KNOWLEDGE BASED MODELLING OF LANDSCAPES

Ralf Tönjes

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.html
E–mail: toenjes@tnt.uni–hannover.de

KEY WORDS: Knowledge Base, Interpretation, Modeling, DEM/DTM, Stereoscopic, Visualization

ABSTRACT:

A knowledge based approach for automatic generation of 3D–landscape models from aerial images is presented. The use of models for visualization tasks results in two requirements: efficient representation and high realism. Efficient representation of 3D–geometry is achieved by polygon meshes. Realism requires that the models meet the expectations of a human observer, who knows e.g. that roads are planar and forest edges possess a height step. The presented knowledge based modeler AIDA employs prior knowledge about the appearance of the objects in the scene to derive object specific constraints for surface reconstruction and to complete partially occluded objects. This requires an image interpretation to assign a semantic to the scene objects. The knowledge is represented explicitly by semantic nets and rules.

1. INTRODUCTION

For visualization of synthetic scenes 3D–models are required from which new simulated views can be computed. Applications such as flight and driving simulators, movie and TV production have a high demand for realistic models. Especially Landscape visualization is becoming an important tool for earth scientists, environmental researchers and civil engineers. Quantity, precision and the kind of models ask for methods that automate the model generation.

The common approach for digital terrain modelling uses stereo matching techniques to recover the height information from aerial images [Ackermann, 1991]. For efficient visualization the height map is subsampled and approximated by a polygon mesh in space. The geometric and photometric fine structure is modelled by projecting the aerial images onto the polygon surfaces.

However, the reconstruction of a 3D–model from its 2D–projections is an inverse and underconstrained problem, which causes model errors:

- The model is incomplete due to occlusions. This applies often to edges of forests and roads passing through forests.
- The sensor resolution limits the level of detail for reconstruction. Especially missing height steps in connection with shadows appear erroneous.
- The mesh approximation of the model geometry does not correspond with the object boundaries resulting in faulty breaklines.

This yields often models that do not meet the expectations of a human observer, who knows that the edges of forests exhibit a height step and roads run continuously. Hence, to improve the realism of the model the presented system uses prior knowledge about the landscape for 3D– reconstruction. To exploit the prototype knowledge about the object classes like roads, forests and grassland an image interpretation is required that assigns a meaning to the image regions. Consecutively the object semantic is used to control the 3D–reconstruction.

To ease the adaptation of the knowledge base for new modelling tasks, the knowledge base has to be formulated explicitly. In general knowledge can be represented by formal logic, fuzzy logic, frames (Clement, 1993; Foresti , 1993), semantic nets (Niemann, 1990), production systems, rule based systems (Matsuyama, 1990; Mc Keown, 1985) and neural nets (Shapiro, 1992). For description of structural relations semantic nets are suited. Hence here, inspired by the work of Niemann et. al. [1990], semantic nets are employed for knowledge representation

The following chapter gives an overview of the system architecture. The third chapter presents the used methods for explicit knowledge representation. In the consecutive chapter the knowledge base is used for image interpretation. Chapter five describes how the symbolic scene description is used to improve the object reconstruction. The paper concludes with a presentation of the results.

