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

# Semantic Modelling of Man-Made Objects by Production Nets

U. Stilla, E. Michaelsen

In: A. Gruen, E.P. Baltsavias, O. Henricsson (Editors)
Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)
pp. 43-52, Basel: Birkhäuser

Ascona Workshop 1997
May 5 - May 9, 1997
Ascona, Switzerland

Ettlingen - 1997

# Semantic Modelling of Man-Made Objects by Production Nets

**Uwe Stilla and Eckart Michaelsen**

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, mich}@gate.fim.fgan.de

**Abstract**

A representational scheme for the analysis of man-made structures in aerial images and maps is described. Knowledge about object structures is represented by a set of productions. The interaction of the productions is depicted by production nets. The approach is discussed in relation to similar representations. Two example nets are given to demonstrate the flexibility and applicability of the approach. The first one is on the automatic 3D structure analysis of suburban scenes in series of aerial images. The second is on the automatic construction of descriptions of complex buildings in vector maps.

## 1 Introduction

Automatic interpretation of urban scenes from aerial images is a difficult task. Such data contain a great variety of man-made structure. Often object edges are partially occluded or are not completely detectable because of low contrast. When performing an image analysis structural context has to be captured by object models. The knowledge of human experts therefore needs to be transformed into notations that have both properties: lucidity, flexibility and simplicity for humans and formal transition descriptions for automatons in order to permit proper semantics and algorithmic complexity assessment. Moreover, the expert might often not be aware of the *common sense knowledge* he or she exploits in performing such a task. Therefore we consider this to be a very complex and ambitious field demanding some theoretical background and formalism.

# 2 Semantic Modelling

## 2.1 Productions: Definition and Example

We use the term *production* for the following structure (Stilla et al., 1996):

$$\text{Production } P_i: \quad \boxed{\underbrace{\text{Input configuration} \ \odot}_{\text{Condition part}} \ \underbrace{\xrightarrow{i} \ \text{Output configuration}}_{\text{Action part}}}$$

The *condition part* contains a predicate $\odot$ which is defined on the attributes of a configuration of objects. It contains logical, topologic, geometric or radiometric knowledge of interest within the domain of concern and exploitable for the task at hand. Examples for relations used in such conditions are adjacency, collinearety, proximity, symmetry, colors, certain measures, or even explicit CAD-model-like knowledge. Usually some of these relations are combined in one condition by logic operatores, and we developed an informal but lucid shorthand notation expressing them like in the following example:

$$\{line, ..., line\} \quad collinear \wedge overlapping$$

Here the condition is defined on a set of objects of the type LINE. It is fulfilled whenever the objects are collinear and overlapping.

The *action part* contains a function $\xrightarrow{i}$ which constructs a new configuration out of a configuration fulfilling the condition part of the production. Often this output configuration only consists of a single object and should be viewed as an interpretation of the input configuration using knowledge of the domain of concern or some standard statistical instruments. Such actions are meant to be an act of extraction of relevant data from an unordered bunch of erroneous and noisy measurements with arbitrary insertions or objects missing, and as such as an act of data reduction. Another possible view is to see it as an inference: If the condition holds on a input configuration, then the new configuration is derived from it. Under the circumstances given in automatic pattern recognition it seems more appropriate to interpret the production as an assertion of a hypothesis about the presence and pose of a certain configuration or single object of concern.

The following example gives the shorthand notation of a *prolongation* production using the condition described above:

$$\{line, ..., line\} \quad collinear \wedge overlapping \quad \stackrel{regression}{\longrightarrow} \quad (longline)$$

where the superscription *regression* stands for the process of calculating a regression straight in the least square euclidean error sum sense and determining the endpoints of the object LONGLINE by projecting the extreme LINE points along the straight on it.

