INFORMATIONSVERARBEITUNG
UND MUSTERERKENNUNG
Eisenstockstr. 12, D-76275 Ettlingen

Building Reconstruction Using Different Views and Context Knowledge

U. Stilla, R. Geibel, K. Jurkiewicz

In: E.P. Baltsavias, et al. (Editors)
3D Reconstruction and Modelling of Topographic Objects
International Archives of Photogrammetry and Remote Sensing
Volume 32, Part 3-4W2, p. 129-136

Joint ISPRS Commission III/IV Workshop
September 17 - September 19, 1997
Stuttgart, Germany

Ettlingen - 1997

BUILDING RECONSTRUCTION USING DIFFERENT VIEWS AND CONTEXT KNOWLEDGE

U. Stilla, R. Geibel, K. Jurkiewicz

Research Institute for Information Processing
and Pattern Recognition (FGAN-FIM)
Eisenstockstr. 12, D-76275 Ettlingen, GERMANY
Ph.: +49 7243 99252, Fax: +49 7243 99229
e-mail: usti@gate.fim.fgan.de

KEY WORDS: Production Net, Blackboard System, Multiple Views, Building Reconstruction, Map Analysis, GIS-Verification

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 organization of subconcepts and production rules. We focus on analyzing multiple aerial images and exploiting map information with production nets. As test data aerial images in the scale 1:6300 and a large scale map (1:5000) of urban area are used. A simple example shows how a building is reconstructed spatially and how the verification uses the map information.

1 INTRODUCTION

The automatic interpretation of complex man-made structures in aerial images is a difficult task. In the field of computer vision this task is often treated with a model based approach. Yet a basic problem is that the ideal structures featured by the models do not always appear in the images in all detail. Often object edges are partially occluded or are not completely detectable because of low contrast. For a tolerant recognition only those object parts should be modeled which will be present with high probability or for a detailed model one will only require a subset of the object parts to be present. However this way the discrimination against distortions and other object classes may be lowered and the risk of confusion with them emerges. One possibility to handle this problem is to provide additional information sources and to model the context. In our research project[1] aerial and satellite images are analyzed by two and three dimensional models using context knowledge.

1.1 Context Knowledge

Modeling an object or a class of objects is based on prior knowledge which is achieved from an example (e.g. an image). Besides the knowledge about the appearance of the object in the image it is often useful to consider additional knowledge about the scene and the exposure situation. Such context knowledge can be provided from different sources:

Surroundings. Objects often have a functional relation to their background. For example, many buildings are situated close to a transport route. Cars are mostly found on roads or parking lots. This context may be exploited by a *background analysis.*

Physical Properties. Depending on the sensor, special features of an object are recorded. While images of IR sensors show objects by their different temperatures, images of TV-cameras discriminate objects by brightness or color. An extension to common color photographs is the recording in multiple spectral ranges. Objects hardly showing in one spectral range can very well be visible in another range. To increase the discriminability of objects of one class from another class, different features of the objects are recorded simultaneously. This context may be exploited by a *multi-sensor analysis.*

Temporal Connection. Objects with mass mostly do not abruptly change their position and their velocity. The position of a moving object in a sequence of images defines a trajectory. This context may by exploited by an *image sequence analysis.*

Spatial Properties. The appearance of spatial objects can vary a lot depending on the view angle and the illumination. The relation between different images of a resting object can be given by a spatial description of the object and exposure situation. To extract those spatial descriptions, several object images are taken from different view points. This context may be exploited by a *3D-analysis.*

[1] This project is funded by the Deutsche Forschungsgemeinschaft and carried out in cooperation with the Institut für Photogrammetrie und Fernerkundung, Universität Karlsruhe

Additional Geometric Descriptions. For man-made objects there often exist detailed geometrical descriptions. For example, urban areas are described by maps, buildings by construction drawings and cars by CAD-models. This context may be exploited by a *map-aided analysis*.

In this paper we focus on a 3D-analysis using different views and exploiting information of a map.

1.2 3D - Reconstruction

Monocular Reconstruction. With certain model assumptions height information can be calculated from a single image. For example McGlone & Shufelt [1994] identify horizontal and vertical lines by vanishing point calculation. The vertical and horizontal attributions are used to constrain the set of possible building hypotheses, and vertical lines are extracted at corners, to estimate structure height. In addition to the shape properties, Lin et al. [1995] also exploit the shadow of buildings to determine the height.

