FORSCHUNGSINSTITUT FÜR
INFORMATIONSVERARBEITUNG
UND MUSTERERKENNUNG
Eisenstockstr. 12, D-76275 Ettlingen

Structural Analysis of Right-angled Building Contours

U. Stilla, E. Michaelsen, K. Jurkiewicz

In: T. Schenk and A. Habib (Editors)
Object Recognition and Scene Classification from Multispectral and Multisensor Pixels
International Archives of Photogrammetry and Remote Sensing
Volume 32, Part 3/1, p. 379-386

ISPRS Commission III
July 6 - July 10, 1998
Columbus, Ohio

Ettlingen - 1998

STRUCTURAL ANALYSIS OF RIGHT-ANGLED BUILDING CONTOURS

U. Stilla, E. Michaelsen, K. Jurkiewicz

Research Institute for Information Processing
and Pattern Recognition (FGAN-FIM)
Eisenstockstr. 12, D-76275 Ettlingen, GERMANY
Ph.: +49 7243 99235, Fax: +49 7243 99229
e-mail: usti@gate.fim.fgan.de

KEY WORDS: Production Net, Blackboard System, Building Recognition, Perceptual Grouping, Map Analysis, Generalization

ABSTRACT:

We present a model-based method for the automatic analysis of structures in aerial images or maps. 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 sub-concepts and production rules. This paper focuses on analyzing right-angled building contours in images and maps. As test data aerial images of vertical and oblique views and a large scale vector map (1:5000) of urban area are used. The modeling, the analysis of different image sources and the generalization of right-angled buildings is demonstrated by the example of image and map data taken from the same urban area of Karlsruhe.

1 INTRODUCTION

Automatic interpretation of complex structures e.g. in aerial images is a difficult task. In order to recognize man-made objects such as buildings, frequently model-based approaches are pursued [Gruen et al., 1997]. Some systems introduce paradigms and methods of the field of artificial intelligence [Ade, 1997]. Methods, which try to "understand" an image typically produce a symbolic description of the image contents. This is done by using structural models, where symbols or object concepts are defined semantically and the relations between these objects are declared.

The description of characteristic buildings or roof shapes by object concepts is difficult, because in urban areas a large variety of buildings exist. The number of object concepts and the complexity of the corresponding formal grammar may be reduced using parameters. If the entire information about the object structure is integrated only in the parameters, the object concepts lose their meaning. For example the object description POLYGON (generic model) provides a less specific semantic contribution to the shape description than the object description RECTANGLE (parametric model).

The contribution at hand describes the capture of certain shapes of buildings by parametric models and the method to recognize these buildings displayed in images and maps.

In correspondence with Marr [1982], vision is understood as "active construction of a symbolic description of images". During the analysis structure hierarchies are built up stepwise by grouping complex structures from less complex structures. This process results in an increasing abstraction level.

This approach has been used with a variety of models to analyze pictorial structures before. We have shown that in 2D analysis complex settlement structures can be composed by simpler structures such as streets, crossings, and buildings [Füger et al., 1992]. Using the example of crossings, it was demonstrated how the analysis can be controlled by additional scene knowledge [Stilla, 1995]. This context knowledge was derived from maps. The model of a gable roof for the detection of buildings was extended to allow a 3D-reconstruction from a stereo pair [Stilla & Jurkiewicz, 1996]. In further work on 3D analysis multiple views were exploited [Stilla et al., 1997] and suburban 3D structures were modeled [Stilla & Michaelsen, 1997]. In that model we assumed a regular structure given by houses of similar size and orientation, equidistantly spaced in rows parallel to a street.

This paper focuses on the symbolic description and analysis of right-angled building contours in monocular views. Instead of using models to recognize roofs by rectangles in the scene or parallelograms in the image we now consider more complex building contours. Models describing such structures seem general enough to suit a larger portion of buildings. On the other hand right angles together with geometric constraints of "pointing at each other" and the existence of supporting straight contours in-between are a strong evidence. Thus, the rise of combinatorial growth inherent in the grouping process seems controllable.

In contrast to features like brightness, color, texture, etc. the mentioned geometric constraint is evident in a wide variety of information sources including maps and images of visual spectral domain or other spectral domains like IR or SAR. Fig. 1 shows some examples of different images of the same building. Depending on the task, images are taken by vertical view (Fig. 1a-d) or oblique view (Fig. 1e-g). Not in all applications metric aerial cameras (Fig. 1b-d) are used (e.g. Fig. 1e,f). Images from other sensors provide completely different physical properties as temperature or height of the objects (e.g. Fig. 1g-h).

