Control of Scene Reconstruction Using Explicit Knowledge

Ralf Tönjes

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

Key Words: Knowledge−Based System, 3D−Reconstruction, Image Understanding, Digital Terrain Models

Abstract

Applications such as landscape planing, environmental monitoring, and flight and driving simulators have a high demand for realistic landscape models. Quantity, precision and the type of models ask for methods which automate the model generation by evaluation of remote sensing data. The presented modelling system AIDA tackles the demand for efficient representation and high realism by integrating a priori knowledge about the appearance of the objects in the scene to derive object specific constraints for 3D–reconstruction. This requires an image interpretation to assign a meaning to the objects in the scene. For explicit representation of the declarative and procedural knowledge a problem–independent formalism based on semantic nets and rules is used. It provides both a data–driven and model–driven control strategy.

1. Introduction

Presently there is a great demand for methods which automate the generation of landscape models from remote sensing data. These models are required for landscape planing, map update, environmental monitoring, and flight and driving simulators. Especially landscape visualization is becoming an important tool for earth scientists, environmental researchers and civil engineers. These tasks require realistic 3D–models.

The common approach to recover the topography from aerial images is to use stereo matching techniques for overlapping aerial images [1][2]. To obtain 3D–models for efficient visualization the height map is subsampled and approximated by a polygon mesh in space. The detailed geometric and photometric properties are modelled by projecting the aerial images onto the polygon surfaces.

This approach often results in models that do not meet the expectations of a human observer who knows that the edges of forests exhibit a height step and that roads run continuously. The faulty breaklines are caused by the mesh approximation which does not correspond with the object boundaries. Furthermore, reconstruction suffers from inaccuracy of the sensor data and occlusions of 3D–objects in their 2D–projections. To overcome these problems the presented modelling system AIDA uses scene interpretation to recognize objects in the scene and to exploit object specific constraints for 3D–reconstruction [3].

To ease the adaptation of the knowledge about the objects as well as the analysis strategy for new modelling tasks, the knowledge base has to be formulated explicitly. In the literature various paradigms for knowledge representation have been suggested using formal logic, fuzzy logic, frames [4][5], semantic nets [6], production systems [7][8] and neural nets [9]. Often knowledge about the objects and the control strategy is structured hierarchically [5][7].

To cope with imprecision and uncertainty of sensor data the analysis systems has to combine bottom–up and top–down image analysis strategies, like SIGMA [7] and ERNEST [6].

Usually the analysis results in many competing hypotheses. To reduce the search space analysis has to consider only the most promising hypotheses. To solve this indexing problem search methods are employed, like the A*–algorithm [6] or production rules [7][8] which exploit heuristics.

Most image analysis systems have been tested for selected examples, like the recognition of airports [8], roads [4][7], and buildings [7] in aerial images. The data sets were small. Furthermore, image analysis should be able to take advantage of existing partial interpretations, i.e. GIS–data.

Since none of the systems met all the requirements for 3D–reconstruction the system AIDA has been developed.

