International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7–4–3 W6, Valladolid, Spain, 3–4 June, 1999

KNOWLEDGE BASED INTERPRETATION OF MULTISENSOR AND MULTITEMPORAL REMOTE SENSING IMAGES

Stefan Growe

Institute of Communication Theory and Signal Processing, University of Hannover
Appelstrasse 9a, D–30167 Hannover, Germany
WWW: http://www.tnt.uni–hannover.de/~growe
E–mail: growe@tnt.uni–hannover.de

KEY WORDS: Knowledge Based Image Interpretation, Semantic Net, Sensor Fusion, Multitemporal Image Analysis.

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. The presented system is able to use specific map knowledge of a geoinformation system (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 within the knowledge base 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 the analysis of multitemporal images is illustrated for the interpretation of an industrial fairground.

KURZFASSUNG

Um die immer größer werdende Menge an Fernerkundungsbildern bearbeiten zu können, werden in zunehmendem Maße effiziente Auswerteverfahren benötigt. Die Kernidee der vorliegenden Arbeit ist es, die Interpretation von multisensoriellen und multitemporalen Luftbildern durch die Nutzung von Vorwissen über die Landschaftsobjekte zu automatisieren. Das vorgestellte System ist in der Lage, spezifisches Kartenwissen eines Geoinformationssystems, Informationen über Sensorabbildungen und über zeitliche Veränderungen der Szenenobjekte für die Auswertung zu nutzen. Das Vorwissen wird explizit in einem semantischen Netz abgelegt. Es wurde ein allgemeines Konzept entwickelt, um innerhalb der Wissensbasis zwischen Objektsemantik und visueller Abbildung in den verschiedenen Sensoren zu unterscheiden, wobei sowohl das physikalische Prinzip des Sensors als auch die Material– und Oberflächeneigenschaften der Objekte berücksichtigt werden. In diesem Beitrag werden die Grundstruktur des Systems und dessen Nutzung für die Sensorfusion auf verschiedenen strukturellen und funktionalen Ebenen erläutert. Beispielhaft werden Ergebnisse für Extraktion von Straßen aus multisensoriellen Bildern präsentiert. Weiterhin wird ein Ansatz für die Analyse von multitemporalen Bildern vorgestellt und am Beispiel der Interpretation eines Messegeländes illustriert.

1. INTRODUCTION

The automatic extraction of objects from aerial images for map updating and environmental monitoring represents a major topic of remote sensing. However, the results of low–level image processing algorithms like edge detectors are in general incomplete, fragmented, and erroneous. To overcome these problems, a scene interpretation is performed which assigns an object semantic to the features segmented in the remote sensing image. Prior knowledge about the objects should be used to constrain the object parameters and to reduce the uncertainty of the interpretation. To increase or decrease the reliability of competing interpretations, structural relationships of the objects could be exploited.

A partial interpretation already exists for most landscapes: the map corresponding to the observed scene. Due to the growing availability of geographic information systems (GIS), the map data can be accessed by computers directly and is therefore usable for the automatic interpretation of aerial images.

For remote sensing, different sensors such as optical, thermal, and radar (SAR) have been developed which collect different image data of the observed scene. The wish to extract more information from the data than it is possible using a single sensor system alone raises the question of sensor fusion. Several parameters influence the data fusion: the different platform locations, the different spectral bands (optical, thermal, or microwave), the sensing geometry (e.g. perspective projection or SAR geometry), the spatial resolution, and the season at image acquisition. State–of–the–art–systems must be able to combine information from different sensors.

Especially for environmental monitoring, it is necessary to investigate images from different acquisition times to study the development of the observation area. The quality of a scene interpretation can be increased by using the information from preceding images. Hence, it becomes possible to distinguish for example between the construction and the dismantling of buildings or between the regeneration and degeneration of moorland areas [Pakzad, 1999]. To realize such a multitemporal analysis, an interpretation system must be able to administrate images from different time instances and to represent and exploit information about possible or at least probable temporal changes.

The leading idea of this work is to automate the evaluation of aerial images of complex scenes using prior knowledge about the object structure, GIS, sensor type, and temporal changes. To ease the adaptation of the analysis system to new requirements and the extension to future tasks, the knowledge is represented explicitly and is separated from system control. Such a so–called knowledge based approach constitutes the focal point of this work.