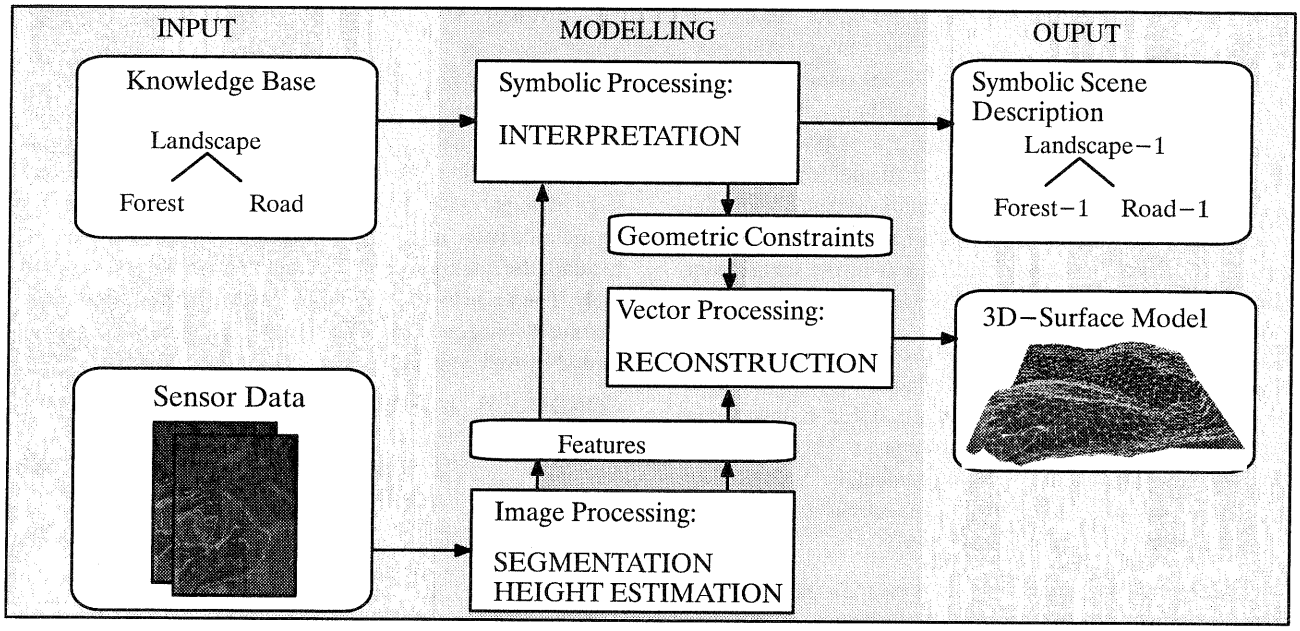

Fig. 1: Architecture of knowledge based modelling system AIDA

2. SYSTEM OVERVIEW

Figure 1 shows the architecture of the knowledge based modeler AIDA. The goal of the modeler is a realistic reconstruction of the observed scene. Input to the modelling are overlapping aerial images and prior knowledge about the objects present in the scene. Modelling consists of three main modules.

Image processing: The overlapping aerial images are rectified in a way that the epipolar line coincides with the image scanline to ease search for homologous points. Consecutively a height map is computed from the stereoscopic image pair. Further line shaped features and regions are segmented in the image.

Symbolic processing: Interpretation uses knowledge about the expected objects to group the features and assign a scene specific semantic to them resulting in a symbolic scene description.

Vector processing: From object semantic geometric constraints are derived to restrict the free parameters of surface reconstruction. The objects are approximated by a surface mesh with overlaid photo texture.

3. KNOWLEDGE BASE

3.1 Types of Knowledge

The a priori knowledge for 3D reconstruction of landscapes from aerial images includes knowledge about

- objects,
- context and task,
- sensors, and
- strategies.

Objects possess attributes and relations to other objects. As attributes geometry (e.g. shape, size, etc.), material (e.g. concrete, sand, etc.), and function can be distinguished.

Objects appear only in special contexts, i.e. forest edges in the context of forests. The task specializes the modelling demands. Both, context and task, reduce the problem domain.

Sensors transform objects into another, here pictorial representation, using geometric and radiometric transform characteristics. Image processing operators can be regarded as sensors that transform images to images. Their representation is not within the scope of this paper. For a representation of image processing knowledge the reader is referred to the system CONNY (Liedtke, 1992).

Strategies state how and in which sequence scene analysis has to proceed. E.g. eminent objects have to be searched for first.

3.2 Knowledge Representation

3.2.1 Objects: Object representation employs frames which contain a collection of attributes, relations, and methods (fig. 2). The relation slot establishes the connection to other objects. The object properties are stored as attribute values. Further the object has methods, i.e. functions, at its disposal to compute the attribute values. There may also be a method available to segment the object in the image data.

Main Road
Relations: is−a: Road part−of−inverse: Road Segment ... Attributes: Width[m]: 10...20 Material: Asphalt ... Methods: Segmentation: RoadExtractionFunction ...

Fig. 2: Example for a frame

3.2.2 Object Relations: Knowledge about structures can be represented efficiently by semantic nets. Semantic nets consist of nodes and edges in–between. Here the nodes of the semantic net represent the scene objects or their sensor specific realization respectively. The nodes are implemented as frames. The edges or links of the semantic net form the relations between the objects. Different relations describe the decomposition of an object into its parts (part–of), the specialization (is–a), and concrete realizations in the image data (con–of). The relations are exploited for object recognition.