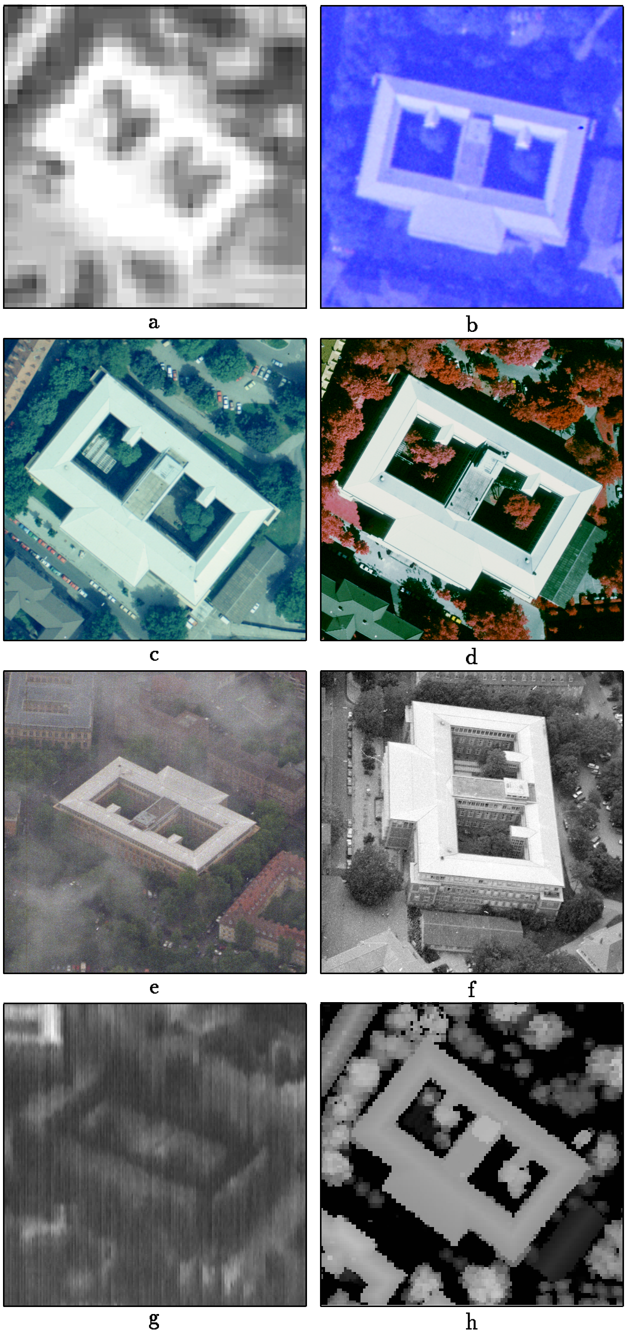

Fig. 1: Different images. a) satellite image KVR-1000 b) aerial image, color, 1:40000, c) aerial image, color, 1:6300, d) CIR-image 1:4000, e) medium format 6x6 cm, f) 35 mm film, g) IR image, h) laser height data [IPG]

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 context 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 geometric variation of the model is given by a set of parameters.

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

2.2 Productions

We describe structural relations of the object models by productions. A production defines how a given configuration of objects is transformed into a single more complex object (or a configuration 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. Such productions operate on sets of objects instead of graphs, strings etc. and define precise and modular semantics [Michaelsen & Stilla, 1998].

2.3 Production Nets

The hierarchical organization of object concepts and productions can be depicted by a production net which – comparable to semantic networks – displays the *part-of* hierarchies of 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.).

2.4 BPI - System

Production nets are preferably implemented in a blackboard architecture in the environment system BPI ([Lütjen, 1986], [Stilla, 1995]). The BPI-system consists of a global data base (blackboard), a set of processing modules (knowledge sources), and a control unit (selection module). The productions are implemented in the processing modules, which test the relations between objects and generate new objects.

Starting with primitive objects the searched target objects are composed step by step by applying the productions. The system works in an accumulating way, this means a replaced initial configuration will not be deleted in the database. Thus all generated partial results remain available during the analysis to pursue different hypotheses. The classical backtracking in search-trees is not necessary [Michaelsen, 1998]. The applied compositions of objects (instances) are recorded by pointers and can be traced back and displayed by a derivation graph.

3 IMAGE ANALYSIS

3.1 Production Net

For the recognition of specified building primitives we use parametric models. Fig. 2 shows an example production net for two basic shapes, called B_UCO_VEX and B_UCO_CAV.

