KNOWLEDGE BASED INTERPRETATION OF AERIAL IMAGES USING MULTIPLE SENSORS

R. Tönjes, C.–E. Liedtke

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung,
Division "Automatic Image Interpretation", Chair Prof. Dr.–Ing. Liedtke,
Universität Hannover, Appelstr. 9A, D–30167 Hannover,
WWW: http://www.tnt.uni–hannover.de/~toenjes, ~liedtke
E–mail: toenjes@tnt.uni–hannover.de, liedtke@

KEY WORDS: Knowledge Base, Interpretation, Scene Analysis, Sensor Fusion, GIS

ABSTRACT

A knowledge based approach for the interpretation of aerial images is presented that combines cues from multiple sensors (visual, infrared, SAR). The sensor fusion is applied at object level. This allows to use prior knowledge to increase the separability of the classes. The prior knowledge is represented explicitly using semantic nets. Interpretation exploits the semantic net to control the sequence of sensor fusion mixing bottom–up and top–down strategies. The presented approach addresses the problem of uncertain and imprecise sensor data by judging the different cues based on possibility theory. Competing interpretations are stored in a search tree. An A*–algorithm selects the most promising, i.e. best judged, interpretation for further investigation.

1. INTRODUCTION

The recognition of land use changes for map updating and environmental and agricultural monitoring represents a major topic of remote sensing. For this task sensors such as optical, thermal, and radar (SAR) have been developed. which collect different image data from the observed scene. The wish to extract more information from the data than is possible from 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 day and season or weather at image acquisition.

Data fusion can take place at pixel or at object level. *Pixel level fusion* processes directly the image data. Prerequisite for the pixel based image fusion is the perfect co–registration of the individual images. The resulting superimposed images provide multispectral vector data per pixel to which a numeric classifier can be applied directly.

The *fusion at object level* extracts features like regions and lines from the different images and combines the result to obtain the most reliable interpretation. The features can be grouped to extract complex structures. Furthermore, the interpretation allows to increase the separability of the classes by exploiting domain knowledge which is related to objects and not to pixels. Hence, here the fusion is applied at object level. The used prior knowledge includes common knowledge about the objects and scene specific knowledge which is provided by a geographic information system (GIS).

In the literature various approaches to 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. To ease the adaptation of the systems to new tasks the domain knowledge should be represented explicitly and be independent from the control of the analysis. Most interpretation systems like SPAM [1] and SIGMA [2] use a hierarchic control and construct the objects incrementally using multiple levels of detail. The system MESSIE [3] models the objects explicitly distinguishing four views: geometry, radiometry, spatial context, and functionality. It employs frames and production rules. ERNEST [4], uses semantic nets to exploit the object structure for interpretation. The presented system AIDA [5] adopts the idea to formulate prior knowledge about the scene objects with semantic nets. In addition the control knowledge is represented explicitly by rules. The system combines cues from different sensors and structural relationships of the objects to increase or decrease the reliability of competing interpretations.