In the literature various approaches to image interpretation and sensor fusion have been presented. Only a few authors try to formalize the representation of the objects and sensors, and the control of the information integration. Most interpretation systems like SPAM (McKeown, 1985) and SIGMA (Matsuyama, 1990) use a hierarchic control and construct the objects incrementally using multiple levels of detail. The system MESSIE (Clement, 1993) models the objects explicitly distinguishing four views: geometry, radiometry, spatial context, and functionality. It employs frames and production rules. In the BPI system (Stilla, 1997) a net of production rules representing a part–of–hierarchy describes the structural prior knowledge. A blackboard realized by an associative memory is used for process communication. Another blackboard–based architecture is suggested by Mees (1998). He distinguishes between strategy knowledge represented by an AND/OR–tree, global knowledge described by sensor–independent fuzzy production rules, and sensor–dependent local knowledge stored in attributed prototypes and image processing operators called local detectors.

ERNEST (Kummert, 1993) uses semantic nets to exploit the object structure for interpretation. The MOSES system extends the ERNEST approach to extract man–made objects from aerial images (Quint, 1997). The presented system AIDA (Liedtke, 1997) adopts the idea to formulate prior knowledge about the scene objects with semantic nets. In addition, the control knowledge is represented explicitly by rules which are selected by an inference engine.

In the following, the system architecture of AIDA 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 in chapter 5.

2. KNOWLEDGE BASED INTERPRETATION SYSTEM

For the automatic interpretation of remote sensing images, the knowledge based system AIDA (Liedtke, 1997; Tönjes, 1999) 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.

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 have a value measured in the data and a range describing the expected attribute value. During instantiation the attribute range of the instance is taken from the corresponding concept and – if possible – is restricted further by the information of instantiated parent nodes. For example, an already detected street segment can constrain the position of the adjacent segment. For both attribute value and attribute range a computation method can 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 relations, this one 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.

For the efficient representation of multiple relations, the minimum and maximum number of edges can be defined in the knowledge base. The minimum quantity describes the number of obligatory relations and the difference to the maximum quantity represents the number of optional relations between objects. In this way, it can be easily modelled that for example a crossroad consists of three up to five intersecting roads. Additionally, for each edge a priority can be defined in order to realize an ordered evaluation of the relations. Edges with high priority are instantiated first. For the application of landscape analysis for example, it can be guaranteed that the streets are extracted prior to the rivers.

Some relations appear exclusively in certain domains. For example roads have always a lane but they have pavements in urban areas only. This fact is taken into consideration by a *domain* dependent relation in the generic model. Fig. 1 shows a simple semantic net for a generic model of a *Road Net* which is defined as a composition of at least one *Road*, illustrated by the set [1, ∞]. A *Road* consists of one or two lanes. Its specialization *Major Road* inherits the properties of *Road* and possesses an additional *Crash Barrier.* For the *part–of* relation between pavement and road the domain *Urban Scene* is defined. Only in urban scenes this relation is valid and the system searches for pavements. All the initial objects *Crash Barrier*, *Lane*, and *Pavement* are represented by a *Stripe–Form* in the image.

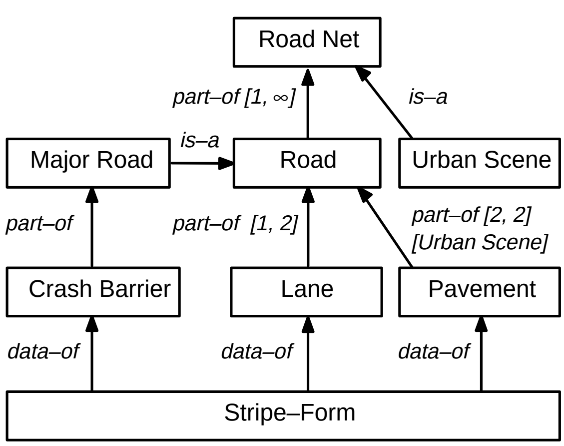

Figure 1. Example for a semantic net: The scene contains at least one *Road*. The *Pavemen*t is defined for the domain *Urban Scene*. The more special concept *Major Road* inherits the properties of *Road*. All objects are represented by a *Stripe–Form* in the image.

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 according to a given strategy. A strategy contains a set of rules out of the rule base. For each valid rule, a priority is defined to determine in which order the rules are tested. The first matching rule is fired. The user can modify the interpretation strategy by changing the priorities and by removing or inserting rules to the current strategy. The default strategy prefers a model–driven interpretation with a data–driven verification of hypotheses. Topological relations are instantiated as soon as possible to realize a spatial reasoning.

