A KNOWLEDGE BASED APPROACH TO SENSOR FUSION APPLIED TO MULTISENSORY AND MULTITEMPORAL IMAGERY *

S. Growe[1], C.–E. Liedtke[1], K. Pakzad[2]

[1]Institute of Communication Theory and Signal Processing
University of Hannover, Appelstr. 9A, D–30167 Hannover, Germany

[2] Institute for Photogrammetry and Engineering Surveys
University of Hannover, Nienburger Straße 1, D–30167 Hannover, Germany

E–Mail: growe|liedtke@tnt.uni–hannover.de, pakzad@ipi.uni–hannover.de

ABSTRACT

The increasing amount of remotely sensed imagery from multiple platforms requires efficient analysis techniques. The leading idea of the presented work is to automate the interpretation of multisensor and multitemporal remote sensing images by the use of common prior knowledge about landscape scenes. In addition the system can use specific map knowledge of a GIS, information about sensor projections and temporal changes of scene objects. The prior knowledge is represented explicitly by a semantic net. A common concept has been developed to distinguish between the semantics of objects and their visual appearance in the different sensors considering the physical principle of the sensor and the material and surface properties of the objects. In this presentation the basic structure of the system and its use for sensor fusion on different structural and functional levels is presented. Results are shown for the extraction of roads from multisensor images. The approach for a multitemporal image analysis is illustrated for the monitoring of moorland areas and the extraction of an industrial fairground.

1.0 INTRODUCTION

The recognition of complex patterns and the understanding of complex scenes from remote sensing images requires in many cases the use of multiple sensors and views taken at different time instances. For this purpose sensors such as optical, thermal, and radar (SAR) are used. In order to automate the processing of these sensor signals new concepts for sensor fusion are needed. There are various parameters which have to be considered during fusion like the different platform locations and orientations, the different spectral bands, the sensing geometry, the spatial resolutions, the weather conditions, the time of the day, the seasons, etc. In addition the reliability can considerably be increased by exploiting prior knowledge about the scene like spatial relations, which might be available from maps or a geographic information system (GIS), functional relations like the relations of buildings, rails, streets etc. in rural, urban or industrial areas or time–relations like the seasons affecting the visual appearance of land use in agricultural areas.

Due to the great variety of scenes to be interpreted a modern system for image analysis should be adaptable to new applications. This flexibility can be achieved by a knowledge based approach where the application dependent knowledge is strictly separated from the control of information processing. In the literature various approaches to image interpretation have been presented. Most interpretation systems like

* Presented at the Fourth International Airborne Remote Sensing Conference and Exhibition/ 21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada, 21–24 June 1999.

SPAM (McKeown, 1985) and SIGMA (Matsuyama, 1990) use a hierarchic control and construct the objects incrementally using multiple levels of detail. Inspired by ERNEST (Kummert, 1993) the presented system AIDA formulates prior knowledge about the scene objects with semantic nets. In the following the system architecture is described and a common concept is presented to distinguish between the semantics of objects and their visual appearance in the different sensors considering the physical principle of the sensor and the material and surface properties of the objects. The necessary extensions to provide a multitemporal image analysis are described and illustrated for two applications.

2.0 KNOWLEDGE BASED INTERPRETATION SYSTEM

For the automatic interpretation of remote sensing images the knowledge based system AIDA [(Liedtke, 1997),(Tönjes, 1998a)] has been developed. The prior knowledge about the objects to be extracted is represented explicitly in a knowledge base. Additional domain specific knowledge like GIS data can be used to strengthen the interpretation process. From the prior knowledge hypotheses about the appearance of the scene objects are generated which are verified in the sensor data. An image processing module extracts features that meet the constraints given by the expectations. It returns the found primitives – like line segments – to the interpretation module which assigns a semantic meaning to them, e.g. *road* or *river.* The system finally generates a symbolic description of the observed scene. In the following the knowledge representation and the control scheme of AIDA is described briefly.

2.1 KNOWLEDGE REPRESENTATION

The knowledge representation is based on semantic nets. Semantic nets are directed acyclic graphs and they consist of nodes and edges in between. The nodes represent the objects expected in the scene while the edges or links of the semantic net form the relations between these objects. Attributes define the properties of nodes and edges.

