Map–based semantic modeling for the extraction of objects from aerial images

Franz Quint, Manfred Sties

Institute for Photogrammetry and Remote Sensing
University of Karlsruhe
D – 76128 Karlsruhe, Germany
Ph.: +49-721-6083092, Fax: +49-721-694568
email: quint@ipf.bau-verm.uni-karlsruhe.de

Abstract

Images taken from satellite or airborne platforms usually do not represent isolated information of man's environment. In most countries, valuable context data are available which may be integrated successfully in the image interpretation procedure. This paper presents the verification phase of a Map Oriented SEmantic image under Standing process[1] (Moses). It is implemented as a model driven process, where semantic networks are used as modeling tools. In a three stage scheme, the models are successively refined and for image analysis an automatically generated semantic network, specialized on the analysis of the underlying scene is used. Digitized topographic maps serve as a principal knowledge source.

1 Introduction

The analysis of aerial and satellite images for exploration and monitoring of our environment gains rising importance. In many cases, it is the only applicable method for solving given problems. As soon as the reliability of the automated processing schemes increases, they are advantageous under cost and time aspects. Therefore, it is not surprising, that the international community supports big efforts for developing efficient and reliable image understanding procedures.

These efforts are encouraged by the fact, that in near future Geo-Data will be available in digital form; however, this is not yet generally the case except for Satellite Remote Sensing Imagery. At the moment many national surveying and cartographic administrations are establishing digital data bases of conventional topographic maps in vector form (see e.g. Harbeck, 1994).

---

[1]The research is funded by the Deutsche Forschungsgemeinschaft (DFG)

Understanding of aerial images is one of the challenging problems in computer vision. Due to its complexity, knowledge based systems have been found to be mandatory since the mid seventies: Agin (1979) developed a knowledge based system for detection of roads and vehicles in aerial images.

A rule-based aerial image understanding system (SPAM) for interpretation of aerial scenes has been presented by McKeown et al. (1985). Several production rules are applied to generate partial interpretations, and then a set of rules for consistency examination is applied.

A blackboard based production system (BPI) is used by Stilla and Hajdu (1994) for a map-aided analysis of structures in aerial images. Based on the knowledge obtained by map analysis, estimations for attribute values of objects are defined. Objects found in the image analysis process are assessed relative to the hypotheses generated by map and object model.

The knowledge based system for aerial image understanding presented by Nicolin and Gabler (1987) is composed of four components: long term memory, short term memory, procedures and control module. In a first, data-driven stage, initial interpretations and assessments are assigned to striking structures in the image. In the second, model-driven stage, hypotheses for not yet classified objects are build using a-priori knowledge and the results of the first stage.

The system SIGMA developed by Matsuyama and Hwang (1990) consists of four loosely coupled reasoning modules. The *Geometric Reasoning Expert* as the central module of the system stores knowledge in frames. Object classes are described with attributes and relations and instantiated objects can generate hypotheses for neighboring objects.

In the system MESSIE (Sandakly, Giraudon, 1994) knowledge is also stored in frames and the communication between the *specialists* takes place using a blackboard mechanism. A specialist starts processing after a query by another specialist (model-driven phase) or when it is triggered by events (data-driven phase).

In this paper the verification part of MOSES (*M*ap *O*riented *SE*mantic image under*S*tanding) is presented. We incorporate knowledge from the map domain for image analysis. Using a semantic network for analysis of the map, a description of the scene in map domain is build. Another semantic network specialized in the analysis of the underlying scene is automatically generated using the scene description in map domain and a generic model in image domain. This later network is used for image analysis.

2 Image and map data

The methodologies for map-based verification and recognition of objects are developed for urban scenes. As a test area, a sector of the urban environment of the city of Karlsruhe was selected, which contains both typical metropolitan densely populated areas as well as extended park and forest areas.

The digital image data were acquired by scanning aerial color photographies of 230 by 230 mm. The aerial images were taken from a flying height of 2 km, at a scale of 1:6300 and scanning of the photography was done with 50 μm spot size and a grey value resolution of 8 bit.

The context information was acquired from a topographic map, the German Topographic Base Map 1:5000. The corresponding sector to the aerial images was acquired by manual digitization of the contours and is stored in the data base of GIS ARC/INFO. To each digitized contour a class is assigned.