Binocular Reconstruction. With two views of a scene and known camera parameters the position of an object point in the 3D-scene can be calculated from two corresponding image points by stereo triangulation. Many authors use simple 2D-structures as edges or corners to determine the 3D-position by a match (e.g. [Haala & Hahn, 1995]). Other authors use more complex 2D-structures to match object parts (e.g. [Stilla & Jurkiewicz, 1996], [Torkar & Bric, 1996]).

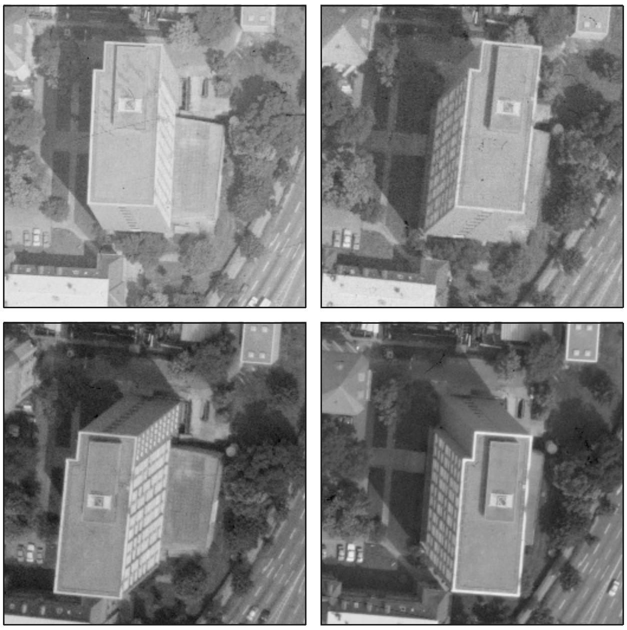

Fig. 1: Different views of a multi-storey building

Multiocular Reconstruction. If aerial images are taken with much end lap and side lap then several views are available for the same section of the scene. Henricsson [1996] uses four images for a building reconstruction. From the most vertical of these images (source image) 2D-contours are extracted and are matched in the other three images.
Also Wiman & Axelsson [1996] use multiple images for the reconstruction of houses. The line elements of the images are projected into the scene and two main directions are extracted. From lines which contribute to these main directions 3D-lines are generated and clustered. From these clustered lines rectangular 3D-planes are constructed.

Some authors make the assumption, that before the reconstruction, a detection of the approximate position has taken place, and corresponding sections of images are selected, each containing a single house (e.g. [Haala & Hahn, 1995], [Henricsson, 1996] [Wiman & Axelsson, 1996]). The detection is done based on additional information sources, e.g. by a digital surface model (DSM) or manually by the user. In contrast to these approaches for a reconstruction we do not presuppose the building position to be known in advance.

1.3 Multiple Views

Several different views of the same scene (Fig. 1) can help to exploit both *nonredundant* as well as *redundant* image contents.

Nonredundant Image Contents. Depending on the angle of view in some images objects may be occluded by other objects. Even in vertical photographs occlusions of smaller buildings by higher ones occur. This problem of occlusion may be reduced if many images with different views are taken into account.
Furthermore, the difference of brightness of differently oriented surfaces can vary with angle of view. Thus edges of a roof, e.g. the ridge, can remain undetected because of a low contrast (see Fig. 2 View1, View4). Using multiple images with different views, also this problem may be reduced if an edge appears at least in two images (see Fig. 2 View2, View3).

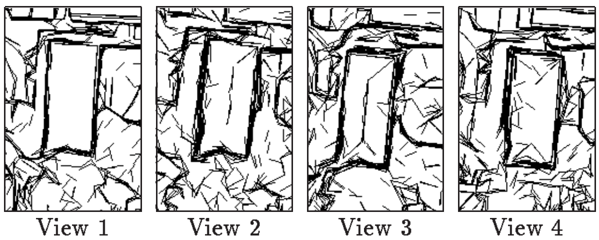

Fig. 2: Edge detection in different views (hip roof)

Redundant Image Contents. Due to image noise and inexactness during processing as well as inaccurate camera parameters, the reconstruction results in deviations from the true scene coordinates. If object points are present in multiple images, relating and summarizing can compensate distortions and improve accuracy.