The *nodes* of the semantic net model the objects of the scene and their representation in the image. Two classes of nodes are distinguished: the *concepts* are generic models of the object and the *instances* are realizations of their corresponding concepts in the observed scene. Thus, the knowledge base which is defined prior to the image analysis is built out of concepts. During interpretation a symbolic scene description is generated consisting of instances. The object properties are described by *attributes* attached to the nodes. They contain an attribute value which is measured bottom–up in the data and a range which represents the expected attribute value. The range is predefined and/or calculated during the interpretation. For each attribute a value and range computation function has to be defined. A judgement function computes the compatibility of the measured value with the expected range.

The relations between the objects are described by *edges* or *links* forming the semantic net. The specialization of objects is described by the *is–a* relation introducing the concept of inheritance. Along the *is–a* link all attributes, edges and functions are inherited to the more special node which can be overwritten locally. Objects are composed of parts represented by the *part–of*–link. Thus the detection of an object can be reduced to the detection of its parts. The transformation of an abstract description into its more concrete representation in the data is modelled by the *concrete–of* relation, abbreviated *con–of*. This relation allows to structure the knowledge in different conceptual layers like for example a scene layer and a sensor layer. Topological relations provide information about the kind and the properties of neighboured objects. Therefore the class of *attributed relations* (*attr–rel*) is introduced. In contrast to other edges this relation has attributes which can be used to constrain the properties of the connected nodes. For example a topological relation *close–to* can be generated which restricts the position of an object to its immediate neighbourhood. The initial concepts which can be extracted directly from the data are connected via the *data–of* link to the primitives segmented by image processing algorithms.

2.2 CONTROL OF THE SCENE ANALYSIS

To make use of the knowledge represented in the semantic net control knowledge is required that states how and in which order scene analysis has to proceed. The control knowledge is represented explicitly by a set of rules. The rule for instantiation for example changes the state of an instance from *hypothesis* to *complete instance*, if all subnodes which are defined as obligatory in the concept net have been instantiated completely. If an obligatory subnode could not be detected, the parent node becomes a *missing instance*.

An inference engine determines the sequence of rule execution. Whenever ambiguous interpretations occur they are treated as competing alternatives and stored in the leaf nodes of a search tree. Each alternative is judged by comparing the measured object properties with the expected ones. The judgement calculus models imprecision by fuzzy sets and considers uncertainties by distinguishing the degrees of necessity and possibility. The judgements of attributes and nodes are fused to a judgement of the whole interpretation state. The best judged alternative is selected for further investigation. Using a mixed top–down and bottom–up strategy the system generates model–driven hypotheses for scene objects and verifies them consecutively in the data. Expectations about scene objects are translated into expected properties of image primitives to be extracted from the sensor data. Suitable image processing algorithms are activated and the semantic net assigns a semantic meaning to the returned primitives.

3.0 KNOWLEDGE BASE FOR THE INTERPRETATION OF REMOTE SENSING IMAGERY

For an object extraction only those features are relevant that can be observed by the sensor and that give a hint for the presence of an object to be extracted. Hence the knowledge base contains only the necessary and visible object classes and properties. The network language described in chapter 2 is used to represent the prior knowledge by a semantic net. In Figure 1 the generic model for the interpretation of remote sensing images is shown. It is divided into the *3D scene domain* and the *2D image domain*. The 3D scene domain splits into the *semantic layer* and the *physical layer.* If a geoinformation system (GIS) is available and applicable an additional *GIS layer* can be defined representing the scene specific knowledge from the GIS. The 2D image domain contains the *sensor layers* adapted to the current sensors and the *data layer.*

For the objects of the *2D image domain* general knowledge about the sensors and methods for the extraction and grouping of image primitives like lines and regions is needed. The primitives are extracted by image processing algorithms and they are stored in the semantic net as instances of the concepts *Line Data* or *Region Data* respectively. Due to fragmentation the lines and regions have to be grouped according to perceptual criteria like continuity, nearness, similarity etc.. A continuous *Stripe* for example is represented in the semantic net by a composition of neighboured *SubStripes*. The sensor layer can be adapted to the current sensor type like SAR, IR or visual camera. For a multisensor analysis the layer is duplicated for each new sensor type to be interpreted assuming that each object can be observed in all the images (see Fig. 1). All information of the 2D image domain is given related to the image coordinate system. As each transformation between image and scene domain is determined by the sensor type and its projection parameters the transformations are modelled explicitly in the semantic net by the concept *Sensor* and its specializations for the different sensor types.

