AIDA: A SYSTEM FOR THE KNOWLEDGE BASED INTERPRETATION OF REMOTE SENSING DATA *

C.–E. Liedtke, J. Bückner, O. Grau, S. Growe, R. Tönjes
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover
Division: Automatic Image Interpretation, Prof. Dr. Ing. C.–E. Liedtke
Appelstr. 9A, D–30167 Hannover, Germany
Phone: +49 511 762–5318, Fax: +49 511 762–5333, WWW: http://www.tnt.uni–hannover.de

ABSTRACT

Because of the increasing amount of remotely sensed imagery there is a growing need for efficient data analysis techniques. Here to automate the image interpretation a knowledge based approach is suggested.

The presented scene interpretation system AIDA uses semantic nets for the explicit representation of the prior knowledge about objects expected in the scene. It exploits the knowledge base to generate a scene description assigning symbolic meanings to the image primitives. The information of a GIS database is used as partial interpretation to produce reliable hypotheses for the expected objects. This initial scene description is verified consecutively in the remote sensing imagery. Multiple sensors can be investigated simultaneously. The explicit knowledge representation eases the adaptation to different tasks.

The system was tested successfully in applications like verification of GIS data, recognition of complex structures, the automatic search of tie points for the registration of remotely sensed images, and the object specific 3D modelling of landscapes and buildings.

1.0 INTRODUCTION

The recognition of land use changes for map updating and environmental and agricultural monitoring represents a major topic of remote sensing. Due to the large amount of acquired data algorithms for the automatic extraction of objects from sensor data are investigated. This contribution suggests a knowledge based approach for image interpretation using semantic nets. The presented scene interpretation system AIDA generates a symbolic scene description which can be used for recognition tasks (Koch 1997) as well as for 3D reconstruction of the detected objects (Grau 1997, Tönjes 1996) or automatic registration of multi sensor data (Growe 1997).

A lot of scene interpretation systems have been developed in the past. They differ in a number of aspects like the control strategies, the knowledge representation or the application domain. Concerning the representation of the scene knowledge a lot of systems for aerial image interpretation like SPAM (McKeown 1985) use rules. Production rules and semantic nets are found in MESSIE (Clément 1993), while ERNEST (Niemann 19990) uses semantic nets only. SIGMA (Matsuyama 1990) is organized in three expert modules using frames and rules. In AIDA the knowledge about the scene objects is formulated in a semantic net, while the control knowledge is represented by rules.

2.0 SYSTEM OVERVIEW

Figure 1 shows the architecture of the knowledge based scene interpretation system AIDA. The system is designed for the interpretation of images, for example remote sensing imagery like aerial images

* Presented at the Third International Airborne Remote Sensing Conference and Exhibition, 7–10 July 1997, Copenhagen, Denmark.