The part–of relation states that the object is composed of parts. Thus object search can be reduced to a more simple task, the detection of its components. Objects linked via cdpart–of appear only in a certain context. Thus these objects are only searched for when the context, i.e. superior object, has been detected. Finally the optpart–of relation points out objects that might be present.

Objects can often be detected based on their geometric or photometric appearance, that can directly be segmented in the image data. This transformation of an abstract concept to a concrete realization is represented by the concrete–of link.

The is–a relation describes a specialization of an object. The specialization inherits automatically all relations and attributes of its more general concept.

The instance–of relation is used durung interpretation and connects instances with their prototypes.

3.2.3 Sensors: Sensors like cameras project the 3D objects onto a two dimensional target. They are sensitive to certain wavelengths. The different radiometric and surface properties of the materials are mirrored by corresponding colours and textures in the sensor image. E.g. the asphalt of roads appears bright in the visual spectrum and dark in SAR (synthetic aperture radar) images. In the semantic net the sensor transformation is modelled by the con–of relation (fig. 3). Propagate methods restrain the expected range of attributes top down. Compute methods obtain the measured value bottom–up from the sensor. To model uncertainties the attributes are described by minimum and maximum values.

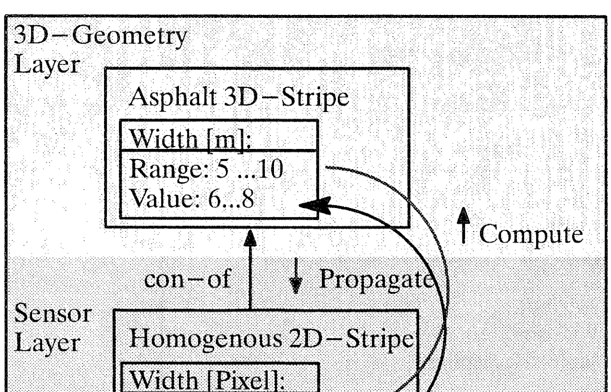

Fig. 3: Representation of sensor transform characteristics

3.2.4 Strategy: Strategy knowledge is represented by rules. These rules exploit the knowledge represented in the semantic net to control interpretation. For the relations exist according rules which propagate new information over the links of the semantic net. A rule is composed of a condition and an action part. The condition checks for a new interpretation state of neighboured nodes in the semantic net. If a situation formulated in the condition is detected, the action is executed to adapt the interpretation state of the focussed node accordingly. The knowledge, that an object is detected when all its parts $n_i \in \mathcal{P}$ are detected, i.e. are complete instances, is represented by following rule:

CONDITION: If state (node n_i) = complete instance
$\forall n_i \in \mathcal{P}\ \mathcal{P} = \{n_i \mid n_i = \text{part-of}(n_0)\}$
ACTION: Then state (node n_0) = complete instance.

Different strategies are represented by various sets of rules.

3.3 Knowledge for Landscape Modelling

Figure 4 shows a simplified semantic net for landscape modelling. The knowledge base distinguishes three conceptual layers. The top layer, called scene layer, describes the scene specific semantic. The middle layer represents the objects based on their 3D–geometry and material. The bottom layer is sensor related and describes the sensor specific photometric and geometric appearance of the objects. If more than one sensor is present the sensor layer is multiplied accordingly. Each layer uses a common appropriate vocabulary. E.g. the attribute size is measured in meter at the 3D–geometry layer and in pixel at the sensor layer (fig. 3).

In the context of landscape modelling roads, forests, and grassland shall be distinguished. The forest is composed of a forest roof and a forest edge which have to be modelled separately. For recognition only the forest roof has to be visible in the image data. Roads appear as homogenous stripes in the aerial images. The initial concepts 'textured 2D–region' and 'homogenous 2D–stripe' posses methods for segmentation of textured regions and homogenous stripes respectively.

4. INTERPRETATION

4.1 Control

The aim of the interpretation is to match the objects of the analyzed scene with the corresponding nodes of the semantic prototype net. Image interpretation exploits the knowledge base to instantiate hypotheses of objects expected in the scene. According to the state of the object recognition three different types of instances are distinguished: hypotheses, partial instances and complete instances. Hypotheses are not yet verified in the sensor data. Partial instances contain all concretes and context independent parts. Complete instances possess all concretes and obligatory parts and context dependent parts.