Starting with the objects CORNER two different object types can be generated by (P1,P2): CORNER_U and CORNER_Z. Objects CORNER_U result from an input configuration of two objects CORNER, pointing at each other and having the same sense of rotation (++). Objects CORNER_Z are constructed from two objects CORNER, pointing at each other and having opposite sense of rotation (+-). From both objects hypotheses are derived for edges of buildings.

When these hypotheses are confirmed by objects LINE which lie in between the vertices of objects CORNER the objects EDGE_U or EDGE_Z arise from productions P3 or P4. Objects EDGE_U and EDGE_Z, which have one common endpoint can be combined in three different ways (+++,++-,+-+). Two objects EDGE_U form an object ANGLE_C (P5), two objects EDGE_Z form an object ANGLE_O (P7) and the combination of objects EDGE_U and EDGE_Z forms an Object ANGLE_S (P6). Objects ANGLE_C or ANGLE_S can be combined with objects EDGE_U or EDGE_Z to construct u-structures with ends pointing towards each other

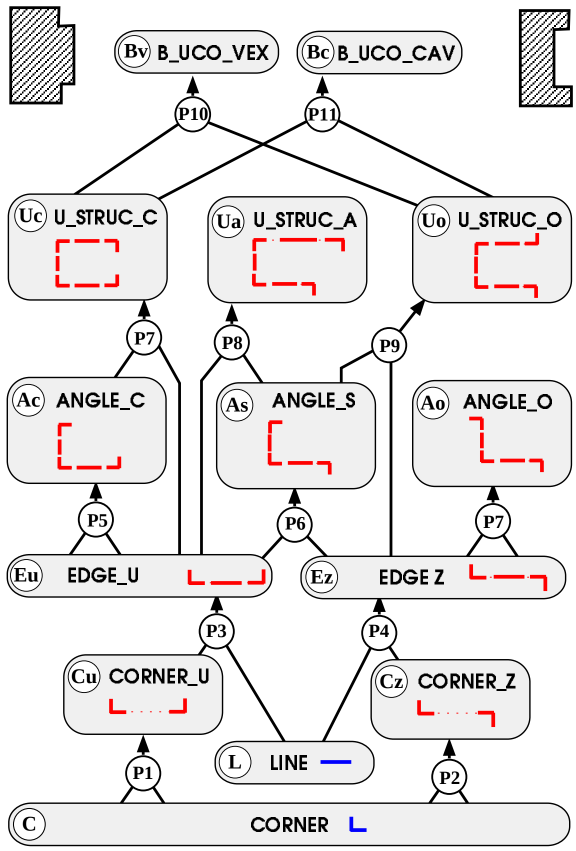

Fig. 2: Production net B_UCO_Cav and B_UCO_Vex (parametric models)

(U_STRUC_C)(closing; ++++), pointing in the same direction (U_STRUC_A)(angled; +++-), pointing in opposite direction (U_STRUC_O)(opening; +--+).

Furthermore we combine objects U_STRUC_C and U_STRUC_O in two different ways (P10, P11) to close the contour. If the area of an object U_STRUC_C is increased by an object U_STRUC_O an object B_UCO_VEX is generated. If the area of object U_STRUC_C is decreased an object B_UCO_CAV is generated.

Further structures can be composed from this basic set of U-structures, angles, and edges. The composition of right-angled shapes with four and six vertices are shown in Fig. 3. As an example for the variability of buildings that can be captured by a simple parametric model we present the concept B_UAA. Some of the possible realizations of such a concept are shown with varying parameter ratios c/a and d/b. L-shaped buildings ($c/a < 0.5$; $d/b < 0.5$) are covered as well as rectangular buildings with additional or missing areas.

3.2 Feature Extraction

Some feature extraction procedure is necessary to transfer the image data into a symbolic description suitable for the production net. One way to do