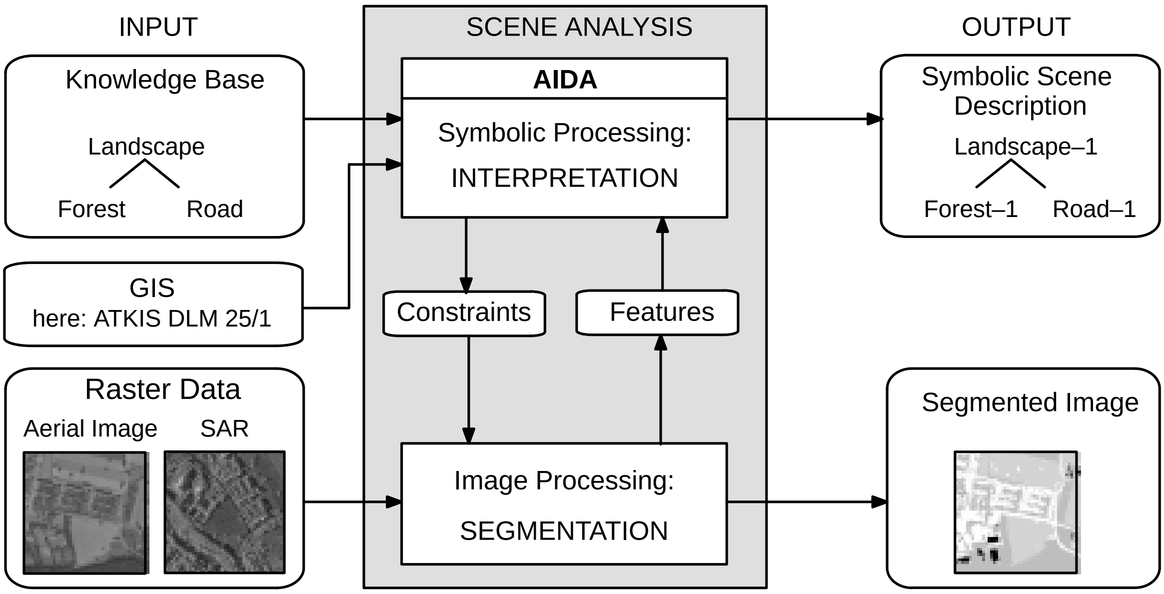

Figure 1. Architecture of the Knowledge Based Scene Interpretation System AIDA

or SAR data. The prior knowledge about the objects to be extracted from the image data is represented explicitly in the knowledge base. Besides this general knowledge about the objects the interpretation takes advantage of a GIS database. We use the German digital landscape model DLM 25/1 of ATKIS (Authoritative Topographic Cartographic Information System) which mirrors the content of the 1:25000 map. It contains the geographic location and some selected properties of the objects.

From the GIS a partial interpretation of the scene is derived. However the GIS my be out–dated. Hence the partial interpretation constitutes a hypothesis which is verified in the sensor data by generating constraints for the features expected in the image. The image processing module extracts features that meet the constraints given by the interpretation module. It returns the found primitives which are again evaluated by the interpretation module resulting in new hypotheses for image features. Using this mixed top–down and bottom–up strategy the system generates a symbolic description of the observed scene.

For the generic description of objects the scene knowledge is formulated in a semantic net and in computation functions. The procedural knowledge employed for control is defined by rules. The execution of the rules is controlled by an inference engine. The rule based formulation results in a flexible system with exchangeable rules. The strategies can easily be adapted to different applications. The next section describes the knowledge representation and the control of interpretation used in AIDA. Some selected applications of AIDA are treated in chapter 4.

3.0 LANGUAGE FOR IMAGE INTERPRETATION

Structural knowledge, like knowledge about the relationships between the objects and their connection to the features apparent in the image data, can be represented efficiently by semantic nets. The application specific knowledge base has to be defined by the user. The system AIDA provides a network language – described below – and a graphical user interface for the formulation of the application knowledge.

3.1 SEMANTIC NET

Semantic nets are a special kind of attributed graphs. They consist of nodes and links connecting them. The knowledge base describes the scene objects, their properties, parts, specializations, and other