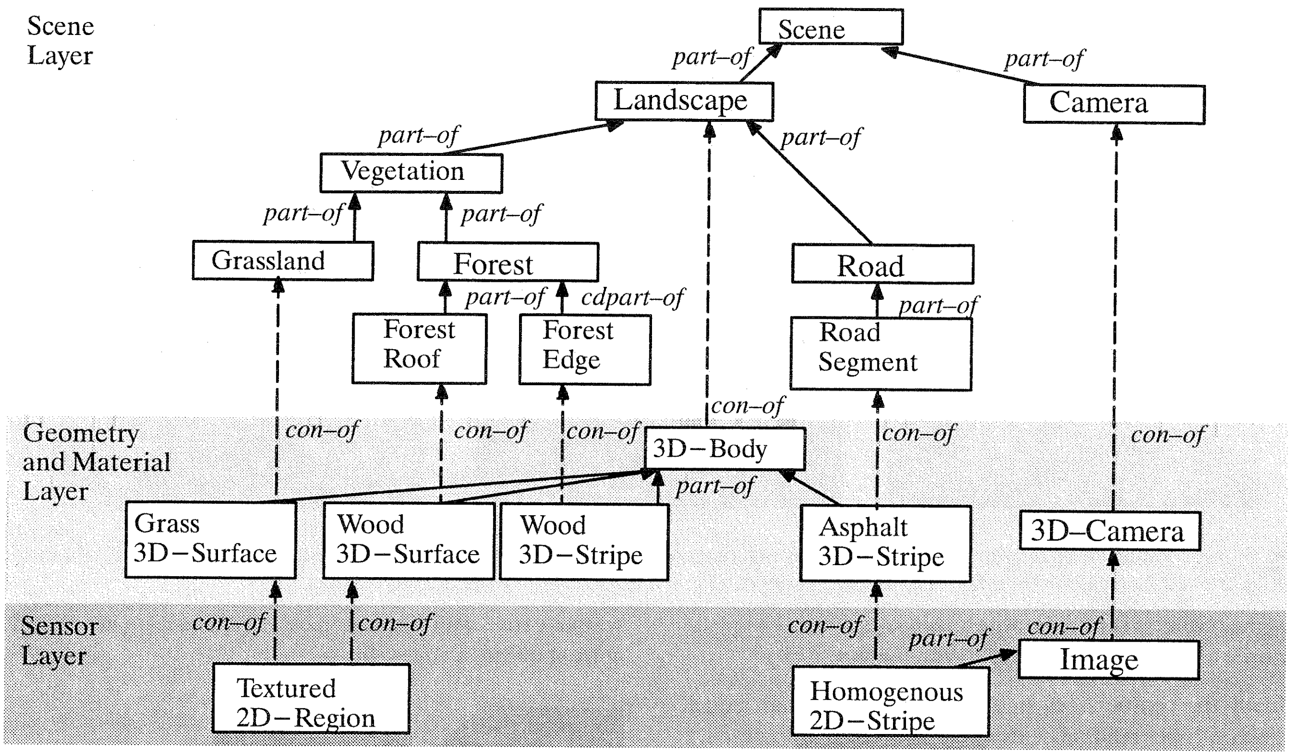

Fig. 4: Simplified semantic net representing prototypes of landscape objects

Interpretation proceeds primarily top down. Hypotheses are propagated from the scene layer to the sensor layer to test them in the sensor data. The propagated hypothesis at the sensor layer calls methods for segmentation of textured regions or homogenous stripes respectively. The result of the verification is returned to the superior concept which consecutively generates new hypotheses.

If the verification returns competing instances each possible interpretation is analyzed separately. For this each possible interpretation is documented by a search node which contains all concepts with their current interpretation state. Each time competing interpretations occur the search node splits into child search nodes. The leaves of the resulting search tree represent the currently competing interpretations. To focus interpretation on promising search nodes they are judged and ranked. The judgement computes the compatibility between expected concept properties and found concept properties by comparing the range and value slots of attributes. An A^*Algorithm selects the best judged interpretation for further investigation.

4.2 Segmentation

The initial concepts in the sensor layer are instantiated by segmentation of images. Presently two different initial concepts with segmentation methods are available: the concept 'textured 2D−region' and 'homogenous 2D−stripe'.

4.2.1 Segmentation of Homogenous Stripes: The extraction of homogenous stripes is based on gradient filters for edge detection with consecutive conditional local ranking to enhance weak contours. Thresholding yields a binary image in which candidates for homogenous stripes show up as long parallel lines. Figure 5a shows the segmentation result.