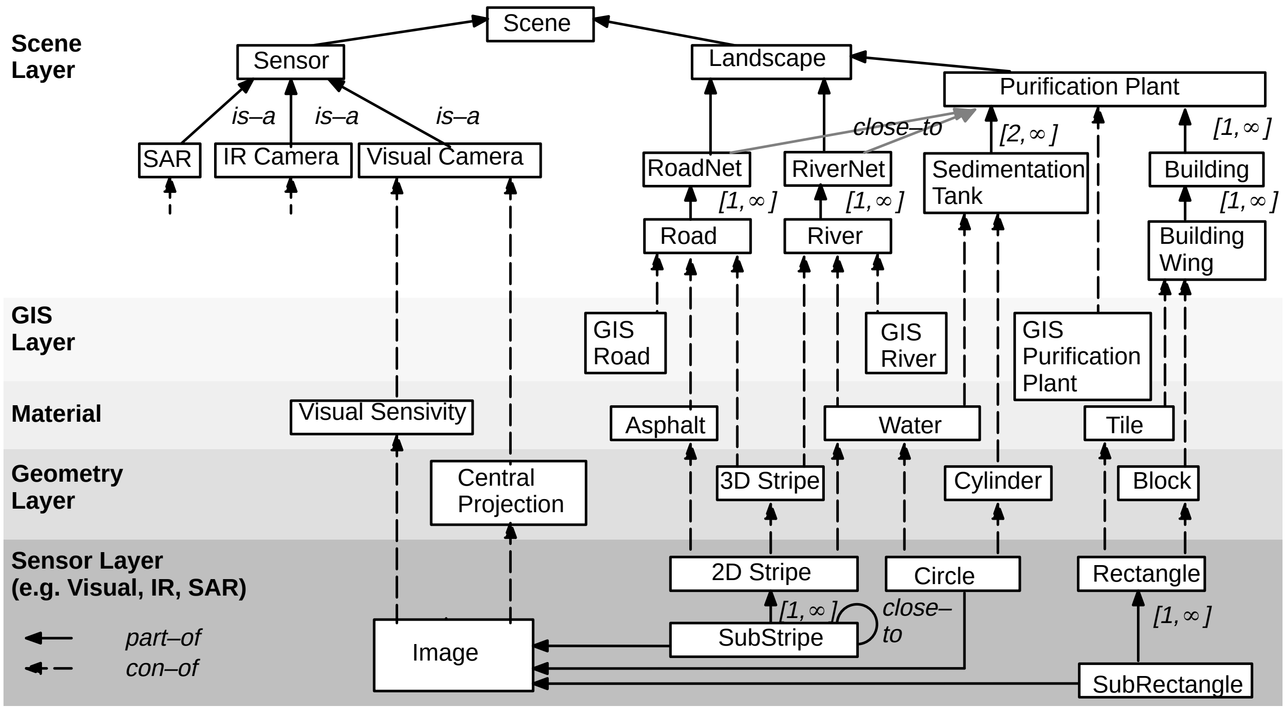

Fig. 1: Simplified semantic net representing the generic landscape model and its relation to the sensor image

2. STRATEGIES FOR SENSOR FUSION

According to the control of the fusion three approaches are distinguished:

Bottom–up fusion: The sensor data is grouped bottom–up. For example, the corresponding pixels or primitives from different sensor images are composed to form a feature vector for classification.

Top–down fusion: The scene is observed by various sensors. Fusion consists of selecting the most appropriate sensor.

Mixed fusion: Analysis proceed mixing successively top–down and bottom–up fusion techniques to accomplish the interpretation. Interpretation can focus on salient objects first and start evaluation with the most reliable sensor.

The mixed fusion is the most adaptive and general for scene analysis and is used in the following.

Sensor fusion has to deal with uncertainty and imprecision of the data. A proposition is uncertain if it can not be classified clearly as true or false. (E.g.: The segmented line is a road with a probability of eighty percent.) A proposition is imprecise if it possesses no accurate value but a range of several values. (E.g.: The road has a width between five and ten meters.) Several schemes have been suggested to represent and combine uncertainty, like possibility theory, bayes nets, and evidence theory, and to model imprecision by fuzzy sets or linguistic variables. The presented approach uses possibility theory [6] to model both uncertainty and imprecision.

According to the mutual dependencies the sensors are redundant or complementary. In the first case the sensory information support each other and can be combined independently. In the second case interpretation depends on all sensors and fails if the evaluation of one sensor does not succeed.

The uncertainty and imprecision are combined accordingly. Possibility theory models uncertainty by the difference between possibility P and necessity N of an interpretation and imprecision by fuzzy sets.

If no knowledge about a proposition e exists the necessity N(e) is zero and the possibility P(e) is one. The same is true for the contrary proposition ¬e because

$$N(\neg e) = 1 - P(e) \tag{1}$$

The difference between possibility and necessity is called uncertainty. The comparison of a proposition, i.e. hypothesis, with the sensor data, i.e. evidence, reduces the uncertainty by increasing the necessity of e or its opposite ¬e.

The joint necessity N(e) and joint possibility P(e) from the cues of complementary sensor information result to:

$$N(e) = \min_i N(e_i) \tag{2}$$

$$P(e) = \min_i P(e_i) \tag{3}$$

The corresponding approach to compute the joint necessity N(e) of redundant sensors from the maximum of the cues fails if the sensory information is in conflict, i.e. one suggests e and the other ¬e. In this case $N(e) + N(\neg e) \le 1$ is not guaranteed. To consider contrary information the cues of redundant sensors are combined similar to Dempster's rule of combination. All combinations of sensor information that support the proposition e are summed up. The sum is normalized by the sum in the denominator of all combinations that are not contradictory. The combination is associative and commutative.

$$N(e) = \frac{N_1(e)\ P_2(e) + N_2(e)\ P_1(e) - N_1(e)\ N_2(e)}{1 - N_1(e)\ N_2(\neg e) - N_1(\neg e)\ N_2(e)}$$

3. KNOWLEDGE REPRESENTATION

The knowledge base has to represent the knowledge about the sensors and the objects with their spatial relationships. The knowledge can be classified into 3D scene domain and 2D image domain knowledge. The latter is sensor related. The sensor coordinate system is referred to the image raster while the scene domain uses a cartographic coordinate system (e.g. Gauss–Krueger). The 3D scene domain can be subdivided into three aspects, the scene specific semantic or functionality (e.g. road), the 3D geometry (e.g. 3D stripe), and the material with its reflectance properties (e.g. asphalt).