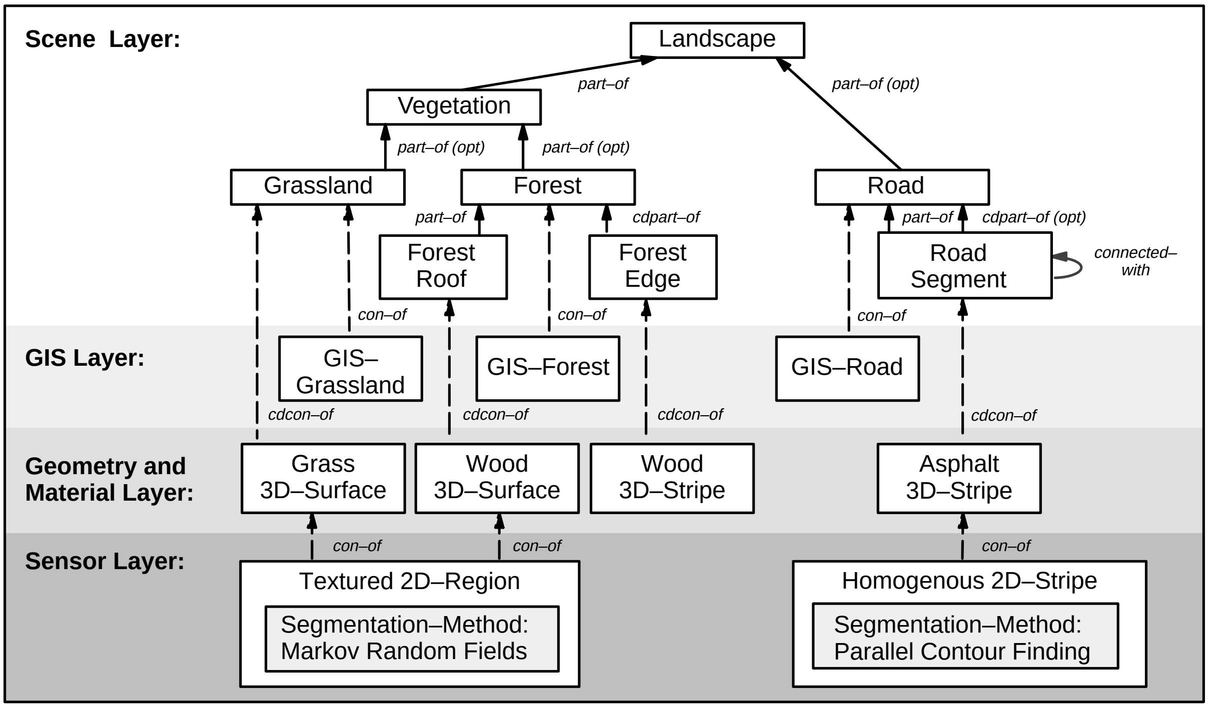

Figure 2. Simplified semantic net representing generic models of landscape objects

relations as a generic model (see figure 2). This semantic net in conjunction with the GIS data is exploited to generate a specific semantic net describing the scene represented by the remotely sensed images. While, for example, the knowledge base contains only one generic model of a road, the scene description includes as many roads as detected in the sensor data.

3.1.1 Nodes

The syntax of the semantic net distinguishes between two types of nodes: concepts and instances. The concepts describe the predefined generic model of the objects as mentioned above. The instances are realizations of the concept C(n) in the observed scene. During interpretation four different states of object recognition are distinguished: hypotheses $I_H(n)$, missing instances $I_M(n)$, partial instances $I_P(n)$ and complete instances $I_C(n)$. Interpretation starts with hypotheses which are initialized with the contents of the concept. After verification in the sensor data hypotheses are successively changed to partial instances and complete instances. Falsification in the sensor data results in missing instances. Complete instances possess all obligatory components of the object. Partial instances describe a predecessor state that contains already all obligatory subnodes that are not context dependent.

3.1.2 Links

The nodes are related by different types of edges or links. The instances are connected via *instance–of* to their concepts. The specialization of an object is described by the *is–a* link. This link type introduces the concept of inheritance. Objects are composed of parts, indicated by the *part–of* link. Parts that are not obligatory are marked as optional by the *part–of (opt)* link. By detecting the single components object search can be reduced to a more simple task. Furthermore some parts may only be detectable if other parts have already been found and thus established a certain context. These context dependent parts are modelled by *cdpart–of*. For example it makes only sense to search for forest edges in the context of forests. Objects can often be detected based on their geometric or photometric appearance in the image data. This transformation of an abstract symbol to its concrete realization is represented by the concrete–of link, abbreviated *con–of*. The con–of link allows to distinguish between different conceptual layers, e.g. scene, GIS, 3D–geometry/material, and sensor (see fig. 2). To control for example the evaluation of different sensor layers a context dependent *cdcon–of* link is introduced. The parallel investigation of multi sensor data, e.g. aerial image and SAR data, may improve the quality of the interpretation result. It can be realized easily by defining a new sensor layer in the semantic net.

The *data–of* link establishes a relation to the features segmented in the image data. A present data–of link indicates that the object was segmented directly in the sensor data by image processing operators. Geometric or photometric relations between objects can be represented explicitly by *attributed–relations*. This link type contains attributes which restrict the attributes of the related object. While the is–a, part–of, and con–of relations propagate information top–down or bottom–up the attributed relation propagates information mainly horizontally. For example a further *Road–Segment* adjacent to an existing one can be hypothesized via the *connected–with* link in fig. 2.