4.2.2 Segmentation of Textured Regions: Forests and grassland of natural terrain are characterized by different textures. The texture of a class k, e.g. forest, is assumed as generated by a stationary ergodic process. The prerequisite of statistical independence allows to compute the probability $P_1(y|k)$ of the texture process from the luminance histogram of all intensities y in the learn region.

Generally the probability for occurrence of a luminance value is not independent from its neighbours. Hence a texture model of second order statistics is used. Local mutual dependencies can be modelled by Gibbs random fields. Couples of neighbored pixels, named cliques, are inspected. Three measures of co–occurrence are computed for each clique:

- the probability $P_2(y|k)$ of common class membership,
- the probability $P_3(y|k)$ of a luminance difference within a region,
- the probability $P_4(y|k)$ of a luminance difference between different regions.

Gibbs random fields describe the joint probability

$$P(\overline{k}|\overline{y}) = \frac{1}{Z}\, e^{-\sum_c V_c(y|k)} \quad ; \; Z = \text{normalizing constant.} \tag{1}$$

$$V_c(y|k) = -\sum_{i=1}^{4} \lambda_i \ln(P_i(y|k)) \tag{2}$$

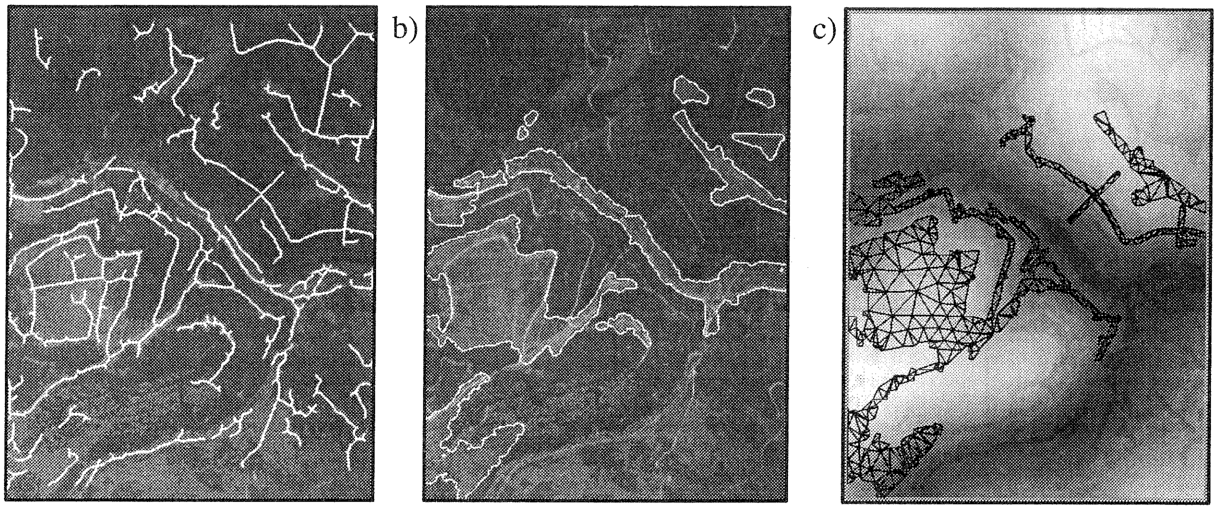

Fig. 5: Aerial image with overlaid segmentation results for a) roads and b) borders between forests and grassland. c) Height map with elevated forests, showing the 3D–surface mesh of selected objects from a) and b) as overlay.

Segmentation is solved iteratively using maximum–a–posteriori estimation (Gimmelfarb, 1991; Tönjes, 1994).

Finally the contours are smoothed using a contour model based on Gibbs random fields which favours smooth contours (Mester, 1988). Figure 5b shows the segmentation results for two classes of textures.

5. RECONSTRUCTION

5.1 Data driven Reconstruction

The initial reconstruction is data driven and employs photogrammetric stereo vision. The correspondence analysis uses normalized cross correlation as cost function for matching of homologous points to determine the height dependent parallax. A Smoothness constraint is exploited by subsequently interpolating continuous regions. Finally the parallax map is transformed to a height map using binocular camera geometry (Koch, 1995).

