STRUCTURAL 3D-ANALYSIS OF URBAN SCENES FROM AERIAL IMAGES

U. Stilla, K. Jurkiewicz

Forschungsinstitut für Informationsverarbeitung und Mustererkennung (FGAN-FIM)
Eisenstockstr. 12, 76275 Ettlingen, GERMANY
E-mail: usti@gate.fim.fgan.de

ISPRS Commission III, Working Group 3

KEY WORDS:

Recognition, Urban, Aerial Image, Three-dimensional Structure Analysis, Production System, Blackboard System

ABSTRACT:

This paper presents a model-based method for the automatic analysis of structures in aerial images. The model of the objects to be recognized is described in the form of a production net. The production net represents a hierarchical organisation of subconcepts and production rules. The production rules are implemented in a blackboard architecture as knowledge sources. The database of the blackboard system is stored in an associative memory. The recognition of spatial objects from an image sequence is illustrated by an example of a simple model for the geometric 3D-reconstruction of a roof. An ISPRS test dataset was used in order to evaluate the efficiency of the analysis system.

1 INTRODUCTION

The automatic interpretation of aerial images by knowledge-based systems has been a subject of research for some time [Nagao & Matsuyama, 1980], [McKeown et al., 1985], [Nicolin & Gabler, 1987]. The research activities in the field of object recognition have received a special impulse from the increased demand of urban scene description. This can be particularly attributed to the development of geographical information systems. But also the availability of additional information by digital maps, spatial image sequences, and distance data has given new impulses to research. Especially with respect to the recognition of man-made objects from large scale images 3D-reconstruction has increasingly gained importance.

In this paper we describe a system for image interpretation and a method for 3D-recognition. The work is part of a research project for map-aided image analysis with two- and three-dimensional models [Stilla et al., 1995b]. The title of this project[1] is: *Analysis of aerial and satellite images for automatic determination of the ground sealing of urban areas.*

In the field of automatic object recognition, knowledge based methods are increasingly used for the analysis and description of aerial imagery. Of particular interest are structure oriented hierarchical methods, which build up structure hierarchies by composing complex structures from less complex structures. Using this approach the analysis proceeds step by step, with constant reference to the patterns being analyzed, producing interim results of increasing degrees of abstraction. Hence following Marr [Marr, 1980], the process of visual recognition is interpreted as the active construction of a symbolic scene description from images.

2 ANALYSIS STRATEGY

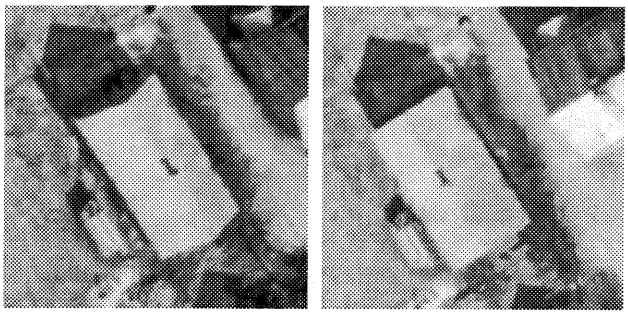
The subject of investigation is the recognition of three-dimensional objects from only two images (Fig 1). It is presupposed that the formulas of projection of points of the scene into the images are known, which is essential for stereotriangulation. In contrast to other research approaches there is no need for epipolar geometry.

In order to test the efficiency of the approach we abstained from including external information such as height information [Haala & Hahn, 1995], digital map data [Quint & Sties, 1995][Stilla, 1995], etc. Due to the lack of external information for establishing hypotheses (e.g. number, position, and orientation of objects) the analysis is carried out bottom up. As on lower abstraction levels it can often not be decided whether an object is part of a target object, alternative interpretations (competing interim results) are pursued in parallel and independently.

The analysis is carried out symbolically by regarding the primitive objects as elements of a set, i.e. preprocessing results involving topological relations between primitive objects are not used. Neighbouring objects in the image are not necessarily neighbouring in space. The actual recognition takes place on the scene level (in space) instead of the sensor level (in image).