The transformation parameters between the coordinate system of the images and the world coordinate system are determined by an adjustment from the ground control coordinates. The transformation parameters allow to project the map information onto the image. Figure 1 shows the digitized contours from the map as an overlay (white lines) on the image of a scene.

Fig. 1: Extract from the aerial image with map information as overlay (white lines)
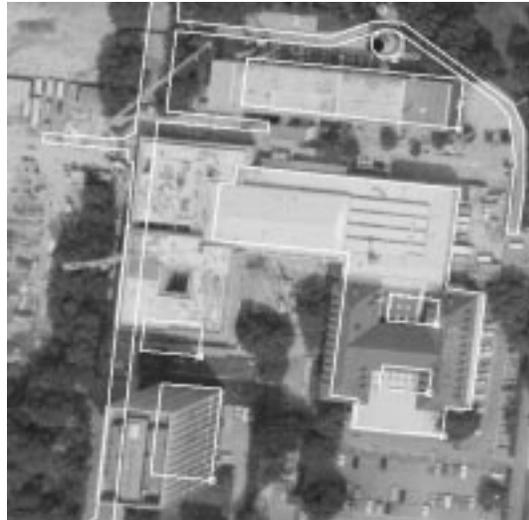
Fig. 2: Differences between the aerial image and the map information (white lines as overlay)

This examples shows, that the object borders acquired out of the map are not at all identical with the contours visible in the image. There are multiple reasons for this effect:

- The map does not include height information. Therefore transformation parameters were only modeled on a 2D-base.

- The perspective distortions in the image, which are extreme especially for tall buildings at the image borders.

- The map shows the ground plans of the buildings, whereas in the images the border lines of the roofs are visible. They necessarily do not fully match.

- Occlusions between objects occur.

- Inaccuracies during the digitization and ground control point determination.

Due to the different acquisition dates of map and image, there are also differences because of the changed reality. These differences are frequent in build-up areas, especially due to activities in construction (Fig. 2). A system for image analysis has to tolerate the differences in representation of the same object, but has to detect the differences due to changed reality.

3 Verification and classification

In general, aerial photography is more economic and faster than the production of new maps. Automatic updating of maps on the basis of aerial images is desirable. As previously shown, the representations of a scene in map and image are different and partly contradictory. However, the map contains knowledge about the scene and can be used as a (incomplete) model for image interpretation. In achieving our goal of image understanding, we define two phases: verification and classification phase.

In the verification phase we rely on the contents of the map: we use the map as a model for image interpretation and verify the map contents. The result of the verification phase is a description of the image, in which for each object of the map, features extracted from the image are measured and an assessment is computed, how good these features match with the model. This description is still incomplete, since it contains only the objects represented in the map.

In the subsequent classification phase, the verification result is used as context for classifying the objects, for which a satisfactory verification could not succeed, and for classifying image primitives not addressed in the first phase.

In this paper we will discuss only the verification phase. The image analysis process for verification of the map contents is conceived as a model driven algorithm. A main component of the model is the map contents to be verified. As a modeling environment, the semantic network system ERNEST (Niemann et al., 1990) is used. For a general discussion of semantic networks, see e.g. Findler (1979). During the analysis process, several models are build and each model is represented in its own semantic network. Because of this unique correspondence between a model and a semantic network, these terms are used synonymously in the following.

4 Usage of map knowledge

The mechanism of inheritance and the hierarchies in semantic networks allow an efficient modeling and use of domain knowledge for pattern recognition tasks.