5.2 Model driven Reconstruction

The model driven reconstruction exploits prior knowledge about object geometry to restrict the parameters for reconstruction. While data driven reconstruction uses uses only a few and general geometric constraints, interpretation offers the facility to exploit object specific geometric properties. Interpretations yields the segmentation of aerial images in various regions, as forests, grassland, and roads. The location of these regions is stored in image masks. Scene reconstruction uses these image masks to apply object specific constraints to the height map obtained by stereoscopic correspondence analysis. The prior knowledge forces a height step between forests and grassland or roads. Further the object semantic controls mesh generation. Roads are approximated by a separate mesh to ensure a continuous course. At the edges of forests a vertical mesh for the height step is inserted (fig. 5c).

The semantics attached to the model parts allow an object specific post processing. This offers the facility to refine the objects artificially by adding details which are invisible to the sensor from computer graphic libraries. E.g. for close views synthetic trees with fine transparent leaf structure are placed in front of forest edges (fig. 6).

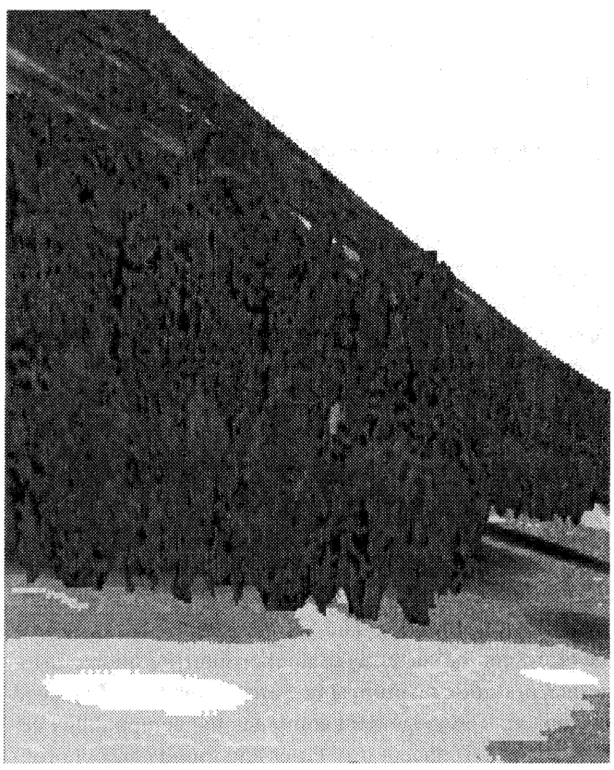

Fig. 6: Close up view of forest edge

6. RESULTS

Figure 7 shows the synthesized view of a landscape model reconstructed from a pair of overlapping aerial images. The model generation considered object semantics: roads are represented by a separate surface mesh and exhibit a continuous course. At the edges of forests a height step was inserted. The model of the Sieber Valley in the Harz (2km x 2km) consists of approximately 13.000 Polygons and a texture map of 2048x2048 pixel. For interactive exploration of the scene in real time the model can be visualized on a graphic computer.