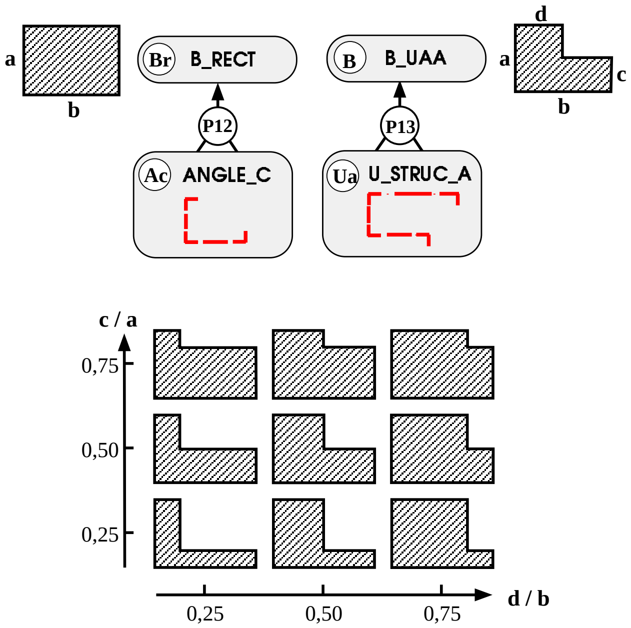

Fig. 3: Right-angled shapes (four and six vertices) and variation of form by two parameters

this 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 of preprocessing is shown in [Stilla et al., 1996]. Short lines which can be grouped by a prolongation are replaced by primitive objects LINE (Fig. 4L). Short lines which can be grouped to corners enclosing a right angle are replaced by primitive objects CORNER (Fig. 4C).

3.3 Monocular vertical views

The production net presented in Fig. 2 was applied on data extracted from the aerial image section shown in Fig. 1c. Partial objects generated during this run are displayed in Fig. 4.

Starting with the primitive objects CORNER (Fig. 4C) and LINE (Fig. 4L), we attempt to construct basic shapes by generating right-angled partial objects. Fig. 4Cu and Fig. 4Cz show the objects CORNER_U and CORNER_Z which provide hypotheses for objects EDGE_U and EDGE_Z. Considering the objects LINE (Fig. 4L) the actual number of generated objects EDGE_U and EDGE_Z (Fig. 4Eu, Ez) is greatly reduced. Further reduction results from the application of productions P5-P7 (see Fig. 4Ac,As) and the productions P7-P9 (see Fig. 4Uc,Uo). On the target level (Fig. 4Bv, Bc) only the building contours remain.

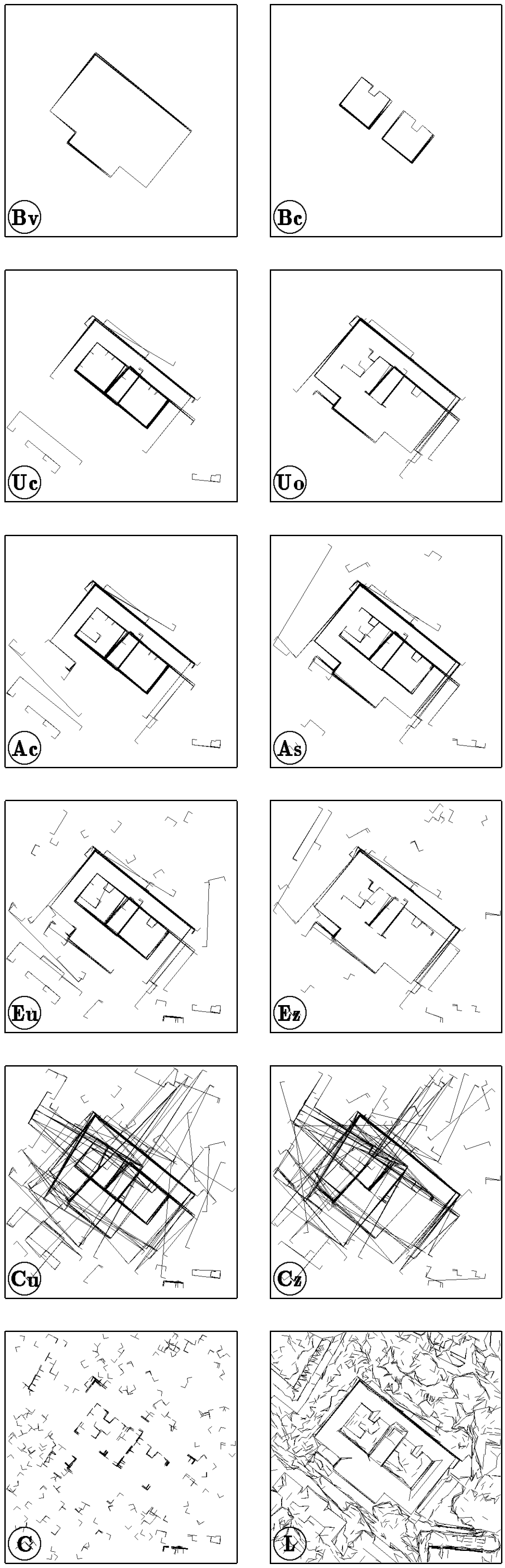

Fig. 4: Intermediate results (vertical view)

3.4 Monocular oblique views

In contrast to images of vertical views, in which horizontal right-angled building contours can be searched directly in detected contour lines of the image, in oblique views the perspective distortion of angles has to be considered. Knowing the horizon, the principal point M and the camera constant c_k these special object structures can be calculated assuming the following simplified conditions (see Fig. 5).

