# RESULTS OF THE TEST ON IMAGE UNDERSTANDING OF ISPRS WORKING GROUP III/3

Monika Sester, Werner Schneider and Dieter Fritsch
Institute of Photogrammetry
Stuttgart University
P.O.B. 106037
70049 Stuttgart / Germany
{monika.sester/werner.schneider/dieter.fritsch}@ifp.uni-stuttgart.de

Commission III, Working Group 3

**KEY WORDS:** data fusion, data integration, model based image interpretation

**ABSTRACT:**

The "Test on Image Understanding" was an initiative started by the ISPRS Working Group III/3 in March 1994. The idea was to get hold of and review the the state of art of image understanding techniques, especially in the domain of automatic recognition and reconstruction of cartographic objects. According to new tendencies in image understanding, as much information as available should be used for the interpretation. In this context, the test provides different data sets of information which is generally available. The additional information includes GIS ground truth, but also a digital surface model, color and stereo aerial imagery. The scales of the imagery vary from 1:5000 to 1:12000, and the corresponding ground pixel sizes are in the range of 0.23 m to 2 m. The tasks slightly vary for the different data sets, but primarily focus on the detection and reconstruction of houses, streets, water bodies, field parcels.

## 1 Introduction and Overview

The given data sets represent realistic information, what should be generally available to date. Thus the overall objective of the test aimed at information interpretation and verification. In particular interests were directed to the methods integrating further (external) information. Therefore the idea behind the test was twofold. On one hand researchers should have the opportunity to test and compare their algorithms on a standardized, realistic test data set. In this way results are getting comparable, and individual approaches or methods can be improved. The second issue was to stimulate research by providing a broad set of different data sources.

Due to the fact that different knowledge sources for the interpretation are available, another fundamental issue of the whole test comes up: in order to evaluate the results and the applicability of the approaches, the underlying strategy and the models used have to be made transparent. This implies to make explicit the knowledge used - namely the object models and the strategy of the algorithms. Concerning the object models, some interesting questions arise, e.g. whether specific object models are needed or or whether they can be generic, whether a 2D-cue of a 3D-object is sufficient for its detection (and for its reconstruction, resp.) ? In this way all the assumptions the program uses should be clearly separated and not hidden in program code.

Having an exact description of the procedures applied and the knowledge involved, extensions to other data sets and also to other objects seem to be possible. Ideally a separate knowledge base consisting of object models and corresponding strategies helps to handle different kinds of problems. In order to clarify this aspect, every participant had to report an extensive description of the approach according to a detailed questionnaire. The tasks concerning the individual data set have not been specified too strictly. The underlying reason was to leave open a broad spectrum of possible operations on the data. In particular, the data set *flat* consists of a stereo image pair and a DEM generated by *shape-from-stereo*. The objects (buildings) contained in the data sets can either be reconstructed by DEM-analysis, by stereo reconstruction or by monoscopic interpretation.

## 2 Test Setup and Responses

The test setup has been reported in Fritsch, Sester & Schenk [1994]. Although there has been a great response to the data (over 300 people have accessed our ftp-server), only a small number of scientists actually handed in their results. The following table gives an overview of the tasks the individual participants of the test solved. The table reveals that the data sets including range data (*flat, suburb*) seemed to be a challenge for most of the interested scientist.

| Participants | data set remstal | | | | data set glandorf | | | | data set *flat* | data set *suburb* |
|---|---|---|---|---|---|---|---|---|---|---|
| | street | field | house | railway | street | field | brook | house | | |
| 1 Stilla | | | | | | | | | x | |
| 2 Weidner | | | | | | | | | x | x |
| 3 Haala | | | | | | | | | x | x |
| 4 Schute | | | | | | | | | x | |
| 5 Fayek | | | | | | | | | x | |
| 6 Löcherbach | | x | | | | | | | | |
| 7 Li | x | | | | | | | | | |
| 8 Lotti | | | | | | | | | x | x |
| 9 Rosin | | | | | | x | | x | | |

In the following the prerequisites, strategies and methods of the individual participants are described in some detail.