Using only general, common sense knowledge will lead to simple semantic networks, which are comparatively easy to build. However, we have to accept the drawback,
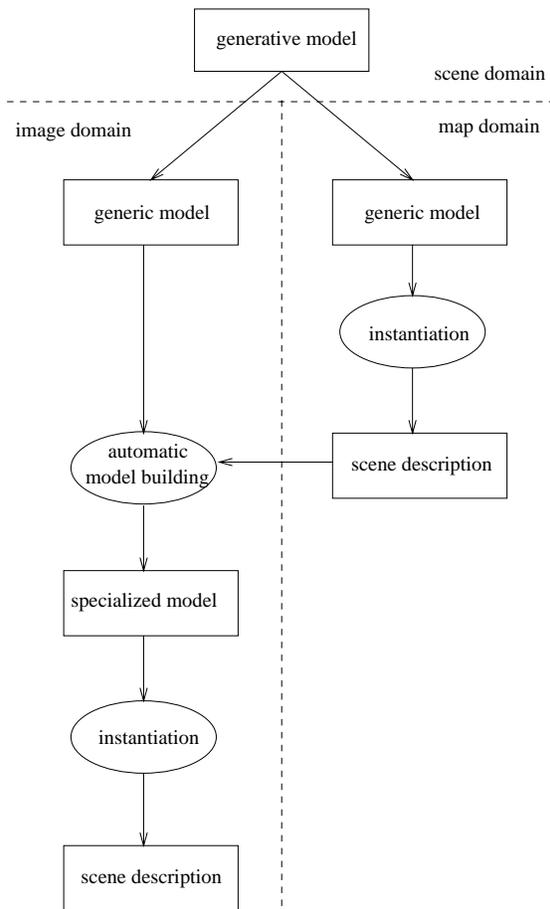
Fig. 3: Structure of the analysis process
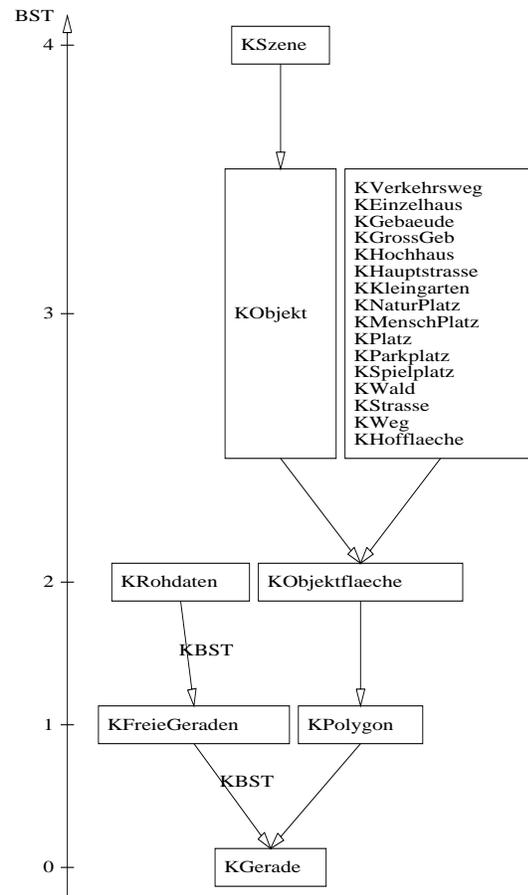
Fig. 4: Part-of hierarchy of the generic model in map domain

that these models in general are not as powerful for the image analysis task as specialized models. Building these requires bigger efforts and they will perform well on the specialized problem they were build for, but generally they will have a poorer ability of generalization.

We present a three stage approach for modeling and use of knowledge: relatively simple, general models are specified by the system designer, whereas specialized models for image analysis are automatically build using the results of map analysis.

The structure of the analysis process is shown in figure 3. The models and the results of an analysis process (scene descriptions) are represented as rectangles, whereas processes are represented as ellipses. The direction of the arrows indicates the direction of data flow.

We distinguish between three domains: the scene domain, the image domain and the map domain. The scene domain is our environment, the real world. The map domain and the image domain comprise representations of the real world in map and image, respectively.

The generative model is a semantic network, which describes our environment in scene domain. The knowledge contained in it is general, common sense knowledge we have about our environment. Objects of the real world are described, as they are perceived by the system designer, regardless of their representation in a map or an image. Since analysis will not be performed in scene domain, there is no need for procedural knowledge in this semantic network.

The generic models in the map and in the image domain describe the mapping of the scene in map and image domain, respectively. Therefore, they roughly contain the same declarative knowledge as the generative model. Both generic models are specializations of the generative model, they reflect, however, particularities of the representations in map and image domain: Since in the map, for example, vehicles are not represented, during the specialization of the generative model to the generic model in map domain, the *part-of* link in the semantic network to the concept representing a vehicle is marked as non existent. In the image, vehicles can be observed and the link to the vehicle-concept is preserved in the generic model in image domain.

Being specializations of the generative model, both generic models describe a general, typical scene and therefore they can be used for analysis of various kinds of scenes. The part-of hierarchy of the generic model in map domain, as drawn by the network analysis tool of ERNEST, is given in figure 4.

