Aerial image understanding using digital map-based semantic models

F. Quint
Institute for Photogrammetry and Remote Sensing
University of Karlsruhe
quint@ipf.bau-verm.uni-karlsruhe.de

Abstract

Our interest is in the task of change detection in aerial images for the purpose of map updating. We are using a model based structural image analysis process implemented in the system MOSES [1]. As a modeling tool semantic networks are used. In a three stage scheme, the models are successively refined and for image analysis an automatically generated semantic network, specialized on the analysis of a concrete scene is used. In this paper an overview of the general architecture of our system and of the control strategy used in the image analysis process is presented.

1 Introduction

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. Different examples of knowledge based systems applied to the task of aerial image understanding can be found e.g. in [1], [4], [5], [6], [11],[12].

Our objective is change detection in aerial images for the purpose of map updating. Knowledge extracted from the maps to be updated is used in our system MOSES (*M*ap *O*riented *SE*mantic image under*S*tanding) [10] for image analysis. The maps in their digital form are analysed and a description of the scene, as far as it is represented in the map, is build. Automatically combining the such gained scene description with a generic model for image analysis results in a model specific for the current scene. This specific model is used for image analysis. For representing the models we use semantic networks (see e.g. [2]) as implemented by the shell for knowledge based analysis ERNEST [7] [3].

2 Map and image

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 color aerial photographies. The size of an image pixel on the ground is $30cm \times 30cm$. The context information is acquired from a topographic map, the German Topographic Base Map 1:5000. The transformation parameters

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

Figure 1: Aerial image (part of the campus of the University of Karlsruhe)

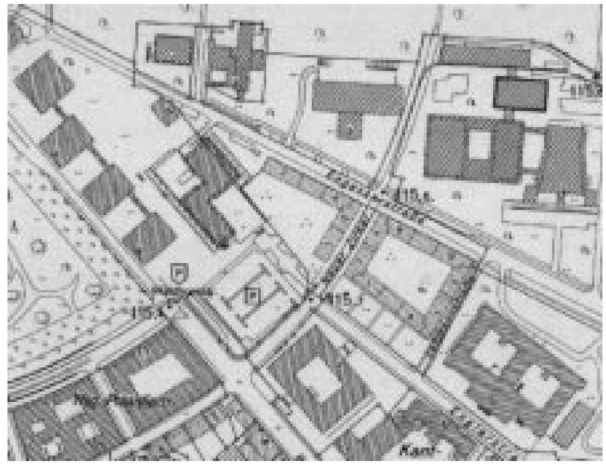

Figure 2: Map of the scene in Fig. 1

between the coordinate system of the images and the world coordinate system are determined by an adjustment from the ground control coordinates.

Image and map represent certain aspects of our environment. Analysing the two representations, conclusions about the real objects in the scene are drawn. Although being representations of the same scene, map and image may represent different properties of the objects. The aerial image of a scene from the campus of the University of Karlsruhe is given in Fig. 1, the corresponding map section is shown in Fig. 2.

The map does not include height information and shows the ground plans of the buildings. In the image the roofs of the buildings are visible. Thus the recognition of the buildings in the aerial image is done by recognizing their roofs. The border lines of the roofs do not fully match with the ground plans of the buildings. The map does not contain any information on the shape and the color of the roofs. Other differences in the representations of the scene by map and image are due to partial or total occlusions, which may occur in the image. In the example in Fig. 1 and Fig. 2 a street in the right portion of the image is totally occluded by trees. Several objects visible in the image (e.g. cars, trees) are not represented in the map.

Due to the different acquisition dates of map and image, there are also differences because of the changed reality. 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. However, the map contains knowledge about the scene and can be used as a (incomplete) model for image interpretation.

3 System architecture

For change detection 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 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. In our system MOSES [10] four models stored in semantic networks are used. Three of these models are scene independent and are specified by the system developer. The fourth is specific for the scene to be analysed and is automatically build. It is the one actually involved in the task of image analysis.

