USE OF BAYESIAN NETWORKS AS JUDGEMENT CALCULUS IN A KNOWLEDGE BASED IMAGE INTERPRETATION SYSTEM

Stefan GROWE, Torsten SCHRÖDER, C.–E. LIEDTKE
University of Hannover, Germany
Institut fuer Theoretische Nachrichtentechnik und Informationsverarbeitung
Appelstr. 9A, D–30167 Hannover
growe@tnt.uni–hannover.de

KEY WORDS: Image Interpretation, Semantic Net, Bayesian Network, Multitemporal Analysis

ABSTRACT

The increasing amount of remotely sensed imagery from multiple platforms requires efficient analysis techniques. The presented image interpretation system tries to automate the analysis of multisensor and multitemporal images by the use of structural, topological, and temporal knowledge about the objects expected in the scene. The knowledge base is formulated by a semantic net. Temporal knowledge about object states and their transitions is represented in a state transition graph which is integrated within the semantic net. The analysis of multitemporal images is improved by the prediction of possible object states derived from the knowledge base. During analysis the system has to deal with uncertainty and imprecision. Competing interpretations have to be judged to succeed with the most promising alternative. For this reason the measured object properties are compared to the expected ones. A probabilistic judgement calculus based on Bayesian networks is presented which uses the rules of belief updating and propagation. The approach integrates the probabilities of object states and their transitions within the judgement procedure. Hence it is well suited for a multitemporal image interpretation. For an example dealing with the detection of an industrial fairground from a set of aerial images the probabilistic judgement is compared with an existing possibilistic approach. It is shown, that the use of Bayesian networks increases the efficiency of the interpretation process.

1 INTRODUCTION

The recognition of complex patterns and the understanding of complex scenes from aerial images is a main issue of remote sensing. Applications like the monitoring of land–use or the update of maps and geoinformation systems (GIS) ask for efficient and automatic analysis techniques. The results of data–driven image processing algorithms (e.g. for segmentation or classification) are in most cases insufficient to distinguish the object classes used in GIS or maps. Hence modern systems for image understanding should use – like a human operator – additional knowledge to support the interpretation.

First of all additional images from different sensors can be used as supplementary knowledge sources. Furthermore GIS data and general prior knowledge about the expected scene objects is suitable to support the image interpretation. For special applications, like the detection of environmental changes, the exploration of images from multiple acquisition times is needed. On the one hand the use of multiple data and knowledge sources represents a great potential to improve the interpretation results significantly. On the other hand it raises the question of sensor fusion, which is a difficult task. Parameters like different platform locations, spectral bands, sensing geometries, spatial resolutions, and acquisition times have to be considered.

Various approaches to image interpretation and sensor fusion have been presented in the literature. The well known systems SPAM (McKeown at al., 1985), SIGMA (Matsuyama and Hwang, 1990), and MESSIE (Clement et al., 1993) represent the first generation of knowledge based systems for the interpretation of aerial images. The use of rules as knowledge representation scheme is widely spread. In the BPI–system (Stilla and Michaelsen, 1997) the rule base is structured in a network describing a part–of–hierarchy of the scene components. Mees and Perneel (1998) suggest the distinction of strategy, global, and sensor–dependent knowledge and represent it in AND/OR–trees, fuzzy production rules, and attributed prototypes with local image processing operators respectively. In ERNEST (Sagerer and Niemann, 1997) the knowledge base is formulated by a semantic net which describes the scene objects and their relations. A well defined network syntax facilitates the automatic reasoning. The MOSES system (Quint, 1997) extends the ERNEST approach to extract man–made objects from aerial images using hints from a map.

Most systems concentrate on the analysis of a single aerial image, some of them are able to exploit two or more images simultaneously, where in most cases a uniform sensor platform is assumed. Only few approaches address the problem of different sensor platforms and multiple acquisition times. The presented system AIDA (Tönjes et al., 1999) tries to formalize the representation of objects, sensors and time. It uses semantic nets to formulate the structural, topological, sensor–dependent, and temporal knowledge about the scene objects. The knowledge base is exploited to generate a symbolic description of the scene observed in one or more images, sometimes from different sensors. Information about object states and its possible changes over time can be integrated within the semantic net in form of a state transition graph. This temporal knowledge is used for the interpretation of multitemporal images to improve the explanation of land–use changes or to detect complex patterns due to a typical behaviour over time observed in the data set.

The image analysis is controlled by a rule–based inference engine, which documents competing scene interpretations in the leaf nodes of a search tree. To optimize the path through this search tree the alternatives are judged and the most promising one is investigated first. For the comparison of the intermediate interpretation results a common judgement calculus is needed which evaluates to which degree the measured object properties match to the expectations derived from the knowledge base.

