# A Technique for 3D Building Reconstruction

**Taejung Kim and Jan-Peter Muller**


## Abstract

*An approach to tackle the problem of three-dimensional (3D) building reconstruction in urban imagery is presented. For 3D building reconstruction, there is a need to combine 2D (such as grouping) and 3D analysis (such as stereo matching). A "good" strategy for the combination is essential for success. A simple but robust combination strategy is proposed. Combination is carried out only after a 2D building detection technique and a 3D height extraction technique are applied completely independently. The 2D building detection technique does not use any information generated from the height extraction technique, nor vice versa. Moreover, any assumptions or conditions derived in the course of 2D building detection or height extraction are not used for combination. 3D building reconstruction is done by interpolating heights into the area covered by 2D building boundaries using the 3D height information. In this way, results from the 2D building detection technique and 3D height extraction technique can be meaningful by themselves. This also can make the process of 3D building reconstruction simple and applicable to a wide range of images. This approach is tested with airborne images, and the results show that 3D building reconstruction can be achieved successfully.*


## Introduction

The rapid emergence of high-resolution spaceborne imagery as well as traditional airborne imagery has created an urgent need for techniques for urban area image understanding. These techniques have numerous applications in urban mapping, urban planning, and other geo-information engineering disciplines and applications. However, the presence of very dense and complex man-made structures in urban area imagery and their distinctive characteristics render many existing approaches for lower-resolution imagery inappropriate. In such cases, new techniques and/or approaches need to be developed.

Among many issues in urban area image understanding, this paper will address the problem of two-dimensional (2D) building detection and 3D building reconstruction. There have been several approaches proposed for automated building extraction. The most popular ones are perceptual grouping (Mohan and Nevatia, 1989), line analysis (Shufelt and McKeown, 1993), and the use of auxiliary information. Auxiliary information can be shadow information, perspective geometry (Huertas and Nevatia, 1988; Herman and Kanade, 1986), building or surface modeling (Maitre and Luo, 1992; Huertas *et al.*, 1996), or knowledge-based systems (Nicolin and Gabler, 1987). However, the nature of real world buildings necessitates the incorporation of 3D analysis in building extraction: building detection performed in 2D cannot fully "understand" buildings in the scene. For such a 3D analysis, stereo matching techniques have been widely used, often in combination with the other approaches mentioned above. Shufelt and McKeown (1993) created building hypotheses from a line-corner analysis, a shadow analysis, and stereo matching, respectively, and combined building hypotheses from each method for 3D building extraction. Cochran and Medioni (1992) refined a depth disparity map from stereo matching by using building extraction output. Maitre and Luo (1992) used surface models to improve stereo reconstruction. Other approaches assume that height information (i.e., DTM or DEM) for the scene is known and use this information for 3D building extraction (Baltsavias *et al.*, 1996; Haala and Hahn, 1996).

However, building extraction (in 3D) still remains difficult because it requires not only good low-level vision techniques, such as edge or line extraction, but also good middle-level or high-level vision techniques, such as cognition and interpretation. It also requires a good strategy for combining 2D building detection results and 3D (or height) information of buildings. In some previous approaches, 3D information was used for better 2D analysis (Haala, 1994) whereas, in other approaches, 2D information was used for better 3D analysis. Both cases addressed the importance of combining 2D and 3D information for better "understanding." However, it seems that the importance of developing an appropriate "combination" strategy was often neglected. Moreover, in some approaches, 2D analysis (or 3D analysis) seems to rely on 3D information (or 2D information) too optimistically. In such cases, if 2D or 3D analysis fails, the other approach may not produce meaningful results at all.

This paper proposes an approach for 3D building reconstruction which uses a combination of a 2D and 3D analysis. In the proposed approach, combining 2D and 3D analysis is done only after each of them is applied completely independently. The 2D building detection technique tries to obtain as accurate 2D building information as possible without the knowledge of height information. The height extraction technique tries to generate as accurate height information as possible without the knowledge of the presence of buildings. In this way, the building detection technique and the height extraction technique can be developed independently, and each technique can create useful information by itself. The 3D reconstruction by combining these two techniques may cope with a wide range of images.