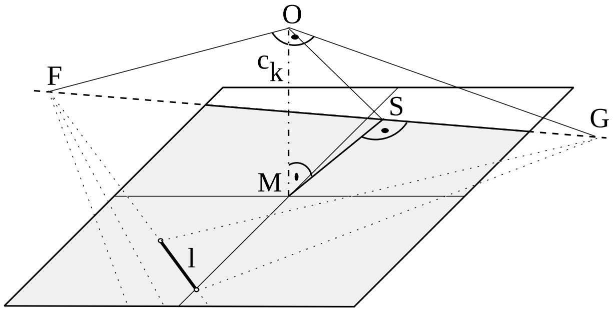

Fig. 5: Horizontal right-angles in oblique view

Suppose the horizon, which does not have to be visible in the image, is given by the line equation $ax + by = c$. The equation can be determined from roll angle and elevation angle. The perpendicular projection S of the principal point M on the horizon is the intersection of the straight line $bx - ay = bM_x - aM_y$ and the horizon. S is the vanishing point of the projection of the camera axis on the horizontal plane in the scene.

Assuming that a straight line l in the image is horizontally oriented in the scene the prolongation of l intersects the horizon in the vanishing point F. This vanishing point is common to all lines in the image running parallel in the scene to line l. The distance $\overline{FS}$ is a measure for the angle between the orientation of l and the projection $\overline{MS}$ of the camera axis on the horizontal plane.

To generate primitive objects CORNER horizontal object contours are searched enclosing a right angle in the scene. Horizontal lines in the scene which are perpendicular to line l run to vanishing point G on the horizon in the image. The right-angled triangles FSO and OSG are similar and thus $\overline{SG} = \overline{SO}^2 / \overline{FS}$ can be calculated. The distance $\overline{SO}$ arises from the right-angled triangle OMS with $\overline{SO}^2 = \overline{MO}^2 + \overline{MS}^2$.

If line l is oriented to S, i.e. $\overline{FS} = 0$, the intersection G can not be calculated. In this case the right-angled lines are oriented parallel to the horizon. If line l is oriented parallel to the horizon F does not exist. In this case the right-angled lines run to S.

