# KNOWLEDGE BASED ROAD EXTRACTION FROM MULTISENSOR IMAGERY


Ralf Tönjes, Stefan Growe

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



Commission III, Working Group 4

**KEY WORDS:** Knowledge Base, Semantic Net, Road Extraction, 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). Here the application of road extraction is described in detail. 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. Results are shown for the detection of roads in urban and agricultural areas exploiting image data from multiple sensors.

### KURZFASSUNG

Im vorliegenden Beitrag wird ein wissensbasierter Ansatz zur Interpretation von Luftbildern vorgestellt, der Merkmale aus unterschiedlichen Sensordaten (visuell, IR, SAR) verarbeiten kann. Dieses wird am Beispiel der automatischen Extraktion von Straßen demonstriert. Die Sensor Fusion wird auf Objekt–Ebene durchgeführt, wodurch es möglich wird, Vorwissen zu nutzen, um die Separierbarkeit der Objektklassen zu erhöhen. Das Vorwissen wird explizit in einem semantischen Netz repräsentiert, was während der Interpretation zur Steuerung der Sensor Fusion genutzt wird. Dabei werden abwechselnd daten– und modellgetriebene Strategien eingesetzt. Die Unsicherheit und Ungenauigkeit der Daten werden durch ein Bewertungskalkül auf Basis der Möglichkeitstheorie berücksichtigt. Konkurrierende Interpretationen werden in einem Suchbaum abgelegt. Ein A*–Algorithmus selektiert die vielversprechendste, d.h. bestbewertete Interpretation, für die vorrangige Bearbeitung. Es werden Ergebnisse zur Extraktion von Straßen in städtischen und ländlichen Gebieten aus multisensoriellen Bilddaten vorgestellt.


## 1. INTRODUCTION

The automatic extraction of road networks 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. Therefore various model based approaches for road extraction were presented in the past. Assumptions about the appearance and properties of roads, like continuity and closeness, are used to improve the results.

The edge–based road finding system RoadF [Zlotnick, 1993] defines road center hypotheses between antiparallel intensity edges and groups them to produce continuous, smooth road seeds. In [Zhu, 1986] a rule based system is used to link antiparallel linear edges to road segments. [Steger, 1997] represents the road network as a graph. Employing the fuzzy set theory the edges of the graph are tested for plausibility. In this way uncertain road segments are eliminated resulting in a complete road network.

Beside the evaluation of a single image some authors integrate additional data to obtain more reliable and accurate results. [Haala, 1992] matches a relational description of a digitized map with the detected road candidates. The multi–scale road extraction system described in [Mayer, 1996] uses images of different resolutions. The roads detected in a large scale image are refined in a small scale image.

The wish to extract more information from the data than is possible from a single image alone raises the question of sensor fusion. Especially the fusion of images from different sensor systems, like a visual, IR, and synthetic aperture radar (SAR) sensor, is a difficult task.

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 separability of the


Presented at the ISPRS Symposium "Object Recognition and Scene Classification from Multispectral and Multisensor Pixels", 6–10 July 1998, Columbus, Ohio, USA