As a proper "combination strategy," this paper proposes a simple method. 3D building reconstruction is achieved by interpolating heights into the area defined by 2D building boundaries using 3D height information. Any assumptions or conditions derived in the course of 2D or 3D analysis are not required for 3D building reconstruction. This makes the pro-

T. Kim is with the Satellite Technology Research Center, Korea Advanced Institute of Science and Technology, 373-1, Kusung-Dong, Yusung-Gu, Taejon, Korea (tjkim@krsc.kaist.ac.kr).

J.-P. Muller is with the Department of Geomatic Engineering, University College London, Gower St., London, WC1E 6BT, U.K. (jpmuller@ps.ucl.ac.uk).

Photogrammetric Engineering & Remote Sensing,
Vol. 64, No. 9, September 1998, pp. 923–930.

0099-1112/98/6409–923$3.00/0

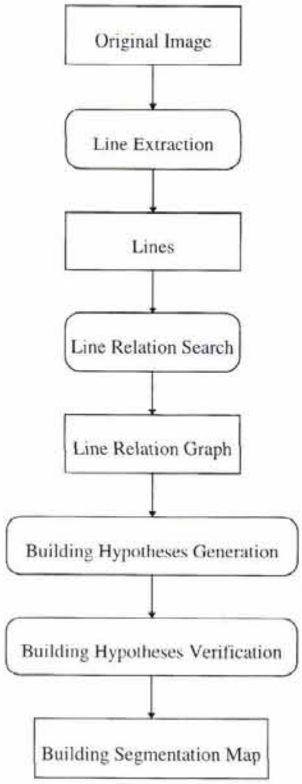

Figure 1. A schematic diagram for 2D building detection.

posed approach applicable to any 2D building boundary map and 3D height data set.

The organization of this paper is as follows: The next section briefly describes the 2D building detection technique and the height extraction technique used. Following that, the 3D building reconstruction technique is explained. Experiments on 3D building reconstruction and their results are then presented. This is followed by discussions and conclusions.

## A 2D Building Detection Technique and a Height Extraction Technique

This section describes the 2D building detection technique and the height extraction technique. These techniques were developed separately and have been reported on elsewhere (Kim and Muller, 1994; Kim and Muller, 1996a). This section describes these techniques briefly.

### A 2D Building Detection Technique

A graph-based technique for 2D building detection has been proposed by Kim and Muller (1994). In this technique, building hypotheses are generated by grouping lines extracted from an image. A graph, constructed from lines and their relationship, is used for grouping. The entire process is performed monoscopically.

Figure 1 summarizes the overall procedure for this building detection technique. First, lines are extracted from an image. There are several steps for line extraction: edge elements are detected by a Canny-Petrou-Kittler (CPK) filter (Canny, 1986; Petrou and Kittler, 1991), a connected edge labeling algorithm is then applied, and two end points of linear elements are searched for with end point templates. A line is defined with two end points. Small lines broken from a long line and parallel lines located closely are merged.

After detecting lines, the relationship between them is examined and stored in a graph. The relationship is classified into positive, neutral, negative, and parallel connections according to the manner in which lines are connected (Kim, 1995). Nodes of a graph consist of lines whereas arcs between nodes represent the relationship between lines. In order to reduce the size of a graph, the relationships of lines whose lengths are shorter than a threshold and whose relative angles with other lines are not near orthogonal are ignored.

Building hypotheses (BHs) are extracted by searching for closed loops in the line relation graph. For the traversal of a graph, a depth-first search algorithm is used. BHs then undergo a verification process. First, similar BHs are merged (Kim and Muller, 1996b). Second, false BHs, which are made from ground-level lines, are removed. This is done by using shadow analysis and perspective geometry (Herman and Kanade, 1986). However, these are optional processes because they require sufficient resolution (0.5m, currently) and knowledge of the direction of illumination and the position of the principal point.

### A Height Extraction Technique

One of the most popular ways of extracting 3D information from stereo image pairs is the use of image matching techniques. However, distinctive characteristics in urban area imagery, such as breaklines, occlusions and shadows of buildings, cause serious problems to conventional stereo matching techniques. Most existing stereo matching algorithms assume surface smoothness or maximum disparity limits. The effects of such characteristics on an area-based stereo matching algorithm were investigated by Kim and Muller (1996a), and it was reported that they produced many isolated regions for the stereo matcher. In such cases, each isolated region required at least one initial estimate for successful matching. A large number of initial estimates were required for matching of urban area imagery. However, results of experiments did not suggest that this was feasible.