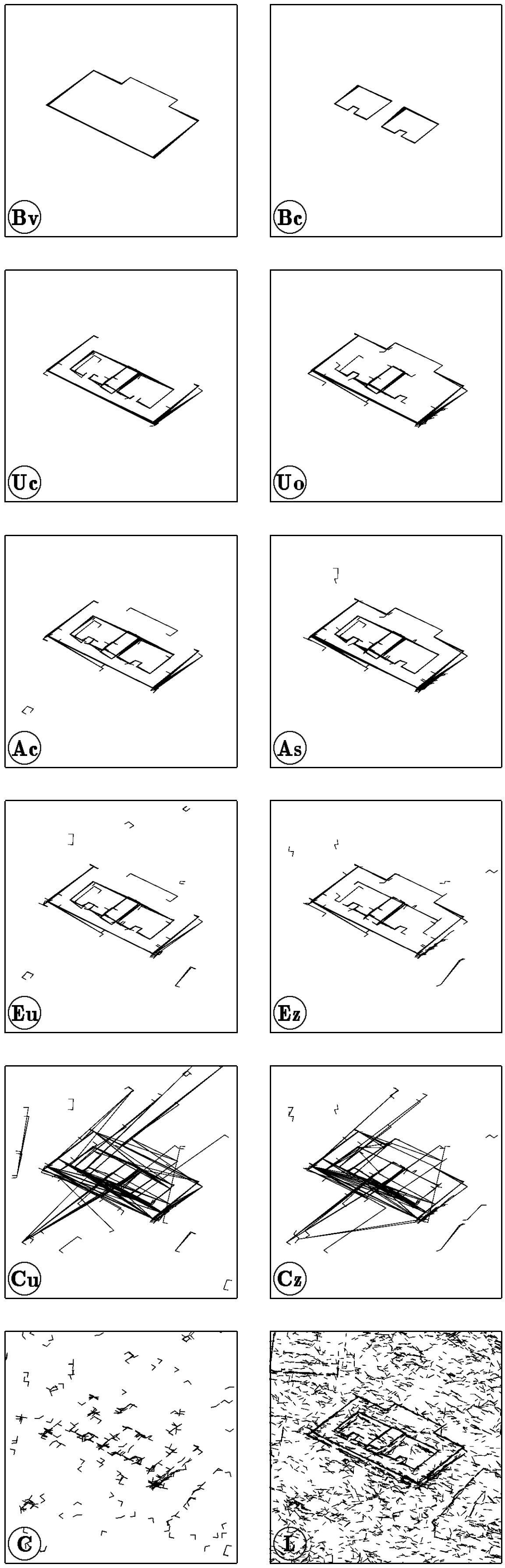

Fig. 6: Intermediate results (oblique view)

This geometric construction uses a pinhole camera model. Thus it is not suitable for distorting wide angle lenses. Near the horizon the calculations tend to be instable and objects located there are neglected. Based on this calculation objects CORNER are generated in the preprocessing stage.

The same production net depicted in Fig. 2 can be used to analyze this set of primitive instances. Exemplarily this has been performed on the section of medium format aerial image with oblique orientation depicted in Fig. 1e. The results are shown in Fig. 6 in the same way as in Fig. 4.

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. Generic models take topological relations into account. Parametric models cover certain parameter defined forms.

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.) First, the topological properties *connectivity*, *closedness*, and *containment* of map-lines (Fig. 7a) are tested by a production net of a generic model. This production net was presented in [Stilla & Michaelsen, 1997]. The aim of the analysis is to separate parts of buildings, to determine encapsulated areas and to group parts of buildings. The output of the analysis is a hierarchical description of the buildings or building complexes. Fig. 7b shows a description on the levels BUILDING(B), CONTOUR(C), and LINE (L).

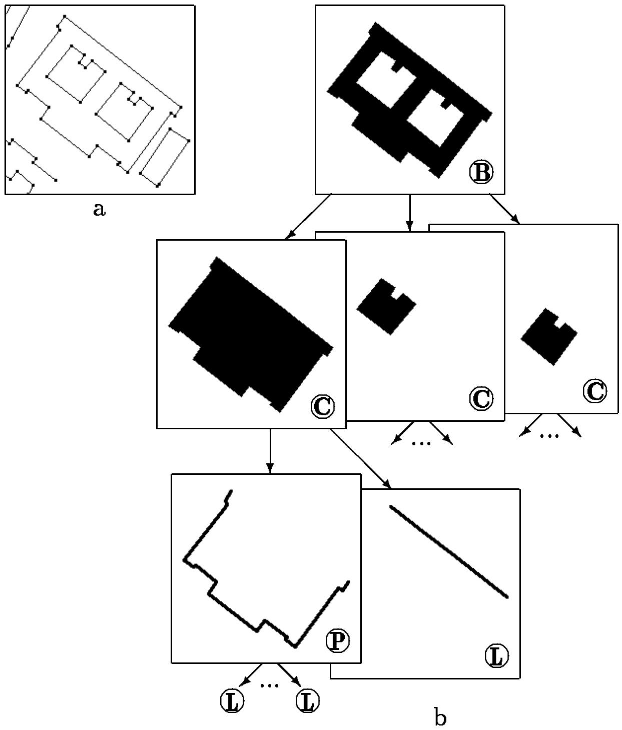

Fig. 7: a) Input data and b) resulting hierarchical object description (generic model) [Stilla & Michaelsen, 1997]

4.2 Parametric model

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. Many building footprints or parts of them have structures which can be described by right angles. For this purpose parametric models of some right-angled basic shapes of building contours are captured in a production net. Buildings fitting into one of the parametric models are associated with the corresponding symbolic concept. Buildings which cannot be assigned, belong to the class 'no right-angled basic shape'.

Using the production net of Fig. 2 to analyze the contours of Fig. 7 both inner polygons of the building can be described by the concept B_UCO_CAV. The outer polygon has 16 Points and can not be described by the concept B_UCO_VEX (8 Points) which is used in Fig. 4Bc and Fig. 6Bc.

5 COMBINING DIFFERENT INFORMATION SOURCES

For some applications such as the multi-sensor analysis or analysis with a map and an image, it is necessary to compare and combine the contour descriptions of objects which where detected in different sources on the symbolic level. Different geometric resolutions of the image or map data however can result in different symbolic descriptions. In order to make a comparison on the symbolic level possible, the structures are generalized following geometric aspects. After that the object descriptions are available for the comparison on different generalization levels. The comparison has to take different grades of generalization into account.