In this contribution a probabilistic judgement calculus for the AIDA system is presented which is based on Bayesian networks. For the interpretation of multitemporal images it is shown that the approach causes a more efficient search compared to an existing judgement approach, if additional information about the probabilities of events is provided. The paper is organized as follows: After a brief introduction in the AIDA system the representation and use of temporal knowledge is described. Thereafter a short excursion into the theory of Bayesian networks is given followed by a discussion how it is used to judge a scene interpretation represented by a semantic net. Finally results are shown for the detection of an industrial fairground from a set of multitemporal images.

2 SYSTEM OVERVIEW

The architecture of the knowledge based image interpretation system AIDA has already been described in numerous publications (e.g. (Tönjes et al., 1999), (Tönjes, 1999b), (Liedtke et al., 1997)). For this reason only a short introduction is given here. The knowledge about expected scene objects is defined prior to the analysis in a separate knowledge base. By exchanging the knowledge base the system can easily be adapted to varying application tasks without modifying the interpretation module itself. This flexibility is the main advantage of a knowledge based approach. From the prior knowledge, hypotheses about the appearance of the scene objects are generated which are verified in the sensor data. Additional domain specific knowledge like GIS data (geographic information system) can be used to strengthen the interpretation process. 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 base is formulated by a semantic net. The *nodes* of the net, called *concepts*, represent generic prototypes of the expected scene objects, like roads, rivers or buildings. Realizations of the concepts detected in the scene during analysis are documented in the semantic net by new nodes called *instances*. The process of their generation is named *instantiation*. While an object is modelled by only one concept in the knowledge base, there might exist several instances of this object in the scene. During interpretation four different states of object recognition are distinguished: *hypotheses*, *partial instances*, *complete instances* and *missing instances*. The object properties are described by *attributes* attached to the concepts. Attributes possess a value derived from measurements in the data and an expected range of values which mirrors the expert knowledge. The expectations are restricted consecutively during analysis due to the current intermediate results. Computation functions are used to determine the attribute values and ranges from the sensor data or other instances at run–time.

The nodes are connected by *edges* to form a semantic network. The edges represent the structural, topological and temporal relations between the objects. The specialization of objects is described by the *is–a* relation along which the more special concept inherits all properties of the more general one. The decomposition of objects in their components is represented by the *part–of* link. Via the *concrete–of* link (abbreviated *con–of*) an abstract description is transformed into its more concrete representation in the data. For example the symbolic term “road” is connected to the primitive “line” to define its geometrical appearance in the image. The concrete–of relation structures the knowledge base into different conceptual layers like for example a symbolic layer, a geometry layer, and a material layer. Topological relations provide information about the kind and the properties of neighbouring objects. Therefore, the class of *attributed relations* (*attr–rel*) is introduced. In contrast to other relations, this one may possess attributes, which are 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. Especially for the representation of temporal knowledge the *temporal relation* is introduced which describes temporal changes of objects. In Chapter 3 the analysis of multitemporal images is discussed in detail.

For the efficient representation of multiple relations, the minimum and maximum number of edges can be defined for a relation. 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. An example for a concept net representing a knowledge base is given later.

2.2 Control of the 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 rules for instantiation for example change the state of an instance from *hypothesis* via *partial instance* 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*. Other rules generate hypotheses in a model–driven or data–driven way.

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: 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.

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 are stored in a search tree. Each node of the search tree (called *search node*) represents a consistent symbolic scene description in form of an instantiated semantic net. To avoid a full search a graph search algorithm (here: a modified A* algorithm (Tönjes et al., 1999)) is used which optimizes the search path through the tree. The algorithm decides in which order the competing alternatives are investigated. Therefore a quality measure is needed that describes the degree of compatibility between the measured object properties and the expectations. Tönjes suggests a possibilistic approach for the judgement of a scene description which considers uncertainty and imprecision of measurements and expectations.

2.2.1 Possibilistic Judgement Calculus. A hypothesis that has not yet been tested in the sensor data is neither right nor wrong. To model this ignorance a proposition e is judged by two measures of belief: the necessity $N(e)$ describes a pessimistic estimation of the belief while the possibility $P(e)$ represents the optimistic value which can be computed from the necessity of the contrary proposition $N(\neg e)$ by Eq. (1). The difference between possibility and necessity is called the ignorance $\Theta(e)$ (Eq. (2)). In the beginning the ignorance is 1, it is reduced consecutively during the interpretation process.

$$P(e) = 1 - N(\neg e) \qquad (1)$$

$$P(e) = N(e) + \Theta(e) \qquad (2)$$

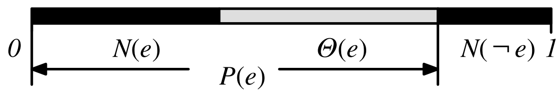

Figure 1: Necessity *N(e)* and Possibility *P(e)*

Imprecision of a proposition, like an imprecise specification of a road width, is modelled by fuzzy sets. Both, the expected range of an attribute value, called the hypothesis H, and the imprecise measurement itself, called the evidence E, are represented by trapezoidal membership functions p_H and p_E respectively as depicted in Fig. 2. In order to judge the compatibility of hypothesis and evidence, the possibility and necessity are determined according to the combination rules defined for fuzzy sets (visualized in Fig. 2). Each attribute of the semantic net is valued in this way.

