A one-eye Stereo System for Semi-Automatic 3D-Building Extraction*

Felicitas Lang, Thomas Löcherbach, Wolfgang Schickler
Institute for Photogrammetry, Bonn University
Nußallee 15, D-53115 Bonn
Tel.: (228) 732902, Fax: (228) 732712
e-mail: thomas@ipb.uni-bonn.de

July 4, 1995

We present a semi-automatic system for building extraction that has been developed at the Institute for Photogrammetry at the University of Bonn. Digitized images are used as data source. Data capture takes place at a simple workstation. Single point measurement is replaced by the measurement of building models. Automation supports the operator and thus increases the system performance. The representation of building models allows a link to CAD, GIS and computer graphics.

1 Motivation

Recently there is a great requirement for three-dimensional building extraction as such data is needed for an increasing number of tasks such as measurement for property management, environmental planning, planning for transportation and rescue organization, simulation studies for flood prediction, emission control or transmitter placement but also for virtual reality. There is a lack of systems which are capable of capturing entire cities in a relatively short time, at a reasonable cost and with appropriate accuracy. For complex city structures fully automatic approaches do not seem to be working really operationally yet. Therefore, the present goal was to have a practicable solution immediately which can be automated stepwise. A mid-term aim is the combination with fully automated approaches and integration of additional data sources. The goal of automation is to improve accuracy and reduce interaction. The basic idea is to replace the measurement of single points by the measurement of models and thus to establish a link to CAD, GIS and computer graphics. The principle is that the interpretation is still performed by the operator and takes place in one single image while the measurement is performed automatically and uses multiple images. We will in the following describe the approach for semi-automatic building extraction developed at the Institute for Photogrammetry at Bonn. This approach was described for the first time in the "Zeitschrift für Photogrammetrie und Fernerkundung 5/93" and has since then been further automated and been applied extensively.
We first describe our representation of buildings, then the approach for building extraction, the performance and finally draw some conclusions.

2 Representation of Buildings

Most real buildings can be modeled by a set of primitives, i. e. basic types of buildings, or by a combination of these primitives. We use two types of building primitives namely parametric and prismatic primitives. Examples for parametric primitives are shown in the left part of fig. 1. They have a fixed structure and variable form and pose parameters, for example length, width, height, position and rotation. Prismatic primitives allow a more general description of buildings with a

*To appear in Geomatics Info Magazine, June 1995

polygonal ground plan and a roof which is assumed to be horizontal. Examples are shown in the right part of fig. 1.
Buildings are internally represented by a boundary representation. By using a constructive solid geometry representation known from computer graphics, more complex buildings can be reconstructed by combining primitives (cf. fig. 2).

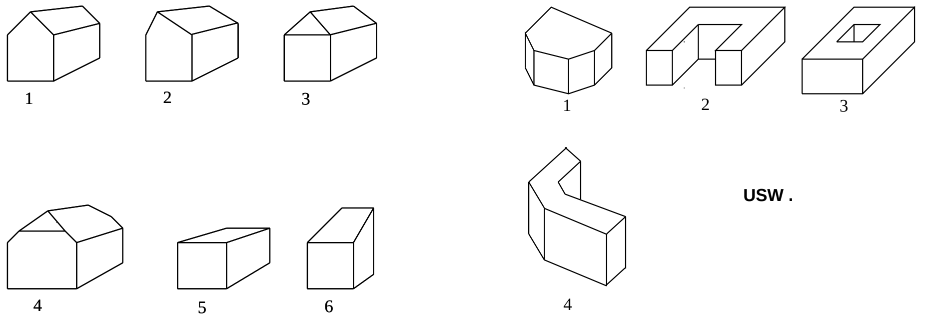

Figure 1: Parametric (on the left) and prismatic (on the right) building models (from Lang and Schickler 1993)