[1]This project is funded by the Deutsche Forschungsgemeinschaft (DFG) and is carried out in cooperation with the Institut für Photogrammetrie und Fernerkundung (IPF), University of Karlsruhe

For the consideration of different recognition tasks it is particularly convenient to use a uniform framework for the analysis. The software tools and special hardware which have already been developed can be used for an analysis. For the description of an object model by a production net we use modular semantics. This description is translated into independent knowledge sources (processing modules), which may be reused for other analysis tasks. For the evaluation we assume that the number of primitive objects describing the image content can possibly be very high. The analysis concept is designed in such a way that in principle it allows a parallelization of processing. Furthermore, the evaluation of several information sources (e.g. multisensor images, spatial image sequences) can easily be integrated.



a b

Fig. 1: Section of ISPRS-Test dataset FLAT
a) left image, b) right image

3 BLACKBOARD-BASED PRODUCTION SYSTEM BPI

For structure analysis of complex scenes the *b*lackboard-based *p*roduction system for *i*mage understanding (BPI) [Lütjen, 1986] is used as framework.

3.1 Production system

A production system typically consists of three basic components: a database, a set of production rules and a control unit. The knowledge about object structures is represented by a set of production rules. A production rule, or production, is a statement in the form:

IF *condition* holds, THEN *action* is appropriate.

The execution of *action* will result in a change of the data contained in the data base of a production system. A control unit controls the overall system activity and has the special task of deciding which production (with satisfied condition part) to *fire* next.

The process of building up more complex structures from less complex structures, using such productions, can be described by a rewriting system. With reference to formal languages the rewriting system may be determined by a Grammar G. Such a formal grammar is defined by a 4-tuple

$$G = (S, V_n, V_t, P),$$

where S is a set of start symbols (target objects), V_n is a set of non-terminal symbols (partial objects) V_t is a set of terminal symbols (primitive objects) and P is a set of rewriting rules (productions). Attributes are assigned to the objects, which represent certain structures. The productions determine how a given set of objects is transferred into a set of more complex objects.

In analogy to string grammars we may say that an image content is parsed (bottom up) by the process of image analysis. Instead of examining concatenation as is done by parsers for string grammars, we examine the topologic or geometric relation of objects in the *condition* part of a production. Therefore, a production rule may be written in the form:

$$\boxed{P_i : X \wedge Y \odot \xrightarrow{i} Z}$$

This means that, if an object of type X and an object of type Y fulfil the relation $\odot$, then an object specific generative function $\xrightarrow{i}$ is carried out which produces an object of type Z. Here the productions describe the *part_of* relations.

Starting with primitive objects, a target object can be composed step by step using the productions repeatedly. Similar to tabular parsing methods (e.g. in Aho & Ullman, 1972) all interim results (partial objects) remain stored in the database during the analysis.

The *general* interaction of productions and the stepwise transfer of objects into objects of a higher abstraction level can be depicted by a production net (Fig. 3). The compositions for the *actual* objects (instances) are recorded with the aid of pointers and can be illustrated by a derivation graph.

After the analysis, derivation graphs of the target objects can be constructed and used to explain the results. Thus, the subset of primitives, which represents the target objects can be determined. If we compare this subset with the set of primitives, we may say that the production net acts like a filter.

3.2 Blackboard Architecture

In the BPI-System the productions are implemented in a blackboard architecture (Newell, 1962; Nii, 1986). Generally, a blackboard architecture consists of a global database (blackboard) and a set of knowledge sources, which communicate only via the blackboard. In BPI the global database is stored in an associative memory. Knowledge sources are constructed as independent object specific processing modules, which examine a *condition* and execute an *action* of a production.

Systems with a blackboard architecture are essentially *data driven*. One or more hypotheses "*part_of* a more complex object" are attached to an object. An object-hypothesis pair (processing element) triggers a processing module to verify the hypothesis. The hypotheses arise from the production net, so that the analysis pro-

ceeds in a goal-directed manner. A control unit, containing a priority ordered queue of processing elements, is added to the blackboard system. Further details concerning the dataflow of the BPI-System are described in previous papers [Lütjen, 1986; Stilla, 1995]

4 OBJECT MODEL

In order to explain the 3D-reconstruction we chose a simple model for the recognition of a house. In many aerial images containing houses only their roofs can be recognized. Thus the houses are actually described by their roofs. This paper only takes houses with simple gabled roofs (ROOF) into account. It is assumed that significant parts of a roof can be described as rectangles in the scene and parallelograms in the image. This applies for many houses in aerial images.

A simple rewriting system for a roof is determined by

$$G_{ROOF} = (\{R\}, \{R, F, P, U, A\}, \{L\}, \{P_1, P_2, P_3, P_4, P_5\}).$$

According to G_{ROOF} and starting with the primitive objects L, the partial objects A, U, P, F, R are composed using the productions P_1-P_5 (see Tab. 2), with object R representing the target object ROOF.