As a possible solution to this problem, pyramidal matching was proposed (Kim and Muller, 1996a). It was shown that this approach could produce automatically a large number of initial estimates and partly overcome the problem of breaklines. This approach used a non-linear least-squares correlation estimation (Otto and Chau, 1989) for matching at each level. The problem of blunder propagation was minimized by using a tile-based filtering technique as a control strategy between levels.

Output points from pyramidal matching are converted into the ground coordinates by applying a camera model. The corresponding height information can be stored in a digital elevation model (DEM).

## A Technique for 3D Building Reconstruction

One of the weakest application areas of the monoscopic approach is, of course, the extraction of height or depth information. The building detection technique described in the previous section suffers from the same drawback. Although some indication of building height is possible by shadow analysis and perspective geometry, the system cannot provide accurate information on building heights.

Stereo matching results usually give accurate height information. However, heights are assigned only at grid-points, and it is difficult to obtain the height of objects, such as roofs, without further processing. Also, there is a problem of interpolation of height for urban areas. In such areas, the surface does not vary in a geo-statistical manner (Davis, 1986). Large areas of the surface are very flat or linearly varying and there are many breaklines. Hence, "kriging" (Davis, 1986), one of the common techniques for height interpolation, cannot be applied. Height interpolation requires external guidance or a proper statistical model of the surface being developed.

The fusion of the results of the 2D building detection technique and the height extraction technique can potentially solve these problems (Kim and Muller, 1996c). In this paper, a simple method for such fusion is proposed. Height information from the height extraction technique can be used to assign height to buildings. BHs derived by the 2D analysis can provide the external guidance of height interpolation so that interpolation only takes place within the boundaries of a BH. Any assumptions or information derived in the course of the 2D and 3D analysis are not used for fusion. This makes the proposed fusion approach applicable to any building boundary maps and 3D height data sets.

Figure 2 shows a schematic diagram of the proposed 3D building reconstruction method. In the figure, a matching list and BHs are produced by the height extraction and building detection techniques, respectively. Each BH provides interpolation boundaries. For each BH, a proper surface model is searched for.

As proper surface models, two surface types — apex and planar — are considered. Figure 3 shows the two models and their surface equations. For a given region, the coefficients of each model equation are estimated through least-squares estimation. The 3D height information is used to set up observation equations for estimation. A surface model with the smaller estimation error is selected. Interpolation is done simply by calculating the corresponding height of a point using the surface model equation.

## Results

The stereo image pair shown in Figure 4 was used for experiments. The 2D building detection system was applied to the left image, and 384 BHs were generated[1]. The height extraction technique was applied to the stereo pair. Twenty-one thousand five hundred and thirty points were matched and, from these, height information was derived. Figure 5 shows a

[1]Many BHs overlap each other.

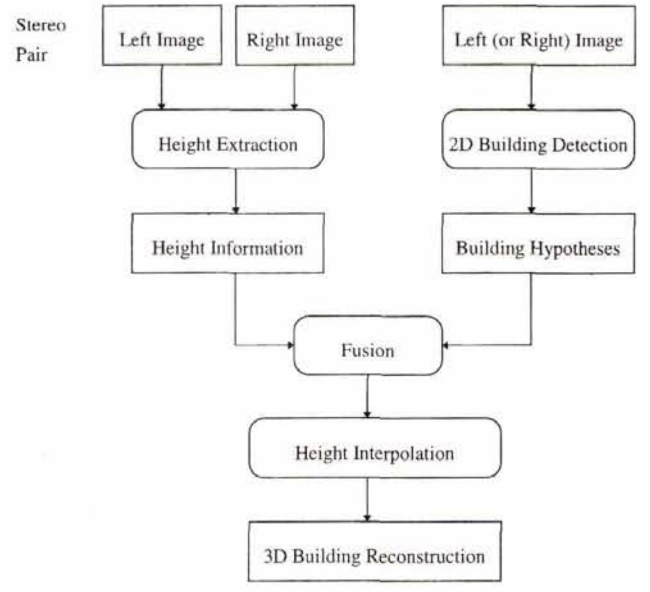


Figure 2. A schematic diagram for 3D building detection.