### 2.2 Semantics of Single Productions

The set of configurations accepted by a *condition* is a well defined mathematical entity. The class of all models in the sense of mathematical semantics (Abramsky et al., 1992) is a more or less technical thing to define. For something well known like an analytical squared error sum minimization in the action part a good deal of meaning can be borrowed from the standards of applied mathematics (Winkler, 1995). In the example above we would propose an optimal estimation for the pose of a long straight contour segment, whose presence is both assumed and evaluated based on the given configuration of primitive measurements. Thus assuming that the semantics of all single productions are given the question arises, whether the semantics of a whole system of productions may be deduced there off.

### 2.3 Control Strategies for Production Systems

Given a finite set of productions and a large database of primitive objects the question arises, in what sequence they are to be applied. This is commonly discussed under the issue of control strategies. Of course the control decisions on the sequence of application (and backtracking) of the productions have a strong impact on the semantics of the system. For the time being we therefore restrict our investigation of semantics to one very important special mode of operation, where the semantics of a whole production system may be deduced from the semantics of all it's productions. This is the special case of exhaustive search. In this case a set of primitve objects belongs to the language of the system *iff there exsists* a sequence of applications using the productions of the system that leads to a set including some target object.

### 2.4 The Production Net

Graph structures are often used to illustrate the knowledge implemented in computer programs. *Semantic nets* are helpful for representing declarative knowledge (Findler, 1979). In general a semantic net consists of nodes and links. For description of a model-scheme nodes represent concepts and links represent relations between concepts (e.g. *part_of* or *concrete_of*). Relations like in Fig. 1a provides an example of a simple semantics.

Another form of knowledge representation are *Petri nets* (Reisig, 1991). A Petri net is a bipartite directed graph with nodes, called places and transitions. A circle represents a place and a bar or box represents a transition (Fig. 1c). An arrow from a place to a transition defines the place to be an input to the transition. Similarly, an output place is indicated by an arrow from the transition to the place. The dynamic aspects of Petri nets are denoted by markings which are assignments of tokens to the places. The execution of a Petri net is controlled by the number and distribution of tokens.

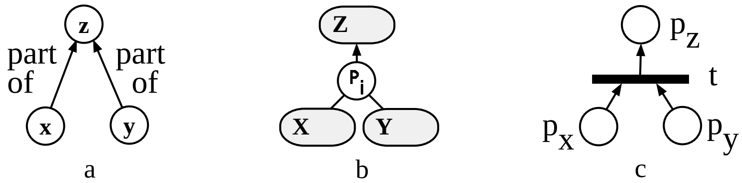


**Fig. 1:** Graphic elements. a) Semantic net, b) Production net, c) Petri net

Similar to Petri nets we use in a *production net* two disjointed sets of nodes, called concepts and productions. Concepts are depicted by ovals and productions are depicted by circles (Fig. 1b). Arrows connect only elements of different sets. An arrow from a production to a concept indicates the generating function and points to the output configuration. Arrows or arcs pointing from concepts to productions are constructed whenever the concepts participate in the input configuration. Often the underlying meaning of such an arc resembles that of a *part-of* link in a semantic net. Other semantic net mechanisms like specialisation-generalisation and inheritance are not implemented in our production nets.

The role of the production net formalism can be seen between the very descriptive semantic net convention that leaves the question of the sequence of application of knowledge open, and the very procedural Petri net scheme that helps to track traps, bottlenecks and looping.

Production nets together with a detailed documentation of all productions used in them including the mathematical definition of each relation and function with all the thresholds and tolerances fixed, give an insight in the semantics of systems using their productions, provided that the control assures exhaustive search. Also an idea is given on the information flow of the system, and the mutual dependencies and independencies of the productions and concepts.

## 3 Application Examples

### 3.1 Example I: Image Analysis of Suburban Areas

Fig. 3a shows a section of a suburban scene taken by a sequence of aerial images. Many edges required for a complete assignment of model parts and image segments do not appear in the images. Others are very prominent but do not fit into the semantics of the model. Nevertheless human observers have no difficulties to recognize objects in suburban areas. Humans possibly use perceptual grouping and consider contextual information from broader areas. That means, knowledge about non local properties of man-made objects in suburban areas has to be considered. Examples are collinearity, parallelity and periodicity. Fig. 3a shows a regular structure given by houses of similar size and same orientation, roughly aequidistantly spaced in rows parallel to a street.