One participant (8) used the stereo image pair as a test for a DEM-generation program ([Lotti & Giraudon 1994]); another one (7) tested his line-tracking program on the data set *remstal* (cf. [Trinder & Li 1995]). In the approach (9) of Fierens & Rosin [1994] GIS data is used to define training regions for a following classification process. Due to the fact that these tasks did not exactly match the scope of the test, they will not be treated in detail here. However, the focus of the evaluation concentrates on the reconstruction of man-made-objects using prior information (participants 1 to 6).

All the results reported back to the data provider are based on totally automatic strategies which involve no interaction of an operator.

## 2.1 Uwe Stilla

**Data Source:** Stereo image pair, data set *flat*

**Object Model:** The generic model describes a building as being composed of two roof parts, namely two rectangular areas in 3D.

**Prior Knowledge:** Prior knowledge is introduced concerning the camera parameters and the thresholds in the extraction and grouping procedure. The common sense knowledge used mainly concerns the scene model, especially the objects in a scene:

- ▷ Buildings are rectangular and have a length l_house (l_house_min < l_house < l_house_max).
- ▷ The two areas of a gabled roof enclose an angle gamma (gamma_min < gamma < gamma_max)

Image related information:

- ▷ Type of primitive objects for structure approximation (Type=LINE)
- ▷ The areas of a roof appear as parallelograms
- ▷ The small angle in a parallelogram is alpha (alpha_min < alpha < alpha_max)
- ▷ The edges of the roof (connected with the gable) have length l_side (l_side > l_side_min)
- ▷ The shorter side of two opposite sides of a parallelogramm must be at least half as long as the longer side

**Strategy:** In a preprocessing step a symbolic description of the images is generated, which consists of a collection of straight lines (LINE). In both images preprocessing and 2D-analysis is carried out independently.

Starting with the object primitives LINE more complex objects ANGLE, U_STRUCTURE, PARALLELOGRAM are constructed by grouping. In lower levels there is no decision yet if an object is part of a target object or not. Thus, a lot of alternative objects are produced.

The 3D-analysis attempts to find pairs of 2D-objects (U_STRUCTURE or PARALLELOGRAM) which are projections of the same 3D surface. This is done by selecting pairs and examining rays originating at the centre of the projection and passing through the vertices of the 2D-objects. The 2D-objects will be called NOT CORRESPONDING if the distance between the rays of pairs of vertices is greater than a given threshold. In 3D-domain more complex objects (gabled roofs) are constructed, if the conditions in space are fulfilled (neighbourhood, location, orientation).

Pseudo code of the program is given by the set of production rules.

```
L  /\ L  (angle-shaped) -> A
A  /\ A  (u-shaped) -> U
U  /\ L  (parallelogram-shaped) -> P
U  /\ U  (corresponding in 3D) -> CA
P  /\ U  (corresponding in 3D)  -> CA
P  /\ P  (corresponding in 3D)  -> CA
CA /\ CA (building an edge in 3D) -> CE

(L)  LINE, (A)  ANGLE, (U)  U_STRUCTURE,
(P)  PARALLELOGRAM
(CA) PART OF ROOF, (CE) HOUSE ROOF
```

More details of the procedure can be found in [Stilla 1995] and [Stilla, Michaelsen & Lütjen 1995].

**Results:** Detection and reconstruction of 14 buildings (from 17 buildings in total).

## 2.2 Uwe Weidner

**Data Source:** Range data, data set *flat*

**Object Model:** The approach bases on generic object models, i.e. that buildings are usually higher then their surrounding topographic surface, that the ground plan of the buildings consists of straight lines and that these straight lines form polygons, which have edges being orthogonal, parallel, and collinear. Furthermore, parametric building models are used, namely rectangular buildings with either flat or symmetrically sloped roofs.

**Prior Knowledge:** The assumption of the minimal size of the buildings and their minimal height is enough to fix control parameters for the subsequent segmentation steps. Buildings are assumed to be separate from each other.

**Strategy:** The strategy consists of two steps, namely the detection of the buildings in the DEM and the reconstruction of a parametric or prismatic geometric description of each detected building.