P_i :	Objects X,Y	$\odot$	$\xrightarrow{x}$	Object Z
P_1 :	L $\wedge$ L	①	$\xrightarrow{1}$	A
P_2 :	A $\wedge$ A	②	$\xrightarrow{2}$	U
P_3 :	U $\wedge$ L	③	$\xrightarrow{3}$	P
P_4 :	(U$\vee$P) $\wedge$ (U$\vee$P)	④	$\xrightarrow{4}$	F
P_5 :	F $\wedge$ F	⑤	$\xrightarrow{5}$	R

①	angle-shaped, $45^o \leq \alpha \leq 135^o$
②	u-shaped
③	parallelogram-shaped
④	corresponding in 3D
⑤	building an egde in 3D, $90^o \leq \gamma \leq 170^o$

Tab. 2: Table of productions

A production net for the object ROOF is depicted in Fig. 3. The analysis distinguishes between a 2D- and 3D-analysis. First, the 2D-analysis is carried out independently in different images.

4.1 2D-Analysis

Starting with the object primitives LINE, the objects ANGLE, U_STRUCTURE and PARALLELOGRAM can be built up by applying the productions (P_1-P_3). Objects ANGLE are built up by connections in pairs of objects LINE (P_1). The constellation of objects LINE enclosing an angle α can be L-shaped or T-shaped (Tab. 2 ①). If two objects ANGLE form a structure like an open parallelogram they are combined to an object U_STRUCTURE (P_2). An object PARALLELOGRAM can be built up if objects U_STRUCTURE and LINE are compatible (P_3). In the subsequent 3D-analysis 2D-objects form the primitives.

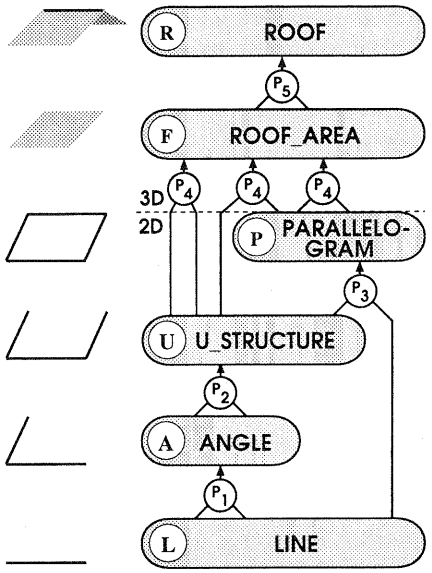


Fig. 3: Production net

4.2 3D-Analysis

The 3D-analysis attempts to find in two different images 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 by passing through the vertices of the 2D-objects. The rays result from inverse mapping of the projection. On ideal conditions rays running through corresponding object vertices of two images will intersect in the 3D-space. Due to image noise, processing errors, and inaccurate camera parameters the rays generally do not intersect. Hence, the minimal distance between the rays is calculated. The 2D-objects will be called *not corresponding* if this distance between the rays of pairs of vertices is greater than a given threshold. Additionally, a model-based plausibility check ist carried out (e.g. regarding the position: object points must not be under the earth's surface).

If 2D-objects of different images correspond, the object ROOF_AREA is generated (P_4). If objects ROOF_AREA are oriented in a way that the surface normals enclose an angle γ that lies within a certain angle interval and if they are located in a way that the vertices are neighbouring, then a target object ROOF is generated (P_5). The angle interval is assumed to be known (Tab. 2 ⑤). After the 3D-analysis is complete the best objects ROOF are selected.

5 PREPROCESSING

In the preprocessing stage a symbolic description of the scanned aerial image is created. The preprocessing is carried out in four steps.

(1) At first the greylevel image (Fig. 4a) is transferred

into a sequence of n_t binary images (Fig. 4b) by n_t thresholds. Possessing previous knowledge of the intensities of objects in greylevel images, a certain distribution of thresholds can be set up. Without this knowledge thresholds are distributed equidistantly between the minimal and maximal greylevel in the image. The number of thresholds n_t is a process parameter (which is usually chosen $n_t \geq 8$).

(2) Then the contour lines of the segments in each binary image are detected (Fig. 4c). Considering the image greylevels to be altitudes of a mountain, we obtain the lines of equal altitudes of this mountain by segmentation and contour tracking. In areas of high gradients we will receive many altitude lines, in areas of low gradients few altitude lines.