2 ANALYSIS STRATEGY

For the analysis of objects in images and maps we propose a structural approach. Structure oriented methods build hierarchies by composing less complex object structures into more complex object structures. The approach proceeds stepwise according to a model and produces intermediate results with an increasing degree of abstraction.

2.1 Model Representation

In Pattern Recognition and Computer Vision the term *model* is often used in different contexts and meanings. Referring to different degrees of freedom within the models we distinguish between specific models and generic models.

Specific models describe objects using a fixed topological structure. These models are further differentiated with respect to geometric constraints, as fixed models, fixed shape models and parametric models. *Fixed Models* are ideal geometric representations for physical objects. They are fixed in position and orientation in reference space. *Fixed Shape Models* have a fixed set of geometrical relations but the global position and orientation is variable. *Parametric Models* permit more transformations than *fixed shape* models with the overall structural complexity of the model remaining fixed. The geometrical variation of the model is given by a set of parameters.

Generic models are more general and describe objects without a fixed structure but using only topological concepts. Objects described by the model can consist of an arbitrary number of parts.

2.2 Production Nets

We describe structural relations of the object models by productions. The productions define how a given set of objects is transformed into a set of more complex objects. In the condition part of a production, geometrical, topological and other relations of objects are examined. If the condition part of a production holds, an object specific generation function is executed to generate a new object.
The organization of object concepts and productions can be depicted by a production net ([Stilla et al., 1996]) which displays the *part_of*-relations between object concepts. Concepts represent object types and define a frame for concrete objects (instances) which are described by their attribute values (e.g. position, orientation, assessment, etc.).

We implement production nets in the environment system BPI. The system architecture and analysis process of BPI is described in previous papers ([Lütjen, 1986], [Stilla, 1995]).

Starting with the primitive objects the searched target objects are composed step by step by applying the productions repeatedly. The applied compositions are recorded with pointers between the composed objects and the composing part objects. Since all intermediate results are stored, the analysis process can be traced in any state.

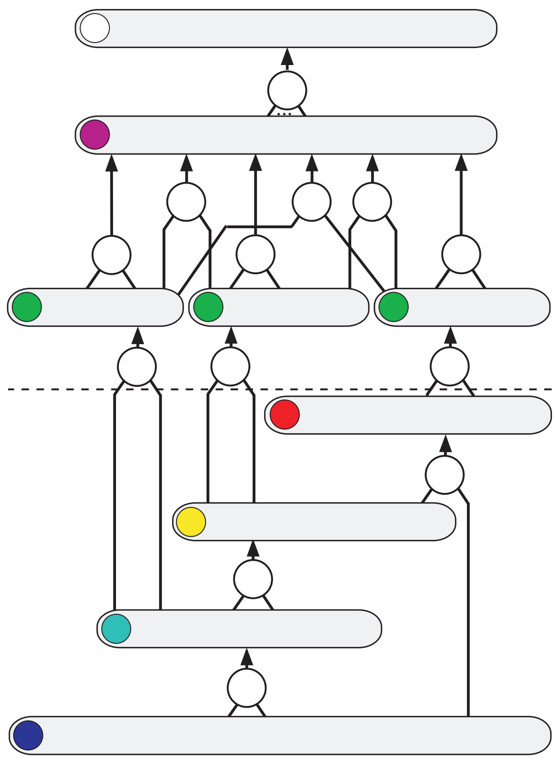

Fig. 3: Production net for the analysis of images (parametric model)

3 IMAGE ANALYSIS

The recognition of buildings from multiple views can be divided into four subtasks: (i) generation of a primitive symbolic image description, (ii) construction of simple 2D-structures, (iii) calculation of the spatial position by stereo triangulation, and (iv) construction of 3D-structures. The structural analysis is carried out based on sets of lines which symbolically describe the image contents. The production net for a simple right-angled building is shown in Fig. 3.

3.1 Preprocessing

In the preprocessing stage the scanned image data are transferred into a symbolic description of the image contents. One procedure to get a symbolic description is the level-slicing method. An image is transferred into a sequence of binary images by multiple thresholding. In each of the resulting binary images the contour lines of the segments are detected and approximated by short straight lines using a dynamic split algorithm. An example with intermediate steps is shown in [Stilla et al., 1996]. Fig. 4 shows the results of the preprocessing applied to the four images of Fig. 1. These short lines are prolongated and stored as primitive objects LINE (see Fig. 6 L).