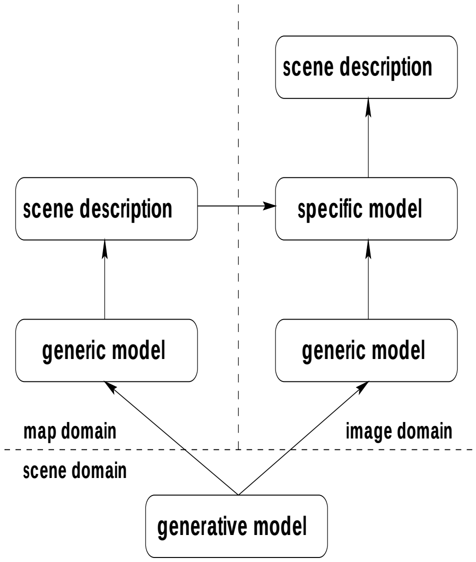

Figure 3: Structure of the analysis process

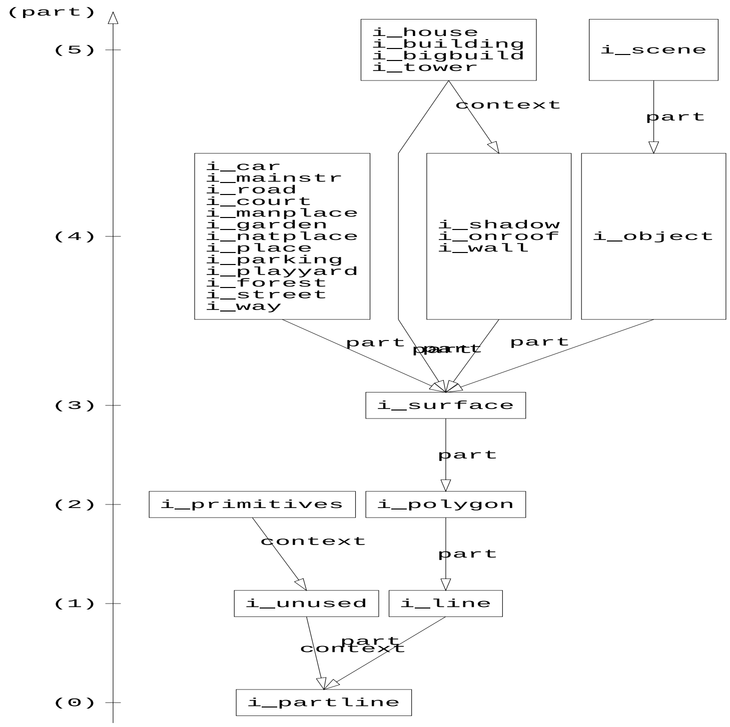

Figure 4: *Part-of* hierarchy of the generic model in the image domain

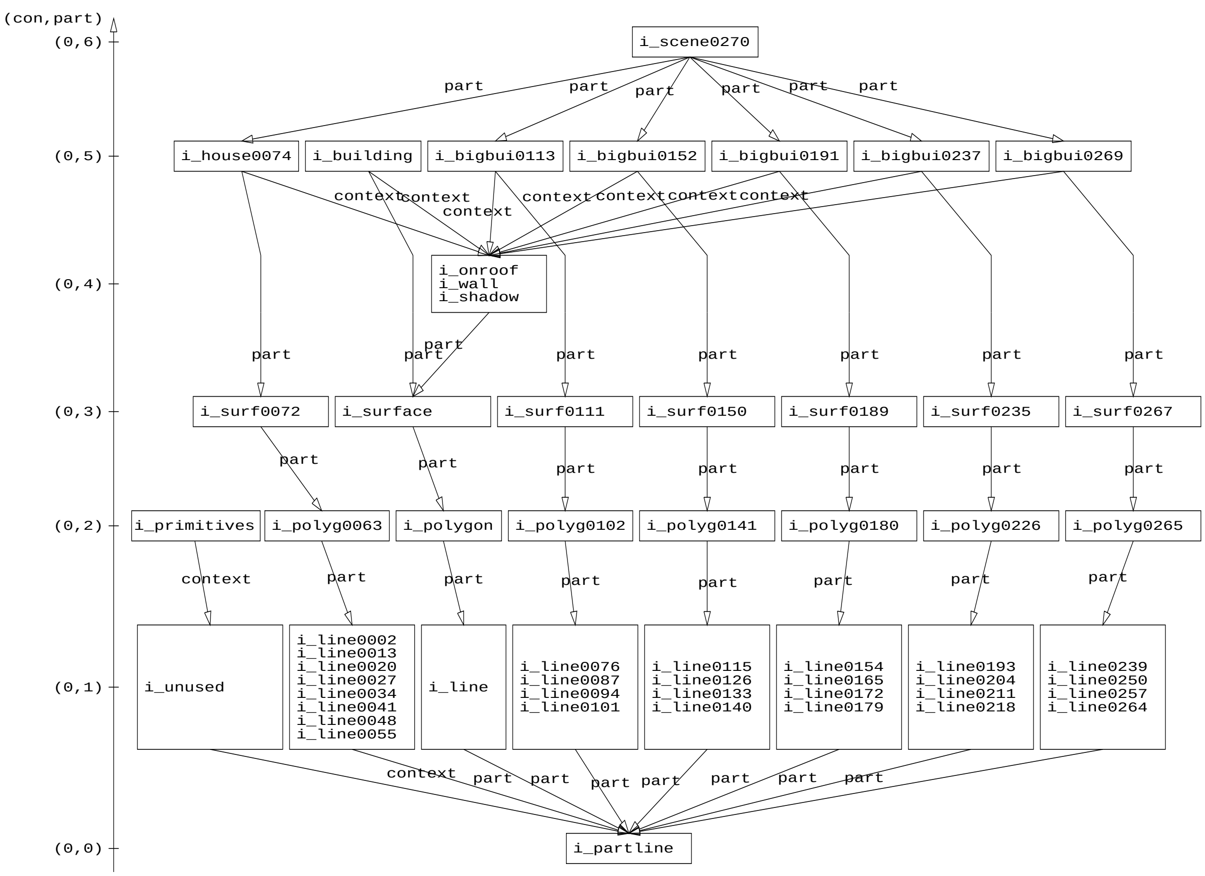

Figure 5: *Part-of* hierarchy of the specific model for image analysis

The structure of the analysis process is given in Fig. 3. 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 of declarative nature and it is general, common sense knowledge we have about our environment. The generative model does not contain methods since this model is never used for analysis purposes. It is used to efficiently store the common properties of the two generic models, taking advantage of the inheritance capabilities of semantic networks.

The generic models in the map and image domain describe the mapping of the scene in the map and image domain, respectively. Besides the common properties inherited from the generative model, they reflect particularities of the representations in the map and image domain. The generic model in map domain does not contain for example the concept *car* as a part of the scene since cars are not represented in the map; the generic model in the image domain however contains this concept. Both generic models contain methods necessary for performing the analysis in their respective domain. These methods comprise functions for feature extraction, for the calculation of attribute values from image or map data and functions for evaluating the preliminary and final analysis results.

We use only the generic model in the map domain for analysis purposes. 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 can be performed with higher reliability in the map domain than in the image domain when using such an unspecific model like the generic model. The analysis process in map domain gives us a complete description of the scene in the limits of the map contents.

The scene description in map domain is automatically combined with the generic model in the image domain resulting in a specific model for analysis in the image domain. The differences between the two models are getting obvious when looking at the *part-of* hierarchy of the generic model in the image domain (Fig. 4) and the *part-of* hierarchy of the specific model (Fig. 5). The name of each concept in the specific model is derived from the name of the concept in the generic model, to which the number of the corresponding instance in the scene description in map domain was appended. Whilst the generic model contains statements of the form:

- buildings are optional parts of the scene, or
- a building may have inner courts,