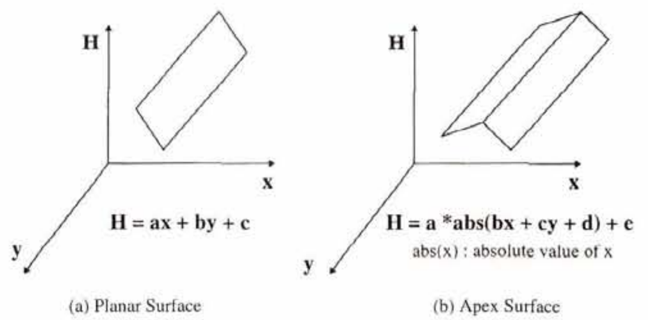


Figure 3. Two surface models (planar and apex).

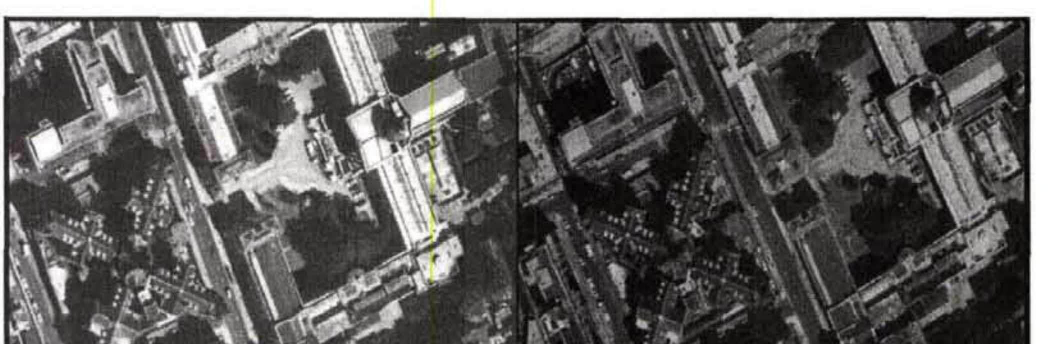

Figure 4. A test image pair, "UCL" image.

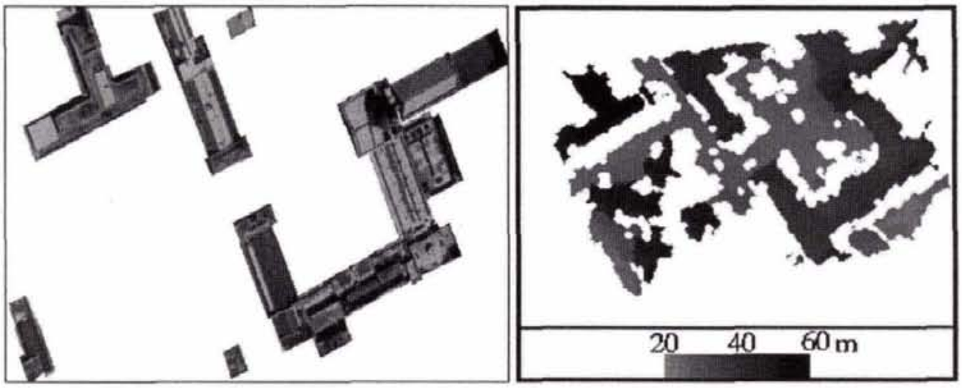


Figure 5. A building boundary map (left) and a DEM (right) for "UCL" images.

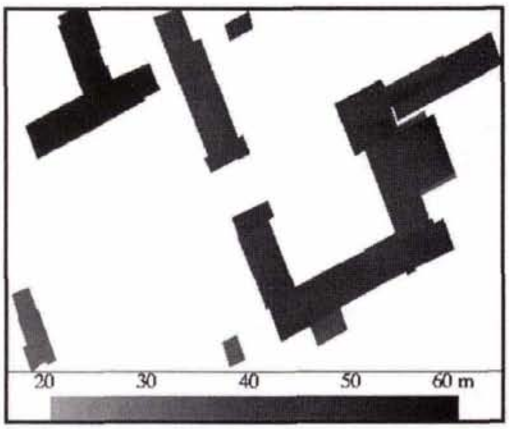


Figure 6. The building height achieved by fusing the building detection results and the height extraction results for "UCL" images.