The detection is achieved by computing an approximation of the topographic surface using morphological filtering. The difference of the approximation and the real surface gives regions of potential buildings, which are filtered again using the above mentioned thresholds concerning size and height. In the second step - the reconstruction - the potential buildings are analyzed according to height and shape. For each region the principal axes are computed, along with width and length. Depending on the height, either a prismatic or a parametric house model is adjusted to the segments using the main axis as ridge of the building. The method is described in detail in [Weidner & Förstner 1995] and [Weidner 1995].

**Results:** The algorithm works fine on the data set where the assumptions are clearly followed. All the 17 buildings could be detected and reconstructed. Investigations on the sampling size of the DEM show that the higher the grid size the better the results are.

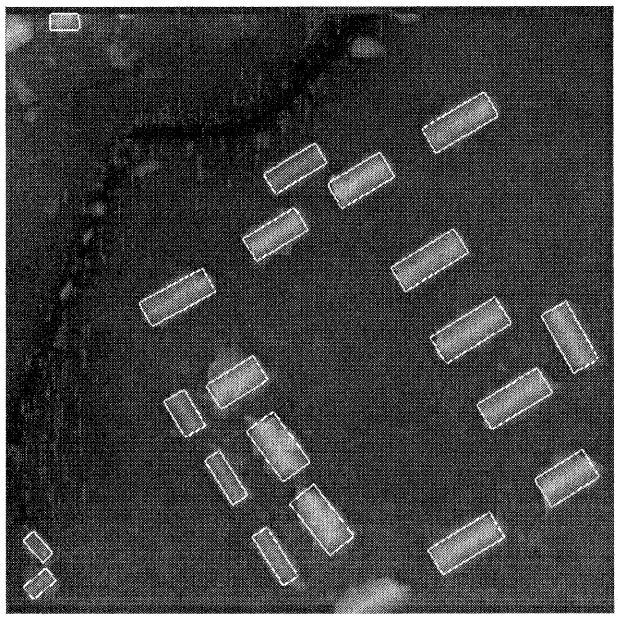

Figure 1: Detection and reconstruction of buildings in data set *flat*

In the second data set (*suburb*) however, the buildings are not that distinct both in height and separation from each other. Therefore the algorithm fails partly to detect all the buildings and is not able to reconstruct them correctly.

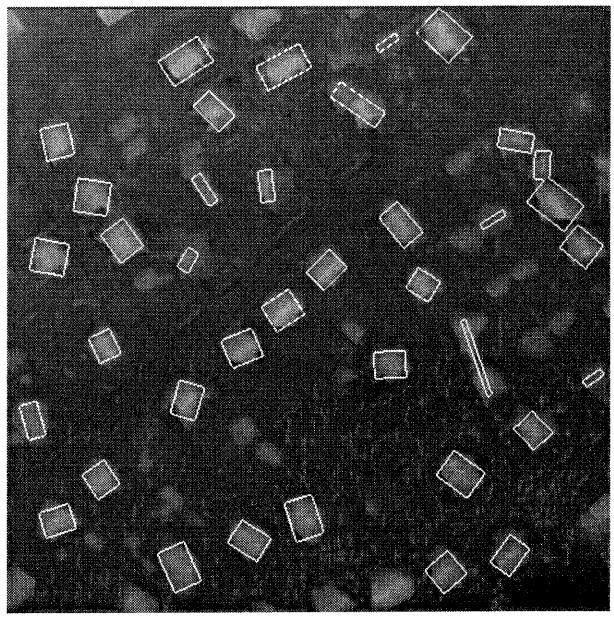

Figure 2: data set *suburb*: although many buildings are detected, most of them could not be reconstructed properly

## 2.3 Norbert Haala

The approach is similar to the previous one in terms of the underlying object model, but differs in the reconstruction phase.

**Data Source:** Stereo image pair and range data, data sets *flat* and *suburb*

**Object Model:** The approach also bases on the generic assumption, that buildings are usually higher then their surrounding topographic surface. As a description parametric building models are used, namely rectangular buildings with either flat or symmetrically sloped roofs.