Whenever ambiguous interpretations occur, for example if more than one suitable image primitive is found for a hypothesis, 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 (Dubois, 1988; Tönjes, 1999). The judgements of attributes and nodes are fused to a judgement of the whole interpretation. The best judged alternative is selected for further investigation.

Starting at the root node of the concept net, 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 the corresponding 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 in a data–driven way. Interpretation stops, if a given goal concept is instantiated completely or no further rule of the current strategy can be fired.

3. KNOWLEDGE BASE FOR THE INTERPRETATION OF REMOTE SENSING IMAGERY

For object extraction, only those features are relevant that can be observed by the sensor and that give a hint for the presence of the object to be extracted. Hence, the knowledge base contains only the necessary and visible object classes and properties. The network language described in chapter 2.1. is used to represent the prior knowledge by a semantic net. In Figure 2 a 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

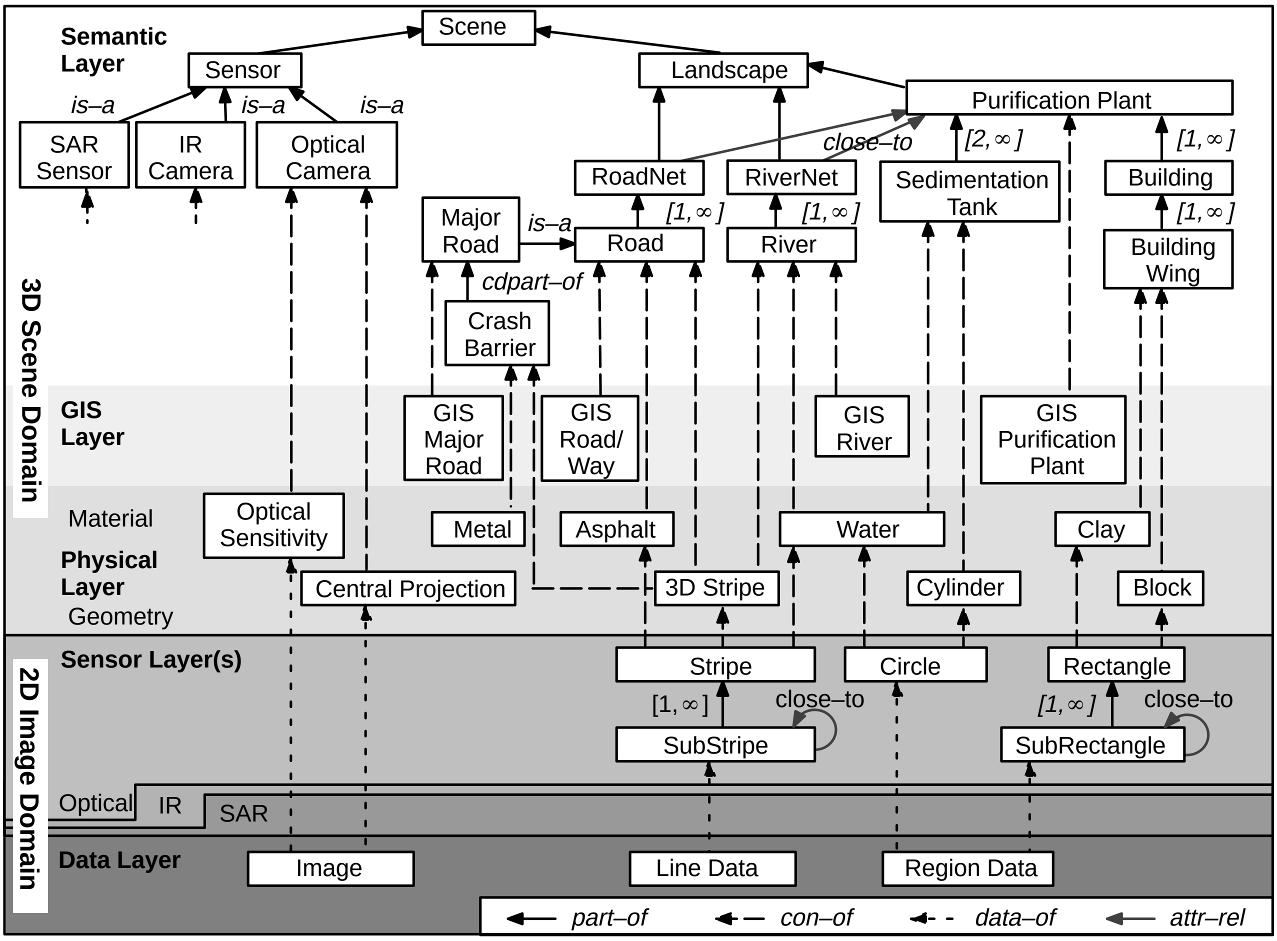

Figure 2. Semantic net representing a generic model of a purification plant and its relation to the image data.