(3) In the next step of preprocessing the contour lines are approximated by short straight lines (Fig. 4d). This process of approximation is done by applying a dynamic split algorithm, which works similar to the algorithm of Ramer [1972]. In order to carry out such an approximation the contour line must have a minimum length. A termination criteria is given by a minimum quality of approximation and a minimal length for line segments.

(4) In the last step the short line segments of the contour sequence are grouped together to long lines. The long lines are stored as primitive objects LINE (Fig. 5, L, left) in the blackboard. They are attributed with length, orientation, coordinates, original image (left/right) and an assessment. The assessment is deduced from the quality of approximation.

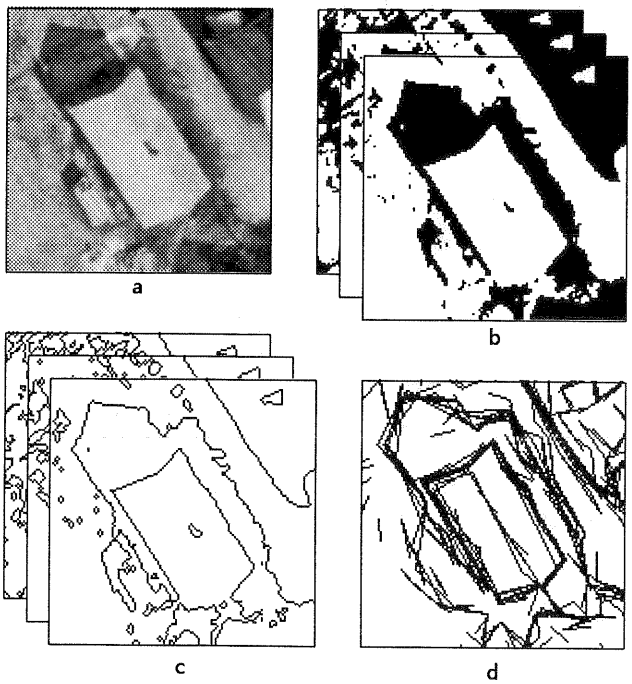


Fig. 4: Preprocessing (section of left image).
a) greylevel image, b) binary image sequence,
c) contour image sequence, d) short lines