3.1.3 Attributes

The object properties relevant for the interpretation are modelled by attributes. An attribute of *Road–Segment* and *Asphalt–3D–Stripe* in fig. 2 might be the width in meters, while the corresponding attribute of *Homogenous–2D–Stripe* is the width in image pixels. Each attribute contains an attribute value which has to be measured in the sensor data or calculated bottom–up from inferior instances. The attribute range represents the expected attribute value. This range is initialized by the generic model, i.e. the attribute range defined for the concept by the user. During interpretation the attribute range is restricted more and more to get a more reliable hypothesis for the object properties. A more sharp attribute range is computed top–down by propagating the information from superior nodes. For example the width of a *Homogenous–2D–Stripe* representing a road segment in an aerial image with known resolution can be estimated by transforming the common width of a road into image pixels. Having detected a first road segment the measured width might serve as more accurate prediction of the width of the adjacent road segment. To model uncertainties the attributes are described by minimum and maximum values and ranges.

A judgement function computes the degree of conformance between predicted attribute range and measured attribute value. It returns a compatibility value and a certainty in the interval [0;1]. The judgement function of the node summarizes all attribute judgements of an instance node using the normalized weighted sum of the attribute compatibility values by default.

3.1.4 Computation Functions

The procedural knowledge for the calculation of attribute ranges and values, for the judgement of attributes and nodes, and for the binding of nodes and image primitives is stored in computation functions. These functions are methods of the corresponding objects and must be able to handle uncertain arguments and results. A *Homogenous–2D–Stripe*, for example, is extracted from the image data by an algorithm searching for parallel contours, while a *Textured–2D–Region* is segmented using Markov Random Fields (see fig. 2). The arguments needed by the functions have to be described generally by a special path grammar. At run time the arguments are filled with the current values by following the defined paths through the semantic net. Because these computation functions are strongly application dependent they have to be defined a priori by the user.

3.2 CONTROL OF INTERPRETATION

The control knowledge, i.e. the knowledge how and in which order scene interpretation has to proceed, is formulated in a set of predefined rules. An inference engine determines the sequence of rule execution. A rule is composed of a condition and an action part. The condition checks for a new interpretation state of neighboured nodes in the semantic net. The action part adapts the interpretation state of the focused node accordingly. Furthermore the attributes are computed based on the new information.

The control strategy is defined by the rules and their associated priorities. According to the number of link types only a small set of rules is required. The predefined rules can be grouped into rules for instantiation, hypothesis propagation, specialization, and binding. If the condition of a rule matches for a node n of the instance net, the inference engine instantiates the rule and adds it to the conflict set. To select one rule the instantiated rules are ranked according to their priority and the position of the nodes in the semantic net. After selection of a rule its action part is executed, which modifies the scene description by establishing new links or changing the state of a node. The rules returns one ore more modified nodes.

By changing the priority of the rules or by inserting a new rule the user is able to create an application specific interpretation strategy. The default strategy generates iteratively hypotheses in a model driven manner propagating knowledge top–down until the lowest nodes of the generic model are reached. These hypotheses are consecutively verified or falsified in the image data.

The current scene description is stored in a search node which contains all concepts and instances with their interpretation state. If a rule returns more than one new node, they have to be regarded as competing interpretations and the search node splits into child search nodes. The leaves of the resulting search tree represent all currently possible interpretations. Each search node is judged by summarizing the judgements of all included instance nodes. An A*–algorithm selects the most promising search node for further investigation. The interpretation process stops if a complete instance of a predefined goal concept was generated.

4.0 APPLICATIONS

The interpretation system AIDA is used for the 3D modelling of buildings from close–range images (Grau 1997) and landscapes from aerial images (Tönjes 1996) where the explicitly formulated knowledge is used for the selection of geometrical constraints for surface reconstruction. Additionally it is employed to verify the GIS data in aerial images and maps (Koch 1997). Two further applications of AIDA, the automatic search of tie points for image registration and the recognition of complex structures, are described in the following.