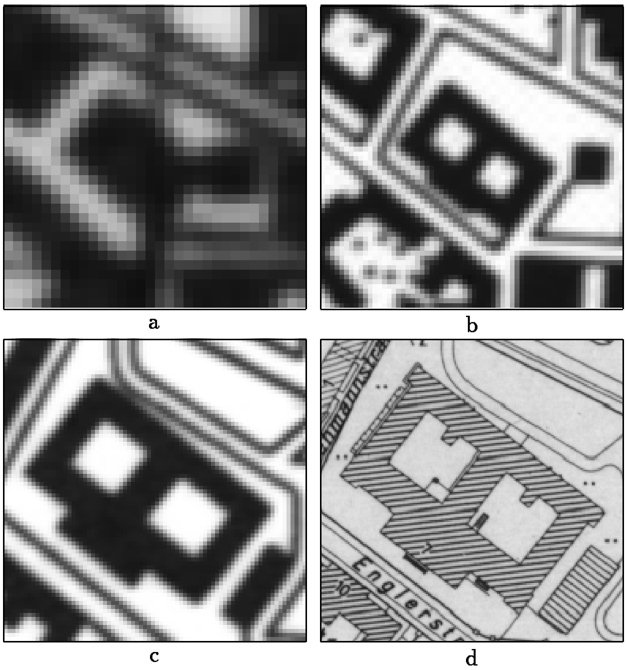

Fig. 8: Topographic maps of different scale.
a) TK 100 (1:100.000), b) TK 50 (1:50.000),
c) TK 25 (1:25.000), d) DGK 5 (1:5000)

5.1 Cartographic Generalization

Producing a small scale map from a large scale map by generalization different partial processes can be distinguished [Hake, 1982]: (i) simplification, (ii) enlargement, (iii) displacement as result of enlargement, (iv) combination, (v) selection and elimination, (vi) classification and typification, (vii) evaluation and emphasizing.

In manual generalization, both the simplification of the shape, form, and map features and the graphic legibility are considered simultaneously. When considering "characteristics and importance" of map entities the process contains a subjective or purpose-dependent component. For example, when producing a city map, the characteristic outline of important buildings or monuments must not be disfigured or even get lost.

In the past three decades several approaches have been developed to automate the generalization of maps. According to the data type being used, procedures of generalization of raster data and vector data are distinguished. Different object types require different generalization methods. Thus methods used for natural objects (e.g. rivers) are, in general, not suitable for artificial objects (e.g. buildings)

An approach for generalizing topographic maps based on vector data was presented by Staufenbiel [1973]. His work focuses on the transformation of buildings from scale dimension 1:5000 into scale dimension 1:25000. These algorithms were developed by the Cartographic Institute of the University of Hannover and are today available as a modular system (CHANGE) in a commercial product (PHOCUS).

In the domain of GIS there is an increasing demand in generalization tools for the derivation of multipurpose databases and map products from a detailed master database. Some of the typical problems in map production are conflict detection and resolution. This conflicts arise from a combination of design factors such as map symbology, output resolution, and proximity of objects. The report of the ICA Workshop on Map Generalization [Mackaness et al., 1997] gives an overview over the current problems and research topics.

5.2 Simplification Of Structures

Common procedures can be found in partial processes of generalization of vector maps and in the kind of structural methods of pattern recognition that we model by production nets. Both procedures reduce complexity of structures by combining structures and replacing them by simpler structures.

In contrast to a cartographic generalization we are only interested in geometric aspects of generalization, i.e. legibility is of no interest. Simple productions may replace a complex right-angled polygon by a less complex right-angled polygon.

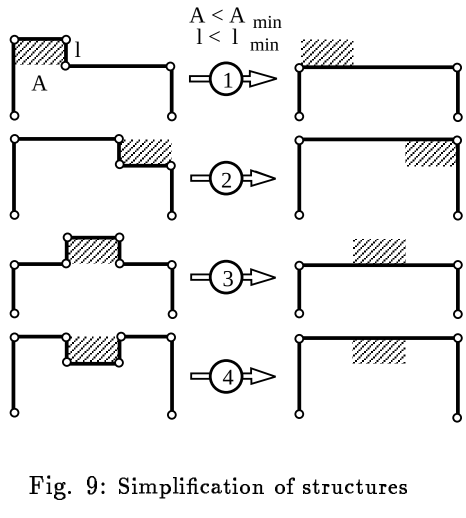

Fig. 9: Simplification of structures

Fig. 9 shows some examples of such productions reducing polygons with six or eight points into polygons with four points. A replacement is carried out, if the hatched area A is smaller than A_{min} and the length l of the short line is smaller than l_{min}. Parameters A_{min} and l_{min} depend on the grade of generalization.