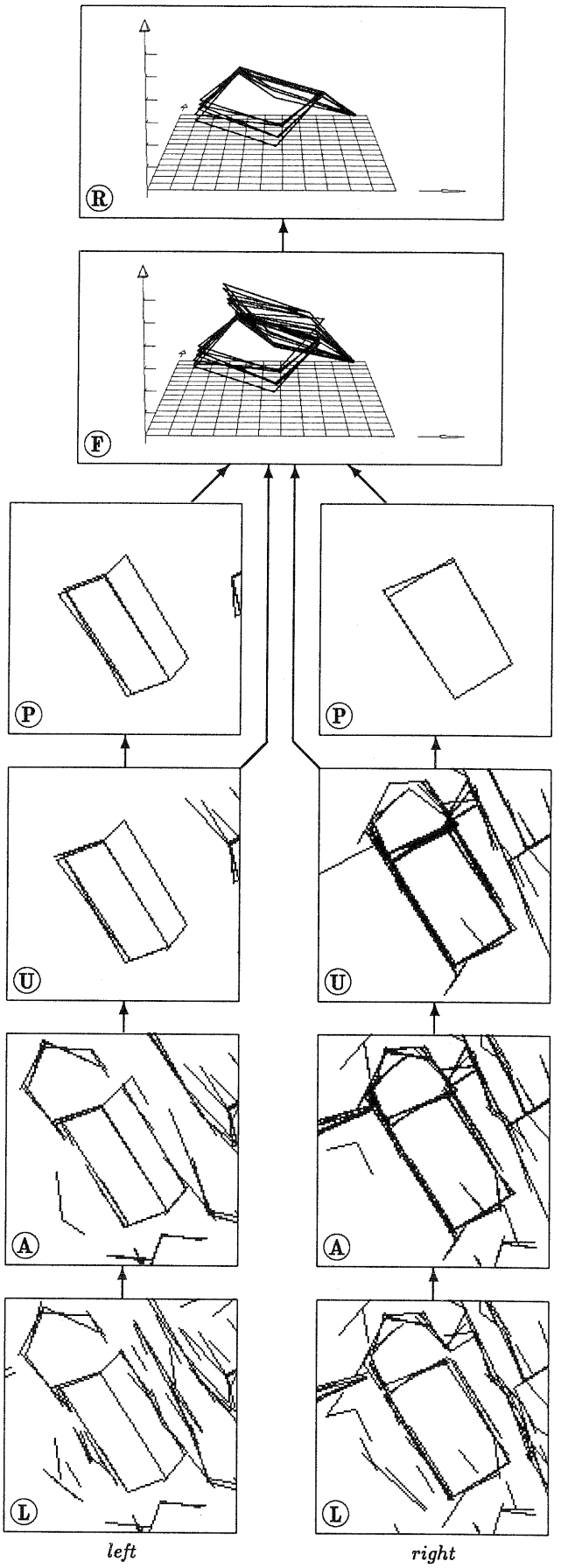


Fig. 5: 3D-Analysis. Interim results displayed by different sets of objects.

6 RESULTS

The image data is taken from a database of aerial images provided via FTP by the ISPRS Working Group III/3, *Test on Image Understanding* [Fritsch & Sester, 1994]. Aerial photographs (f=153.19 mm, scale 1:4000) of an urban area were used as test data. The scanned b/w images (stereo-pair) have a resolution of 60 μm per pixel. The image size is 1000 $*$ 1000 pixels. The images are in epipolar geometry. Camera parameters and projection formulas are given.

In the preprocessing stage for both images about 40000 short lines and 6000 primitive objects LINE are generated. Using the productions of the production net illustrated in Fig. 3 the scene analysis was carried out. As all generated objects remain stored in the blackboard, the process of analysis can be examined easily. For interactive selection and visualisation of object sets (interim results) an explanation component is used.

In order to demonstrate the recognition of a single house all objects generated within the scene section shown in Fig. 1a,b are illustrated in Fig. 5. Starting with the objects LINE in both the left and the right *channel* the stepwise *composition* of compatible object combinations can be traced up to the objects ROOF. On each stage the image structures are subjected to additional geometrical constraints by applying productions. The chaining of productions in a production net results in logical AND-operations of constraints. Tracing the subimages (L) to (P) in Fig. 5 we realize that parallelogram-shaped image structures are filtered out of the primitiv objects LINE.

Fig. 5 shows that both in the left and in the right *channel* the number of objects PARALLELOGRAM is smaller than the number of the objects LINE. Due to the combination of the objects PARALLELOGRAM and U_STRUCTURE of the right and the left channel (P_4 in Tab. 2 and Fig. 3) the number of the objects (ROOF_AREA) is much bigger than the number of objects (PARALLELOGRAM). However, the number of objects (ROOF) is smaller than the number of objects (ROOF_AREA).

Thus, the structures displayed in Fig. 5 (R) meet the geometrical relations of the stated productions within the limits of tolerance set by the parameters. The displayed objects ROOF cumulate in a small region of the scene and are a clear indication of a house. The object ROOF with the best assessment stands representatively for the house's roof of the scene.

This object is displayed in Fig. 6 (top). The previous objects which have built up the target object can be obtained by the derivation graph. A part of this derivation graph is displayed graphically in Fig. 6. The result of the analysis of the *ISPRS* dataset *FLAT* is shown in Fig. 7. For a 3D-visualisation of the recognized buildings we have assumed a flat terrain and to this plane the left image was mapped.