Figure 7. A perspective view of buildings, "UCL" image (height exaggeration factor 1.5).

building boundary map generated from the BHs and a DEM generated from the height information. Seven-thousand eight-hundred and forty points out of 21,530 matching points were found to lie inside the region covered by the 384 BHs, and these were used to model the surface. Two-hundred ninety-three BHs were modeled as planar surfaces and 21 BHs as apex surfaces. Seventy BHs were not modeled as there were insufficient matching points to model them[2]. Figure 6 shows the resulting building heights and Figure 7 shows a perspective view of these buildings.

The performance of the building detection and 3D building reconstruction is summarized in Table 1. Performance is analyzed in terms of building detection coverage, error of omission, error of commission, building detection accuracy, height extraction coverage, and height accuracy. Building detection accuracy is defined as building detection coverage minus the error of commission. For the test scene shown in Figure 3, building detection coverage is 61.23 percent. This low coverage is due to the detection failure on the very complicated X-shape buildings in the scene. Building heights have an RMS error of 5.4m, whereas the original pyramidal matching output has an RMS error of 5.1m[3] (Kim and Muller,

[2]In the current implementation, there is a limit to the minimum number of height points for surface modeling. The limit varies according to the dimension of images.

[3]Near breaklines, small errors in the $x$ and $y$ directions result in huge errors in height.

TABLE 1. PERFORMANCE OF 3D BUILDING RECONSTRUCTION

| | Test Scene 1 (Figures 4–7) | Test Scene 2 (Figures 8–11) | Test Scene 3 (Figures 12–15) |
|---|---|---|---|
| Building Detection Coverage (%) | 61.23 | 71.64 | 98.76 |
| Error of Omission (%) | 38.77 | 28.36 | 1.24 |
| Error of Commission (%) | 12.19 | 5.91 | 6.46 |
| Building Detection Accuracy (%) | 49.04 | 65.73 | 92.30 |
| Height Extraction Coverage (%) | 45.7 | 91.2 | 86.6 |
| Building Height Accuracy (RMS error, m) | 5.4 | 0.986 | NA |

1996a). Some additional errors were introduced. This is because the building detection output itself had some errors, and these errors propagated to the surface modeling.

Figure 8 shows another test image pair. The building detection technique was applied to the left image, and 35 BHs were generated. After applying the height extraction technique, height information for 18,448 points was obtained. The results for each technique are shown in Figure 9. The fusion of the results from each technique was applied. Twenty-four BHs were modeled as planar surfaces and 5 BH as apexes. Six BHs were not modeled. Figure 10 shows the results of surface modeling. Building roofs were successfully modeled with apex and planar surfaces. A perspective view of the buildings is shown in Figure 11. The performance analysis is summarized in Table 1. Because this test scene is a suburban scene and buildings are not as complex as those in the previous test scene, the building detection and the 3D reconstruction show improved performance.

Figures 12 through 15 show results of 3D building reconstruction from another test data set. This scene contains even simpler and larger industrial buildings than the previous two cases; hence, building detection has improved performance. Some building roofs in the scene have noise-like patterns and some have little texture. These caused holes in the DEM shown in Figure 13. Because ground reference height data were not available, building height accuracy was not analyzed.

## Discussion and Conclusions

This paper describes the work carried out for 3D building reconstruction. For successful 3D building reconstruction, the importance of an appropriate "combination strategy" of the

Figure 8. Another test image pair, "Avenches residential" scene. (courtesy of ETH Zurich)