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 neighbouring *SubStripes*. The sensor layer can be adapted to the current sensor type like SAR, IR or optical sensor. 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. 2). 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 (see Fig. 2). The symbolic objects are specified more concretely by their geometry and material. In 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.

4. INTERPRETATION OF MULTISENSOR IMAGES

The automatic analysis of multisensor images 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

Figure 3. Sensor Fusion demonstrated on the aerial view of a purification plant: Rejected (thin lines) and accepted (wide lines) road features from (a) optical and (b) infrared image with (c) fusion result.

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, optical) are activated.

For the application of road extraction, the advantages of a multisensor image analysis are illustrated in Fig. 3. Using only the aerial image (a) or the infrared image (b) yields 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 (Tönjes, 1998).

5. INTERPRETATION OF MULTITEMPORAL IMAGES

Currently, the system is being extended for the interpretation of multitemporal images. Applications like change detection and environmental 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. Preliminary results are shown for the extraction of an industrial fairground.

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. The latest scene interpretation represented in an instantiated semantic net is therefore regarded as a kind of GIS and it is used to guide the analysis of the current image. The objects found in the last image are verified in the current one but changes are difficult to detect and to explain. However, 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 transition graph in a semantic net the states are represented by concept nodes which are connected by a new relation: the *temporal relation* (see Fig. 4). 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 be also specified in the temporal relation. 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. 2). 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 time. Figure 4 shows a simple example of a state transition graph consisting of three states. For the different transitions priorities and transition times are defined.

To exploit the temporal knowledge, a time stamp is attached to each instance of the semantic net which documents the time of its instantiation. Thus, it will be possible to filter time slices out of the semantic net. The possible time stamps are given by the

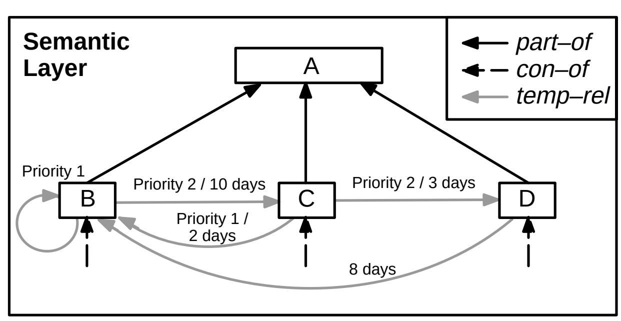

Figure 4. State transition graph represented by concepts of a semantic net. To each temporal relation a priority and a transition time can be assigned.

acquisition times of images to be interpreted. Images acquired simultaneously – eventually from different sensors – are collected in a *sensor group*. All sensor groups are sorted in their chronological order to ease the proper administration of the image sequence.

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 selected many times, 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 rejected in the current image. For continuous monitoring, the time stamps of the instances can be used to remove the old nodes of t_1.

In the example of Fig. 5, a successor state for the complete instance of $B(t_1)$ is determined which can be reached within a time step of 14 days. According to the knowledge base of Fig. 4, the states *B*, *C*, and *D* are possible. The successor *B* can be reached either via the loop *B–B* or via the path *B–C–B*, but identical solutions are considered only once. The node *C* is reached via the transition *B–C* and the successor *D* by following the path *B–C–D* omitting the intermediate state *C*. All three possibilities are treated as competing alternatives and the search tree splits into three leaf nodes. The system prefers the best judged node according to the possibilistic approach mentioned in chapter 2.2. If the judgements are similar the alternative of the highest transition priority is chosen.