Since analysis in image and in map domain is desired, procedural knowledge is present for both generic models. Part of the procedural knowledge are functions for feature extraction, for the calculation of attribute values from image or map data and functions for calculation of confidence values. The procedural knowledge is specific for map domain and for image domain, but it is not scene-specific.

Although both generic models can be used for analyzing the scene in the respective domain, we use only the generic model in map domain for analysis. Because map contours are available in digital form, fault tolerant feature extraction procedures are not necessary. It is expected, that building of the scene description in map domain can be performed with higher reliability than the one in image domain using the unspecific generic model.

The instantiation process in map domain gives us a complete description of the scene in the limits of the map contents. Using this description and the generic model in image domain, we automatically build a new semantic network, which is a specialized model of the concrete scene to be analyzed. Whilst the generic model contains statements of the form:

- buildings are optional parts of the scene, or

- a building may have inner courts,

the specialized model of the actual scene to be analyzed contains precise information about the objects of the scene, e.g.:
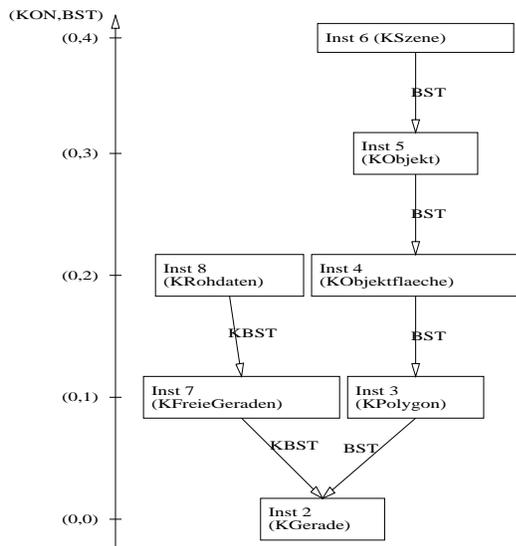
Fig. 5: Intermediate scene description graph
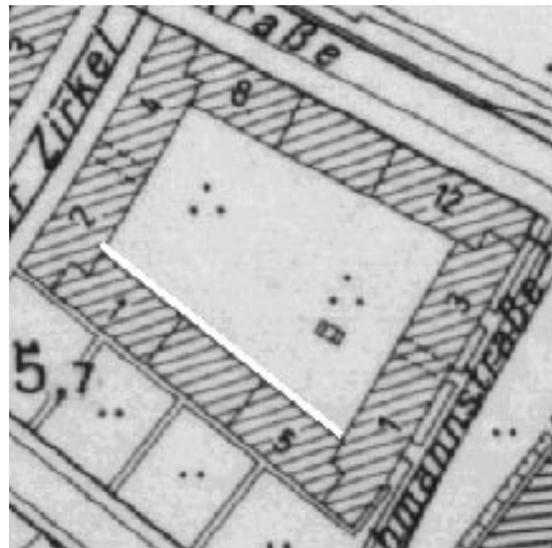
Fig. 6: First instantiated line as overlay in the map

- building number 5 is rectanglarly shaped and the coordinates of corner points are such and such.

The specialized model is now used for image analysis. Because of the concrete modeling of the expected objects in the scene, using this model for image analysis will give more reliable results than using the generic model. Being automatically generated, expense for the system designer is not increased.

For verification of the map objects in the image, we do not compare the results of an analysis process in map domain with those of an analysis process in image domain, where both processes are performed independently, but we use the knowledge obtained from map analysis for the image analysis process.

5 Map analysis

We present some fundamental aspects of the analysis process (instantiation) in our semantic network by an example of map analysis.