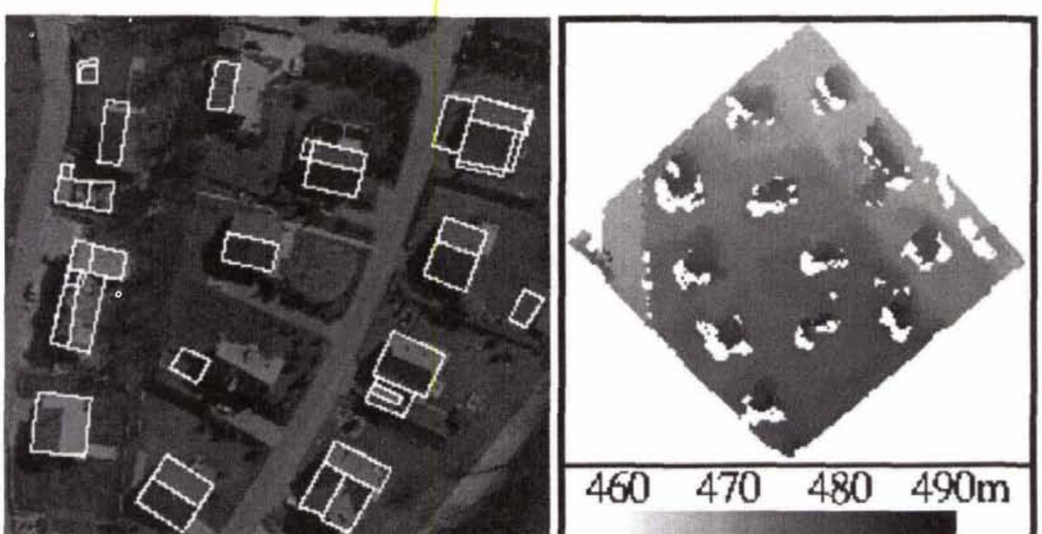


Figure 9. A building boundary map and a DEM for "Avenches residential" scene.

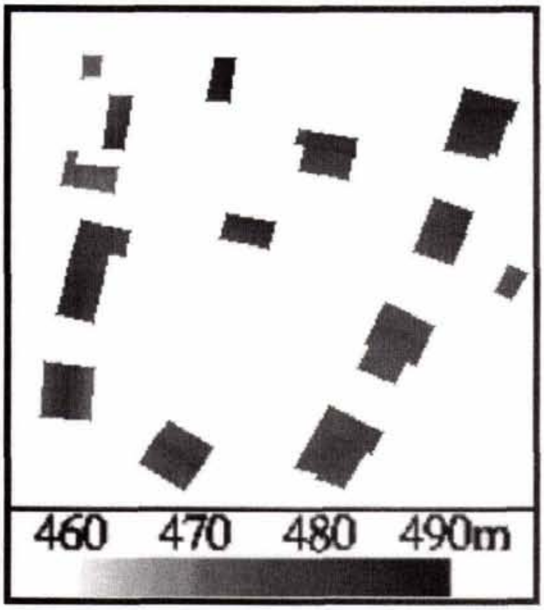


Figure 10. A building height map for "Avenches residential" scene.

Figure 11. A perspective view of buildings for "Avenches residential" scene.

two techniques is emphasized. This paper combines the 2D and 3D information by interpolating height into the area defined by 2D building boundaries using 3D height information. The results of experiments shown in the previous section seem to support that, through this, 3D buildings are successfully reconstructed.

The examples shown include complicated urban buildings, isolated suburban houses, and large industrial buildings. All of these were reconstructed. Due to the lack of an available data source, 3D reconstruction from high resolution spaceborne images could not be tested. Nevertheless, previous analysis of 2D building detection with DD5 images and airborne images with 1-m pixel spacing (Kim and Muller, 1995) indicates that 3D reconstruction from high resolution spaceborne images is feasible.

Compared with other approaches, the proposed approach uses the simplest combination strategy. Other approaches may delineate more detailed roof structures through the use of comprehensive building CAD models (Haala and Hahn, 1996). Others may achieve improvement in building detection coverage through the use of reference height data or color analysis (Henricsson *et al.*, 1996). The results in this paper show the limit of 3D building reconstruction without using these additional data sets or information.

There are a few things to consider. The 2D and 3D analysis techniques may need to be further improved, in particular, for high resolution spaceborne images with 1-m pixel spacing. Height information near breaklines needs to be refined. Surface models may need further consideration. A more sophisticated surface model may be introduced to handle domes or more complicated building roofs.