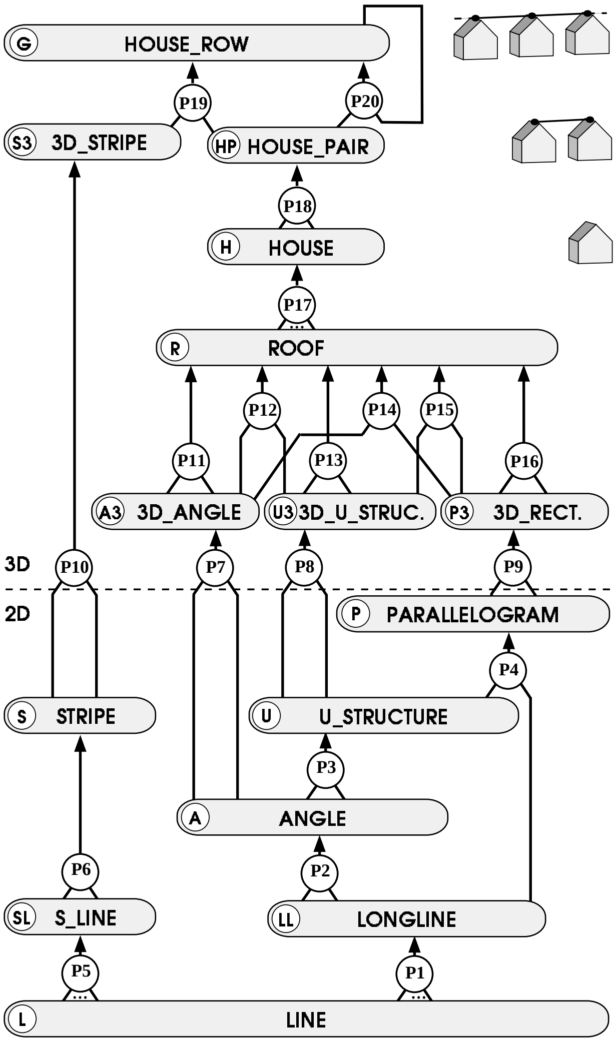


**Fig. 2:** Production net HOUSE_ROW

Fig. 2 shows a production net meant to capture such knowledge in it's semantics. Many topologic and geometric constraints are naturally formulated in 3D. Examples are the mutual amplification of evidence between house rows and parallel road-shaped stripes (P19), the grouping of houses (P18 and P20), the clustering of roof hints to house hypothesis (P17), and the construction of roofs from pairs of objects 3D_RECTANGLE, 3D_U_STRUCTURE, and 3D_ANGLE (P11-P16). Many objects are excluded from further analysis due to 3D constraints.

Other knowledge about scenes may already have been exploited on the level of 2D image analysis. Some man-made objects like streets show long straight contours. This property is invariant under a perspective projection. As long contours are often partially occluded the production P5 is designed to prolongate and to bridge over gaps of considerable extention. Production P5 tolerates much larger gaps than production P1.

Production P2-P4 use adjacency in image space. These fail frequently, because adjacency is not invariant under perspective projection. Whereas production P6 proved rather stable, although it exploits parallelity which is also not an invariant. This is due to the special perspective of aerial photography and will not hold for groundbased or tilted views. Productions P7 to P10 invert the projection from the scene to the image. Productions P7 to P9 construct scene rays from image points and intersect them, whereas production P10 intersects planes in the scene resulting from long straight line segments in the image.

A production net similar to Fig. 2 has been presented in a previous paper (Stilla & Jurkiewicz, 1996). The net has been extended by concidering streets as context which allows us to be more tolerant in accepting roofs.

Fig. 3d and 3e show the best objects of type HOUSE_ROW and corresponding contextual object 3D_STRIPE resulting from two aerial images. Fig. 3b and 3c show some intermediate results.