4.1 AUTOMATIC SEARCH OF TIE POINTS FOR IMAGE REGISTRATION

Prerequisite for the registration of multi sensor and multi temporal remotely sensed imagery in a common cartographic coordinate system is the detection of corresponding points in the sensor data and the map. For precise rectification the tie points must be distributed equally and detected accurately. The proposed approach assumes a coarse orientation of the image to be registered known from parameters like flight height, course, and image resolution. The exact image to map registration is performed in a blockwise manner. For each image block a GIS database, here ATKIS, is asked for objects suitable to represent a tie point. Here, roads are chosen which are grouped to crossroads by the interpretation system AIDA since crossroads themselves are not included in the used GIS. However, due to the flexible knowledge representation the system can be extended easily to other object classes representing a possible tie point.

The simplified knowledge base is shown in fig. 3a. The semantic net consists of three conceptual layers describing the scene, the GIS and the sensor specific objects respectively. If the aerial image layer is exchanged by a SAR layer, the registration of SAR images can be performed without costly adaptations.

Interpretation starts by adding the GIS data returned from the database into the initial scene description resulting in instances of *GIS Data Object*. Generating top–down hypotheses for *Tie Point,*

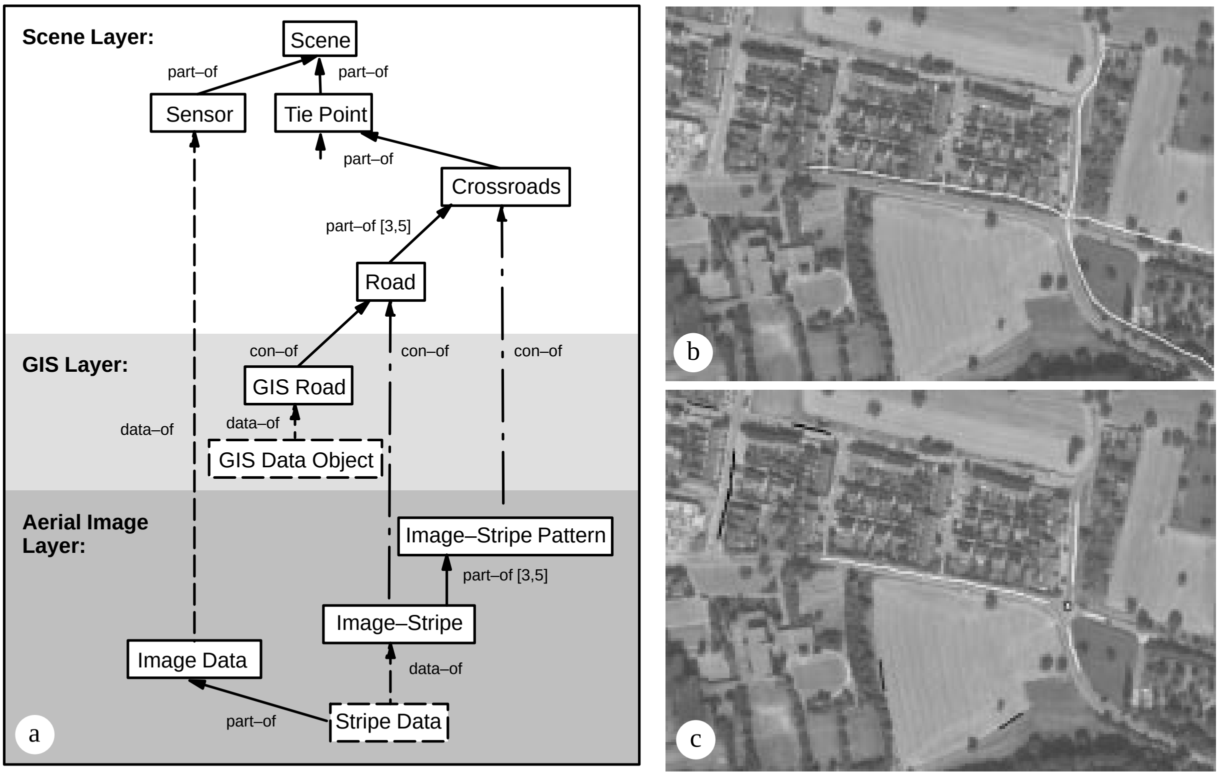

Figure 3. Knowledge base for tie point search (a) with expected (b) and detected roads (c)