Figure 12. Another test image pair, "Avenches industrial" scene. (courtesy of ETH Zurich)

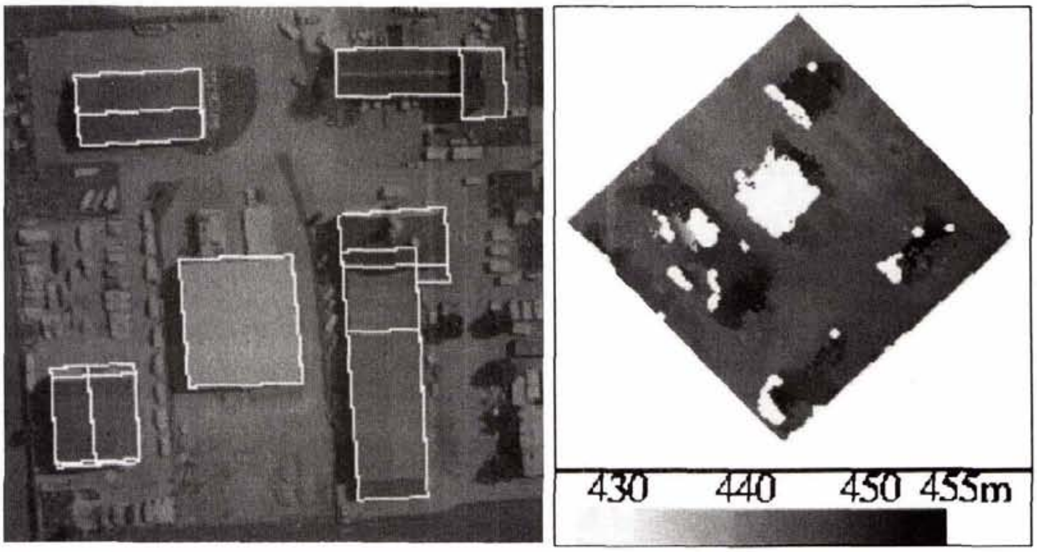


Figure 13. A building boundary map (15 BHS) and a DEM for "Avenches industrial" scene.

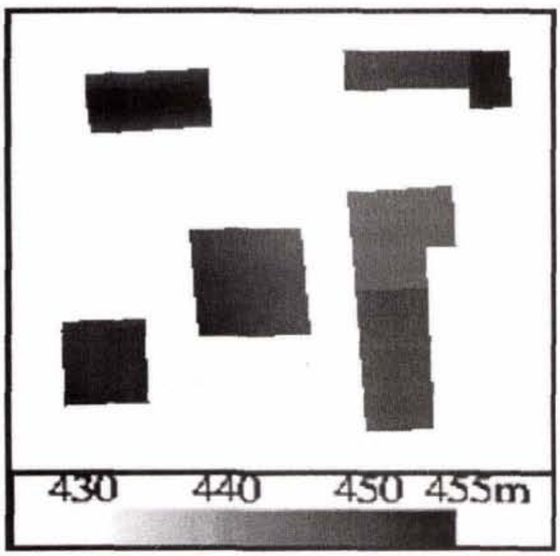


Figure 14. A building height map for "Avenches industrial" scene.

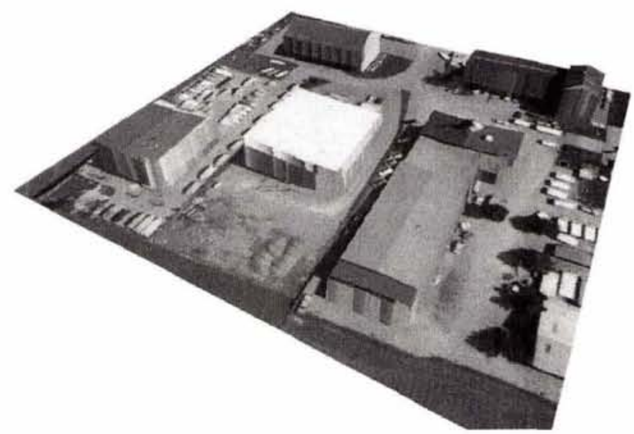

Figure 15. A perspective view of buildings, "Avenches industrial" scene.