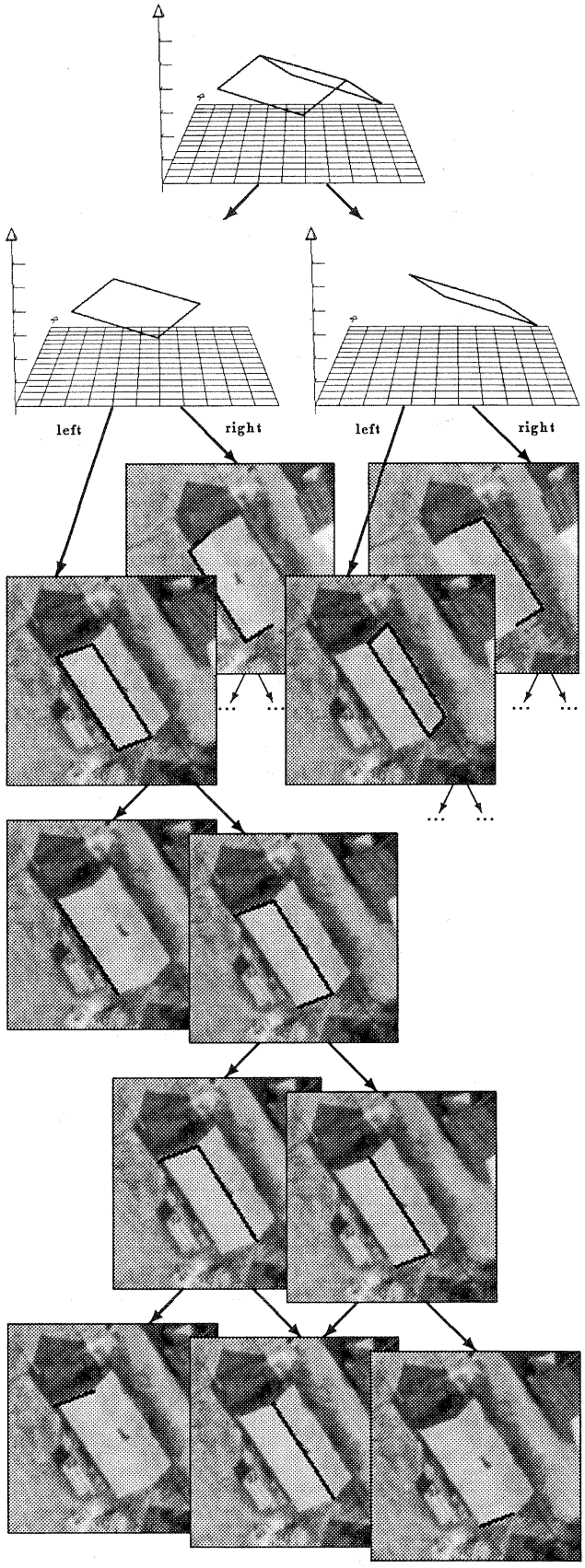


Fig. 6: Part of a derivation graph

7 DISCUSSION

The results of Fig. 7 show, that in principle the recognition of simple gabled roofs is possible using the presented production net. The scene section displayed by the left and right image of the dataset *FLAT* contains 18 houses. Out of these 14 houses have been detected by the objects ROOF. For one house the indication was provided by the objects ROOF_AREA. Three houses were not detected. Two of these undetected houses (Fig. 7-e7,-f8) could not be built up, because they did not fit the model. They do not possess a simple gabled roof. For one house (Fig. 7-a9) the object ROOF could not be generated, because due to the weak contrast in the right image the corresponding objects U_STRUCTURE could not be built up.

In order to build up a target object, all partial objects required by the model have to be contained in the set of primitive objects. If a required object is missing caused by a distortion, the target object may not be generated. Therefore the preprocessing parameters have to be chosen such that the set of primitive objects contains the required objects with high certainty. In this case a lot of primitive objects will be generated, which do not belong to target objects. The subsequent active geometrical constraints in our production net are so strong that coincidentally generated objects (distortions) will not build up target objects.

Due to the detailed modelling it may not be presumed, that all parts required for the detection, are actually available; incomplete structures also have to be considered. In the present production net both objects PARALLELOGRAM and objects U_STRUCTURE, which are used for the construction of the object ROOF_AREA, are considered (P_4). The objects U_STRUCTURE are the required parts and the objects PARALLELOGRAM are the optional parts of the model. If stronger distortions are to be tolerated the number of required parts has to be reduced. In further studies we will have to examine what effects a *loosening* of geometrical constraints will have on the analysis (computation effort) and its results (detection rate).

Looking at Fig. 5 one could get the impression that the execution of productions takes place *in layers*. This means that first all objects ANGLE are generated from objects LINE (P_1), then all objects U_STRUCTURE are generated from the objects ANGLE (P_2) etc. However, this is not the case because the analysis is data-driven. The order of execution of productions depends on the assessment of objects and takes place in varying hierarchical stages. Nevertheless the exceptional case of processing in layers (breadth-first search) may be forced (model-driven) using the object assessment corresponding to their hierarchical stage.

In this paper we have assumed that no external information is available. For the present tasks the processing order of the productions is of no importance, because all the productions with satisfied condition part are carried out (complete search).