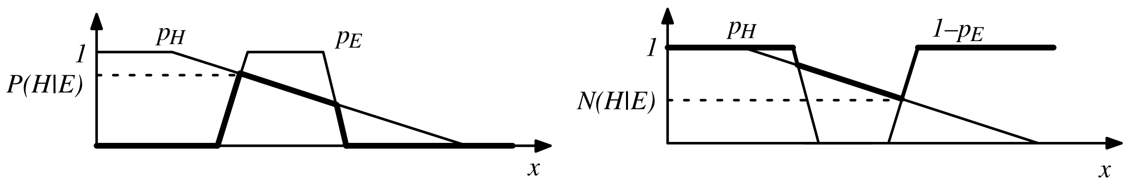

Figure 2: Computation of possibility P and necessity N given the hypothesis H and the evidence E

The judgement of a node, i.e. an instance, is derived by fusing the judgements of its attributes and its current subnodes. The necessity and possibility values of complementary information, like attributes and object parts, are combined by a weighted geometric mean. Redundant information, like evidence from multiple sensors regarding the same object, are fused using Dempster's rule of combination. For a more detailed description see (Tönjes, 1999 and 1999b). All node judgements of the semantic net are subsumed to an overall judgement of the scene description represented by the current node of the search tree. The possibility of such a search node defines an optimistic estimation of the interpretation quality. It is used by the system control to decide which alternative is investigated next.

The possibilistic approach supports the integration of uncertainty and imprecision within the interpretation process. The evidence found in the sensor data is aggregated strictly bottom–up to a final judgement of the scene description. Consequently a set of competing interpretations, that differ solely in model–driven hypotheses and possess therefore the same evidence, obtain identical values of merit. Hence, the candidate for the further interpretation has to be selected randomly. But in many cases an expert has prior knowledge which alternative is more probable. A plausible strategy would be to prefer the most probable alternative. For this reason in Chapter 4 a new judgement calculus based on Bayesian networks is presented which considers those prior known probabilities.

3 MULTITEMPORAL ANALYSIS

Applications like environmental monitoring and change detection require the evaluation of images from different acquisition times. Multitemporal images are also needed for the recognition of complex patterns, that are characterized by a typical temporal behaviour. Change detection can be carried out at pixel or at object level. Approaches that detect differences at pixel level require a perfect co–registration of the data sets. Furthermore they are limited to the comparison of images from the same sensor platform and they are very sensitive to variations in illumination, weather conditions, and perspective of view. Here, the recognition of changed object semantics is aspired. The scene description derived from the preceding image is used as prior knowledge for the interpretation of the current image. 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 awarenesses of the last interpretation are transformed unchanged into model–driven hypotheses to guide the analysis of the current image. But in many cases more reliable hypotheses can be generated, if additional temporal knowledge is used. Assuming biannual observations, a construction site as an example will probably not be observed at the same location again, because the construction has been finished meanwhile. To take advantage of temporal knowledge, it has to be represented appropriately in the knowledge base so that it can be exploited automatically during analysis.

3.1 Representation of Temporal Knowledge

In this scope temporal knowledge is understood as the knowledge about possible (or probable) transitions between different object classes over time. It is represented in a *state transition graph* which is integrated seamlessly within the semantic net. The states itself are modelled by concept nodes, the state transitions are defined by a new relation called *temporal relation*, which describes the temporal order of the states. For each state s_i both, a relative duration d_i and an absolute starting date t_{0_i} can be specified. To consider the uncertainty of the knowledge, time intervals are used for all temporal declarations. It is possible to define a prior probability $P(s_i)$ for each state, which represents the relative frequency of its occurrence. By default the probabilities are postulated to be equally distributed. For each state transition connecting two states s_i and s_j its duration d_{ij} and its conditional probability $P(s_j|s_i)$ can be defined also. Temporal relations are established exclusively between objects with a symbolic meaning, because no general statements about temporal changes of geometric objects, for example, can be made. In contrast to hierarchical relations like *part–of* and *con–of*, the start and end node of a temporal relation may be identical – forming a loop – to represent that the state stays unchanged over time. The mentioned representation scheme combines aspects of classical state transition graphs and markov chains known from the system theory, temporal constraint networks (Dechter et al., 1991) used in AI, and planning or activity networks known from the operations research field.