**Prior Knowledge:** The minimal size of the buildings is 10 $m^2$ and their minimal height is 3 m.

**Strategy:** This approach bases on a fusion of stereo-image and range data. The strategy consists of two steps, namely the detection of the buildings in the DEM and the reconstruction of a parametric geometric description of each detected building in the stereo images.

The detection is achieved by computing regions of interest in the DEM. To this end, regions being higher than their surrounding are extracted. Regions of a certain height difference and size are investigated in the following reconstruction step. The reconstruction is performed in the stereo image. In both images straight lines are extracted and matched to form 3D-lines. The matching makes use of the approximate parallax given by the DEM. In the final step a parametric building model is approximated to this set of 3D-lines. The building with the best fit in terms of minimal errors is chosen to represent its correct reconstruction. Another approach aims at a segmentation of the DEM alone, namely extracting 3D-lines directly from the range data set and fitting the parametric model to these lines.

A detailed description can be found in [Haala 1995] and in [Haala & Hahn 1995].

**Results:** The algorithm was tested both on the *flat* and the *suburb* data set. In both cases, good results could be achieved. In the data set *flat* all 17 buildings could be detected and reconstructed, in the data set *suburb* **30 from 38**

were detected and reconstructed (cf. Figure 3). The second strategy which relies on a segmentation of the range data set only gives correct results for the *flat* data set, in the *suburb* data, the segmentation is not strong enough to feed the parametric model correctly (cf. also Weidner).

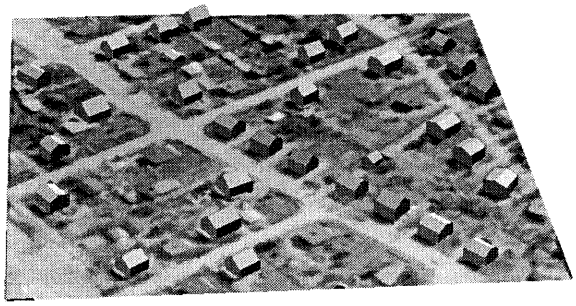

Figure 3: Perspective view of the detected and reconstructed buildings in the data set *suburb*

## 2.4 Klamer Schutte

**Data Source:** Single image, data set *flat*

**Object Model:** Buildings are modelled as polyhedral objects, what results in a polygonal description, in terms of the projected faces. For each object in the database a set of aspects is generated. Depending on the light source shadow regions are defined. Each aspect is defined by a collection of constraints concerning regions, contours, corners and their relations.

**Prior Knowledge:**

- ▷ Camera position and parameters,
- ▷ light position and direction,
- ▷ approximate size of buildings (xsize between 5 and 14 m, ysize between 10 and 30 m, h1 between 1 and 20 m, h2 between 1 and 10 m),
- ▷ noise in images (gaussian noise with $\sigma = 10$),
- ▷ generic parametric object descriptions which are transformed into aspects interactively.

**Strategy:** Object hypothesis are generated in a relaxation step, where the regions in the segmented image are matched to the aspects from the data base. This step is followed by an estimation and verification process, to exactly fit the objects to the image data. A detailed description of the approach is given in [Schutte 1994].

**Results:** Not all buildings present in the image are found. This is mainly caused by errors in the segmentation of the image, which cannot be recovered. Another cause is that the models used are quite simple (i.e. not including windows is the roof.) However, even if these models were incorporated, the segmentation of the image for such buildings proves to be extremely difficult.

For some buildings, multiple hypotheses are found. This is a 'feature' of the system used. The reason is that sometimes hypotheses are generated for 3 segments found (2 roof + shadow), and sometimes for less.

Some buildings found do not match exactly. This on one hand due to segmentation errors. Missed or incorrect boundaries will result in wrong object parameters. On the other hand it is due to the fact that not enough segments are found. If, for example, the shadow is not found, it is impossible to estimate the height h1 of the house. Since only one image is used, the z parameter is not very accurate.