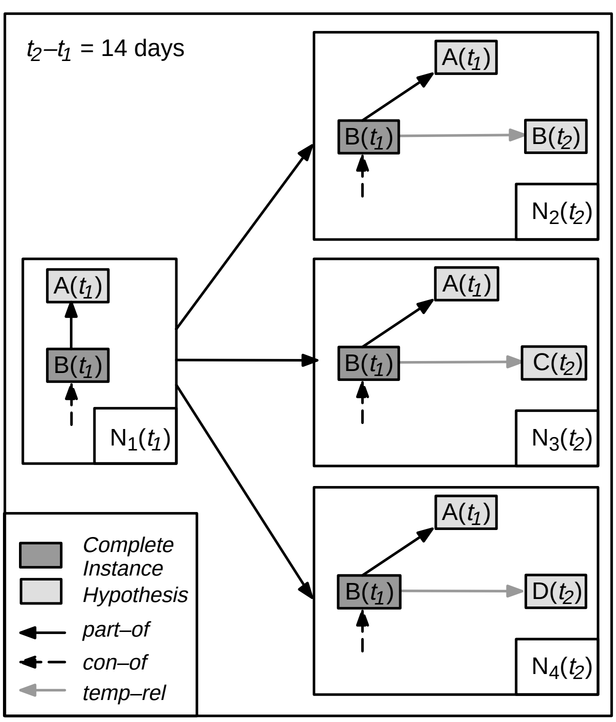

Figure 5. Search tree of the interpretation process according to the knowledge base in Fig. 4: Assuming a time step of 14 days the possible successor states of *B* are *B*, *C*, or *D*. Hence, the search tree splits into three leaf nodes $N_2(t_2)$ to $N_4(t_2)$.

5.2. 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. The special use of this industrial area can not be detected from one time instance alone. However, during several weeks of the year some unnormal activity can be observed: exhibition booths are constructed, crowds of people visit the site, and the booths are dismantled again. In aerial images the different phases can be recognized by full or empty parking lots, or vehicles or people on the fairground respectively. This knowledge can be exploited for the automatic extraction of a fairground and formulated in a semantic net for a multitemporal image analysis (see Fig. 6). The different states of a fairground are represented by the concepts *Fair Idle, Fair Construction*, *Fair Active*, and *Fair Dismantling*. The construction, active and dismantling phase are transient compared to the state *Fair Idle*. Therefore, transition times of four to eight days are defined for the corresponding temporal relations. Additionally, the node *Fair Idle* is looped back to itself.

The analysis starts with the first image of the sequence looking for an *Industrial Area*. In the given example, the system searches for at least three halls and one parking lot. These objects are represented in the image by regions of special geometric and radiometric properties. Halls for example are in most cases right-angled polygons. To verify the hypotheses suitable image processing algorithms are activated. Segmented regions that meet best the expectations are chosen, others are rejected. 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 (*Fair Idle* to *Fair Dismantling*) and the search tree splits into four leaf nodes. Each hypothesis is tested in the image data. Normally any cars are prohibited on the fairground. But during the construction or dismantling phase there are trucks near the halls which keep the equipment for the booths. Hence, the system searches for small bright rectangles close to the halls. An active fair can be recognized by parking lots filled with cars and – if the image resolution is sufficient – by persons walking on the fairground (see images in Fig. 6).

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 *Fair Idle* follows immediately after *Fair Active* omitting the dismantling