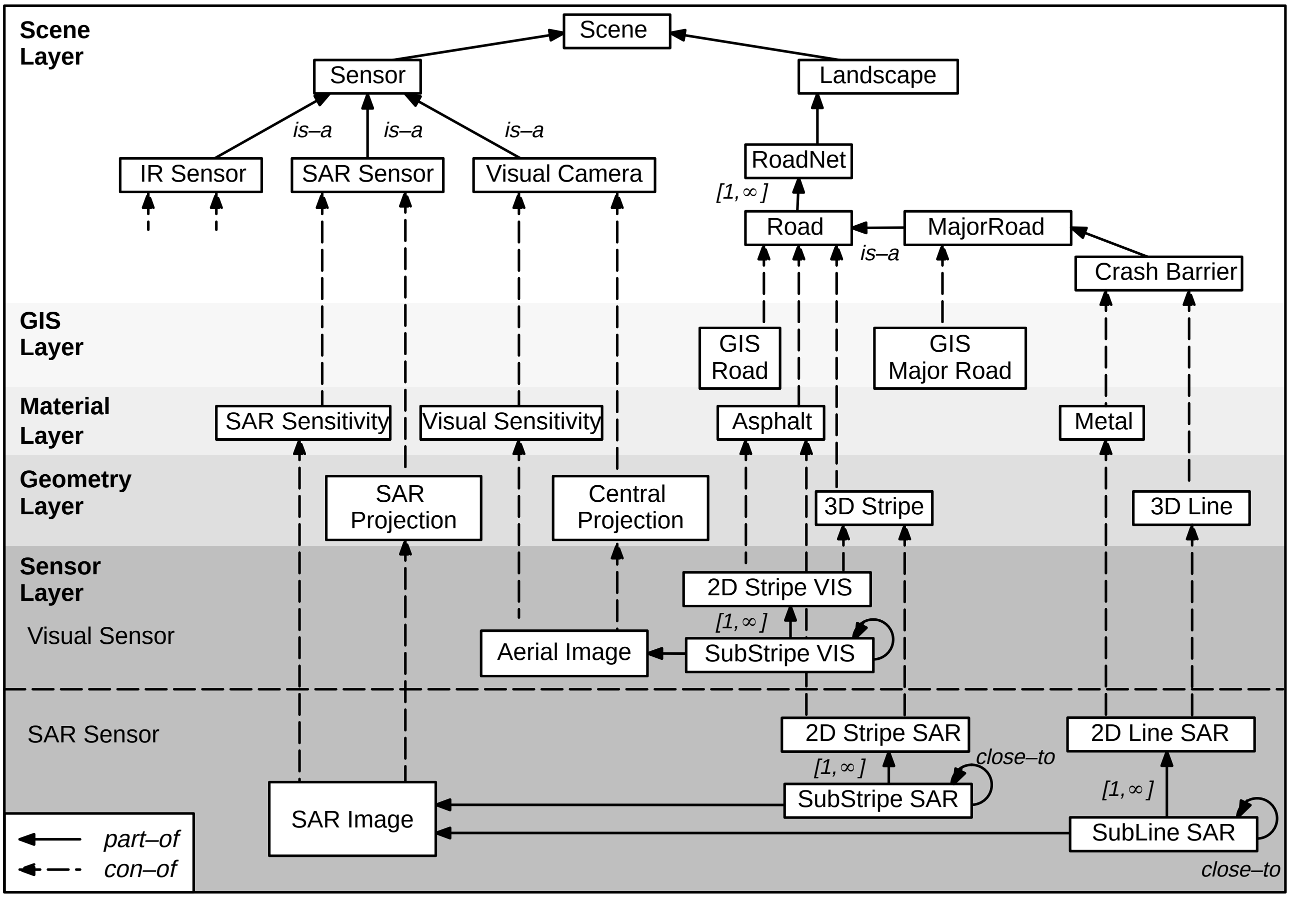


Fig. 1 Simplified semantic net for the extraction of roads and major roads including the relation to the multisensor image data

classes can be increased by exploiting domain knowledge which is related to objects and not to pixels. This task requires an image interpretation which assigns symbolic meanings to the segmented objects.

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. 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 [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. ERNEST [Niemann, 1990] uses semantic nets to exploit the object structure for interpretation.

The presented knowledge based image interpretation 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. The system combines cues from different sensors and structural relationships of the objects to increase or decrease the reliability of competing interpretations. Hence, the fusion is applied at object level. Additionally scene specific knowledge provided by a geographical information system (GIS) can be integrated in the interpretation process. In the field of remote sensing the system is used for the 3D reconstruction of landscapes [Tönjes, 1997] and for the detection of control points for image registration [Growe, 1997]. In this paper the application of road detection using multiple sensors is described.

## 2. 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 or UTM). 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. The specialization of an object is described by the *is–a* link introducing the concept of inheritance. Objects are composed of parts indicated by the *part–of* link. Thus the detection of a complex structure is simplified to the search for its parts. The transformation of an abstract object to its more concrete realization is represented by the concrete–of link, abbreviated *con–of*.