## 2.5 Reda Fayek

**Data Source:** Range data, data set *flat*

**Object Model:** This approach does not base on an explicit building model, but on the idea, that man-made objects consist of planar faces and are of a certain size. For a detailed description see e.g. [Fayek & Wong 1994]. The generic building model consists of one or more vertical (or near-vertical) surface patches, together with the neighboring surface patches. Groupings of these patches have to have a certain size.

**Prior Knowledge:** This approach starts with a triangular mesh of the raw range data. The general strategy is to iteratively coarsen the given mesh while preserving topographic details of the original data. The mesh coarsening requires certain preset parameters controlling the allowable surface approximation errors.

- ▷ only surface patches larger than a certain minimal size are allowed to initiate a possible man-made surface entry. The threshold is set to 5 $m^2$.
- ▷ A potential building has to consist of a collection of patches with a given minimal size. This size threshold is set to 300 $m^2$.
- ▷ Man-made structures are separated from the background by the simple strategy of being enclosed by a larger nearly horizontal patch corresponding to the background.

**Strategy:**

- ▷ Triangulation of the range map.
- ▷ Segmentation of range map into nearly planar patches.
- ▷ Categorization of patches according to their slope.
- ▷ Instantiation of suitable patches for man-made structures.
- ▷ Growing of initiated models and final validation.

**Results:** The procedure principally aims at a detection and recognition of the man-made structures. Thus the result proves a good localization of the houses, hovever only a rather poor reconstruction.

The results on the data set *flat* are given in terms of the 3D-coordinates of the center of gravity of the building, the number of nearly-planar patches it consists of, the total outside surface of the building, and the total surface area covered by the it. 13 buildings could be detected properly, 4 buildings were reconstructed as being aggregated in 2 pairs. The author points out that for fast recognition even a coarser mesh is sufficient.

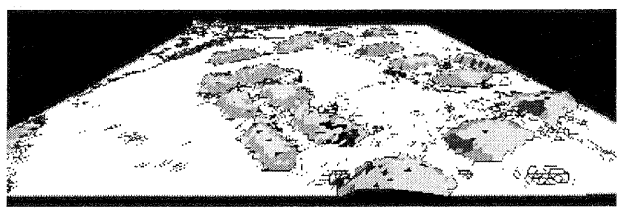

Figure 4: Detected buildings

### 2.6 Thomas Löcherbach

This work presents a method which adjusts prior GIS-information to given image data. In an adjustment process, the GIS-data is fit to the edges extracted from the images. The aim is the reconstruction of the geometry of land-use parcels and their classification.

**Data Source:** Multispectral image and GIS-information (field boundaries)

**Object Model:** The object model contains 3D polygons, representing the geometry of the land-use units. The radiometric part consists of one feature vector per object, e.g. the field mean, within-field variance, or a field histogram.

The geometry of the image model is represented by a 2D polygon network. The radiometric part contains the assumption of homogeneous features within a given object and a feature edge model describing the transition between two neighboring fields along the land-use boundary.

The observations of the adjustment process are the intensity values of the images and the map coordinates as prior information. The correct object coordinates, the transformation between image and object space, and the feature vectors per fields are derived in an estimation procedure.

**Prior Knowledge:** The GIS-data is used as prior information.

**Strategy:** Aim of the procedure is the estimation of the geometry of the land-use parcels. A field is assumed to be a homogeneous area. To set up the observation equations the transition of the feature of one field to the feature of the neighbouring field is modelled. Therefore the image is partitioned into regions along each boundary. From the pixels within one region along a boundary the position of the boundary and the features of the areas on both sides of the edge may be estimated. If fields are large compared to the pixel size, the pixels in the center of the field may be used to estimate the radiometry, not the geometry of the boundaries. The procedure is an iterative process, where each iteration step results in a new boundary position.

Differences between map and image may have several reasons: they may be shifted to some degree, there might be additional boundaries in the map which do not exist (or are not visible) in the image, or there are boundaries which are not contained in the map. The procedure aims at a reconstruction of the elements in the map, thus the 3rd case is not treated here. Detailed information on the method can be found in [Löcherbach 1994]

**Results:** The experiments reveal that the shifts between the data sets can be adapted very well (conf. Figure 5, upper). If, however, large displacements occur, further modelling would be necessary, e.g. to impose constraints concerning the parallelism of paths (conf. Figure 5, lower).

