	649

Table 2: CSG trees and primitives in OEDEKOVEN test area (status April, 1996). Op(erator)1: expert. Op(erator)2: non-expert.

1372 buildings or building-blocks (CSG trees) have been extracted with an average amount of 2.8 primitives per CSG tree (cf. Table 3.2). The area of operator 2 (Op2) contains a higher amount of blocks of flats, (terraced) housing estates, and few single buildings.

3.3 Performance

The gross and net times for data acquisition are given in total and for each operator separate in Table 3.3. The gross time contains the modeling time, the internal navigation and the external navigation and organization. The modeling time contains the form adaptation, the specification of operations, the measurement of homologous points and for complex buildings a 3D visualization. The times per building primitive are given as mean values. Due to some very complex buildings the mean value for the modeling (50.3 seconds) is higher than the median value. The median modeling time per building primitive is below 40 seconds and about 75% of all primitives are modeled in a time below 60 seconds per primitive. The internal navigation time contains the local navigation through the pyramid and the selection of primitives. The short navigation time of 8.8 seconds per primitive indicates the optimality of this acquisition step. The global navigation and organization (31.5 seconds) covers the navigation through the project area, checks of completeness, editing and 3D visualization. The good modeling and local navigation times are due to the well developed and further improved semi-automation (one-eye stereo), whereas global navigation and organization can not yet be assisted by many automated procedures and require still substantial interactive efforts.

The gross times of this test are about 25% shorter compared to an earlier version of the system [Löc95] tested on comparable image material, but with much less

Time/primitive [seconds]	Total	Op1	Op2
Modeling time	50.3	51.0	49.8
Local navigation	8.8	4.9	11.0
Global navigation, organization	31.5	22.7	34.6
Gross time	90.6	78.6	95.4

Table 3: Average acquisition times OEDEKOVEN (status April, 1996) for all primitives (3853) and the two operators (Op1, Op2).

buildings. The higher performance here is in addition combined with a more detailed building acquisition! The experienced operator (Op1) showed higher performance in the global navigation and organization, the inexperienced operator (Op2) reached the same performance for the modeling, after one week training only. This shows the potential of this system, for usage as a modeling tool by a non-photogrammetrists in practice.

For 481 CSG trees with single primitives a modeling time of 47.5 seconds per primitive and a local navigation time of 12.0 seconds per primitive were measured, which did not differ significantly from the total times given in Table 3.3. This means that independent from the complexity of the building we are observing a more or less constant modeling time per primitive.

The accuracy of the system depends on many factors, like image scale, pixel size, orientation, film processing, scanning, selection and measurement of models and homologous points, the definition uncertainty of building corners, and the generalization level. A preliminary accuracy check of few buildings based on the differences of measured roof point coordinates and ground truth yielded an external accuracy of $\sigma_{X,Y} \approx 25$ cm and $\sigma_Z \approx 35$ cm. This corresponds to the accuracy of analytical photogrammetric methods.

4 Conclusions

We have enabled non-experts to acquire 3D descriptions of complex buildings in a robust and efficient manner. The extended CSG structure has opened the connection to CAD and GIS systems.

Further improvements on accuracy and speed of the modeling can be reached by simultaneous multi-image matching of building structures (e.g. the B-rep of a CSG tree). In recent work it has been shown, e.g. [LS93], that the edges of a single primitive could be matched to automatically extracted edges from the images. Based on our experiences we will further automate the project management including navigation. Currently we are investigating in the learning of regularities of 3D building structures in order to enhance automatic building recognition and interpretation.

We regard the concept one-eye stereo system as the most promising method for the future to acquire complex 3D buildings in a fast and reliable way. Our design enables us to easily integrate CSG primitives for other objects, e.g. vegetation, in order to model a complete real 3D scene consisting of buildings, streets and vegetation.

Acknowledgments

The inspiration and support by Prof. Dr. Armin B. Cremers and Prof. Dr.-Ing. Wolfgang Förstner are gratefully acknowledged. We would also like to thank for the assistance by Andreas Barden, Hardo Müller and Olaf Schmidt. The aerial images have been kindly provided by the LVA Nordrhein-Westfalen.

This research is supported in part by BMBF/ DLR under Grant 01 M 3018 F/6 and in part by BMBF/ DARA GmbH under Grant 50 TT 9536. The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, expressed or implied of those agencies.

References

[BDR96] M. Beines, R. Decker, and B. Ruff. Raumbezogene Basisdaten zur Unterstützung der Funknetzplanung in urbanen Gebieten. *GIS*, 1, 1996.

[BKL+95] C. Braun, T. H. Kolbe, F. Lang, W. Schickler, V. Steinhage, A. B. Cremers, W. Förstner, and L. Plümer. Models for Photogrammetric Building Reconstruction. *Computer & Graphics*, 19(1), 1995.

[Hof89] C. M. Hoffmann. *Geometric and Solid Modeling.* Morgan Kaufmann, Palo Alto, CA, U.S.A., 1989.

[Löc95] T. Löcherbach. System Performance for Semiautomatic Building Reconstruction. In *Second Course in Digital Photogrammetry*, Institute of Photogrammetry, University of Bonn, Bonn, Germany, February 1995.

[LS93] F. Lang and W. Schickler. Semiautomatische 3D-Gebäudeerfassung aus digitalen Bildern. *Zeitschrift für Photogrammetrie und Fernerkundung*, 5:186 – 193, 1993.

[SFH92] P. Suetens, P. Fua, and A. J. Hanson. Computational Strategies for Object Recognition. *ACM Computing Surveys*, 24(1):5 – 63, March 1992.