The object properties are described by attributes attached to the nodes. They contain an attribute value which is measured bottom–up in the data and a range which represents the expected attribute value. The range is predefined and/or calculated during the interpretation. For each attribute a value and range computation function has to be defined. A judgement function computes the compatibility between expected range and measured value.

Figure 1 shows a simplified semantic net for the extraction of roads. 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.

According to the knowledge base shown in Fig. 1 a *RoadNet* consists of more than one *Road* which is modelled by the minimum and maximum quantity in the *part–of* link (*[1, ∞]*). Each *Road* is represented in the GIS by the corresponding database object. Furthermore the road's geometry is described as *3D Stripe* while its material is modelled by the node *Asphalt*. During the interpretation the knowledge about geometry and material of an object is used to generate more accurate hypotheses about its appearance in the image data. For example in an aerial image a road is supposed to be a bright stripe whereas it is expected to be a dark stripe in a SAR image.

In the sensor data a road is represented by a *2D Stripe*. Because of erroneous and incomplete segmentation results this stripe is composed of several *SubStripes* which are located close to each other. This knowledge is modelled explicitly in the semantic net by an attributed relation called *close–to*. By a spatial inference process via the *close–to* link the system is able to track the segmented image primitives bridging small gaps.

For the extraction of major roads the inheritance mechanism is used. The node *Major Road* is a specialization of *Road*. Therefore it inherits the attributes and subnodes of its parent node. The GIS representation *GIS Road* is overwritten by the node *Major GIS Road* because major roads are described by a different object class in the GIS. The geometrical and physical properties stay the same. Only the dimensions are changed, like the expected width and length of the *3D Stripe*. In addition a major road possesses a metallic crash barrier as obligatory part which appears clearly as a bright line in a SAR image. The combination of a wide stripe in the visual and/or SAR sensor with a thin bright line in the SAR image justifies the instantiation of a major road.

## 3. CONTROL OF SENSOR FUSION

The analysis exploits the generic model represented as semantic net to control the extraction of the objects from the sensor data. In principle three approaches to sensor fusion can be 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. 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.

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 the possibility theory [Dubois, 1988] to model both uncertainty and imprecision.

### Modelling of Uncertainty

To judge the uncertainty of a proposition we define a measure of belief. It maps all propositions into the interval [0, 1]. A hypothesis that has not yet been tested in the sensor data is neither necessary right nor wrong. To model this ignorance the interval [0, 1] is divided into the three intervals necessity N(e), necessity N(¬e) of the contrary proposition, and ignorance Θ (fig. 2). If no knowledge about a proposition e exists both necessity N(e) and the necessity of the contrary proposition N(¬e) are zero. (The necessity measure in the sense of Dubois and Prade [Dubois, 1988] assumes that the elementary focal propositions can be ordered in a hierarchy of inclusion, i.e. are consonant). The possibility P(e) of a proposition is given by:

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

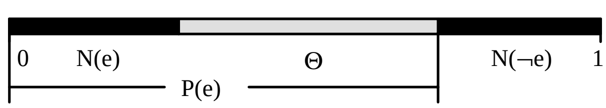


Fig. 2: Necessity N(e) and Possibility P(e)

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.

### Modelling of Imprecision

To model the imprecision of a proposition fuzzy sets in the sense of Zadeh [Zadeh, 1979] are employed. They describe the membership of a value x to a set, e.g. hypothesis H, with a membership function in the interval [0, 1] (fig. 3). A certain membership value a is interpreted as possibility p(a) that a proposition "x possesses the Value a" (e.g. the road width is 4 m) is true for the assumption "x is H" (e.g. the road is small).