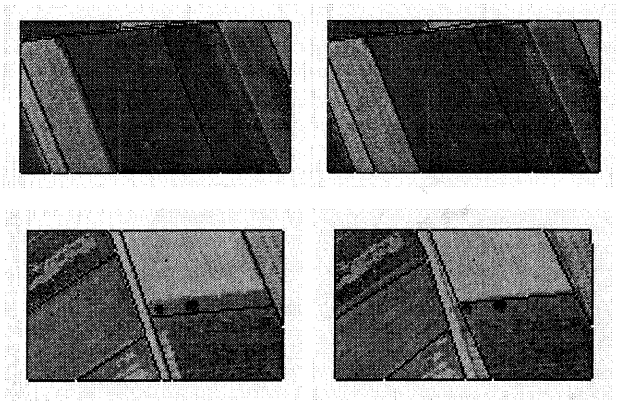

Figure 5: Initial maps (left) and corresponding estimated maps (right) superimposed on original image

## 3 Evaluation

Since most of the participants concentrated on the data set *flat* only these results will be presented and compared in detail. The evaluation bases on a reference data set which was measured manually on a digital stereo workstation. The accuracy of the manual measurement can be expected to be in the range of 0.20 to 0.30 m due to measurement and definition uncertainty. Thus the results of the participants can be compared against this reference. The buildings were represented simply by the 3D-coordinates of their roofs, i.e. each building is described by 6 coordinates. The comparison is done based on the differences in x,y, and z-coordinates between the manual measurement and the individual reports of the participants.

The following table gives the mean value of the differences (RMS) in the coordinates of the roofs of the buildings[1]:

| Participant | $\sigma_x[m]$ | $\sigma_y[m]$ | $\sigma_z[m]$ |
|---|---|---|---|
| Weidner | 1.60 | 1.29 | 0.52 |
| Stilla | 0.41 | 0.37 | 0.99 |
| Schutte | 1.38 | 0.97 | 4.09 |
| Haala (DEM-Image-Fusion) | 0.30 | 0.36 | 0.61 |
| Haala (DEM) | 0.61 | 0.42 | 1.09 |

The maximal RMS of all participants lie in the range of 0.4 to 1.5 m. These figures are higher than the expected values, which is probably due to the definition uncertainty of the building. The positional accuracy (in x and y) of methods using image data is higher than those using solely the DEM.

The following general conclusions can be drawn from the test:

- The problem of detection and recognition can be solved for the building objects from the range data alone - provided that the data is dense enough compared to the object size (cf. Weidner and Fayek). Range data is very suitable for localization, especially when objects distinctively emerge from background. Even if not, generic models help to increase the reliability of object recognition (e.g. the model that buildings consist of nearly planar surfaces, or the fit of a parametric building model to the data set).

[1] The results of Fayek are not included, since his aim was the detection and not the reconstruction of the buildings.

- ▷ Reconstruction requires more detailed information both on the data but also on the model side.
- ▷ All participants restricted themselves to simple building models (parametric or prismatic) except for Fayek. The generality of his model however is not strong enough with respect to the weak range data. The approach seems very promising when more detailed information is available (higher sampling rate of range data).
- ▷ A combination of all available information proves to deliver very reliable results (cf. Haala). In the given data set the DEM was derived by a matching process from the stereo images, thus the DEM is certainly less accurate that the original image information. Furthermore the sampling rate in the DEM is lower than the pixel size in the images. There are approaches to introduce building hypothesis into the matching process in order to derive a more accurate DEM (cf. [Maitre & Luo 1992] or [Kim & Muller 1995]).

## 4 Conclusion

The integration of prior information, especially GIS-data has not been exploited to the extent possible and epected. The only approach relying on prior GIS-information is of Löcherbach. He restricts himself to 2D-objects, although in principle the object-model can be given in 3D. This issue is of particular importance especially for the revision of existing databases (e.g. national databases like ATKIS).