Figure 3 shows a semantic net for the detection of an industrial fairground. From a single aerial image a number of halls and parking lots can be recognized, sufficient to classify the site as an industrial area. For the decision, whether it is a fairground or not, a complete cycle of inactivity, construction of booths, fair activity, and dismantling of booths has to be observed in a sequence of multitemporal images. The mentioned cycle is represented in the semantic net. Each of the four states can be recognized by specific features. During the construction and dismantling phase the parking lots are empty, but there are a lot of trucks next to the halls, that keep the equipment for the booths. During a fair the fairground itself is free of cars, but the parking lots are crowded. If the image resolution is high enough, visitors can be detected walking on the fairground. If no fair takes place, there are few or even no cars neither on the parking lots nor on the fairground. Typical durations are defined for the states: The inactivity might last the whole year represented by the interval [1, 365]. For the other phases durations of 2 to 10 days are assumed. Absolute starting dates are not defined, because there are no seasonal restrictions for fairs. The state transitions in this example can occur from one day to the next. Hence their durations are specified by one day. For the states and state transitions prior probabilities are estimated due to general experiences. The inactivity state is the most probable of the four. The conditional probabilities express that this state is stable, while the others are more or less transient. An absolute precision of the probabilities is not necessary (even not possible).

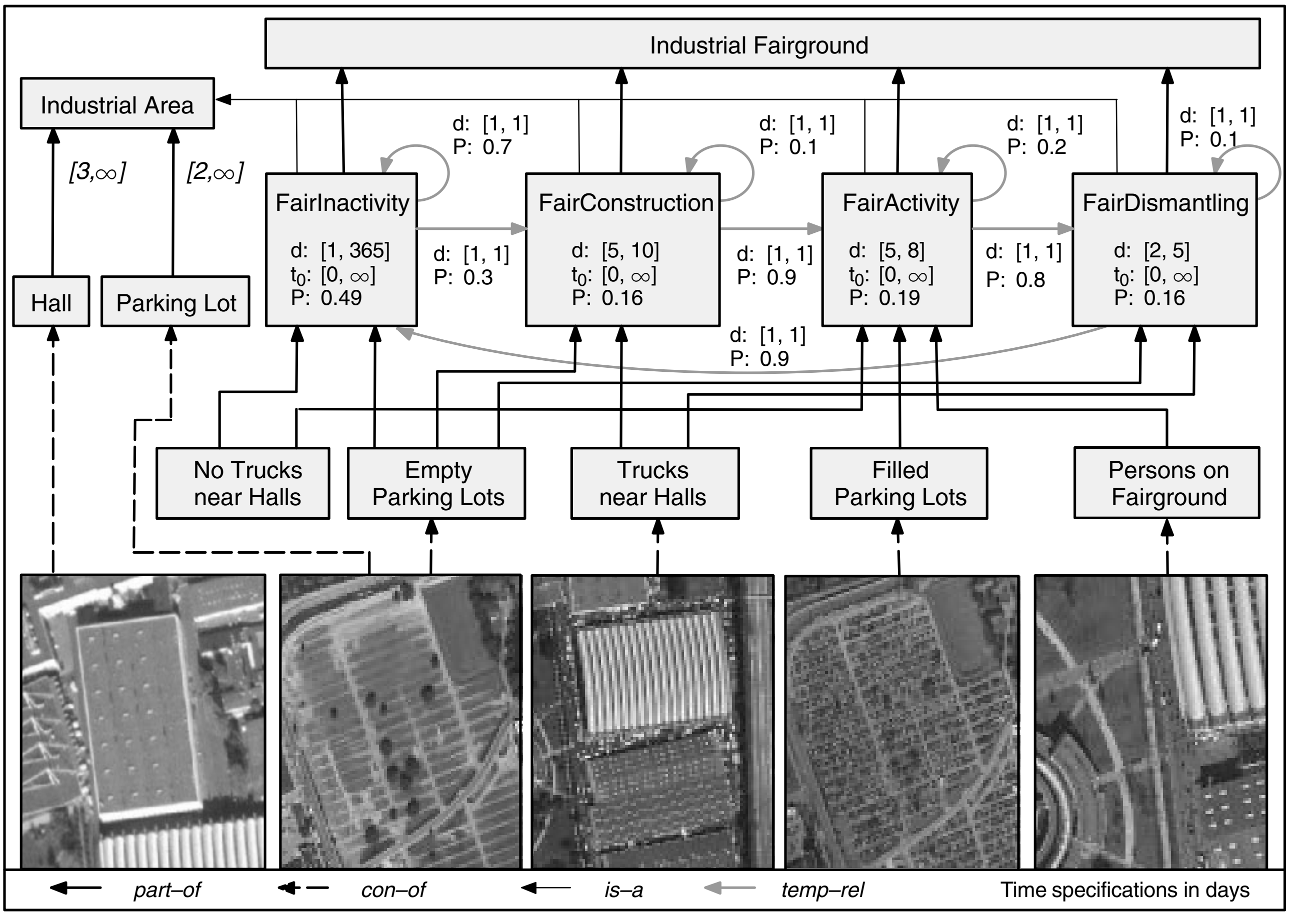

Figure 3: Semantic net for the detection of an industrial fairground with integrated state transition graph.

The relative proportions are sufficient to enable the system to favour the more probable solutions during analysis (see Chapter 4). For a consistent network it has to be considered, that all state probabilities and the transition probabilities of each state sum up to 1.