Instantiation is an alternate top-down and bottom-up processing. It starts top-down: The target concept, which in our case is *KSzene* (Fig. 4) and represents the scene to be analyzed, is submitted for instantiation. Since *KSzene* depends on other concepts, its parts, a final instance of it can be created only when all its parts are instantiated. The *part-of* hierarchy is tracked down until a basic concept is found, which can be instantiated by directly computing its attributes from available data. In the following bottom-up phase the concepts, for which previously it was not possible, are now instantiated. An intermediate scene description graph showing the situation after the first cycle is given in figure 5. In figure 6, the line corresponding to instance no. 2 (*KGerade*) of figure 5 is overlayed in white onto the map.
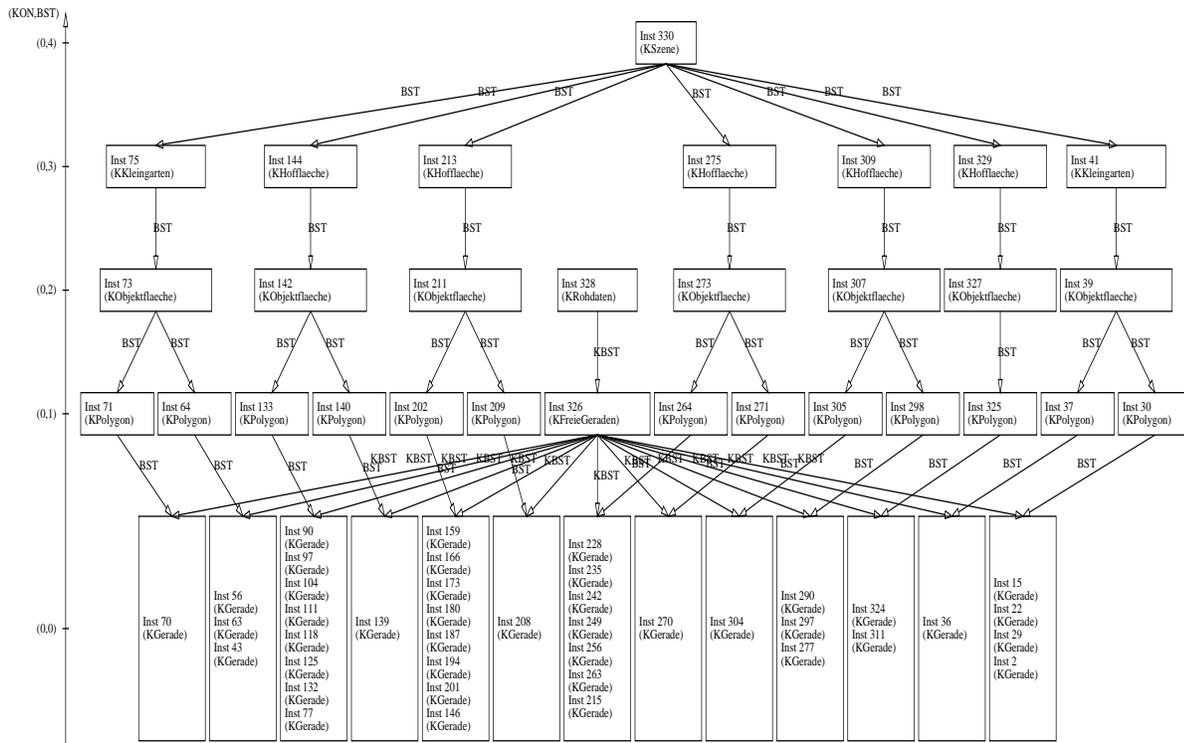
Fig. 7: Scene description graph after map analysis

In top-down phase, constraints are propagated. For example, the coordinates of the next line to complete the contour of the courtyard (Fig. 6) must satisfy the constraints imposed by instance no. 2.

Instantiation continues until no more digitized contours are available. The resulting scene description graph contains the knowledge extracted from the map in a structured fashion. Figure 7 shows an example of such a scene description graph at the end of a map analysis process. This scene description graph is used for automatically building the specialized model for image interpretation.

6 Specialized model for image analysis

The program for automatically building the specialized model for scene analysis relies only on the supposition, that the generic model in map domain and the generic model in image domain are two semantic networks, which originate from the same parent, the generative model. Beginning with the target concept of the scene analysis process, for each concept in generic model in image domain, instances of corresponding concepts in the generic model in map domain are searched in the scene description and for each correspondence found, a new concept in the specialized model is created. Using the part-of hierarchy of the network and of the scene description graph, this is done recursively.