The knowledge about inherent and sensor independent properties of objects are represented in the *3D scene domain* which is subdivided into the physical, the GIS and the semantic layer. The physical layer contains the geometric and radiometric properties as basis for the sensor specific projection. Hence it forms the interface to the sensor layer(s). The semantic layer represents the most abstract layer where the scene objects with their symbolic meanings are stored. The semantic net eases the formulation of hierarchical and topological relations between objects. Thus it is possible to describe complex objects like a purification plant as a composition of sedimentation tanks and buildings close to a road and a river where the cleaned water is drained off. The symbolic objects are specified more concrete by their geometry and material. In

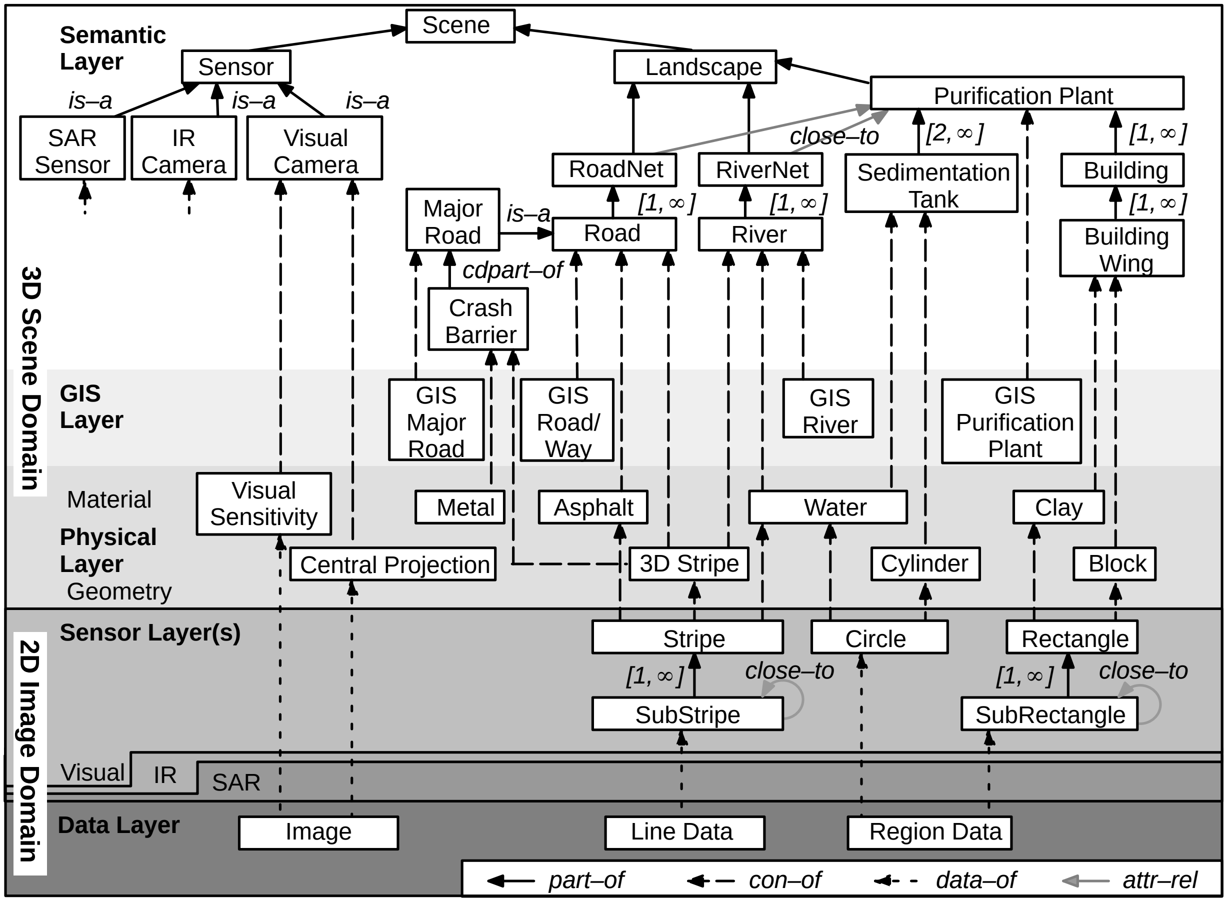

Figure 1. Semantic net representing a generic model of a purification plant and its relation to the image data

conjunction with the known sensor type the geometrical and radiometrical appearance of the objects in the image can be predicted. This prediction can be improved if GIS data of the observation area is available. Though the GIS may be out of date it represents a partial interpretation of the scene providing semantic information. Hence the GIS objects are connected directly with the objects of the semantic layer (Fig. 1).