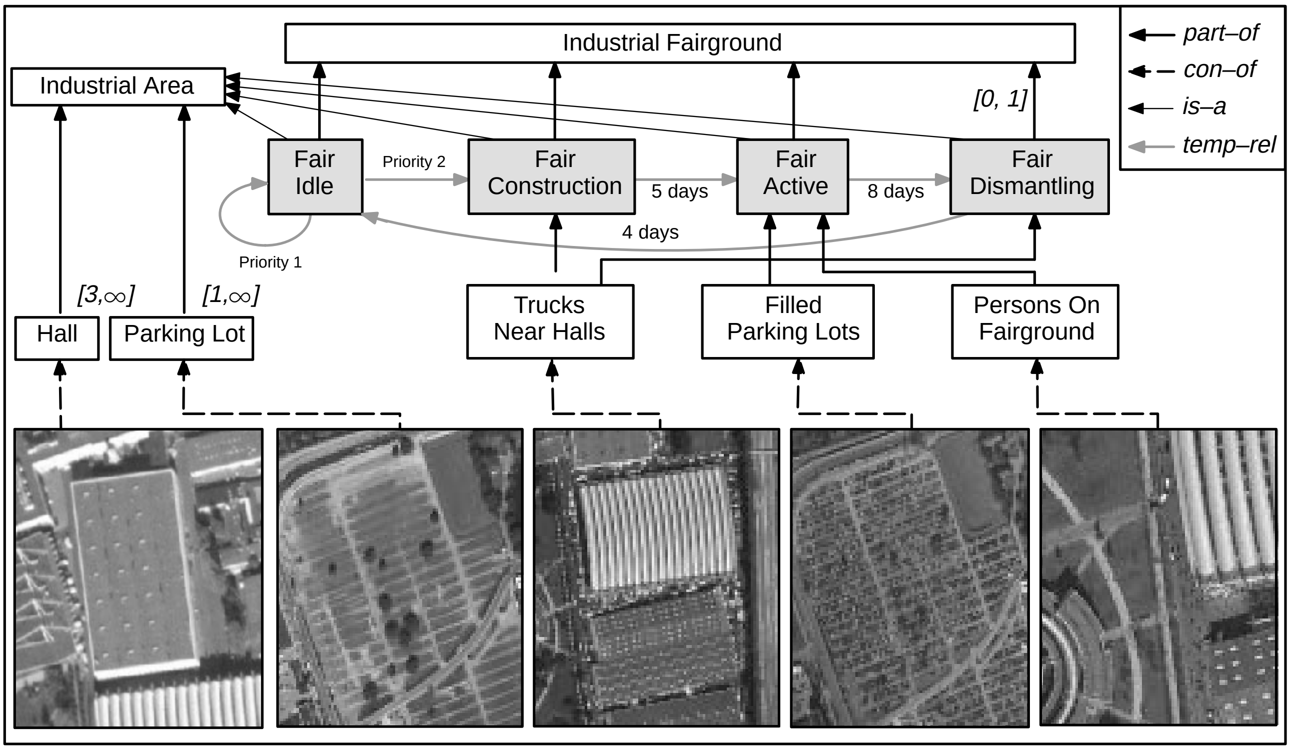

Figure 6. Simplified semantic net for the extraction of an industrial fairground introducing temporal relations. The more concrete representations of the objects (physical and sensor layer of the semantic net) are illustrated here by corresponding image regions.

phase. Having found hints for all obligatory states, a complete instance of *Industrial Fairground* can be generated and the interpretation goal is reached.

5.3. Results

The presented approach is currently being tested for a sequence of five aerial images (three colour, two greyscale images) of the Hannover fairground – the future Expo 2000 exhibition ground. The images cover the years 1995 to 1998 and show a construction/dismantling phase in 1997 and an ongoing fair in 1998. Furthermore, the construction of three new halls can be observed. Unfortunately, no continuous sequence of aerial images exists which depicts all phases of a single fair. But the given images are suitable to simulate the whole cycle. They were coregistered and resampled in resolution pyramids of 0.5 to two meters per pixel to permit the segmentation of both large halls and small vehicles with minimal processing effort.

The multitemporal features described in chapter 5.1. were integrated in the AIDA system and the semantic net illustrated in Fig. 6 was implemented. Additionally, a number of special image processing algorithms are necessary to realize the aspired application. Halls and parking lots have to be segmented, trucks, persons and cars must be detected on the fairground and the parking lots respectively. Figure 7 shows preliminary results for a colour image of 1997.

The aerial image in Fig. 7a was classified at a resolution of two meters per pixel using a Maximum–Likelihood operator which exploits all available image bands. The classifier considers the classes of the 3x3–neighbourhood to modify the prior probability of the current pixel. This results in a more homogenous classification result and suppresses small noisy regions. The training regions were defined manually in the first image of the sequence and stayed the same for the following images. Figure 7b shows the classification result for the class *Hall* (marked grey). From all candidates the semantic net chose the white ones to be a hall using features like area, elongateness, compactness, luminance value, and variance of the corresponding image region. The expected feature values were defined prior to the analysis in the attributes of the concept net according to the human experience.

To verify the states *Fair Construction* and *Fair Dismantling,* the system looks for trucks on the fairground which appear as small bright rectangles in the image. The characteristic width and length of a truck is stored (in meters) in the semantic net, which is transformed into pixel units and used as expectation. To limit the segmentation spatially to the immediate neighbourhood of the halls, the detected halls are used to define a valid search area in the current image of 0.5 m/pixel resolution. Assuming an accurate segmentation of the halls, it can be avoided to confuse trucks with the skylights of the halls. For a part of Fig. 7a the search area is shown in Fig. 7c as a dark region, the detected trucks are marked white. If the number of detected trucks is larger than a given threshold defined in the semantic net, the hypothesis *Trucks near Halls* is regarded as verified. Consecutively, the construction or dismantling phase can be instantiated. A final distinction between construction and dismantling becomes possible, only if the preceding state of the fairground is known.