the specific model of the actual scene to be analyzed contains precise information about the objects of the scene, e.g.:

- (big)building number 0113 is rectangularly shaped and the coordinates of corner points are such and such.

The specific model is now used for image analysis. Because of the concrete modeling of the predicted objects in the scene, we expect to obtain more reliable results using the specific model for image analysis than using the generic model. Since the specific model is automatically generated, expense for the system designer is not increased. The result of the image analysis process is a description of the image, in which for each object of the map, features extracted from the image are measured and their match with the model is evaluated.

4 Image analysis

Before the model based image analysis is started, primitives are extracted from the image data. At present time, we are using line segments as primitives and assume that these primitives sufficiently describe the image for our purposes. The procedure used for extracting the primitives is described in [9].

Analysis in our approach can be interpreted as a model driven search for specified structures represented by a subset of primitives. The analysis starts by creating a modified concept for the analysis goal. In the case of Fig. 5, the analysis goal is the concept *i_scene0270*, representing the whole scene. A modified concept is a preliminary result before the correspondence between a concept and an instance is actually established. It reflects constraints for the concept that have been determined out of the context of the current analysis state. An instance is a particular manifestation of a concept in the sensor data. The process of creating or changing modified concepts is called expansion; establishing a correspondence between a concept and an instance is called instantiation. Starting with the analysis goal and following the hierarchy in the semantic network, stepwise the concepts on lower hierarchical levels are expanded until a concept on the lowest level is reached. For this concept instantiation can now be performed.

Expansion phase and instantiation phase alternate in the analysis process. The control module determines according to the current state of analysis by means of task independent rules which phase to follow next. If the necessary conditions for both phases are fulfilled, instantiation has priority. An instantiation can be performed when for all parts of a concept instances have already been found. Thus, our method is a structural, hierarchical method, where in a step by step process complex structures are composed from less complex structures.