## References


Baltsavias, E., S. Mason, and D. Stallmann, 1996. Use of DTMs/DSMs and Orthoimages to Support Building Extraction, *Automated Extraction of Man-Made Objects from Aerial and Space Images* (A. Gruen, O. Kuebler, and P. Agouris, editors), Birkhauser, pp. 199–210.

Canny, J., 1986. A Computational Approach to Edge Detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(6): 679–697.

Cochran, S.D., and G. Medioni, 1992. 3-D Surface Description from Binocular Stereo, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(10):981–994.

Davis, J.C., 1996. *Statistical Data Analysis in Geology, Second Edition*, John Wiley & Sons, pp. 239–248 and 383–405.

Haala, N., 1994. Detection of Building by Fusion of Range and Image Data, *Proceedings, ISPRS Commission III Symposium*, 30(Part 3/1):341–346.

Haala, N., and M. Hahn, 1996. Data Fusion for the Detection and Reconstruction of Buildings, *Automated Extraction of Man-Made Objects from Aerial and Space Images* (A. Gruen, O. Kuebler, and P. Agouris, editors), pp. 211–220.

Henricssion, O., F. Bignone, W. Willuhn, F. Ade, O. Kubler, E. Baltsavias, S. Mason, and A. Gruen, 1996. Project AMOBE: Strategies, Current Status and Future Work, *International Archives of Photogrammtery and Remote Sensing*, 31(B3):321–330.

Herman, M., and T. Kanade, 1986. The 3D Mosaic Scene Understanding System, *From Pixels to Predicates* (A.P. Pentland, editor), Ablex Pub. Corp., pp. 322–358.

Huertas, A., and R. Nevatia, 1988. Detecting Buildings in Aerial Images, *Computer Vision, Graphics, and Image Processing*, 41:131–152.

Huertas, A., M. Bejanin, and R. Nevatia, 1996. Model Registration and Validation, *Automated Extraction of Man-Made Objects from Aerial and Space Images* (A. Gruen, O. Kuebler, and P. Agouris, editors), Birkhauser, pp. 33–42.

Kim, T., 1995. *Automated 3D Modelling of Buildings from Aerial and Space Imagery Using Image Understanding Techniques*, PhD Thesis, University of London.

Kim, T., and J-P. Muller, 1994. Automated Building Height Extraction and Building Detection from High Resolution Aerial and Space Imagery, *Proceedings, The 4th IAPR International Workshop on Machine Vision Applications*, Kawasaki, Japan, 13–15 December, pp. 364–367.

———, 1995. Automated Building Height Estimation and Object Extraction from Multi-Resolution Imagery, *Proceedings of SPIE Conference on "Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II,"* Orlando, Florida, 19–21 April, 2486:267–276.

———, 1996a. Automated Urban Area Building Extraction from High Resolution Stereo Imagery, *Image and Vision Computing*, 14(1996):115–130.

———, 1996b. Building Extraction and Verification from Spaceborne and Aerial Imagery Using Image Understanding Fusion Technique, *Automated Extraction of Man-Made Objects from Aerial and Space Images*, (A. Gruen, O. Kuebler, and P. Agouris, editors), Birkhauser, pp. 221–230.

Maitre, H., and W. Luo, 1992. Using Models to Improve Stereo Reconstruction, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):269–277.

Mohan, R., and R. Nevatia, 1989. Using Perceptual Organization to Extract 3D Structures, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139.

Nicolin, B., and R. Gabler, 1987. A Knowledge-Based System for the Analysis of Aerial Images, *IEEE Trans. on Geoscience and Remote Sensing*, GE-25(3):317–329.

Otto, G.P., and T.K. Chau, 1989. A Region-Growing Algorithm for Matching of Terrain Images, *Image and Vision Computing*, 7(2): 83–94.

Petrou, M., and J. Kittler, 1991. Optimal Edge Detectors for Ramp Edges, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(5):483–491.

Shufelt, J., and D.M. McKeown, 1993. Fusion of Monocular Cues to Detect Man-Made Structure in Aerial Imagery, *Computer Vision, Graphics, and Image Processing: Image Understanding*, 57(3): 307–330.

(Received 10 March 1997; revised and accepted 18 February 1998; revised 27 February 1998)