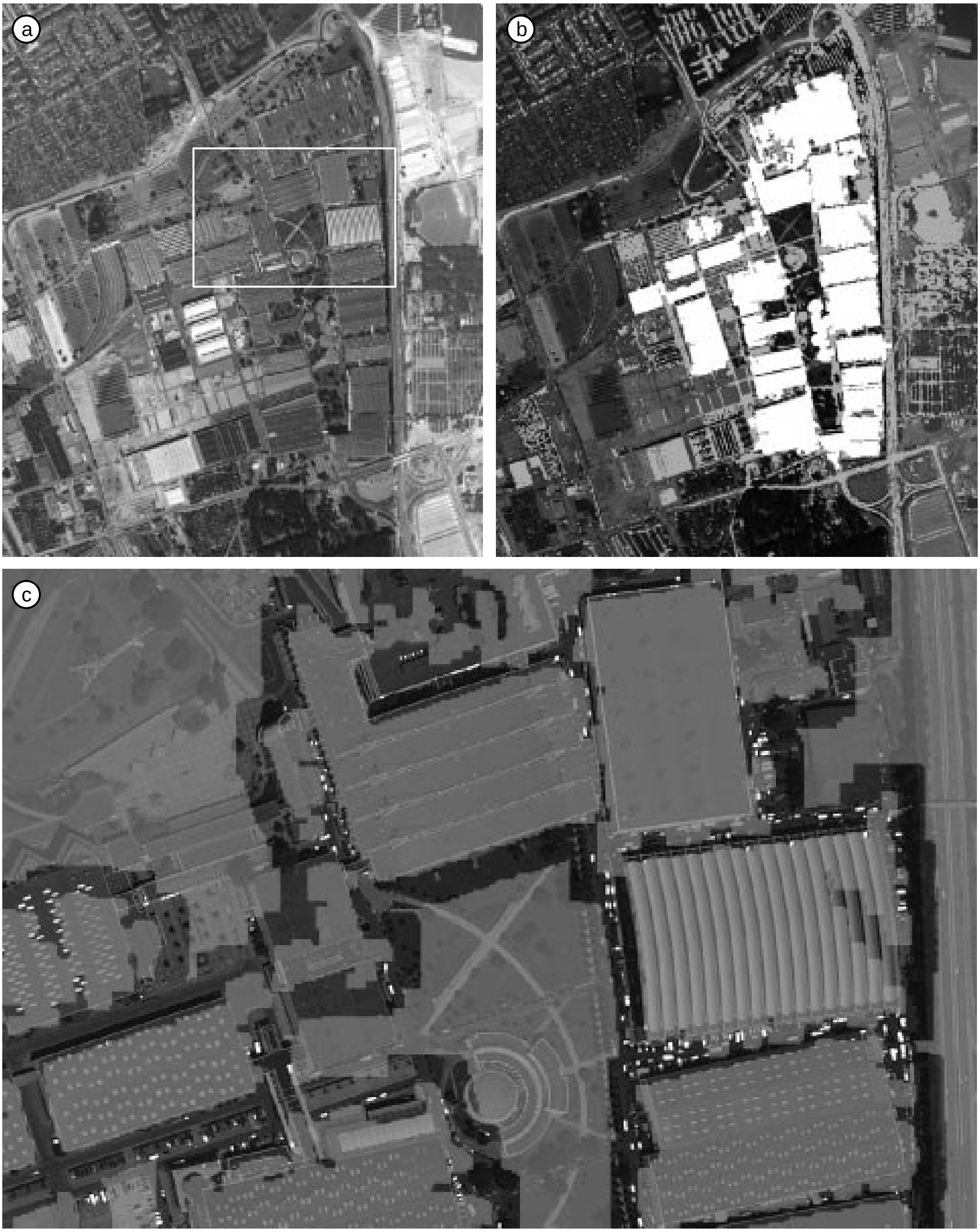

Figure 7. (a) Aerial image of the Hannover fairground, (b) Classification result for the class "Hall" (grey) and selected hall candidates (white), (c) Detail marked in (a): Search area for trucks on the fairground derived from the found halls (dark) and extracted trucks (white).