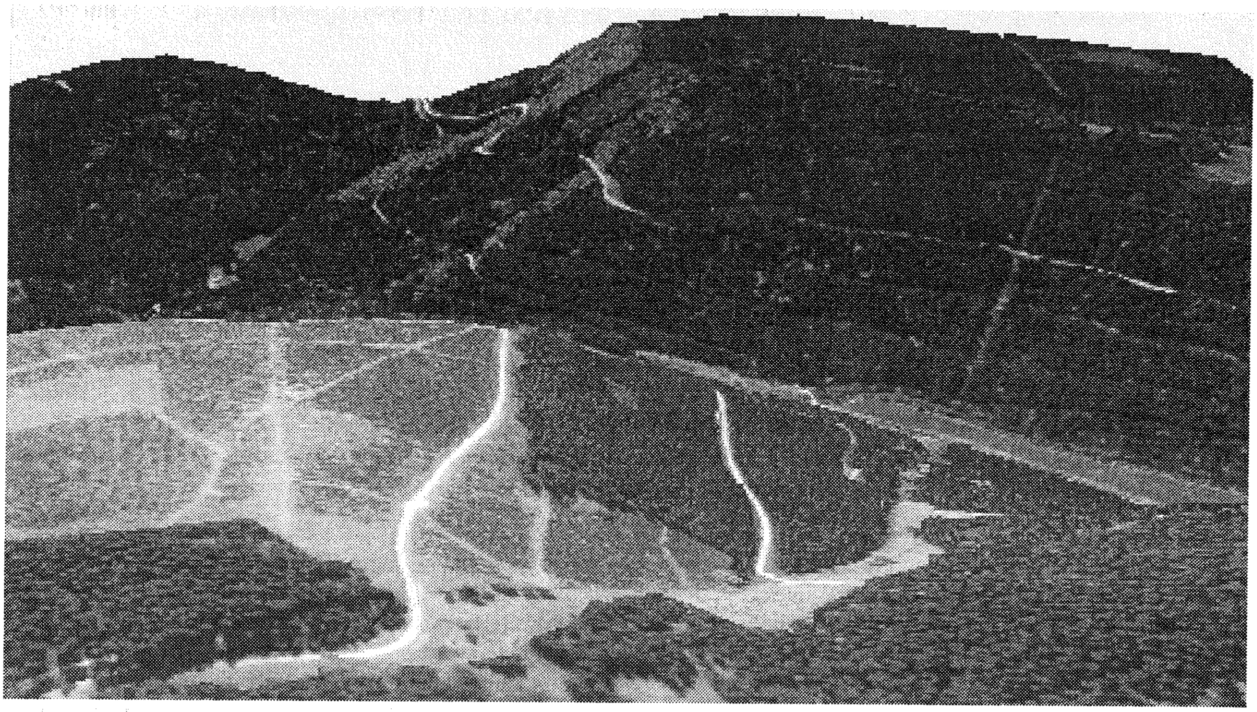

Fig. 7: Synthesized view of landscape model with continuous roads and elevated forests

7. CONCLUSION

The presented system exploits prior knowledge about the scene to improve the realism of the model. The explicit knowledge representation with semantic nets and rules eases the adaptation of the knowledge base to new tasks. The advantage of the system is that the knowledge can constrain the model parameters and select object specific surface primitives. Occluded object parts and lost details due to image resolution are added to obtain a consistent model. Modelling takes care of what is important for a realistic impression of a human observer, e.g. planar roads and height steps at forest edges.

Further work will focus on the development of the control for interpretation and exploit multiple sensors and especially prior interpretations of the scene represented in the german topographic and cartographic information system ATKIS.

REFERENCES

Ackermann, F., Krzystek, P., "MATCH−T: Automatic Mensuration of Digital Elevation Models", *Proceedings of Technical Seminar of the Sociedad Espanola de Cartografia Fotogrammetria y Teledeteccion*, pp.67−73, Barcelona, 12th April 1991.

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.

Foresti, G.L., Murino, V., Regazzoni, C.S., Vernazza, G., "Distributed spatial reasoning for multisensory image interpretation", *Signal Processing*, 32, pp. 217−255, 1993.

Gimmelfarb, G. L., Zalesny, A. V., "Low−Level Bayesian Segmentation of Piecewise−Homogenous Noisy and Textured Images", *International Journal of Imaging Systems and technology*, Vol. 3, pp. 227−243, 1991.

Koch, R., "3−D Surface Reconstruction from Stereoscopic Image Sequences", *ICCV'95*, Boston, June 1995.

Lemmens, M., "A survey on Stereo Matching Techniques", *ISPRS Conference*, Kyoto, Japan, 1988.

Liedtke, C.−E., Grau, O., Growe, S., "Use of Explicit Knowledge for the Reconstruction of 3−D Object Geometry". *6th International Conference Computer Analysis of Images and Patterns CAIP'95*. 6.−8. Sept. 1995 Prag.

Liedtke, C.−E., Blömer, A., "Architecture of the Knowledge Based Configuration System for Image Analysis CONNY", *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, The Hague, Vol. I, pp. 375−378, IEEE Computer Society Press, September 1992.

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

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.

Mester, R., Franke, U., "Statistical model based image segmentation using region growing, contour relaxation and classification", *Proc. of SPIE Conf. on Visual Communication and Image Processing*, Cambridge Mass., Vol. 1001, November 1988.

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.

Shapiro, S. C. (Ed.), "Encyclopedia of Artificial Intelligence", *John Wiley & Sons*, New York, 1992.

Tönjes, R., "Realistic Landscape Modelling with High Level of Detail", *IEEE International Conference on Image Processing*, Austin, Texas, 13−16 November 1994.