Fig. 4: Short lines from preprocessing

3.2 2D-Analysis

At first the 2D-analysis is carried out independently in all different images. Starting with the objects LINE, the objects ANGLE, U_STRUCTURE and PARALLELOGRAM can be built up by applying the productions P1-P3 (Fig. 3). Objects ANGLE are constructed of pairs of objects LINE (P1). If two objects ANGLE form a structure like an open parallelogram they are combined into an object U_STRUCTURE (P2). An object PARALLELOGRAM can be assembled if objects U_STRUCTURE and LINE are compatible (P3).

3.3 Stereotriangulation

The 3D-analysis attempts to find in two different images pairs of 2D-objects which are projections of the same 3D-structure. 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 rays result from inverse mapping of projections. The 2D-objects will be called *corresponding* if the distance between the rays of pairs of vertices is smaller than a given threshold. If we use n images instead of only two images in general n(n-1)/2 pairs can correspond. For an object visible in four images we get six calculations for the position in space. Typically these calculations do not result into exactly equal coordinates but they scatter for example because of inaccurate camera parameters. The inverse projection is done by Productions P4, P5 and P6 for the 2D-objects ANGLE, U_STRUCTURE and PARALLELOGRAM. If 2D-objects of two different images correspond then the resulting 3D-objects are generated. Objects ANGLE, U_STRUCTURE and PARALLELOGRAM are combined into objects 3D_ANGLE, 3D_U_STRUCTURE and 3D_RECTANGLE respectivly.
Unlike earlier implementations [Stilla & Jurkiewicz, 1996], here the objects 3D_ANGLE were also constructed. This step provides greater flexibility when building up complex structures despite distortions in some images.

3.4 3D-Analysis

If two objects 3D_RECTANGLE are oriented in such a way that the surface normals enclose a right angle and if they are located in a way that the vertices are neighbouring, then a target object CUBOID_EDGE is generated (P12). Instead of combining two objects 3D_RECTANGLE, an object 3D_RECTANGLE can also be combined with an object 3D_U_STRUCTURE (P11) or an object 3D_ANGLE (P10). Further combinations are two objects 3D_U_STRUCTURE (P9), an object 3D_U_STRUCTURE and an object 3D_ANGLE (P8) or two objects 3D_ANGLE (P7). Some possible combinations of two objects 3D_ANGLE are given in Fig. 5.

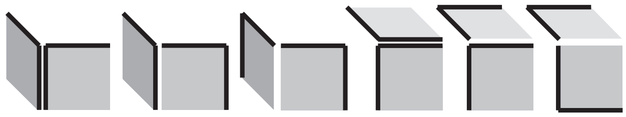

Fig. 5: Combinations of objects 3D_ANGLE (P7)

In the composition of two objects 3D_ANGLE, 3D_U_STRUCTURE or 3D_RECTANGLE into an object CUBOID_EDGE the positions of the vertices will usually not match exactly. This divergence from perfect fi called *composition gap* is stored as one of the attributes of the generated object CUBOID_EDGE.
Complex objects (3D_RECTANGLE) get a better basic assessment because their probability of correct positioning is higher than that of simple ones (3D_ANGLE).

In 3D-space there are certain places where the objects CUBOID_EDGE cumulate. Production P13 examines these clusters using the attributes position gap and assessment and generates a representative object BUILDING.

Fig. 6 shows the intermediate results of the image analysis in the four different views. The bottom row displays all primitive objects LINE (L). Going up there follow the objects ANGLE (A), U_STRUCTURE (U) and PARALLELOGRAM (P).