In the expansion phase, constraints can be propagated bottom-up (data-driven) or top-down (model-driven). Bottom-up propagated constraints restrict the search for correspondences and hence contribute to the efficiency of the analysis. Top-down propagated constraints can be interpreted as hypotheses generated by a complex object for its parts. By checking these hypotheses, the complex object is revised. After each expansion or instantiation, an evaluation of the involved modified concepts and instances takes place.

Interpreting analysis as a search, one can represent this process graphically by a search tree. The nodes of the tree represent the current state of analysis and they contain complete information about the analysis cycle from the root node to the current node. With progressing analysis the instances and modified concepts in the nodes emulate the semantic network in the knowledge base. If a correspondence between a concept and several instances is possible, the search tree is splitted: for each hypothesis a new node as successor of the current node is created.

The analysis process continues with that leaf node of the search tree, which is considered to be the best according to a task dependent evaluation. It is know that the problem of finding an optimal path in the search tree can be solved by the A^*-algorithm [8]. Its application is possible if one can evaluate the path from the root node to the current node and can give an estimate for the valuation of the path from the current node to the (not yet known) terminal node containing the solution.

One faces the problem to estimate the valuation of the path from the current node to the solution node (future successful path). The valuation has to offer a common ground for both the comparison of paths developed to the same level and for the comparison of paths advanced in the building of the solution with paths abandoned earlier. The valuation is difficult to establish since one has very little knowledge about the future successful path: at a given state of analysis, it is not known how many nodes the path from the actual node to the solution node will contain.

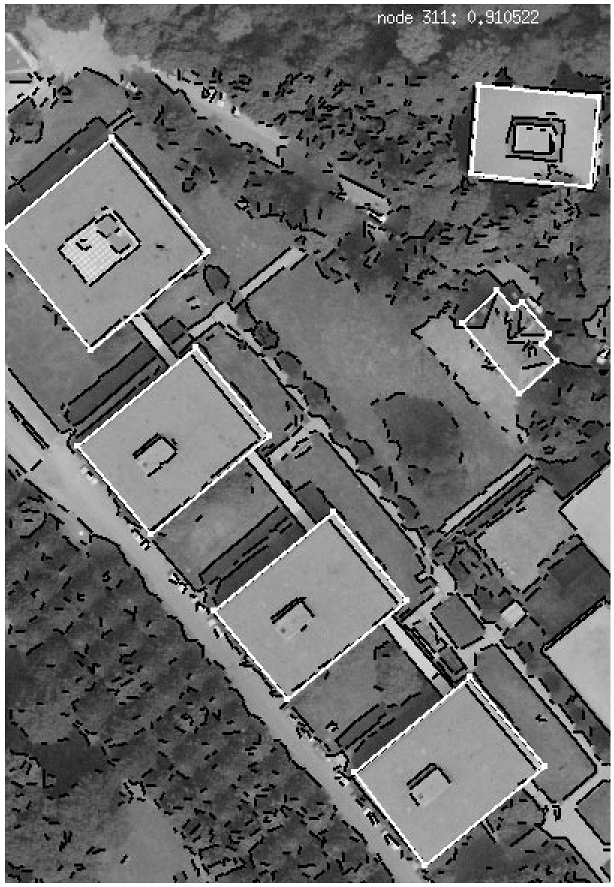

Figure 6: Result of the model based verification. Dark lines: image primitives; white lines: solution.