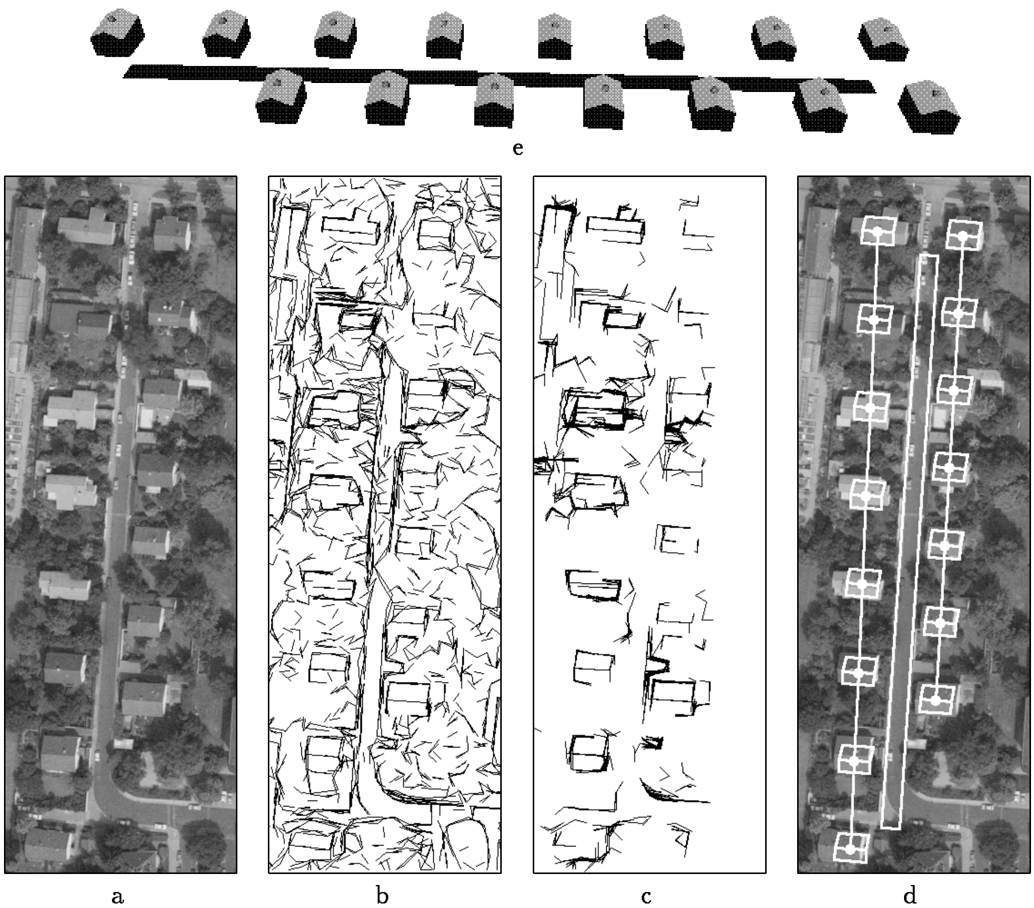


**Fig. 3:** a) Section of an aerial image, b) objects LINE, c) objects 3D_ANGLE, d) best objects HOUSE_ROW projected to a), e) 3D visualisation of best objects HOUSE_ROW

### 3.2 Example II: Map Analysis with Generic Building Models

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

One task of an image analysis is the *extention* 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 an image analysis (e.g. restricting search).

Another task of an image analysis is the *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. Both for the image analysis using the map as well as for the map update a hierarchical description of the map is suitable.

Such descriptions are generated by a map analysis using parametric and generic models. A large scale vector map is used which is organized in several layers each of which contains a different class of objects (e.g. streets, buildings, etc.) One of these layers contains large buildings with their parts. The task of map analysis is to seperate building parts, to determine enclosed areas, and to group building parts. Fig. 4a shows a production net of a generic model for the analysis of complex buildings examining the topological properties: *connectivity*, *closedness*, and *containment*.