3.2 Use of Temporal Knowledge

During interpretation the temporal knowledge is used to predict successor states for already detected objects. Knowing the amount of elapsed time $\Delta t = t_2 - t_1$ between two consecutive images the state transition diagram is exploited to determine possible states of an object in t_2 based on its state in t_1. To avoid, that a possible candidate is omitted accidently, the prediction is done in an optimistic way. Starting in the state s_i detected for the time–stamp t_1 possible successor states $\{s_j\}$ are determined according to the defined transitions. For each candidate s_j the earliest start $t_{s_j}^{\min}$ is calculated from the minimum transition time $d_{ij}^{\min}$ (Eq. (3)). The latest end $t_{e_j}^{\max}$ is defined by the maximum remaining time d_i^{rest} of state s_i, the maximum transition time $d_{ij}^{\max}$, and the maximum duration $d_j^{\max}$ of state s_j. (Eq. (4)). State s_j is a possible successor of s_i, if the current time–stamp t_2 lies inside the interval $[t_{s_j}^{\min} \quad t_{e_j}^{\max}]$ and additionally inside $[t_{0_j}^{\min} \quad t_{0_j}^{\max} + d_j^{\max}]$, i.e. the interval of the permitted occurrence of s_j. The cases, that a state stays the same or a possible successor is skipped due to a large time difference, is considered also.

$$t_{s_j}^{\min} = t_1 + d_{ij}^{\min} \tag{3}$$

$$t_{e_j}^{\max} = t_1 + d_i^{rest} + d_{ij}^{\max} + d_j^{\max} \tag{4}$$

Assuming, that for example the state *FairInactivity* was detected for a region in an image dated March 1, possible successor states for this region at March 10 are (according to the temporal knowledge in Figure 3): *FairInactivity*, if the state stays unchanged, *FairConstruction*, due to its maximum duration of 10 days, and *FairActivity*, if the construction phase was shorter than 8 days and therefore not observed. The state *FairDismantling* can not be reached within 9 days. The system AIDA generates hypotheses for each alternative and tries to verify them in a model–driven manner. For the decision, which solution is investigated first, the competing hypotheses have to be judged and compared. As mentioned in Chapter 2.2.1 the evidence introduced at this stage of interpretation is equal for all alternatives, they differ in new model–driven hypotheses only. A judgement calculus that considers exclusively the evidence, like the possibilistic one described earlier, produces identical judgements for all solutions. Based on the conditional probabilities of the state transitions it would be most promising, for the given example, to prefer the successor state *FairInactivity* because it is the most probable one. In order to integrate prior probabilities of object states and their transitions within the judgement of a scene description, a probabilistic approach is suggested, which is described in the following.

4 PROBABILISTIC JUDGEMENT CALCULUS

The developed probabilistic judgement approach transforms the semantic net, i.e. the scene interpretation, into a Bayesian network, from which a measure of belief is derived. This value is used to select the best alternative for further investigations. After a short theoretical excursion to Bayesian networks their use for the judgement of semantic nets is described.

4.1 Theory of Bayesian Networks

Bayesian networks are directed acyclic graphs where each node represents a random variable and the edges in between are quantified by conditional probabilities. The structure of a Bayesian network encodes the dependency relations between the variables in the network. As the edges are established through causal relations pointing from cause to effect, the network provides an intuitive tool to model multiple interdependencies. Bayesian networks have become popular over the last years because it is not only possible to reason from measurements in a bottom–up fashion towards the most likely interpretation of the observed data, but also top–down from a hypothesis towards measurements to be expected. The theory of Bayesian networks is described in detail in (Pearl, 1988).

Each node of a Bayesian network models a discrete random variable with a finite number of different values. Initially the belief of each node is assumed to be equally distributed, i.e. each value of the underlying random variable is given the same probability. As soon as evidence is introduced into the net, for example by a certain observation in the data, the belief of the corresponding random variable changes. The probability of the observed value becomes 1 while the probabilities of the other values are reduced to 0. According to the causal dependencies the beliefs of related nodes are influenced, too. The evidence is propagated through the whole network according to a dedicated algorithm distinguishing messages from inferior and superior nodes. This propagation process is also known as *belief update*. The belief $BEL(x)$ of a node X is given by Eq. (5), where $\lambda(x)$ denotes the diagnostic support from the m child nodes and $\pi(x)$ the causal support from the n parent nodes (Fig. 4a). The term α normalizes the vector, so that the sum of the components, i.e. the probabilities of the individual values, becomes 1. Starting at the leaf nodes of the net the belief and subsequently the λ– and π–messages sent to the neighbouring nodes are computed recursively until an equilibrium is reached. Special techniques exist to cope with loops in the Bayesian network, for example the method of conditioning (Pearl, 1988).

$$
\begin{aligned}
BEL(x) &= \alpha \cdot \lambda(x) \cdot \pi(x) \\
\lambda(x) &= \prod_j \lambda_{Y_j}(x) && \text{with : } Y_j \text{ child nodes of } X \\
\pi(x) &= \sum_{u_1,\dots,u_n} \left[P(x|u_1, \dots, u_n) \prod_i \pi_X(u_i) \right] && \text{with : } u_i \text{ parent nodes of } X
\end{aligned}
\tag{5}
$$

In Figure 4b a propagation for a simple Bayesian network is illustrated. The evidence introduced in node E is distributed bottom–up and top–down until all nodes are up–to–date. An increasing number of observations reduces the ignorance represented by equally distributed probabilities. Finally the focussed node values are used for classification or decision making.