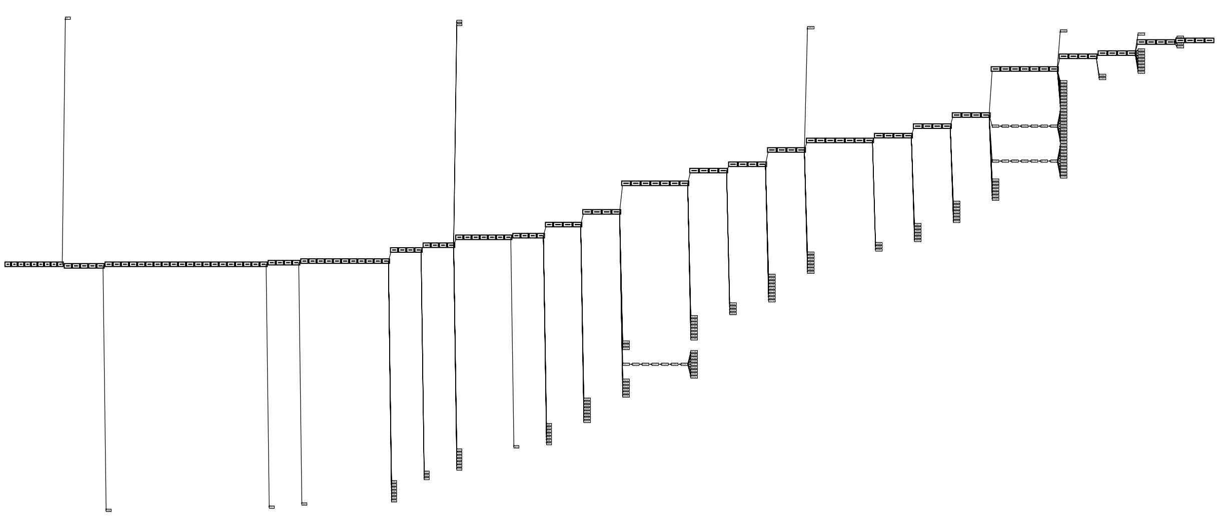

Figure 7: Search tree build during the verification of the objects in Fig. 6

Due to noisy image data it is also not known, how many instances will have to be created until the solution node is reached since for example a line segment in the model may be broken into several line segments in the image.

In our image analysis problem we relate the valuation of the search path to the analysis goal. The valuation of the analysis goal is calculated considering the valuations of the instances and modified concepts already created and the estimates for the valuations of the instances and modified concepts which will be created in the future successful path. The computed valuations are measures of our subjective belief in the model fidelity and certainty of the instances and modified concepts. The valuations are propagated in our structural hierarchical model and combined to result in the valuation of the analysis goal by using Dempster-Shafer's theory of evidence.

The solution of the verification for some of the buildings of the scene in Fig.1 (image) and Fig. 2 (map) is given in Fig. 6. The displayed dark lines are the primitives extracted from of the image; the instances belonging to the solution node form the white polygons. The computed valuation for the solution node amounted in this example to 0.91 in a range between 0 and 1. The tree build by the A^*-algorithm to find the solution for the objects in Fig. 6 is represented in Fig. 7.

5 Conclusion and future work

For the purpose of change detection we cooperatively use large scale aerial images and maps. The maps are used to build a specific model for the scene to be analysed. Because of the concrete modeling of the predicted objects in the scene, more reliable results can be obtained than in the case of using generic models. However, expense for the system designer is not increased since the specific models for various scenes are automatically generated.

In the current development stage of the system only the objects represented in the map are verified. We are extending our system to recognize objects in the image, which are not represented in the map and for which a specific model is thus not available.

References

[1] J.P. Agin. Knowledge–based detection and classification of vehicles and other objects in aerial images. In *Proceedings of the DARPA Image Understanding Workshop*, pages 66–71, Palo Alto, CA, April 1979.

[2] N.V. Findler. *Associative Networks*. Academic Press, Orlando, 1979.

[3] F. Kummert, H. Niemann, R. Prechtel, and G. Sagerer. Control and explanation in a signal understanding environment. *Signal Processing*, 32:111–145, 1993.

[4] T. Matsuyama and V. Hwang. *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Advances in Computer Vision and Machine Intelligence. Plenum Press, New York, London, 1990.

[5] D.M. McKeown, W.A. Harvey, and J. McDermott. Rule based interpretation of aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(5):570–585, September 1985.

[6] B. Nicolin and R. Gabler. A knowledge–based system for the analysis of aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 25(3):317–329, Mai 1987.

[7] H. Niemann, G. Sagerer, S. Schröder, and F. Kummert. ERNEST: A Semantic Network System for Pattern Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):883–905, September 1990.

[8] N. Nilsson. *Principles of artificial intelligence.* Springer-Verlag, Berlin, 1982.

[9] F. Quint and H.-P. Bähr. Feature extraction for map based image interpretation. In X. Shi, D. Du, and W. Gao, editors, *Third Internartional Colloquium of LIESMARS: Integration, Automation and Intelligence in Photogrammetry, Remote Sensing and GIS*, pages 1–8, Wuhan, China, October 1994.

[10] F. Quint and M. Sties. Map–based semantic modeling for the extraction of objects from aerial images. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 307–316, Basel, 1995. Birkhäuser.

[11] F. Sandakly and G. Giraudon. Multispecialist system for 3D scene analysis. In A. Cohn, editor, *11th European Conference on Artificial Intelligence, ECAI 94*, pages 771–775. John Wiley & Sons, Ltd., 1994.

[12] U. Stilla and A. Hajdu. Map-aided structural analysis of aerial images. In H. Ebner, C. Heipke, and K. Eder, editors, *Proceedings of the ISPRS Commision III Symposium*, München, 1994.