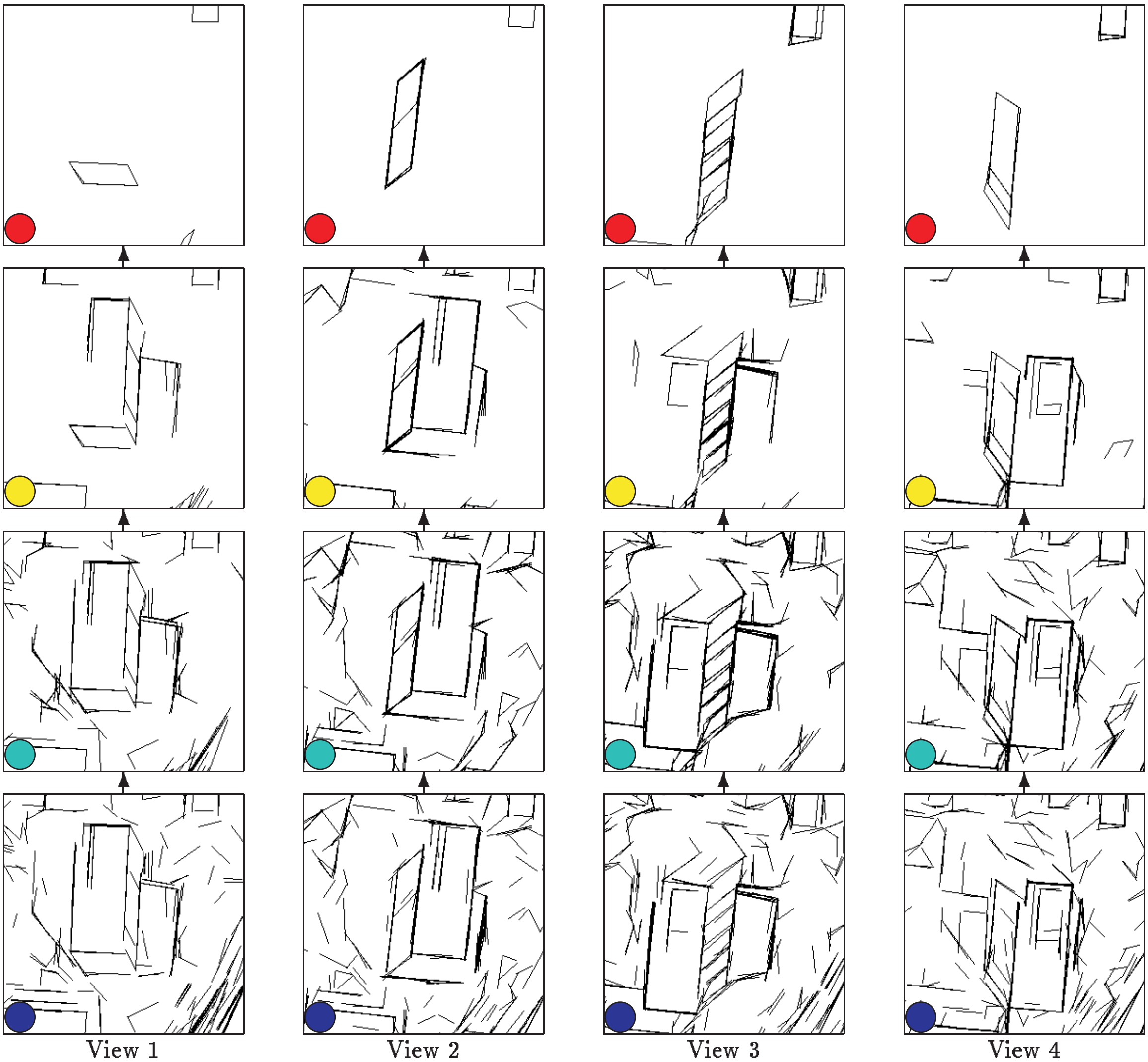

Fig. 6: Intermediate results of different views displayed by the sets of objects LINE (L), ANGLE (A), U_STRUCTURE (U) and PARALLELOGRAM (P)

Fig. 7a displays a cluster of 1889 objects CUBOID_EDGE and 7b their representative. Fig. 7c displays a differently oriented cluster of 451 objects CUBOID_EDGE with the representative in 7d. Fig. 7e shows the relative position of both representatives.

4 MAP ANALYSIS

Within the field of knowledge acquisition for Geographic Information Systems (GIS) there are different tasks for image analysis to deal with. In some cases we can assume that the GIS already contains a scene description given by a map.