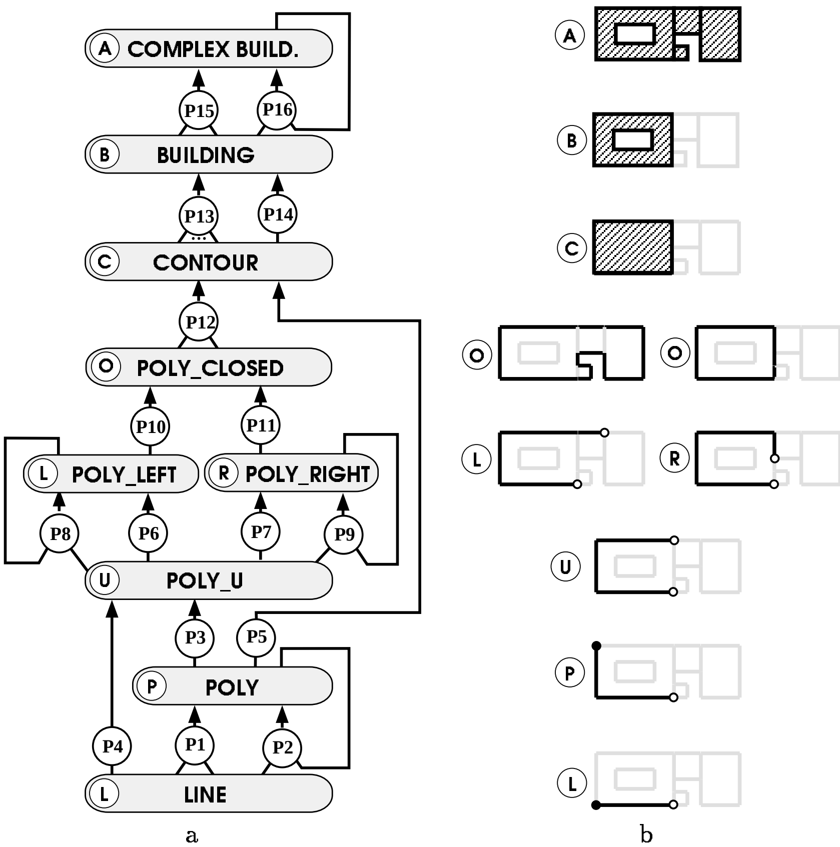


**Fig. 4:** a) Production net COMPLEX_BUILDING, b) Examples of objects

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 both endpoints of a polygon are branch points, an object unbranched polygon (POLY_U) is produced of it (P3). An object LINE having two branch points by itself is copied into an object POLY_U by production P4.

Based on the set of unbranched polygons POLY_U we search for those polygons enclosing minimal delimitable areas (meshs). For that purpose objects POLY_U are connected so that objects POLY_LEFT follow one path with maximal continuation angles (P6, P8), and objects POLY_RIGHT follow another path with minimal continuation angles (P7, P9).