The knowledge about the object structure and its relationship to the sensor specific appearance is represented efficiently by semantic nets. Semantic nets consist of nodes and edges in between. The edges or links of the semantic net form the relations between the objects.

Figure 1 shows a simplified semantic net for landscape analysis. The different aspects of the domain knowledge are modelled by conceptual layers, namely the scene, geometry, material, and sensor layer. If more than one sensor is available the sensor layer is duplicated (e.g. visual, infrared, and SAR layer). The GIS can be regarded as a symbolic sensor that is directly connected to the top scene layer.

The layer specific level of abstraction is modelled by the concrete–of link, abbreviated *con–of*. The decomposition of an object into its parts is described by the *part–of* link. The specialization of an object is described by the *is–a* link. that introduces the concept of inheritance. Spatial relations between objects are represented explicitly by *attributed–relations* like the *close–to* link.

4. INTERPRETATION

In AIDA a problem independent control mechanism exploits the semantics of the network language to control the interpretation. It uses the generic model in figure 1 to generate, i.e. instantiate, a specific semantic net describing the scene observed by the remotely sensed images. While, for example, the knowledge base contains only one generic model of a road, the scene description includes as many roads as detected in the sensor data. The control knowledge, i.e. the knowledge how and in which order scene interpretation has to proceed, is formulated in a set of predefined rules. An inference engine determines the sequence of rule execution. If competing interpretations occur they are judged using possibility theory. An A*–algorithm selects the most promising interpretation for further investigation.

The interpretation distinguishes the following types of sensor fusion:

Sensor selection: The object can be extracted completely using only one sensor. For example, rivers show up clearly in infrared images (fig. 2b) due to their cold temperatures.

Composite feature: This fusion type exploits several con–of links to combine redundant sensors. The extraction of the feature from only one sensor is erroneous like the road extraction from the visual sensor or infrared sensor alone. Hence the extraction combines the measured feature properties to improve the road detection (see fig. 2).

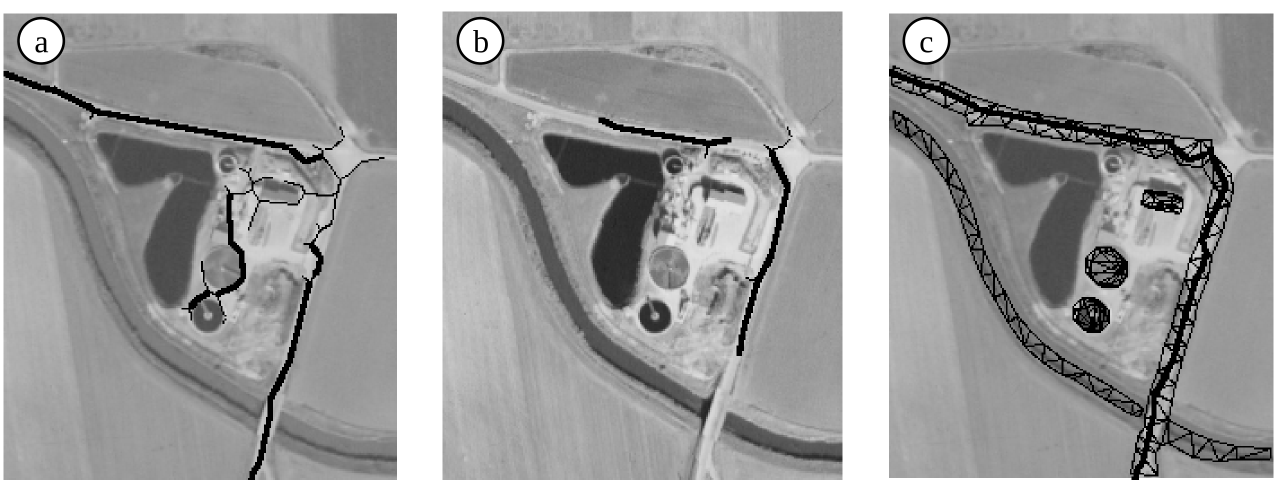

Fig. 2: Rejected (thin line) and accepted (wide line) road features from a) visual and b) infrared image with c) fusion result. Each object (road, river, building, sedimentation tank) is approximated by a polygon mesh to model geometry.