One possible task of image analysis is the *extension* of the map by extracting additional descriptions or interpretations. Examining the building heights, roof shapes or determining the usage of terrain are some examples. In this case we assume the map to be accurate. The map information can be used as prior knowledge for image analysis (e.g. restricting search).

Another task of image analysis is *change detection* for updating the map. In this case we presume the map not to be up to date and attempt to find changes by image analysis. A direct comparison of map and image is often not possible. This becomes especially evident when the map and the images have different scales.

For both tasks described above the map information is first transformed by a map analysis. The aim of this map analysis is to produce a hierarchical description of the map contents adequate for the actual task. For this purpose we use both generic and parametric models.

The footprints of buildings in urban and suburban areas in European towns show a great variety of geometric forms. Generic models only take topological relations into account. Parametric models cover only

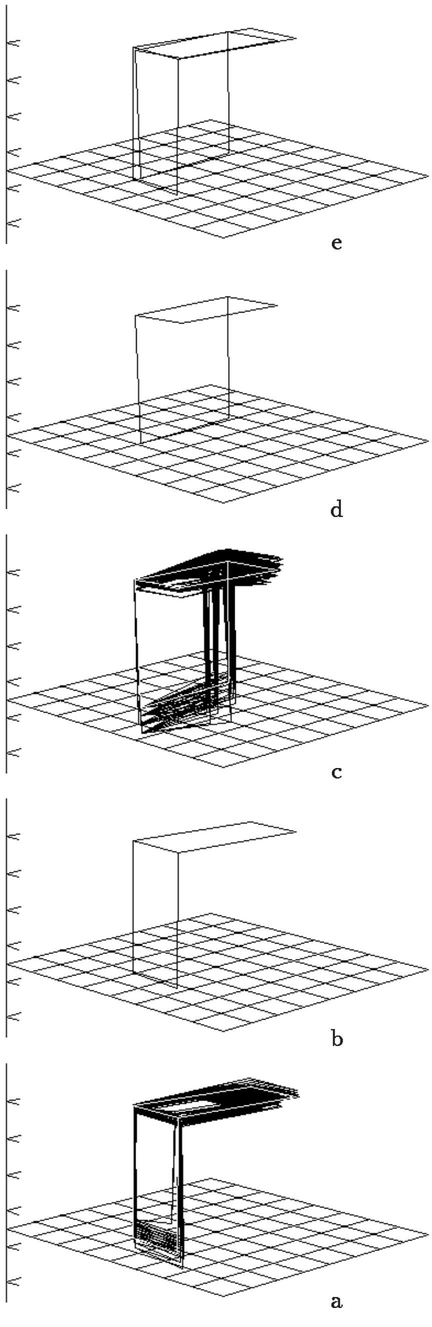

Fig. 7: Results of 3D-analysis

certain parameter defined forms. Thus we first perform an analysis with generic models. In order to gain additional knowledge for scene understanding, the objects which have been separated and whose topological structure has been recognized are then subjected to a geometrical analysis with parametric models. Even if for an area the parametric model search ends up void we gain additional information: we then know that those objects present in this area differ in structure from all the parametric models actually available.

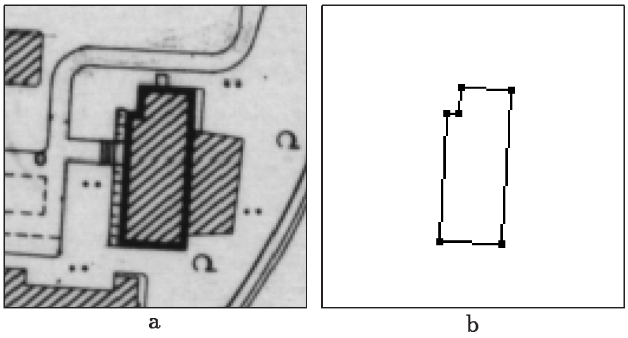

Fig. 8: a) scanned map, b) digitized map (layer: high building)

4.1 Generic model

We use a large scale (1:5000) vector map which is organized in several layers each of which contains a different class of objects (e.g. streets, buildings, etc.) Some of these layers contain buildings with their parts. The task of map analysis is to separate building parts, to determine enclosed areas, and to group building parts. Fig. 9 shows a section of a production net of a generic model for the analysis of complex buildings examining the topological properties *connectivity*, *closedness*, and *containment*. It is part of a more general production net described in [Stilla & Michaelsen, 1997] required for more complex topological structures.