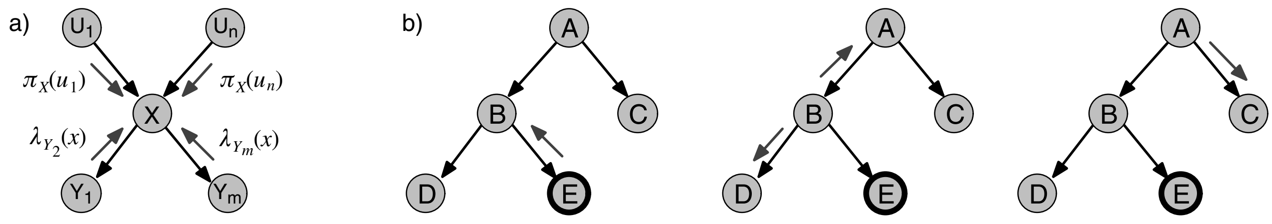

Figure 4: a) Diagnostic and causal support in Bayesian networks. b) Propagation of evidence introduced in node *E*.

4.2 Judgement of Competing Scene Interpretations

In order to use Bayesian networks for the judgement of a semantic net, a transformation between both types has to be defined that states how the nodes, attributes, and edges of the semantic net are mapped to Bayes nodes and edges. In the presented approach all instances and hypotheses of the semantic net are interpreted as Bayes nodes. Each Bayes node X models a binary random variable and thus possesses a two dimensional belief vector $BEL(X) = (P(x), P(\neg x))$ representing the probabilities for the verification and falsification of the event X. The node attributes of the semantic net, which mainly introduce the evidence, are transformed to special nodes of the Bayesian net that send exclusively λ–messages and are not influenced by π–messages of their parent nodes. *Part–of–* and *con–of* relations are mapped to Bayes links inverting the direction of the edge, because the concretizations and parts are understood as diagnostic support for their parents. *Temporal* relations are mapped unchanged to an edge of the Bayesian network. *Attributed* relations are not modelled by a Bayes link, but the contained attributes are considered as normal node attributes during the propagation process.

After mapping the semantic net to a Bayesian network, the root nodes are initialized by a top–down π–message, which is the ignorance vector $(0.5, 0.5)$ by default. If a prior probability is known for the node, like for example for object states as part of the temporal knowledge, this value is used instead. Consequently a predefined belief is assigned to these nodes from the beginning, which is an important way to prefer more probable hypotheses during the analysis. After initialization the evidence, represented by the attribute values measured so far, is introduced into the Bayesian network and propagated according to the mentioned algorithm. Similar to the possibilistic judgement approach the degree of compatibility between attribute value and range is determined and used as λ–message of the Bayes node representing the attribute. Hence, the two corresponding fuzzy sets μ_E and μ_H are superimposed and the normalized ratio of intersecting and total area is calculated (s. Eq. (6)). If a node possesses multiple attributes, the individual λ–values are combined by a weighted geometric mean according to Eq. (7).

$$\lambda_i(x) = \left(\frac{A_1}{A_1 + A_2}, \frac{A_2}{A_1 + A_2}\right) \tag{6}$$

$$\lambda(x) = \sqrt[I]{\prod_{i=1}^{I} \lambda_i(x)^{\frac{w_i}{W}}} \quad \text{with : } W = \frac{1}{I}\sum_{i=1}^{I} w_i \tag{7}$$

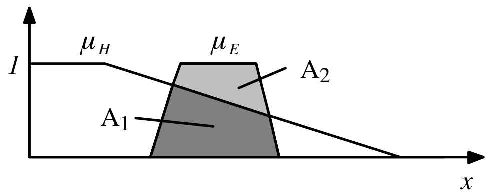

Figure 5: Computation of the diagnostic support *λ(x)* from attribute values *(μ_E)* and attribute ranges *(μ_H)*.

The evidence is propagated through the Bayesian network considering the conditional probabilities attached to the Bayes links. Here, the state transition probabilities of the temporal knowledge are taken into account. The nodes of the semantic layer are used to derive an overall judgement of the scene description. Their belief values are again combined by a geometric mean. The essential differences to the possibilistic approach mentioned in Chapter 2.2.1 are: The information is propagated both, bottom–up and top–down through the network. Therefore the belief of an object part, for example, influences the belief of the remaining parts and vice versa. Prior probabilities of objects, like temporal states, are considered within the judgement procedure via a corresponding π–message. This enables the system to prefer the more probable solution, even if the evidence is identical for all alternatives. The same effect cause the defined conditional probabilities of the temporal state transitions. The benefit of the Bayesian approach is illustrated in the following example.