Crossroads, *Road*, and finally *GIS–Road* the GIS data is connected to an instance of *GIS–Road*. Because crossroads consists of at least three roads the system instantiates further roads until no *GIS Data Object* can be bound any more. The intersecting point of all roads, taken as tie point, is calculated from the 3D coordinates known from the GIS. The aim of the tie point matching is to find the corresponding pixel representing the crossroads in the sensor data. Therefore the single roads have to be detected in the image.

Generating a hypothesis for an *Image–Stripe* the 3D position of the first road is projected into the image coordinate system using the assumed coarse image orientation (see fig. 3b). This prediction restricts the search space for the following segmentation process significantly. Furthermore, the knowledge about the photometric appearance of the objects in the specified sensor is used for the configuration of the image processing algorithms. The segmentation module returns one or more road candidates meeting best the expectations of the system. Competing interpretations are handled by the mentioned A*–algorithm (see chapter 3.2). Consecutively all roads are verified in the image data resulting in a complete instance of *Image–Stripe*, they are grouped to an *Image–Stripe–Pattern*, and the intersection of the polylines is computed which represents the image point corresponding to the crossroads. Figure 3c shows the segmented (black) and the selected (white) road candidates as well as the computed 2D tie point. The found pair of corresponding points is one of the control points needed later for the automatic rectification process.

For road segmentation we developed an algorithm which searches for parallel contours in the gradient image. As mentioned above it is configured at run–time by the expectation of the intermediate scene description, i.e. the predicted position of the road, road width, expected luminance value, and a search depth for the contour tracking. At first the gradient magnitude and direction, quantized into eight values, is computed by a Sobel operator. The gradient image is searched for neighboured peaks which have a distance less than the given road width. If the corresponding grey level values in–between are inside the given luminance range, the candidates for a double contour are taken as starting points for the following tracking

Figure 4. Segmented roads in an aerial image (left) and the corresponding SAR image (right)

algorithm. The parallel contours are tracked perpendicular to the gradient direction. The algorithm tries to generate contiguous double contours as long as possible by investigating the adjacent pixels for both end points. Using a best search with predefined search depth the optimal path through the gradient image is computed considering the constraints like road width and luminance. The central axis of the found road is extracted, short gaps between adjacent road segments are closed, and the resulting road candidates are approximated by a polyline. These polylines are returned to the interpretation system which selects one or more candidates as representation of a hypothesized image stripe. Segmentation results for an aerial image and a SAR image of the same area are shown in fig. 4.

4.2 RECOGNITION OF COMPLEX STRUCTURES

Many complex structures like purification plants or airports can not be detected in the aerial image directly because they are composed of several parts which are related in a specific way. A purification plant, for example, consisting of houses and sedimentation tanks, has usually a road access and is located close to a river to drain off cleaned water. This knowledge can be formulated explicitly in a semantic net using