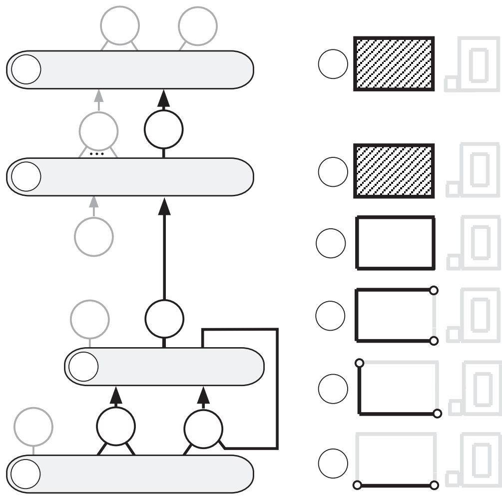

Fig. 9: Section of a production net for the analysis of a map (generic model)

Beginning with objects LINE, objects POLY are composed if two lines have a common endpoint which is not a branch point (P1). Such polygons can be prolonged by production P2 with objects LINE. If polygons can be closed without branch points, production P5 directly creates an object CONTOUR. Production P14 generates objects BUILDING from building contours, wich neither lie inside other contours nor contain other contours.

4.2 Parametric model

Many building footprints have parts which can be described by right angles. A production net for the recognition of two simple building footprints is depicted in Fig. 10. Beginning with objects LINE at first objects CORNER are composed (P1) if they make a right angle. If two objects CORNER can be composed such that they make a rectangle then an object RECTANGLE is generated by production P2. If two objects CORNER can be composed into a right-angled convex form then an open rectangle (object RECTANGLE_O) is generated. Such an object RECTANGLE_O can be combined with another CORNER into an object BUILDING_CO by production P4.

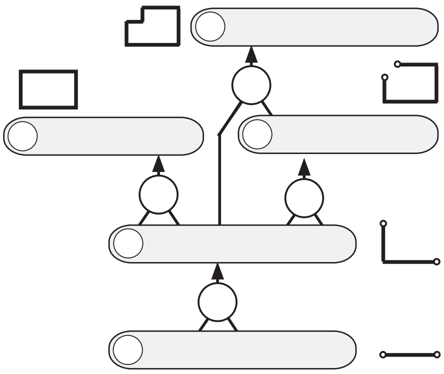

Fig. 10: Production net for the analysis of a map (parametric model)

A section of a scanned map shows a multi-storey building (Fig. 8a). The corresponding vector map symbolically describes the building in a map layer by a set of lines (Fig. 8b). These lines are analyzed applying the productions of the net in Fig. 9. The fact that the lines form a simply connected closed polygon with no inner yards and not contained in another polygon tells us that the object beeing recognized is of topologically simple form. By the analysis following the productions of the net in Fig. 10 we then learn that this building has the form of a rectangle with one open corner.

5 VERIFICATION

When updating maps for areas already mapped cartographically we need not construct all the buildings and traffic ways from scratch out of images. Instead we can initally go back to existing maps and only need to inspect the objects already mapped cartographically.

Since the position and horizontal extension of the investigated objects are known we can restrict our study in the first part of the work to those ranges of the 3D scene and the corresponding 2D sections. Hence the map provides certain expectations and we have to test if the expected objects are present.

In the case of confirmation of the entire object the hierarchical composedness of the parts can then be inspected down to the position of the involved lines.

Beginning with global attributes on high abstraction levels (e.g. center of gravity, area) contour elements on the low abstraction levels are compared. To relate the contour elements the ground plan of the 3D-objects is used.

The distance measure used here is the normalized Euclidean distance on the vectors made of the endpoints of the best fitting subsets of the contour elements. An object resulting from the map analysis is said to be verified by the image analysis if this distance remains small enough. The quality of the verification result is given by the distance value.