If external information is available, this information may influence the models or can be used indirectly for the control of the analysis. In the BPI-system the control of analysis (top-down) is possible by defining expectations [Stilla, 1995]. Expectations may be set up using internal information (interpretation of interim results and context knowledge) or external information (additional scene information) and will be automatically applied during the analysis. Expectation ranges may by formulated for any attributes (e.g. position, orientation, angle, area, length ratio, etc.). In the search process, the focus of attention (computation power) can be directed to the corresponding attribute ranges (e.g. area of interest). Using for example terrain data (DTM in Fig. 8) as an external information source, expectations from height data may be easily defined for the position of the objects to be generated.

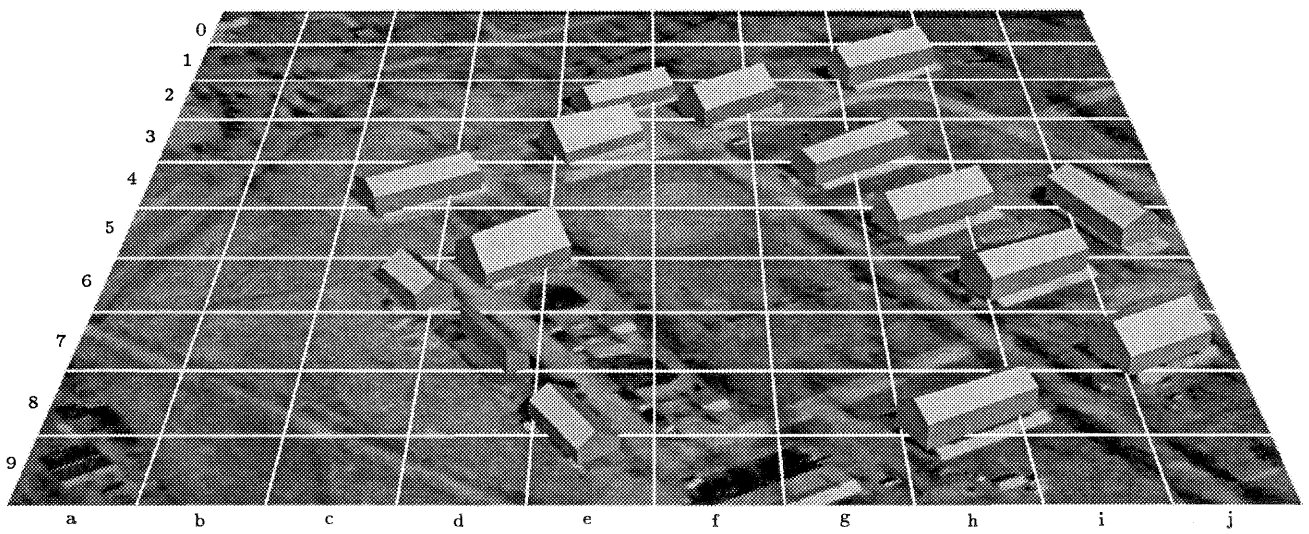


Fig. 7: Result of ISPRS-test dataset FLAT. 3D-visualisation of reconstructed buildings together with left image