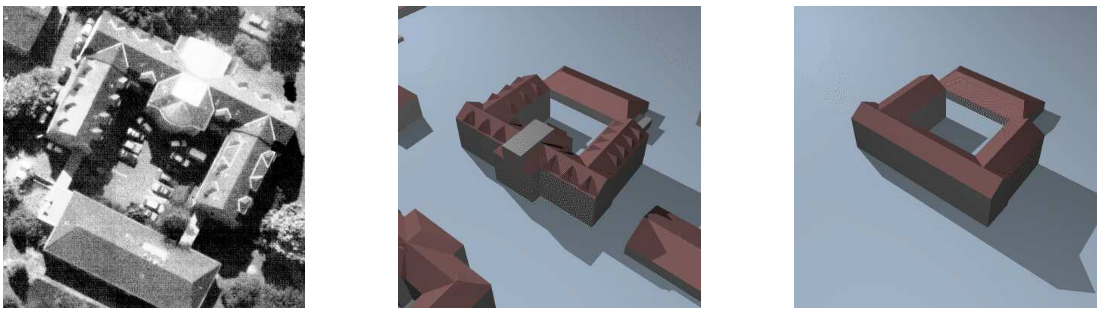

Figure 2: Buildings can be composed of building primitives and can be captured with different degrees of generalization

3 Building Extraction

The data capture process takes place at a Unix-workstation. Input data are digital or digitized images including their orientation parameters, an image pyramid and a file containing edges that have been extracted by an edge operator. The pyramid, i. e. a scale transformation of the images, is needed to enable zooming and panning and to handle the large amount of data. The edges both serve to support the operator and to automatically estimate parameters.
Buildings are 3-dimensional objects while an image is only 2-dimensional. Usually at least 2 images are needed to extract 3-dimensional coordinates. Using building primitives which include for example rectangularity constraints enable us to measure 3-dimensional buildings from one single image except for an unknown scale. From 2 images the full 3-dimensional shape can be recovered. By using more than 2 images accuracy can be improved due to the redundancy. The data capture consists of four steps:

1. The operator selects the appropriate type of model, either one of the parametric primitives or the prismatic type (cf. middle part of fig. 3).

2. The second step still is an interactive step: In case of measuring parametric models the operator can move and change the building, i. e. adapt its form and pose parameters until the model fits the image (right part of fig. 3). A mouse is used as measuring device. During this process the building primitive is projected into the image in real time. For buildings with a more complex ground plan the prismatic model is well suited. In the prismatic mode, the extracted edges are superimposed on the image. The operator now can measure the ground plan of the building either by selecting edges or by drawing edges into the images. Edges are intersected automatically, so that a closed polygon is constructed. Finally the operator has to determine the height of the building by shifting the polygon to the ground plane of the building approximately.

3. As soon as the operator has measured the building roughly in one image either using the parametric or the prismatic mode he or she can trigger the automatic part of the process. We are not restricted to a stereo model but can use any number of images with arbitrary viewing directions. Image edges extracted with subpixel accuracy serve as primary observations. In a first step the model that has been measured in one image is moved along the epipolar lines in the other images. A 1-dimensional automatic clustering is used to find the best fit between model and image data. In this way the unknown scale is determined which can not be determined from one image without additional information.

4. In a final step a robust estimation is performed automatically to estimate the unknown form and pose parameters of the building simultaneously from all images. In this way an optimal fit is found. This estimation has been described earlier in the April 1993 issue of this journal.