5 RESULTS: EXTRACTION OF AN INDUSTRIAL FAIRGROUND

To validate the capabilities to a multitemporal image analysis of the AIDA system the mentioned example of the industrial fairground was chosen. The knowledge base illustrated in Fig. 3 was implemented including the necessary image processing algorithms to extract halls and parking lots. The application was tested for a set of aerial images of the Hannover fairground. The images, dated from 1995 to 1998, cover all states of the site (inactivity, activity, and construction/dismantling). Unfortunately, no continuous image sequence exists which depicts all phases of a single fair, but the given images are suitable to simulate the whole cycle by manipulating the time–stamps accordingly.

The analysis starts with the first image of the sequence looking for an industrial fairground. The system searches for the obligatory parts *Hall* and *ParkingLot*. Halls are recognized by right–angled polygons in elevation data, which is derived by stereo or is given by a DEM including buildings and vegetation. A hall candidate is accepted, if the region meets predefined expectations about shape, area, compactness, and neighbourhood to other halls. Parking lots are characterized by clusters of parallel lines representing the individual lanes. Only those clusters are selected to represent a parking lot that lie outside the fairground area surrounded by the halls. After the detection of at least three halls and two parking lots the *IndustrialArea* is instantiated completely. As the interpretation goal is to find a fairground, the system proceeds and tries to replace the *IndustrialFairground* by a more special concept. There are four possible specializations (*FairInactivity* to *FairDismantling*) and the search tree splits into five leaf nodes which are judged individually. The possibilistic judgement

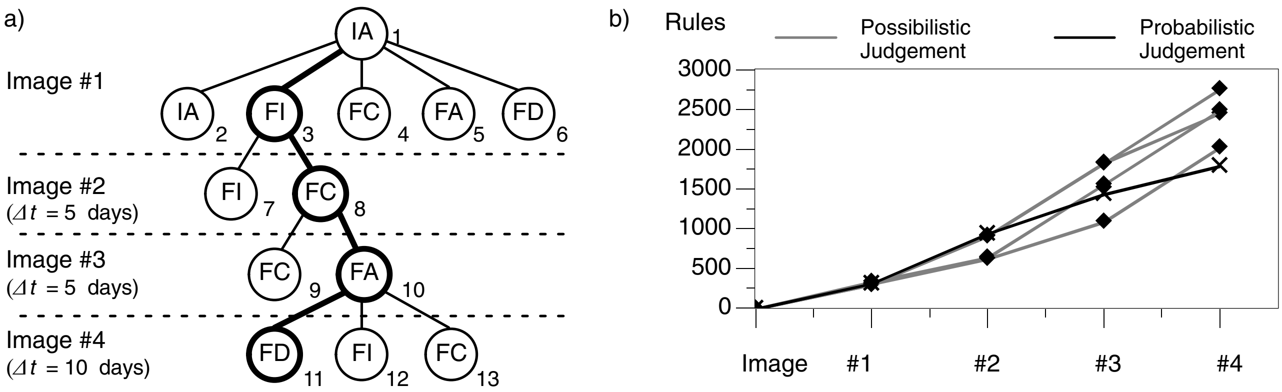

Figure 6: a) Constructed search tree representing the possible alternatives, the correct path is marked bold. (IA = *IndustrialArea*, FI=*FairInactivity*, FC=*FairConstruction*, FA=*FairActivity*, FD=*FairDismantling*). b) Number of activated inference rules during analysis using the possibilistic and probabilistic judgement approach respectively.

approach yields identical values, while the probabilistic approach favours the solution containing the most probable state *FairInactivity*. The hypothesis is tested in the data by verifying, whether the parking lots are empty and whether no trucks can be found next to the halls. For this purpose, regions of interest are derived from the already detected parking lots and halls respectively. Inside these regions vehicles, represented by small rectangular spots of a predefined size and luminance, are counted. The number of detected vehicles decides about emptiness and fullness. Finally for the first image the state *FairInactivity* can be verified. In order to instantiate the concept *Fairground* the other states are still missing. Thus, the system continues with the second image dated five days later. Hypotheses for the successor state of the site are generated according to the temporal knowledge. Within five days only *FairInactivity* and *FairConstruction* can be reached in the state transition diagram. The hypotheses are tested consecutively in the image data until the latter is instantiated because of the detection of trucks near the halls. The process repeats for the third and fourth image and the states *FairActivity* and *FairDismantling* are verified. As all necessary states were detected in the image sequence the concept *Fairground* can be instantiated completely. The goal is reached and the analysis stops. The whole interpretation process is illustrated in Fig. 6a. The constructed search tree consists of 13 search nodes, the final scene description contains 888 instances describing all the halls and parking lots detected in the four given images. The use of the temporal knowledge and the prediction of possible successor states restricts the search space, so that the analysis becomes more efficient.