4.0 INTERPRETATION OF MULTISENSORY IMAGES

The automatic analysis of multisensor imagery requires the fusion of sensor data. The presented concept to separate strictly the sensor independent knowledge of the 3D scene domain from the sensor dependent knowledge in the 2D image domain eases the integration and simultaneous interpretation of images from multiple sensors. New sensor types can be introduced by simply defining another specialization of the *Sensor* node with the corresponding geometrical and radiometrical transformations. According to the images to be interpreted the different sensor layers (SAR, IR, Visual) are activated.

For the application of road extraction the advantages of a multisensor image analysis are illustrated in Fig. 2. Using only the aerial image (a) or the infrared image (b) yields to fragmented results. If both images are analyzed simultaneously the gaps can be closed. In those areas where both images provide a hint for a road segment the reliability of the interpretation is increased. In other areas the information from the images complement each other. Other examples for the fusion of multisensor images are given in (Toenjes, 1998b).

Figure 2. Sensor Fusion demonstrated on the aerial view of a purification platn: Rejected (thin lines) and accepted (wide lines) road features from (a) visual and (b) infrared image with (c) fusion result.

5.0 INTERPRETATION OF MULTITEMPORAL IMAGES

Currently the system is being extended for the interpretation of multitemporal images. Applications like change detection and monitoring require the analysis of images from different acquisition times. By comparing the current image with the latest interpretation derived from the preceding image land use changes and new constructions can be detected. In the following the necessary extensions to a multitemporal analysis with the system AIDA are described and illustrated with help of two applications.

5.1 EXTENSION OF THE KNOWLEDGE BASED SYSTEM

The easiest way to generate a prediction for the current image from an existing scene interpretation is to assume that nothing has changed during the elapsed time. But in many cases humans have knowledge about possible or at least probable temporal changes. Hence the knowledge about possible state transitions between two time steps should be exploited in order to increase the reliability of the scene interpretation.

Temporal changes can be formulated in a so called *state transition graph* where the nodes represent the temporal states and the edges model the state transitions. To integrate the graph in a semantic net the states are represented by concept nodes which are connected by a new relation: the *temporal relation*. For each temporal relation a priority can be defined in order to sort the possible successor states by decreasing probability. As states can either be stable or transient the corresponding state transitions differ in their transition time which can also be specified for the temporal relation. For the exploitation of the temporal knowledge a time stamp is attached to each node of the semantic net which documents the time of its instantiation. As normally no knowledge about the temporal changes of geometrical objects or materials is available the state transition diagram is part of the semantic layer (compare Fig. 1). In contrast to hierarchical relations like *part–of* or *con–of* the start and end node of temporal relations may be identical – forming a loop – to represent that the state stays unchanged over the time.

During the interpretation process the state transition diagram is used by a new inference rule. Analysis starts with the first image of the given sequence marked with time stamp t_1. If a state of the state transition diagram can be instantiated completely, the temporal knowledge is used to hypothesize one or more possible successors of this state for the next image in the chronological order (time stamp t_2). The system selects all successor states that can be reached within the elapsed time $t_2–t_1$ according to the transition times defined in the temporal relations. States which are multiple selected due to loops in the transition diagram are eliminated. The possible successor states are sorted by decreasing priority so that the most probable state is investigated first. All hypotheses are treated as competing alternatives represented in separate leaf nodes of the search tree (see chapter 2.2). Starting with the alternative of the highest priority the hypotheses for the successor state are either verified or falsified in the current image. For continuous monitoring the time stamps of the instances can be used to remove the old nodes of t_1.

5.2 MONITORING OF MOORLAND AREAS

The monitoring of moorland areas is an example for a multitemporal analysis because it shows that some moorland states can not be interpreted correcly without the knowledge of state transitions. Since the probable state transitions in moorland areas are limited due to restrictions by law and nature it is possible to describe and include them into the concept net by means of temporal relations.