Productions like this can be performed using different grades of generalization resulting in a sequence of symbolic descriptions for different scales. Fig. 10a-d gives an example for three steps of generalization which reduce the number of polygon points from 32 to 12.

The structure shown in Fig. 10b (24 points) consist of an outer polygon B_UCO_VEX and two inner polygons B_UCO_CAV. This exactly matches the description (Fig. 10e) obtained from the vertical image (Fig. 4Bc,Bv) or oblique image (Fig. 6Bc,Bv) described before.

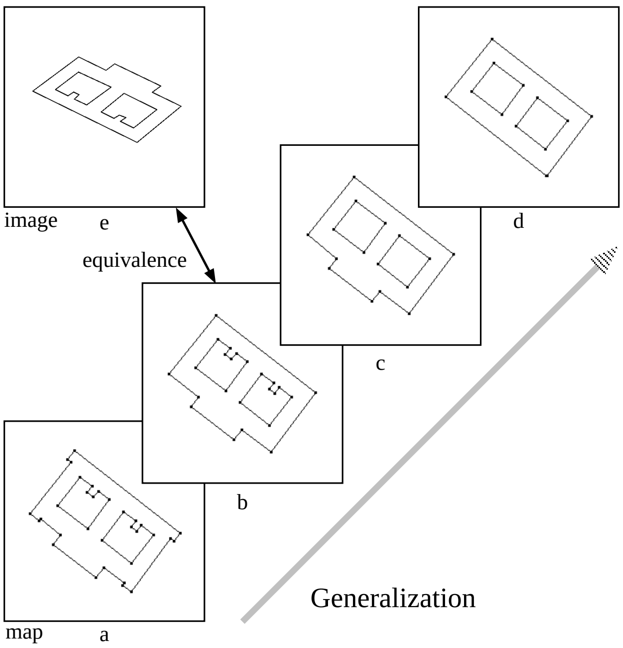

Fig. 10: Generalization of the map in different scale dimensions and comparison with the image

6 CONCLUSION

This contribution demonstrates that the structures of different images and maps can be represented using production nets. Productions and production nets have been presented as a suitable tool for several image and map analyzes tasks. We demonstrated this by giving an example net which handles complex right angle buildings of a specific type and running it on different information sources describing the same scene. Problems with different view angles can be handled using some simple geometric method. The symbolic description may differ depending on the scale used. We propose to handle this by modeling the process of generalization with productions.

7 REFERENCES

Ade F (1997) The role of artificial intelligence in the reconstruction of man-made objects from aerial images. In: Gruen A, Baltsavias EP, Henricsson O (eds) 23-32

Füger H, Jurkiewicz K, Lütjen K, Stilla U (1992) Ein wissensbasiertes System für die automatische Bildanalyse. ISPRS, XVIIth Congress, International archives of photogrammetry and remote sensing, Vol 29, Commission III, 167-172

Gruen A, Baltsavias EP, Henricsson O (1997) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser

Hake G (1982) Kartographie I. Berlin: de Gruyter

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

Mackaness WA, Weibel R, Buttenfield BP (1997) Report of the 1997 ICA Workshop on Map Generalization. Gävle, Sweden. http://www.geo.unizh.ch/ICA/

Marr D (1982) Vision. San Francisco: Freeman

Michaelsen E (1998) Über Koordinaten Grammatiken zur Bildverarbeitung und Szenenanalyse. Thesis

Michaelsen E, Stilla U (1998) Remarks on the notation of coordinate grammars. In: Amin A et al. (eds) Advances in Pattern Recognition: Joint IAPR Int. Workshop SSPR'98 and SPR'98. Springer: Berlin, 421-428

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

Staufenbiel W (1973) Zur Automation der Generalisierung topographischer Karten mit besonderer Berücksichtigung großmaßstäbiger Gebäudedarstellungen. Wissenschaftliche Arbeiten der Universität Hannover, Nr. 51

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

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 (1997) Semantic modelling of man-made objects by production nets. In: Gruen A, Baltsavias EP, Henricsson O (eds) 43-52

Stilla U, Geibel R, Jurkiewicz K (1997) Building reconstruction using different views and context knowledge. In: Baltsavias et al. (eds) ISPRS Workshop, Vol. 32, Part 3-4W2 129-136