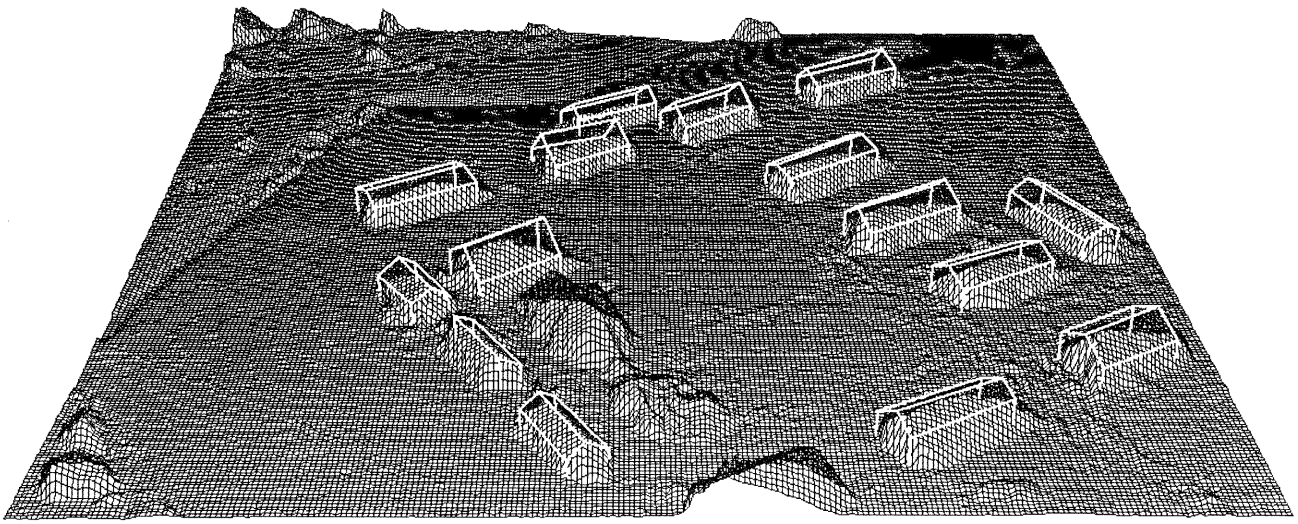


Fig. 8: Perspective view of reconstructed buildings together with DTM. The DTM was not used for the analysis !

8 REFERENCES

Aho A, Ullmann JD (1972) The theory of parsing, translation and compiling. London: Prentice-Hall

Bähr HP, Quint F, Stilla U (1995) Modellbasierte Verfahren zur Kartenfortführung. ZPF, 63(6): 224-234

Fritsch D, Sester M (1994) Test on image understanding. In: Ebner H, Heipke C, Eder K (eds) Spatial information from digital photogrammetry and computer vision. International Archives of Photogrammetry and Remote Sensing, Vol. 30, Part 3/1, 243-248

Füger H, Stein G, Stilla U (1994) Multi-populations evolution strategies for structural image analysis. IEEE Conference on Evolutionary Computation (ICEC'94), Orlando, Vol I, 229-234

Grimson WEL (1990) Object recognition by computer: The role of geometric constraints. Cambridge: MIT Press

Haala N, Hahn M (1995) Data fusion for the detection and reconstruction of buildings. In: Gruen A, Kuebler O, Agouris P (eds) Automatic extraction of man-made objects from aerial and space images, 211-230. Basel: Birkhäuser.

Lütjen K (1986) BPI: Ein Blackboard-basiertes Produktionssystem für die automatische Bildauswertung. In: Hartmann G (ed) Mustererkennung 1986, 8. DAGM-Symposium. Berlin: Springer, 164-168

Marr D (1980) Vision. San Francisco: Freeman

McKeown DM, Harvey WA, McDermott (1985) Rule-based Interpretation of aerial imagery. IEEE PAMI, 7: 570-585

Nagao M, Matsuyama T (1980) A structural analysis of complex aerial Photographs. New York: Plenum

Nicolin & Gabler (1987) A knowledge-based system for the analysis of aerial images. IEEE Trans. on geosc. and rem. sens., Vol. GE-25: 317-329

Newell A (1962) Some problems of basic organization in problem-solving programs. In: Yovits MC, Jacobi GT, Goldstein GD (eds) Proceedings second conference on self-organizing systems, Spartan Books, 393-423

Nii HP (1986) Blackboard systems. AI Magazine, 7: 38-53, 82-106

Quint F, Sties M (1995) Map-based semantic modeling for the extraction of objects from aerial images. In: Gruen A, Kuebler O, Agouris P (eds) Automatic extraction of man-made objects from aerial and space images, 307-316. Basel: Birkhäuser.

Ramer U (1972) An iterative procedure for the polygonal approximation of plane curves. CGIP, 1: 244-256

Stilla U, Jurkiewicz K (1991) Objektklassifikation mit einem blackboardorientierten Inferenzmechanismus. Ettlingen: FIM/FGAN, FIM-Bericht Nr. 230

Stilla U (1995) Map-aided structural analysis of aerial images. ISPRS Journal of Photogrammetry and Remote Sensing, 50(4): 3-10

Stilla U, Michaelsen E, Lütjen K (1995a) Structural 3D-analysis of aerial images with a blackboard-based production system. In: Gruen A, Kuebler O, Agouris P (eds) Automatic extraction of man-made objects from aerial and space images, 53-62. Basel: Birkhäuser.

Stilla U, Quint F, Sties M (1995b) Analyse von Luft- und Satellitenbildern zur automatischen Ermittlung der Bodenversiegelung städtischer Siedlungsbereiche: Zwischenbericht II. Ettlingen/Karlsruhe: FIM-FGAN/IPF-Universität