The relevant moorland areas are divided into five different states of land use: *Forest*, *Grassland*, *Area of Degeneration*, *Area of Regeneration* and *Area of Peat Working*. The prior knowledge about the states and their structural description are formulated in a concept net (see Fig. 3). In the *scene layer* the different land use states are described with their obligatory parts. The nodes in the *sensor layer* represent the depiction of the scene layer nodes in aerial images using structural and radiometric features. Every node of the sensor layer has access to a special *segment analysis operator* which verifies the meaning of the node for a particular moor segment. The states *Area of Regeneration* and *Area of Degeneration* have partially a similar structural description, also the states *Area of Regeneration* and *Grassland*. Their distinction in aerial images from one acquisition time alone is very difficult or in some cases impossible (especially in images without color information). It is necessary to use aerial images from different acquisition times.

Because for the most moorland areas in Germany a biotope mapping exists, a corresponding label image is used to describe the segments of time stamp t_1. Based on this initial interpretation for every node describing a moor segment by a particular state several hypotheses are created along the temporal relations in order to predict the meaning of the moor segments from acquisition time t_2. The verification of the different state transition hypotheses are processed according to the defined priority values. In case of an acceptable verification result the other competitive hypotheses are not verified anymore.

In Fig. 4 an example is shown for two aerial images of a moorland area near Hannover. The reduction of the search space for the possible successor states leads to a correct interpretation of the segments. For segment 1 a transition from *Area of Peat Working* to *Area of Regeneration* is stated using the

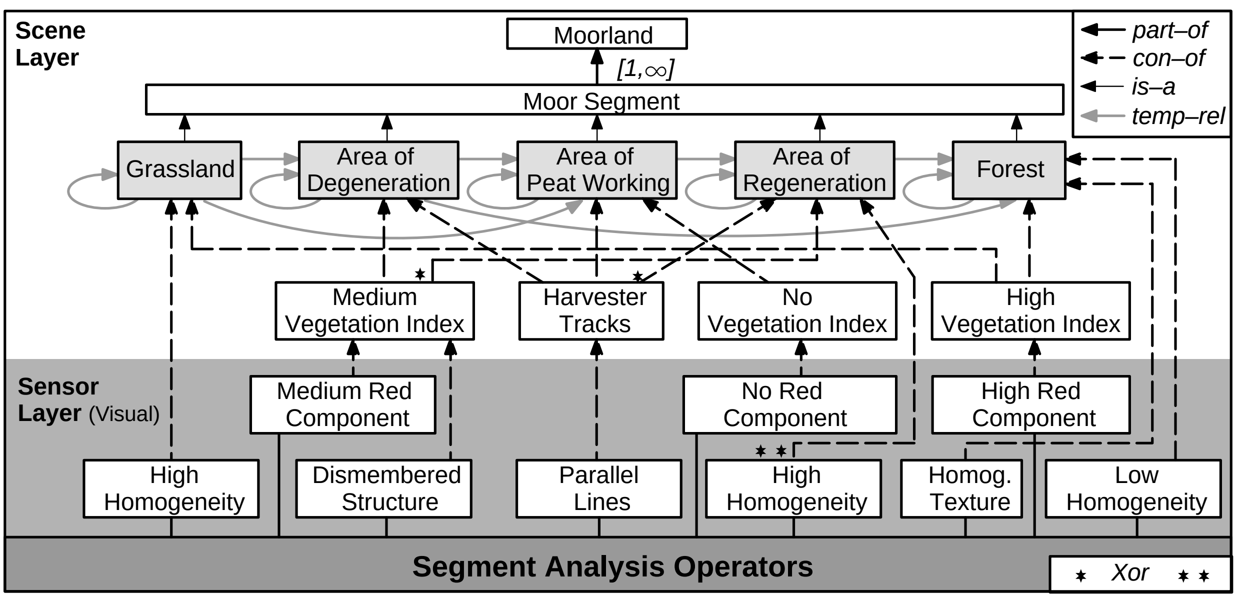

Figure 3. Semantic net and state transition graph for the interpretation of moorlands