The combination rules that are defined for fuzzy sets allow to judge imprecise attributes. The possibility and necessity of an imprecise hypothesis H for a given imprecise measurement E are depicted in fig. 3 and result to

$$P(H|E) \;=\; \sup_{x \in X} \; \min(p_H(x), p_E(x)) \tag{2}$$

$$N(H|E) \;=\; \inf_{x \in X} \; \max(p_H(x), 1 - p_E(x)) \tag{3}$$

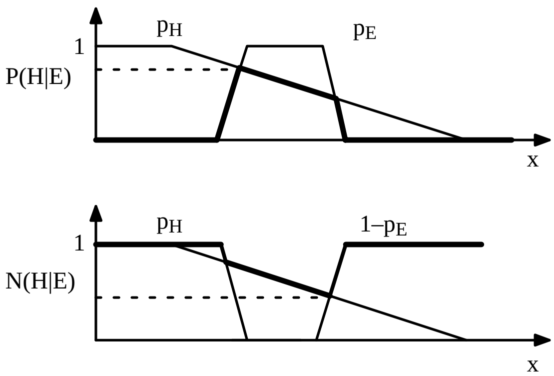


Fig. 3 Computation of possibility P and necessity N of a hypothesis H for a given evidence E

**Information Integration**

The decision, whether for example a segmented line is a road, is based on several attributes, like length, curvature, etc., and the evaluation of several complementary object parts or sensors. The joint necessity N(e) and joint possibility P(e) from the cues of complementary sensor information result to:

$$N(e) \;=\; \min_{i} \; N(e_i) \tag{4}$$

$$P(e) \;=\; \min_{i} \; P(e_i) \tag{5}$$

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 $\neg$e. In this case $N(e) + N(\neg e) \leq 1$ 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 (black rectangles in fig. 4). The sum is normalized by the sum in the denominator of all combinations that are not contradictory (all but white rectangles in fig. 4). The combination is associative

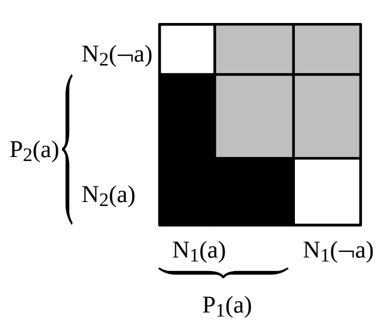


Fig. 4: Combination of two sources of different certainty

and commutative. Hence it can be written for two sensors without loss of generality:

$$N(e) \;=\; \frac{N_1(e)\; P_2(e) \;+\; N_2(e)\; P_1(e) \;-\; N_1(e)\; N_2(e)}{1 \;-\; N_1(e)\; N_2(\neg e) \;-\; N_1(\neg e)\; N_2(e)} \tag{6}$$

The belief measures are used to judge the different competing fusion results and to select the most possible interpretation $P^i$ for further investigation that fulfills

$$P^i(e) \;>\; P^j(e) \qquad \forall \;\; i \neq j \tag{7}$$

## 4. KNOWLEDGE BASED ROAD EXTRACTION

The detection of roads performs in a model driven manner and is described exemplarily for an aerial image. In a first step the available sensors and images are instantiated. If a GIS is available, the roads of the observed area are extracted from the database to constrain the expected object locations. Here the German ATKIS DLM 25 is exploited which contains the properties and the geometry of objects like roads, rivers, forests, and settlements.

The roads extracted from the GIS constitute a hypothesis that does not always match exactly with the roads in the sensor image. The hypothesis is propagated top–down via the geometry and material layer down to the sensor layer converting the geometry attributes from world to image coordinates. For example the width measured in meter is transformed into pixel (Fig. 5). Accordingly the material specific properties, e.g. reflectance and roughness, are transformed into the sensor specific photometric appearance, e.g. brightness and texture. Here the expected brightness of asphalt in the aerial image is computed.

The hypothesis *2DStripeVIS–1* can be tested in the image data. The expectation about the brightness, width and orientation of the stripe can be used to select an appropriate segmentation algorithm and adapt its parameters. We use the algorithms suggested in [Mayer, 1996] and [Bückner, 1998]. Depending on the segmentation one or more extracted stripes from the aerial image are required for one expected *2DStripeVIS*. For this reason the *2DStripeVIS* in the generic model is composed of several sub stripes. The extraction of the *2DStripeVIS–1* starts with the search for the first *SubStripeVIS–1* within a search area given by the expectations from the GIS. From the segmented stripes the best match, e.g. the longest within the search area, is used. Consecutively adjacent substripes are searched for. The expected location of their start point is further constrained by a maximum distance and an angle for the difference in orientation. The search area is described by a cone as depicted in figure 6. This enables the system to bridge small gaps due to fragmented segmentation results.