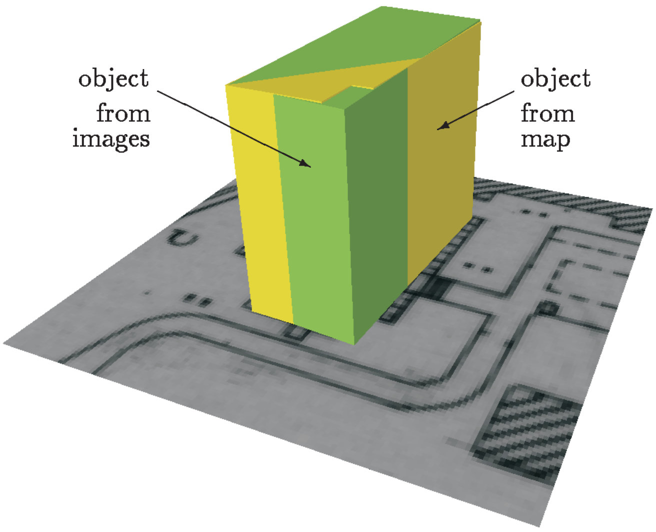

Fig. 11: Comparison of map and image result

In the case of nonconfirmation it is examined if there are other objects at the considered place or hints for such objects.

The example in Fig. 11 shows the result of the image analysis (dark cuboid-like body) and the building resulting from the map analysis (bright body) in comparison. The two visible walls and the top area of both bodies intersect. The distance measure is 1.93 m.

The object Building from the map analysis and the result of the image analysis intersect. The big distance of nearly 2 m is partially caused by the deviation of the building from the model in Fig. 3 with rectangular form.

The heights calculated by the image analysis ranges from 35.1 m to 35.9 m, the real height of the building being 34.86 m.

6 CONCLUSION

This contribution demonstrates that the structures of the image and the map can be represented using production nets. The evaluation of multiple views reduces the sensitivity of the 3D-reconstruction to distortions. For this advantage we have a higher computational effort on the other hand.

With the presented procedure the effort required for the generation of the 2D-objects increases linearly with the number of views used. The effort for the stereo triangulations needed increases quadratically with the number of images used if the examined objects appear in all images. Thus using four images instead of two, the effort maximally increases by a factor of six.

7 REFERENCES

Haala N, Hahn M (1995) Data fusion for thre detection and reconstruction of buildings. In: Gruen A, Kuebler O, Agouris P (eds) Automatic extraction of man-made objects from aerial and space images, 211-220. Basel: Birkhäuser.

Henricsson O (1996) Analysis of image structures using color attributes and similarity relations. Zürich: ETH, Institute of Geodesy and Photogrammetry, Diss.

Lin C, Huertas A, Nevatia R (1995) Detection of buildings from monocular images. In: Gruen A, Kuebler O, Agouris P (eds) Automatic extraction of man-made objects from aerial and space images, 125-134. 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

McGlone JC, Shufelt JA (1994) Projective and object space geometry for monocular building extraction. IEEE Comp. Soc. Conf. on CVPR, 54-61

Quint F (1996) Reconstruction of structured objects in monocular aerial images using context information. In: Leberl F, Kalliany R, Gruber M (eds) Mapping buildings, roads and other man-made structures from images. Wien: Oldenburg, 213-228

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 (1995) 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. Basel: Birkhäuser, 53-62

Stilla U, Jurkiewicz K (1996) Structural 3D-analysis of urban scenes from aerial images. ISPRS, XVIIIth Congress, International archives of photogrammetry and remote sensing, Vol. 31, Part B3, 832-838

Stilla U, Michaelsen E, Lütjen K (1996) Automatic extraction of buildings from aerial images. In: Leberl F, Kalliany R, Gruber M (eds) Mapping buildings, roads and other man-made structures from images. Wien: Oldenburg, 229-244

Stilla U, Michaelsen E (1997) Semantic modelling of man-made objects by production nets. In: Gruen A, Baltsavias EP, Henricsson O (eds) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser, 43-52

Torkar D, Bric V (1996) On automatic restitution of building from aerial stereo images. In: Leberl F, Kalliany R, Gruber M (eds) Mapping buildings, roads and other man-made structures from images. Wien: Oldenburg, 245-257

Willuhn W (1997) Integrating models for object recognition. Technical Report 175, CV-Group, Communication Technology Lab, ETH Zürich

Wiman H, Axelsson P (1996) Finding 3D-structures in multiple aerial images using lines and regions. ISPRS, XVIIIth Congress, International archives of photogrammetry and remote sensing, Vol. 31, Part B3, 953-959