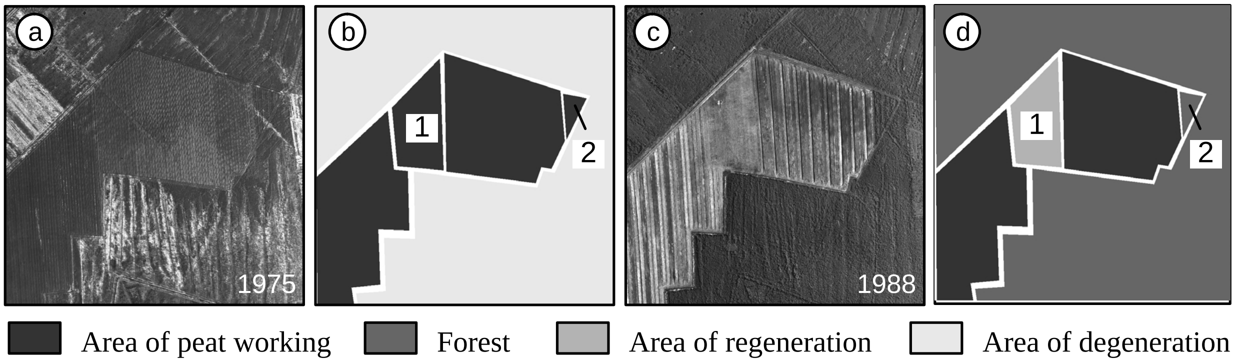

Figure 4. Aerial images of a moorland area dated (a) 1975 and (c) 1988 with corresponding interpretation results

knowledge about the previous land use of the segment. Without using this prior information the system could not distinguish the state *Grassland* and *Area of Regeneration* because both states are characterized by a *High Homogeneity.* For segment 2 the land use class changed from *Area of Peat Working* to *Forest* although there is no direct state transition between them represented in the semantic net. Due to the elapsed time of 13 years the state of regeneration was not observed. But using the knowledge about mean transitions times the system also generates a hypothesis for *Forest* as direct successor of *Area of Peat Working* which could be verified successfully for this example.

5.3 EXTRACTION OF AN INDUSTRIAL FAIRGROUND

An industrial fairground is an example for a complex structure detectable by a multitemporal image interpretation only. Using a single image it would be classified as an industrial area consisting of a number of halls. But during several weeks of the year some unnormal activity can be observed: exhibition booths are constructed, visitors throng to the site, and the booths are dismantled again. This knowledge can be exploited for the automatic extraction of a fairground and formulated in a semantic net (see Fig. 5). The different states of a fairground are represented by the concepts *FairIdle, FairConstruction, FairActive,* and *FairDismantling*. The construction, active and dismantling phase are transient compared to the state *FairIdle*. Therefore transition times of four to eight days are defined for the corresponding temporal relations. Additionally the node *FairIdle* is looped back to itself.

The analysis starts with the first image looking for an *Industrial Area*. In the given example the system searches for at least three halls and one parking lot. If the *Industrial Area* can be instantiated completely the system tries to refine the interpretation by exchanging the *Industrial Area* by a more special concept. There are four possible specializations (*FairIdle* to *FairDismantling*) and the search tree splits into four leaf nodes. Each hypothesis is tested in the image data. A construction or dismantling phase is characterized by trucks near the halls which keep the equipment for the booths. Hence the system searches for bright rectangles close to the halls. An active fair can be recognized by parking lots filled with cars and – if the image accuracy is sufficient – by persons walking on the fairground.

If one of the four states can be verified the temporal inference is activated. The system switches to the next image in the sequence and generates hypotheses for the successor state. According to the elapsed time and considering the transition times all possible successors are determined. If for example the time step between the two images was two weeks, it is possible that *FairIdle* follows immediately after *FairActive* omitting the dismantling phase. All hypothesized successor states are represented in separate leaf nodes and are treated as competing alternatives. Having found hints for all obligatory states a complete instance of *Industrial Fairground* can be generated and the interpretation goal is reached. The presented approach is currently being tested for a sequence of five aerial images of the Hannover fairground – the future Expo 2000 exhibition ground.