The propagation of this topological constraint is modelled by the attributed relation *close–to* with its attributes distance and angle. The *close–to* relation propagates the topological constraint to a second *SubStripeVIS–2* (see fig. 5) Again, if a stripe primitive in the aerial image matches, the expectation is verified and the *SubStripeVIS* is instantiated. This procedure is repeated until no further adjacent substripe is found.

If more than one segmented stripe matches the expectations the search node, i.e. current interpretation, splits into competing

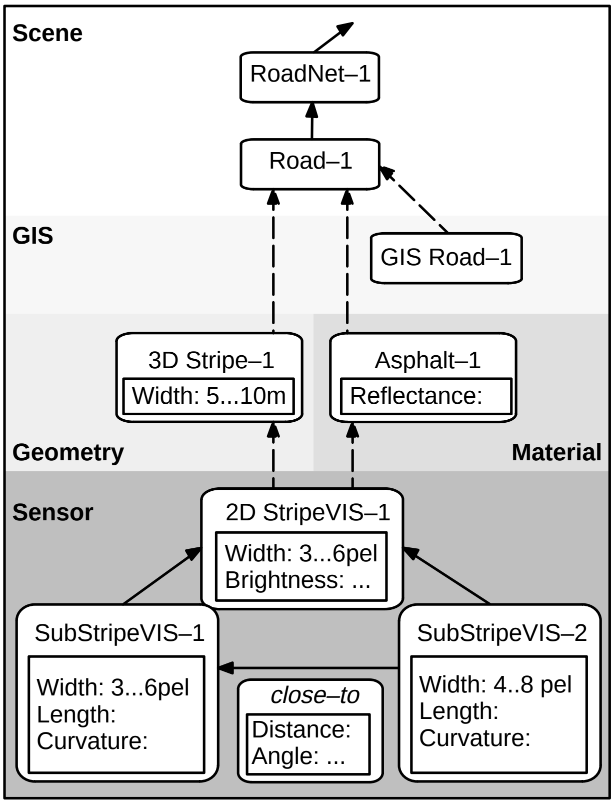


Fig. 5: Instantiation of the semantic net for the detection of roads

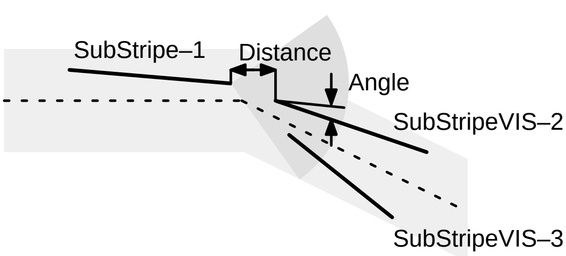


Fig. 6: Spatial inference for adjacent road segments

interpretations for each possible segmented stripe. The search nodes are judged applying the possibility theory mentioned above and the most promising search node is investigated further. The extracted *2DStripeVIS–1* causes the instantiation of *3DStripe–1* that models the continuous course of the road in 3D.

If there are multiple sensor images available, sensor fusion at object level is applied. The following types of sensor fusion have been realized successfully with AIDA:

**Sensor selection:** The object can be extracted completely using only one sensor. For example, metallic crash barriers show up clearly in SAR signals (fig. 8c) due to their high reflectance. Therefore they are extracted in the SAR image only. The same meets for the extraction of rivers which are significant in infrared images because of their cold temperature (fig. 7b). This knowledge is modelled in the semantic net by defining a concrete representation of the scene object in the specified sensor layer like *3D LineSAR* as subnode of *Crash Barrier* in the example of fig. 1.

**Composite feature:** The extraction of the feature from only one sensor might be erroneous like the road extraction from the visual sensor or infrared sensor alone. Hence the extraction combines the measured feature properties of different sensors to improve the road detection (see fig. 7 and 8). The contribution of each sensor is weighted and fused using Dempster's rule of combination. To achieve this behaviour in each sensor layer a representation of the scene object has to be defined and linked via the *con–of* relation.