5.4. Future Work

To improve the extraction of halls, the segmented regions are planned to be splitted in compact regions and approximated by right–angled polygons fitted to the contours in the image. This will also yield a more accurate detection of the trucks. For the extraction of parking lots, cars, and persons, image processing operators and their interface to the AIDA system have to be implemented. Concerning the semantic net, the knowledge base for the fairground example has to be completed including the definition of computation and judgement methods. The strategy for the multitemporal image analysis will be tested in detail and improved, if necessary.

Additionally, a concept will be developed to allow the monitoring of landuse changes and detection of new constructions using again temporal relations to model possible state transitions. This will be tested for the interpretation and monitoring of moorland areas near Hannover.

Currently, the uncertainty and vagueness of the data is handled within the semantic net by a possibilistic judgement approach. It is planned to develop a second judgement calculus based on a probabilistic belief network (Bayesian net), which exploits the nodes and edges of the semantic net. Thereafter, a comparison of the two judgement approaches will be carried out.

To get more accurate segmentation results, a self–adaptive image processing module based on agents is currently developed. This system will select, configure and adapt iteratively an appropriate image processing operator according to a task description derived from the expectations and constraints of the semantic net. Finally, the segmentation results matching best with the given task description will be returned to the semantic net.

6. CONCLUSIONS

A knowledge based scene interpretation system called AIDA was presented, which uses semantic nets, rules, and computational 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 extraction of complex structures like an industrial fairground. The temporal knowledge is represented in a state transition graph and integrated in the semantic net. A new interpretation strategy generates hypotheses for the successor state of an object in the next image, which are verified in the sensor data. The first results show that the knowledge based scene interpretation is a promising approach for the analysis of multisensor and multitemporal images.

REFERENCES

Clement, V., Giraudon, G., Houzelle, S. , Sadakly, F., 1993. ”Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture”, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 31, No 4, pp. 779–791.

Dubois, D., Prade, H., 1988. ”Possibility Theory: An Approach to Computerized Processing of Uncertainty”, Plenum Press, New York and London, 263 p.

F. Kummert, H. Niemann, R. Prechtel, G. Sagerer, 1993. ”Control and explanation in a signal understanding environment”, *Signal Processing*, Vol. 32, No. 1–2, pp. 111–146.

Liedtke, C.–E., Bückner, J., Grau, O., Growe, S., Tönjes, R., 1997. ”AIDA: A System for the Knowledge Based Interpretation of Remote Sensing Data”, *3rd. Int. Airborne Remote Sensing Conference and Exhibition*, Copenhagen, Denmark, pp. 313–320.

Matsuyama, T., Hwang, V.S.–S., 1990. ”SIGMA : A Knowledge–Based Aerial Image Understanding System”, *Plenum Press*, New York, 277 p.

McKeown, D., Wilson, A., McDermott, J., 1985. ”Rule–Based Interpretation of Aerial Imagery”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–7, No. 5, pp. 570–585.

Mees, W., Perneel, C., 1998. ”Advances in computer assisted image interpretation”, *Informatica – International Journal of Computing and Informatics*, Vol. 22(2), pp. 231–243.

Pakzad, K., Bückner, J., Growe, S., 1999. ”Knowledge Based Moorland Interpretation using a Hybrid System for Image Analysis”, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3–2W5, Munich, Germany (to be published).

Quint, F., 1997. ”MOSES: A structural approach to aerial image understanding”, In: A. Gruen, E. Baltsavias, and O.P. Henricsson (editors): *Automatic Extraction of Man–Made Objects from Aerial and Space Images (II)*, Birkhäuser, Basel, pp. 323–332.

Stilla, U., Michaelsen, E., 1997. ”Semantic modelling of man–made objects by production nets”, in A. Gruen, E. Baltsavias, and O.P. Henricsson (editors): *Automatic Extraction of Man–Made Objects from Aerial and Space Images (II)*, Birkhäuser, Basel, pp. 43–52.

Tönjes, R., Growe, S., 1998. ”Knowledge Based Road Extraction from Multisensor Imagery”, In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3/1 , Columbus, Ohio, USA, pp. 387–393.

Tönjes R., Growe, S., Bückner, J., Liedtke, C.–E., 1999. ”Knowledge Based Interpretation of Remote Sensing Images Using Semantic Nets”, *Special ISPRS Commision III issue of Photogrammetric Engineering & Remote Sensing*, July 1999 (to be published).