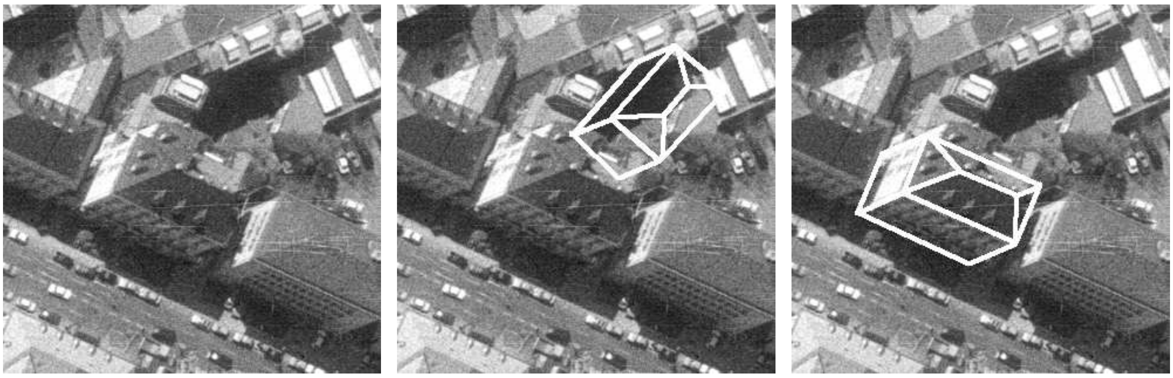

Figure 3: Measuring of a building in the parametric mode (lines have been thickened for better illustration)

Output data can be seen on front page and in fig. 2 which also shows how buildings can be measured with different degrees of generalization.

4 Performance

We have also investigated the performance of the system. Two main aspects of performance are the time that is needed for data capture and the accuracy of the output data.
Time performance was tested on a data set of a city center with rather dense and complexly structured buildings. The image scale was 1:16000. 9 minutes approximately were needed per ha when measuring in a high resolution, i. e. with a low degree of generalization. In a very low resolution, when blocks of buildings are merged to one block with a flat roof, this time went down to approximately 3 minutes per ha. Part of this time was used for panning, zooming and visualization, which can be reduced by an optimized handling of large data sets.
The internal accuracy of the estimation of pose and form parameters was in the range of half a decimeter for position and a few decimeters for height. This was better than the accuracy that could

be expected with an analytical plotter. The images for this test had a scale of 1:12000 and were scanned with 15 microns. The high accuracy is mainly due to the high redundancy when using 6 images in this test. This test, however, still is based on a small sample size.
An extensive test has been performed with an interactive version of the program. That means step 4, the robust estimation was omitted and step 3 was replaced by a manual measurement of one homologous point in 2 images. For this test images with an image scale 1:16000, scanned with 15 microns were used. A number of buildings was measured by 9 different operators. In this case the accuracy was somewhat lower than the accuracy that could be achieved with an analytical plotter. This is mainly due to the pixelsize of 24 microns in object space.

5 Summary

We presented a semi-automatic approach for model-based building extraction from digital images, as images seem to be the appropriate basis for acquisition. The principle of the approach is that the interpretation still is done by the operator and takes place in one single image, while the measurement is performed automatically and uses multiple images. Buildings are modeled by a boundary representation of parametric and prismatic primitives. Parametric primitives use pose- and form-parameters to describe buildings with a fixed shape. Prismatic primitives allow a more general description of buildings with a polygonal ground plan. Input data for the building extraction are at least digitized images including their orientation parameters. The measurement is done at a workstation. During the data capture process the operator has to measure a building roughly in one image. One-dimensional matching between building model and image edges along the epipolar lines is used to automatically determine the terrain height. A robust estimation is applied to automatically estimate the form- and pose-parameters of the buildings to find the best fit. Image edges that have been extracted in a pre-processing step with subpixel accuracy serve as observations.
There are several new aspects of this approach:

- single point measurement is replaced by the measurement of objects;
- compared to the measurement of single points, the number of interactions is considerably reduced by measuring models.
- most types of buildings can be measured by a combination of building primitives (cf. fig. 2);
- no stereo viewing is necessary;
- digital images are used, so that image processing techniques can be integrated and further automation is possible;
- occlusions can be handled by using building models with a predefined shape;
- the underlying data structure allows a connection to CAD, GIS and computer graphics;
- in principle one image is sufficient if the terrain height is known;
- multiple images with arbitrary views may be used simultaneously;

In this way the accuracy is improved considerably while interaction is reduced.