**Composite object:** The object is composed of several parts which can be extracted from different sensors. The major road in figure 8 consists of two carriageways and a crash barrier in between. The complex task of detecting a major road is simplified to the extraction of the wide carriageways from the visual and/or SAR image and the crash barrier from the SAR image alone (fig. 8d).

**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 SAR image. In figure 9 only those segmented dark stripes in the aerial image are interpreted as roads which are supported by parallel bright lines in the SAR image.

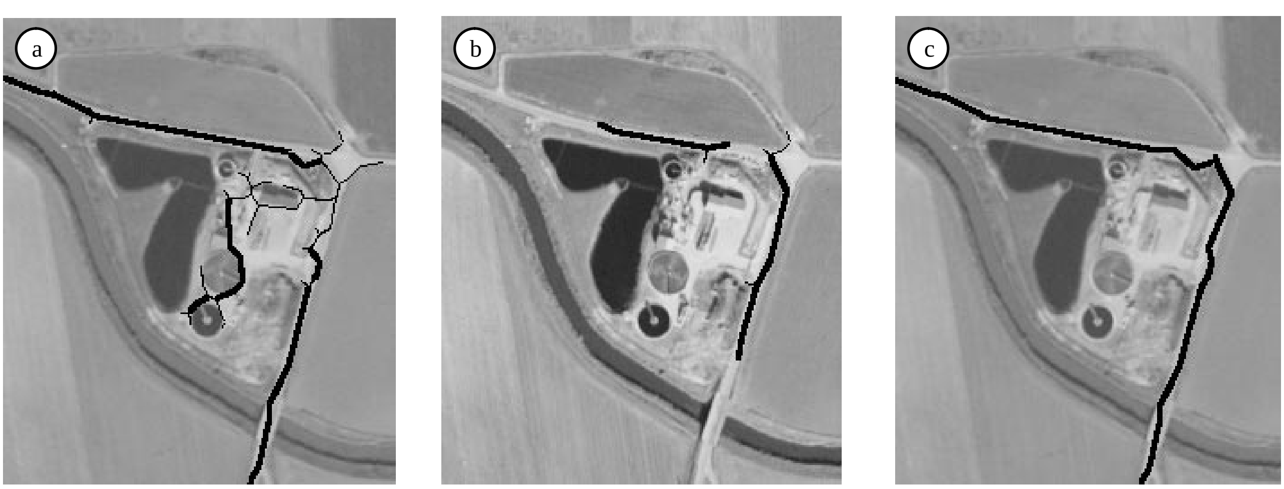


Fig. 7: Rejected (thin lines) and accepted (wide lines) road features from (a) visual and (b) infrared image with (c) fusion result.