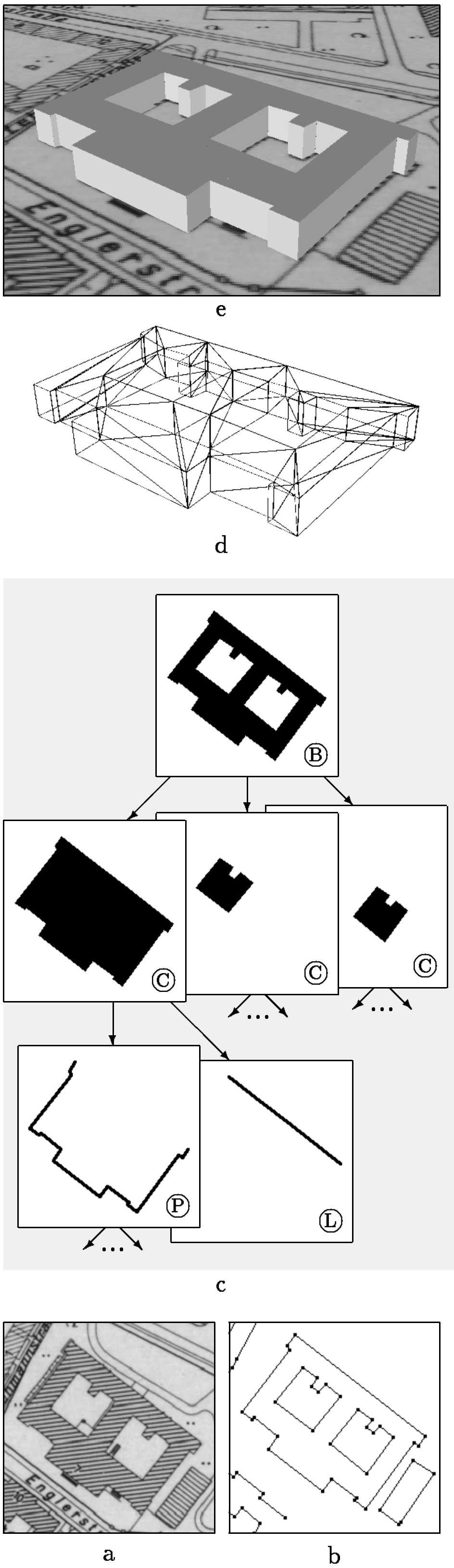


**Fig. 5:** Example of map analysis.

If such a path closes a polygon by productions P10 and P11, objects POLY_CLOSED are created. Comparing the areas of objects POLY_CLOSED pairwise determines the object CONTOUR with the smaller area (P12). If polygons can be closed without branch points, production P5 directly creates an object CONTOUR. If there are one or more contours inside another contour, they are combined into an object BUILDING (P13). Production P14 generates from building contours, which do neither lie inside other contours nor contain other contours objects BUILDING. Adjacent objects BUILDING are combined into an object COMPLEX_BUILDINGS (P15) and are eventually extended by other adjacent buildings (P16).

An example of map analysis using the presented production net COMPLEX_BUILDING is given in Fig. 5. A section of a scanned map shows a building which has two interior yards and a non convex shape (Fig. 5a). The corresponding vector map symbolically describes the building in a map layer by a set of lines (Fig. 5b). These lines are analyzed by applying the productions of the net. Because the building is not part of a complex building and therefore has no branch points, only productions P1, P2, P5 and P13 were applied to compose the objects POLY Ⓟ, CONTOUR Ⓒ and BUILDING Ⓑ.

Using the derivation graph of the building (Fig. 5c) a hierarchical description can be given in the description levels Ⓑ, Ⓒ, and Ⓛ. The features of the building are described on different levels of detail, for example on level Ⓑ with attributes: Building area, bounding box, number of interior yards, center of gravity, axis of inertia, or on level Ⓒ with: Area of parts, perimeter, center of gravity, or on level Ⓛ with: Coordinates of contour lines.

In the field of 3D object recognition and reconstruction from images, it might be helpful to use a map for generating a simple 3D-model. When we assume a standard height of buildings for this class, the 2D-description of a map layer can easily be extended to a rough 3D-description (wireframe model) by prismatic objects.

For the modelling of projections occlusion has to be considered. Therefore the wireframe model is transformed into a surface model using an automatic triangulation (Fig. 5d). Fig. 5e shows the rendered 3D-object with the scanned map underlayed. The result on a bigger section of the city map of Karlsruhe is given in Fig. 6.