For the detection and reconstruction of buildings using DEM in combination with stereo seems to be a very promising way (see also [Collins, Hanson, Riseman & Schultz 1995]).

In summary, the integrated interpretation of data of different data sources is only in the beginning. To date, still many researchers rely on a single data source, which is analyzed with specific strategies. This might be due to the fact, that this variety of different information sources has not been available until recently, and many people are not yet aware of it. With the availability of new sensors however (e.g. laser scanners) and also the direct availability of high resolution image data (e.g. airborne or space borne digital line scanners with resolutions from some decimeter to some m) many diverse information sources will be ready for use in the near future. The importance of data fusion has also been clearly pointed out at the ISPRS workshop in Stuttgart in November 1995 [Fritsch, Sester & Hahn 1995]. Especially for the extraction of man-made objects great profits can be expected.

## References


Collins, R., Hanson, A., Riseman, E. & Schultz, H. [1995], Automatic extraction of buildings and terrain from aerial images, *in* Grün, Kübler & Agouris [1995], pp. 169–178.

Fayek, R. & Wong, A. [1994], Triangular mesh model for natural terrain, *in* 'Proc. of the SPIE Intelligent Robots and Computer Vision XIII', Boston.

Fierens, F. & Rosin, P. [1994], Filtering remote sensing data in the spatial and feature domains, *in* 'Conf. on Image and Signal processing for Remote Sensing', Vol. 2315, SPIE, pp. 472–482.

Fritsch, D., Sester, M. & Hahn, M. [1995], Minutes of the joint workshop on integrated acquisition and interpretation of photogrammetric data, Technical Report 1995/2, Institut für Photogrammetrie, Stuttgart.

Fritsch, D., Sester, M. & Schenk, T. [1994], Test on image understanding, *in* H. Ebner, C. Heipke & K. Eder, eds, 'Spatial Information from Digital Photogrammetry and Computer Vision', Vol. 30/3, ISPRS, Munich, Germany, pp. 243–248.

Grün, A., Kübler, O. & Agouris, P., eds [1995], *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, Boston, Berlin.

Haala, N. [1995], 3D building reconstruction using linear edge segments, *in* D. Fritsch & D. Hobbie, eds, 'Photogrammetric Week '95', Herbert Wichmann Verlag, Heidelberg, pp. 19–28.

Haala, N. & Hahn, M. [1995], Data fusion for the detection and reconstruction of buildings, *in* Grün et al. [1995], pp. 211–220.

Kim, T. & Muller, J.-P. [1995], Building extraction and verfification from spaceborne and aerial imagery using image understanding fusion techniques, *in* Grün et al. [1995], pp. 221–230.

Löcherbach, T. [1994], Reconstruction of land-use units for the integration of gis and remote sensing data, *in* H. Ebner, C. Heipke & K. Eder, eds, 'Spatial Information from Digital Photogrammetry and Computer Vision', Vol. 30/3, ISPRS, Munich, Germany.

Lotti, J.-L. & Giraudon, G. [1994], Adaptive window algorithm for aerial image stereo, *in* '12th International Conference on Pattern Recognition', IAPR, IEEE, Jerusalem, pp. 701–703.

Maitre, H. & Luo, W. [1992], 'Using models to improve stereo reconstruction', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 269–277.

Schutte, K. [1994], Knowledge Based Object Recognition of Man-Made Objects, PhD thesis, University of Twente, Enschede, Netherlands.

Stilla, U. [1995], 'Map-aided structural analysis of aerial images', *Journal of Photogrammetry and Remote Sensing*.

Stilla, U., Michaelsen, E. & Lütjen, K. [1995], Structural 3d-analysis of aerial images with a blackboard-based productionsystem, *in* Grün et al. [1995].

Trinder, J. & Li, H. [1995], Semi-automatic feature extraction by snakes, *in* Grün et al. [1995], pp. 95–104.

Weidner, U. [1995], Building extraction from digital elevation models, Technical report, Institut für Photogrammetrie, Bonn.

Weidner, U. & Förstner, W. [1995], 'Towards automatic building extraction from high resolution digital elevation models', *ISPRS Journal* **50**(4), 38–49.