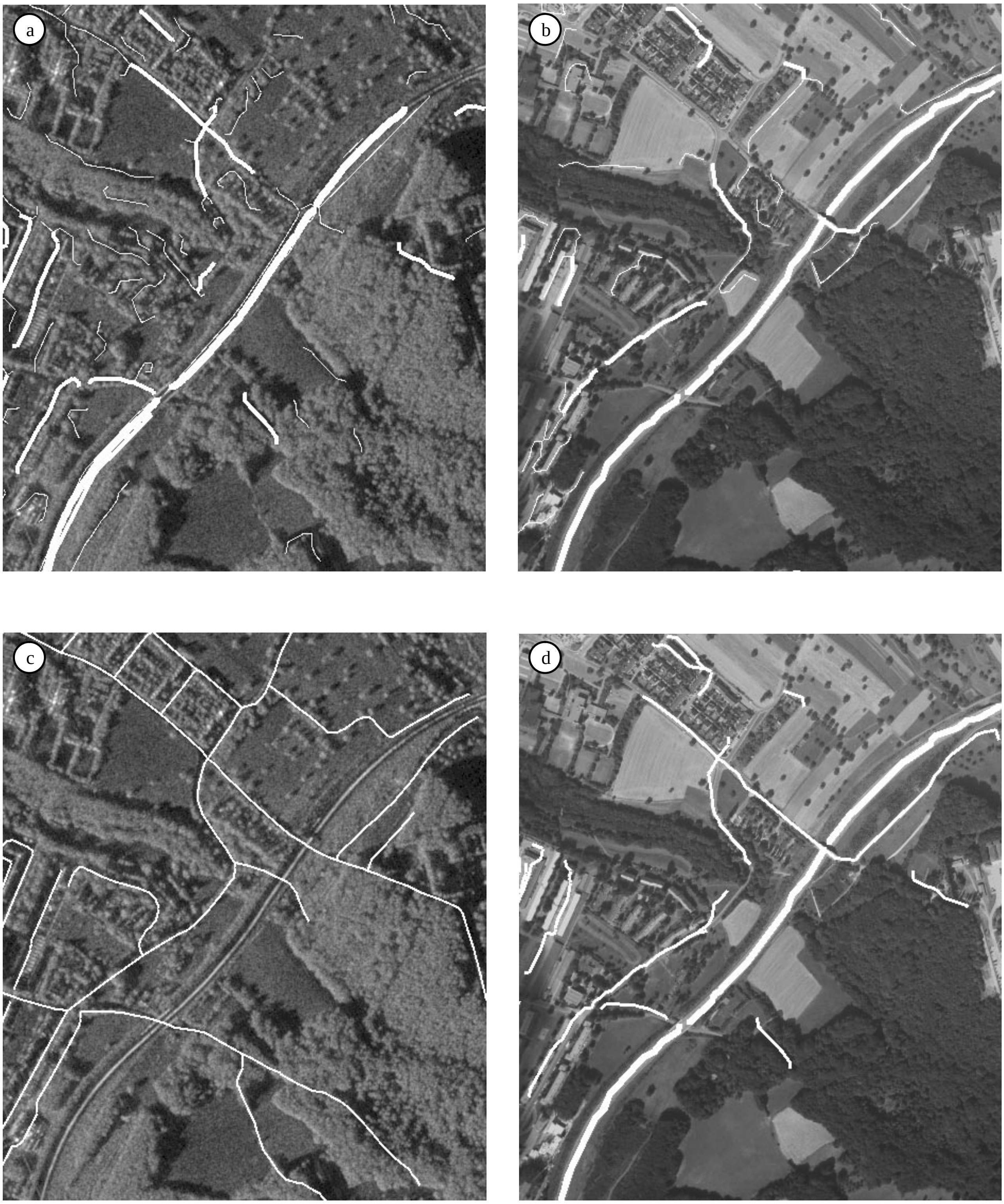


Fig. 8: Segmented (thin lines) and accepted road candidates (wide lines) in (a) SAR image and (b) the corresponding aerial image. (c) Used GIS data for the object class *Road* and (d) fusion result. The major road is emphasized.