The attributes and procedural knowledge attached to the newly created concepts

(KON,BST)

(0,4)

BSzene 330

BST BST BST BST BST BST BST

(0,3)

| BKleinga0041 | BKleinga0075 | BHofflae0144 | BHofflae0213 | BHofflae0275 | BHofflae0309 | BHofflae0329 |

BST BST BST BST BST BST BST

(0,2)

| BObjektf0039 | BObjektf0073 | BObjektf0142 | BObjektf0211 | BObjektf0273 | BObjektf0307 | BObjektf0327 |

BST BST BST BST BST BST BST

(0,1)

| BPolygon0030 | BPolygon0064 | BPolygon0133 | BPolygon0202 | BPolygon0264 | BPolygon0298 | BPolygon0325 |

BST BST BST BST BST BST BST

(0,0)

BGerade0002
BGerade0015
BGerade0022
BGerade0029

BGerade0043
BGerade0056
BGerade0063

BGerade0077
BGerade0090
BGerade0097
BGerade0104
BGerade0111
BGerade0118
BGerade0125
BGerade0132

BGerade0146
BGerade0159
BGerade0166
BGerade0173
BGerade0180
BGerade0187
BGerade0194
BGerade0201

BGerade0215
BGerade0228
BGerade0235
BGerade0242
BGerade0249
BGerade0256
BGerade0263

BGerade0277
BGerade0290
BGerade0297

BGerade0311
BGerade0324

Fig. 8: Part-of hierarchy of the specialized model for image analysis

are gained from the generic model in image domain. The values of the (mainly geometric) attributes of the instances build during map analysis are transformed from map domain to image domain and stored as constraints of the corresponding attributes in the concepts of the specialized model. These constraints lead to a substantial reduction of search space during image analysis.

The part-of hierarchy of the specialized model for image analysis of the scene with the description graph in figure 7 is given in figure 8. The names of the concepts are derived from the names of the concepts in the generic model, to which the number of the instance in scene description was appended. One can observe the correspondence between the part-of hierarchy of the newly created semantic network and the scene description graph after map analysis.

Because of an implementation detail, not all instances in the scene description have a correspondence in the specialized model. It is a characteristic of the semantic network ERNRST, that calculation of attributes is done during instantiation. And even if no proper attribute values satisfying the constraints are found, the instance remains in the resulting description graph. However, these instances are marked as not corresponding. This is the case e.g. for instance no. 70 *KGerade* or no. 71 *KPolygon* in figure 7. For these instances no corresponding concepts in the specialized model are build.

7 Summary and conclusions

We presented a new scheme for image analysis using contextual knowledge. The interpretation task is divided in two phases: verification and classification phase. As a modeling tool, we use semantic networks. Starting with a very general model (generative model), we build two models describing the scene in map and image domain (generic models). The model in map domain is used for analysis of the map. The resulting scene description graph is combined with the generic model in image domain and a new semantic network, specialized in the analysis of the concrete scene is built.

This procedure is very effective, since map knowledge is automatically incorporated in the analysis process. The resulting specialized model is tailored for image analysis of the underlying scene. Presently, only line segments are used in the image analysis process. We plan to integrate region-based features for better grouping of line segments and experiment with different assessment procedures in the generic model in image domain.

References

Agin, J. (1979) *Knowledge based detection and classification of vehicles and other objects in aerial images*, Proceedings of the DARPA Image Understanding Workshop, Palo Alto, CA, pp. 66-71.

Findler, N.V. (ed.) (1979) *Associative Networks*, Academic Press, Orlando.

Harbeck, R. (1994) *Das Geoinformationssystem* ATKIS *und seine Nutzung in Wirtschaft und Verwaltung*, Landesvermessungsamt Nordrhein–Westfalen, Bonn.

Matsuyama, T., Hwang, V. (1990) SIGMA: *A knowledge based aerial image understanding system*, Plenum Press, New York, London.

McKeown, D., Harvey, W. and McDermott, J. (1985) *Rule based interpretation of aerial imagery*, IEEE–PAMI, Vol. 7, No. 5, pp. 570–585.

Nicolin, B., Gabler, R. (1987) *A knowledge based system for the analysis of aerial images*, IEEE Transactions on Geoscience and Remote Sensing, Vol. 25, No. 3, pp. 317–329.

Niemann, H., Sagerer, G., Schröder, S., Kummert, F. (1990) ERNEST: *A semantic network system for pattern understanding*, IEEE–PAMI, Vol. 12, No. 9, pp. 883–905.

Sandakly, F., Giraudon, G. (1994) *Multispecialist system for 3D scene analysis*, Proceedings of ECAI 94, J. Wiley & Sons Ltd., pp. 771–775.

Stilla, U., Hajdu, A. (1994) *Map-aided structural analysis of aerial images*, Proceedings of the ISPRS Commission III Symposium, München, Vol. 30, Part 3/2, pp. 769–775.