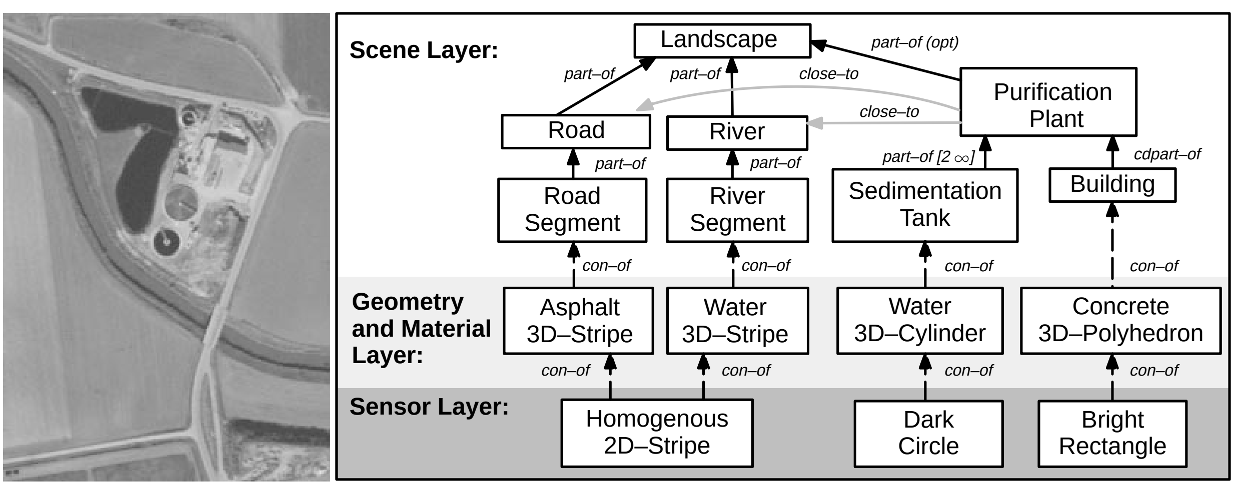

Figure 5. Aerial Image of a purification plant and the corresponding semantic net

part–of hierarchies and attributed relations (see fig. 5). The complex task of detecting a purification plant is simplified to the detection of objects like buildings and rivers which must have a specified constellation.

After the instantiation of roads and rivers spatial reasoning via the *close–to* link generates a hypothesis of a *Purification Plant* which has to be located next to a river. Thus the search area for the purification plant can be restricted. Successively hypotheses for two *Sedimentation Tanks* and *3D–Water–Cylinders* are built resulting in the search for dark circles in the aerial image. Buildings, represented by bright polygons in the image, are detected in the same way. The *Purification Plant* can be instantiated completely if all obligatory components have been found.

5.0 CONCLUSIONS

The scene interpretation system AIDA was presented, which uses semantic nets, rules, and computation functions to represent the declarative and the procedural knowledge needed for the interpretation process. Controlled by an adaptable interpretation strategy the knowledge base is exploited to derive a symbolic description of the observed scene in form of an instantiated semantic net. In remote sensing applications the information of a GIS database is used as initial interpretation increasing the reliability of the generated hypotheses.

The system is employed for the 3D reconstruction of buildings and landscapes as well as for the automatic image registration, the recognition of complex structures, and the verification of GIS data and maps. The first results show that the knowledge based scene interpretation is a promising approach in the field of image understanding suitable to solve the problems addressed.

6.0 REFERENCES

- Clément, V., Giraudon, G., Houzelle, S., Sandakly, F., ”Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 4, July 1993.
- Grau, O., ”A Scene Analysis System for the Generation of 3–D Models”, In *Proceedings of the International Conference on Recent Advances in 3–D Digital Imaging and Modeling*, Ottawa, Ontario Canada, May 1997.
- Growe, S, Tönjes, R, ”A Knowledge Based Approach to Automatic Image Registration”, submitted for *International Conference on Image Processing*, Santa Barbara, California, October 1997.
- Koch, H., Pakzad, K., Tönjes, R., ”Knowledge Based Interpretation of Aerial Images and Maps Using a Digital Landscape Model as Partial Interpretation”, In *Proceedings of the Semantic Modeling Workshop SMATI*, Bonn, Germany, May 1997.
- Matsuyama, T., Hwang, V.S.–S., ”SIGMA : A Knowledge–Based Aerial Image Understanding System”, *Plenum Press*, New York 1990.
- McKeown, D. M. Jr., Harvey, W. A. Jr., McDermott, J., ”Rule–Based Interpretation of Aerial Imagery”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–7, No. 5, pp. 570–585, Sept. 1985.
- Niemann, H., Sagerer, G., Schröder, S., Kummert, F., ”ERNEST: A Semantic Network System for Pattern Understanding”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 9, Sept. 1990.
- Tönjes, R.**,** ”Control of Scene Reconstruction Using Explicit Knowledge”, In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, pp. 15–20, Nov. 1996.