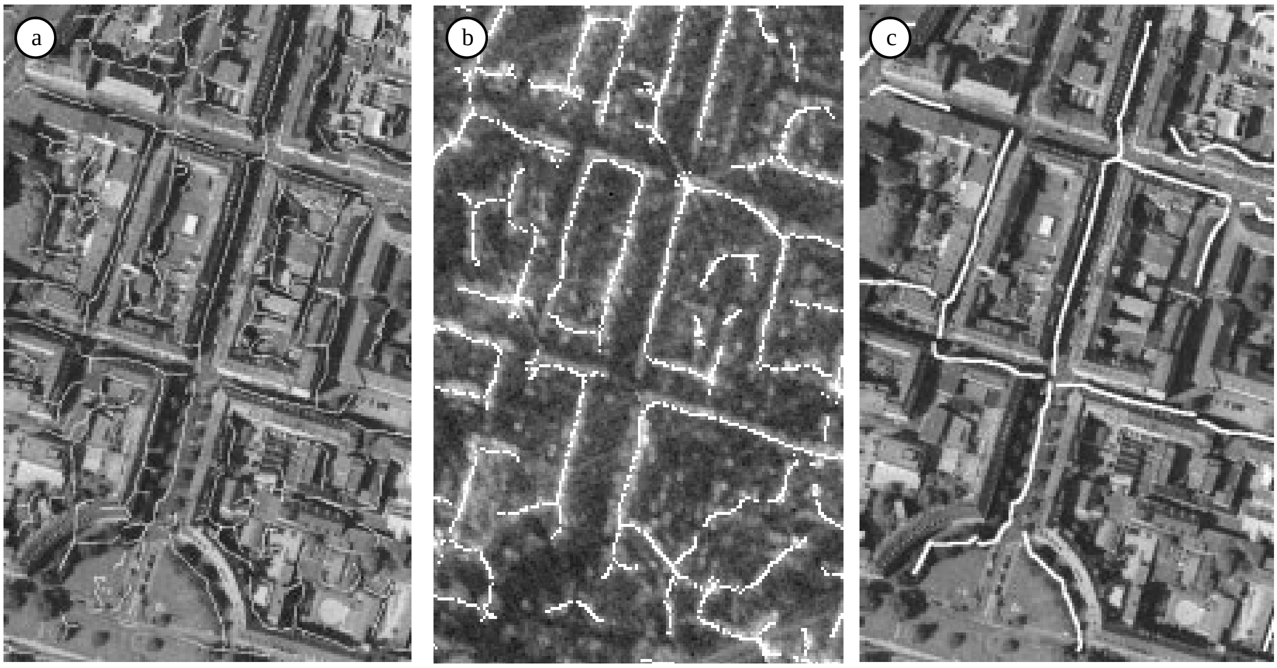


Fig. 9: The segmented lines, i.e. road candidates, from the visual image (a) must be accompanied by parallel lines as hint for buildings in the SAR image (b) to verify the road hypothesis (c)

## 5. CONCLUSIONS

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 which is represented here explicitly using semantic nets. Interpretation exploits the semantic net to control the sequence of sensor fusion mixing bottom–up and top–down strategies. For the application of road detection the cues from multiple sensors are combined to refine and confirm the results extracted from a single sensor. If available the information of a GIS is used to constrain the expectations about the object properties.

## REFERENCES

[Bückner, 1998] Bückner, J., ”Model based Street Extraction for the Registration and Interpretation of Remote Sensing Data”, submitted for the ISPRS Commission IV Symposium on *GIS– Between Vision and Applications*, Stuttgart, Germany, September 7–10, 1998

[Clement, 1993] 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

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

[Growe, 1997], Growe, S., Tönjes, R., ”A Knowledge Based Approach to Automatic Image Registration”, Proc. of *International Conference on Image Processing*, Vol. III, Santa Barbara, USA, 1997

[Haala, 1992] Haala, N., Vosselman, G., ”Recognition of Road and River Patterns by Relational Matching”, *Int. Archives of Photogrammetry and Remote Sensing*, 17. Congress of ISPRS, Vol. 29, Part B3, Com. III, pp. 969–975, 1992

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

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

[Mayer, 1996] Mayer, H., Steger, C., ”A New Approach for Line Extraction and its Integration in a Multi–Scale, Multi–Abstraction–Level Road Extraction System”, IAPR TC 7, Workshop: *Methods for Extraction and Mapping Buildings, Roads and other Man–Made Structures from Images*, 1996

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

[Niemann, 1990] 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

[Steger, 1997] Carsten Steger, Helmut Mayer, Bernd Radig, ”The Role of Grouping for Road Extraction”, in: *Automatic Extraction of Man–Made Objects from Aerial and Space Images (II)*, Armin Gruen, Emmanuel Baltsavias, Olof Henricsson (Editors), pp. 245–256, Birkhäuser Verlag , 1997

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

[Zadeh, 1979] Zadeh, L.A., ”A Theory of Approximate Reasoning”, Machine Intelligence Vol. 9, pp 149–194, 1979.

[Zhu, 1986] Zhu, M.–L., Yeh, P.–S., ”Automatic Road Network Detection on Aerial Photographs”, *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, pp. 34–40, Miami Beach, USA, 1986

[Zlotnick, 1993] Zlotnik, A., Carnine, P. D., ”Finding Road Seeds in Aerial Images”, CVGIP: Image Understanding, Vol. 57, No. 2, pp. 243–260, March 1993