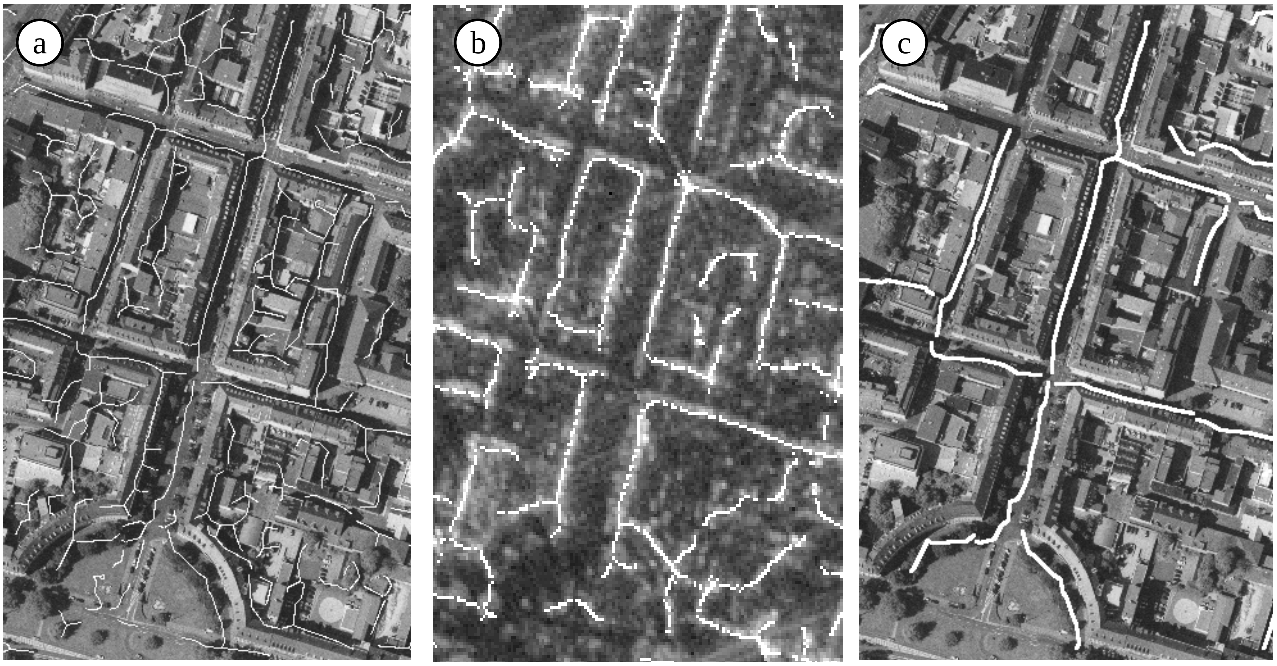

Fig. 3: The segmented lines, i.e. road candidates, from the visual image (a) must be accompanied by parallel lines as hint for buildings in the SAR image (b) to verify the road hypothesis (c)

Composite object: The object is composed of several complementary parts, indicated by *part–of* links, which can be extracted from different sensors. The purification plant in figure 2c consists of sedimentation tanks and buildings (fig. 1). The complex task of detecting a purification plant is simplified to the extraction of the building from the visual and the sedimentation tank from the infrared image. Furthermore, the plant has a road access and is located *close–to* a river to drain off cleaned water.

Composite context: The object may be only detectable in a certain context. For example, the roads in urban areas are usually accompanied by building rows along their sides which show up as bright lines in a SAR image. In figure 3 only those segmented dark stripes in the aerial image are interpreted as roads which are supported by parallel bright lines in the SAR image.

If a GIS is available the object location can be constrained further. However, the GIS may be out of date and incomplete. Hence the GIS is used to hypothesize an initial scene description to be tested in the remote sensing data. The use of a GIS is described in [7].

5. CONCLUSION

A knowledge based approach for the interpretation of aerial images from multiple sensors was presented. The sensor fusion is applied at object level. This allows to use prior knowledge to increase the separability of the classes. The prior knowledge is represented explicitly using semantic nets. Interpretation exploits the semantic net to control the sequence of sensor fusion mixing bottom–up and top–down strategies.

REFERENCES

[1] **McKeown, et al.**, ”Rule−Based Interpretation of Aerial Imagery”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI−7, No. 5, pp. 570−585, Sept. 1985.

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

[3] **Clement, V., Giraudon, G., Houzelle, S. , Sadakly, F.** ,”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, July 1993.

[4] **Niemann, H., Sagerer, G., Schröder, S., Kummert, F.**, ”ERNEST: A Semantic Network System for Pattern Understanding”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 9, pp. 883−905, Sept. 1990.

[5] **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.

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

[7] **Tönjes, R**., ”3D Reconstruction of Objects from Aerial Images Using a GIS”, Workshop on Theoretical and Practical Aspects of Surface Reconstruction and 3−D Object Extraction Haifa, Israel, September 9−11, 1997.