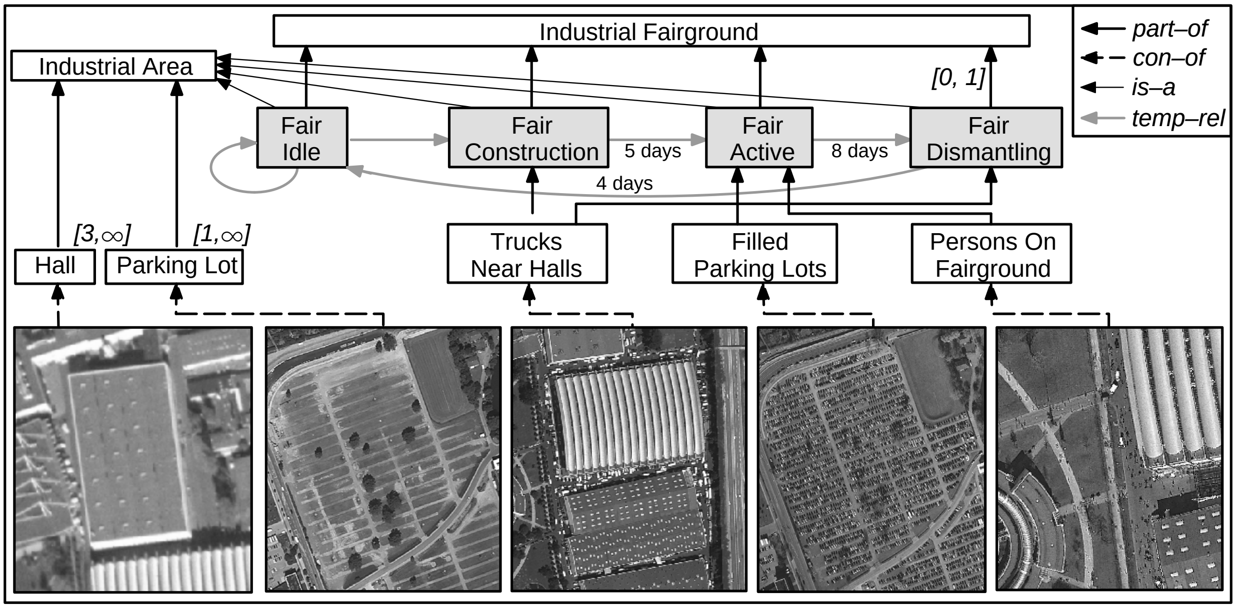

Figure 5. Simplified semantic net for the extraction of an industrial fairground introducing temporal relations

6.0 CONCLUSIONS

A knowledge based scene interpretation system called AIDA was presented, which uses semantic nets, rules, and computation methods to represent the knowledge needed for the interpretation of remote sensing images. 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. If available the information of a GIS database is used as partial interpretation increasing the reliability of the generated hypotheses. The system is employed for the automatic recognition of complex structures from multisensor images. Currently extensions are made in order to provide a multitemporal analysis. The use of knowledge about temporal changes improves the generation of hypotheses for succeeding time instances and allows for example the detailed interpretation of moorland areas or the extraction of complex structures like an industrial fairground. The knowledge based scene interpretation is a promising approach in the field of image understanding suitable to solve the problems addressed.

7.0 REFERENCES

[Kummert, 1993], F. Kummert, H. Niemann, R. Prechtel, G. Sagerer.: ”Control and explanation in a signal understanding environment”, *Signal Processing*, Vol. 32, No. 1–2, May 1993.

[Liedtke, 1997], Liedtke, C.–E., Bückner, J., Grau, O., Growe, S., Tönjes, R.: ”AIDA: A System for the Knowledge Based Interpretation of Remote Sensing Data”, *3rd. Int.. Airborne Remote Sensing Conference and Exhibition*, Copenhagen, Denmark, July 1997.

[Matsuyama, 1990], Matsuyama, T., Hwang, V.S.–S., ”SIGMA : A Knowledge–Based Aerial Image Understanding System”, *Plenum Press*, New York 1990.

[McKeown, 1985], 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.

[Toenjes, 1998a], Toenjes, R., ”Wissensbasierte Interpretation und 3D–Rekonstruktion von Landschaftsszenen aus Luftbildern”, *Dissertation*, University of Hannover, 1998.

[Toenjes, 1998b], Toenjes, R., Growe, S., ”Knowledge Based Road Extraction from Multisensor Imagery”, *ISPRS Symposium ”Object Recognition and Scene Classification from Multispectral and Multisensor Pixels”*, Columbus, Ohio, USA, July 1998.