5.1 Possibilistic vs. Probabilistic Judgement

In order to compare the two presented judgement approaches the described image interpretation process was performed using both methods. The strict separation of knowledge representation and system control permits the exchange of the judgement calculus without any modifications of the knowledge base.

In both cases the correct interpretation result has been reached after having generated 13 search nodes, but the efficiency differs considerably. Fig. 6b shows the accumulated number of inference steps, each represented by the number of activations of inference rules, needed for the detection of the *Fairground*. The possibilistic approach does not use the prior probabilities of states and state transitions. Hence, the search node to be investigated is chosen by random, whenever the evidence of the alternatives and therefore their judgement is identical. The randomness causes the total number of activated rules to vary between 2028 and 2766 rules. In contrast the probabilistic approach is deterministic and requires constantly 1793 rule activations, a reduction by up to 35%. For the interpretation of the first image (*FairInactivity*) both methods need roughly 300 rule activations. During the analysis of the second image, the Bayesian approach favours the more probable solution of an unchanged state, and therefore follows erroneously the path to search node n_7 until the observations made in the image cause the rejection of this hypothesis. The correct state *FairConstruction* (search node n_8) is found after 943 rule activations. If the possibilistic approach selects the correct search node n_8 by random search, only 622 rule activations are needed to reach this intermediate result. During the analysis of the third image the probabilistic method focuses immediately on the correct node n_{10}, so that 493 additional inference steps are sufficient (total 1436) to verify the state *FairActivity*. The other method needs between 469 and 921 rules (accumulated 1091 to 2766 rules) dependent on the order of investigated search nodes. For the interpretation of the final image the possibilistic approach fires 623 to 937 rules compared to 357 rules using the Bayesian network, which again prefers the most probable state transition from *FairActivity* to *FairDismantling*.

For the given example, the presented probabilistic judgement calculus takes advantage from the temporal knowledge introducing additional information about the probabilities of object states and their transitions. In the absence of such probabilities the judgement using Bayesian networks produces results comparable to the possibilistic method at higher computational costs due to the more complex propagation algorithm. In such cases the more robust and simple possibilistic judgement should be chosen.

6 CONCLUSIONS

In this contribution the use of the knowledge based image interpretation system AIDA for the analysis of multitemporal remote sensing images was presented. General knowledge about scene objects, their structure, and their appearance in the sensor data is stored in a semantic net. Additional temporal information about object states, their duration, their probability of occurrence, and knowledge about possible state transitions is represented by a state transition graph integrated within the semantic net. The system exploits the temporal knowledge to predict possible successor states of scene objects for the current image based on the object's state in the preceding image of the multitemporal sequence. Thus, the search space is reduced which accelerates the interpretation process.

A probabilistic judgement calculus was suggested in order to compare competing scene descriptions and to select the most promising alternative. The semantic net of instances is transformed to a Bayesian network and the measurements in the sensor data are introduced and propagated through the network. An overall judgement is derived from the belief values of the topmost Bayes nodes. An advantage of the approach is the capability to consider the prior probabilities given by the temporal knowledge. In cases, where the evidence is identical for several alternatives, the most probable solution is judged best and therefore preferred in the ongoing analysis.

The system was tested successfully for the detection of an industrial fairground in a set of four multitemporal aerial images. The fairground can be recognized, because the cycle of inactivity, construction, activity, and dismantling was observed consecutively in the set of images. For the given example the probabilistic judgement approach was compared to an existing possibilistic method. It was shown, that the exploitation of prior probabilities increases the efficiency of the interpretation process. The number of necessary inference steps is reduced by up to 35%.

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.

Dechter, R., Meiri, I., Pearl, J., 1991. Temporal constraint networks, Artificial Intelligence, Vol. 49, pp. 61–95.

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

Growe, S., 1999. Knowledge Based Interpretation of Multisensor and Multitemporal Remote Sensing Images, IAPRS, Vol. 32, Part 7–4–3W6, Valladolid, Spain, pp. 130–138.

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.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, San Francisco, 552 p.

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

Sagerer, G., Niemann, H., 1997: Semantic Networks for Understanding Scences, Advances in Computer Vision and Machine Intelligence, Plenum Press, New York and London.

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

Tönjes, R., Growe, S., Bückner, J., Liedtke, C.–E., 1999. Knowledge Based Interpretation of Remote Sensing Images Using Semantic Nets, Photogrammetric Engineering & Remote Sensing, Vol. 65, No. 7, pp. 811–821.

Tönjes, R., 1999b. Wissensbasierte Interpretation und 3D–Rekonstruktion von Landschaftsszenen aus Luftbildern, Fortschritt–Berichte VDI, Reihe 10, Nr. 575, VDI–Verlag, Düsseldorf