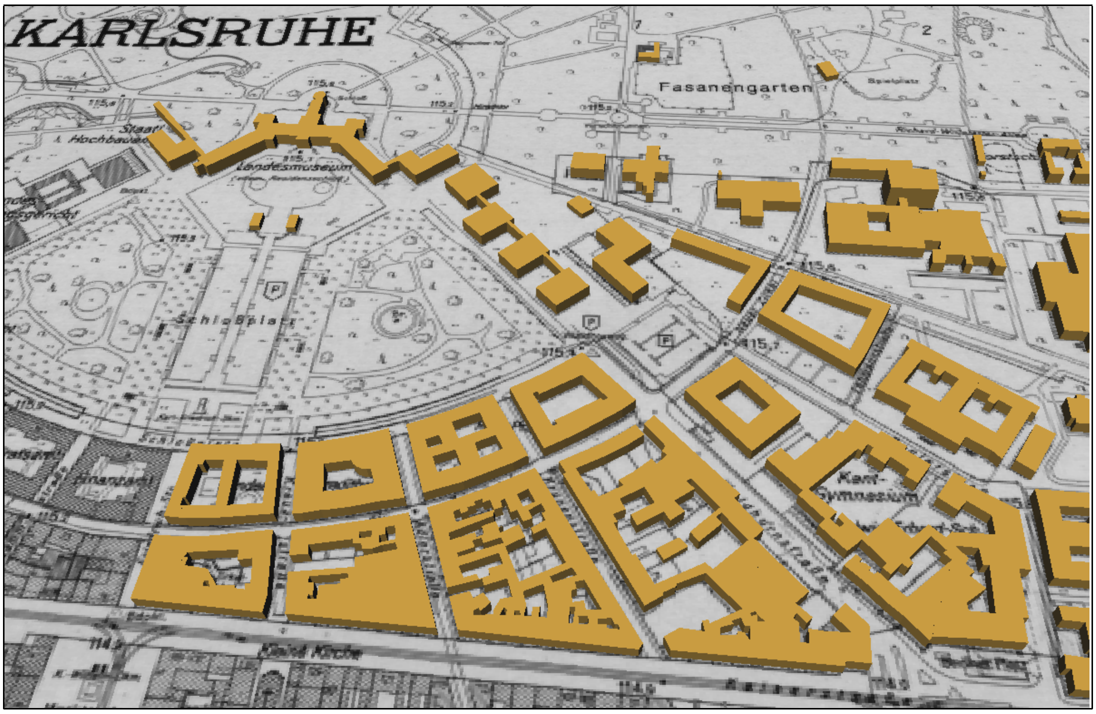


**Fig. 6:** Result of map analysis.

## 4 Conclusion

Some generalizations can be drawn from the examples presented: Production nets for image analysis tasks will have to cope with numerous alternatives. Exhaustive search may not be tractable with reasonable effort. For the time being only special cases can be considered in the domain of strict semantics.

First automatic tasks on the images alone are tractable, provided the nets are small, the productions are simple and the underlying models are of restrictive and parametric type. Second automatic tasks, with generic models of sufficient generality, are tractable if no segments are missing, not too many alternatives of linking and strong local constrains are given. An example is the analysis of vector maps.

Interactive demonstration of constructibility with production nets is always possible. The existence of a construction may be demonstrated by human experts using the production net together with tools for visualization of intermediate results and for the explanation by derivation graphs. The next step would be to look for evaluation criterions that make control decisions possible and prune the search tree down to a tractable size without losing the correct reduction with considerable probability. It is very difficult to prove that a certain target object is *not* constructible from a set of terminal objects without performing exhaustive search. Thus the assessment of the discrimination remains unclear.

We conjecture, that the problems mentioned above are not approach-specific for production nets. However, we suggest that semantic modelling of man-made objects by production nets does increase the awareness of these problems, because apart from the declarative aspects procedural aspects become explicit, too.

## References

Abramsky S, Gabbay DM, Maibaum TSE (1992) *Handbook of logic in computer science.* Oxford: Clarendon

Gruen A, Kuebler O, Agouris P (1996) *Automatic extraction of man-made objects from aerial and space images.* Basel: Birkhäuser

Findler NV (1979) *Associative networks.* New York: Academic Press

Niemann H, Sagerer G, Schröder S, Kummert F (1990) *ERNEST: A semantic network system for pattern understanding.* IEEE PAMI, 12: 883-905

Quint F (1996) *Recognition 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

Reisig W (1991) Petri nets. New York: Springer

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, 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

Winkler G (1995) *Image analysis, random fields and dynamic Monte Carlo methods.* Berlin: Springer