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

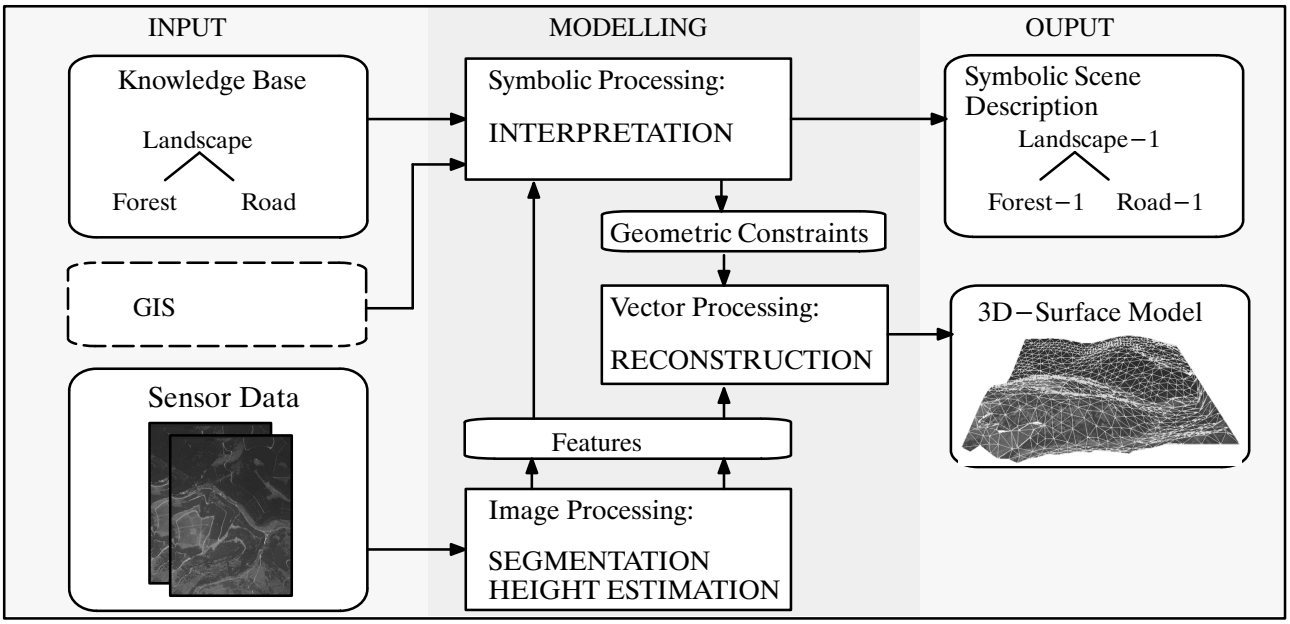

Fig. 1: Architecture of the knowledge based modelling system AIDA

2. System Overview

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

Image processing: The overlapping aerial images are rectified in such a way that the epipolar line coincides with the image scanline to ease the search for homologous points. Consecutively, a height map is computed from the stereoscopic image pair. Furthermore, 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 scene–specific semantics to them resulting in a symbolic scene description. The interpretation algorithm is designed to take advantage of available GIS data. The connection to a GIS data base is under construction.

Vector processing: From object semantics 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 Paradigms

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

- objects,
- tasks,
- sensors, and
- strategies.

3.1.1 Objects: Knowledge about structures can be represented efficiently by semantic nets. Semantic nets consist of nodes and edges between them. Here, the nodes of the semantic net represent the objects in the scene or their sensor specific realization respectively. The nodes are implemented as frames which contain a collection of attributes. Furthermore, an object has methods, i.e. functions, assigned to it to compute the attribute values. Additionally, an object can possess a method to segment the object in the image data.

The edges or links of the semantic net form the relations between the objects. Different relations describe the decomposition of an object into its obligatory (part–of) and optional parts (optpart–of), the specialization of an object (is–a), the concrete sensor specific realizations (con–of), and the feature in the segmented image data (data–of). The relations are exploited for object recognition. 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.

Geometric relations are represented explicitly, only if a strategy needs them to propagate interpretation via that relation. Relationships which only restrict the value are represented as attributes.

3.1.2 Task: The task specializes the modelling demands and reduces the problem domain.

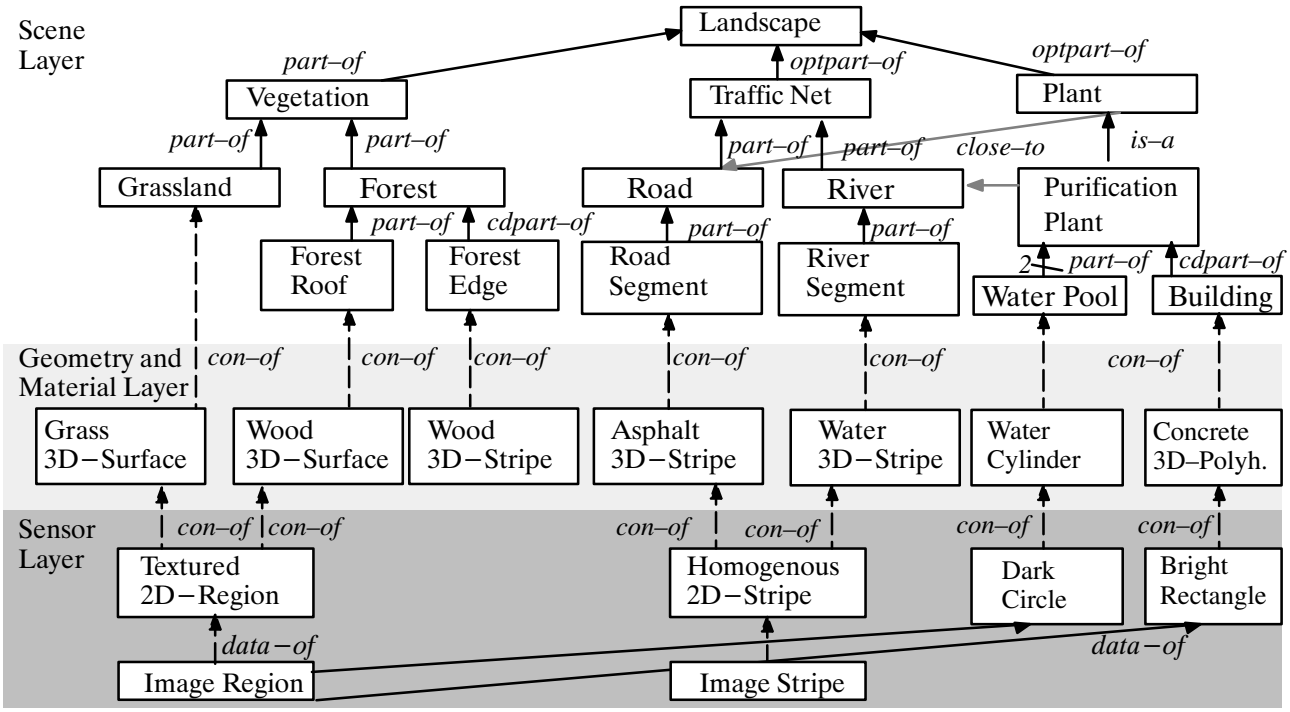

Fig. 2: Simplified semantic net representing prototypes of landscape objects

3.1.3 Sensors: Sensors like cameras transform objects from a spatial into another, here pictorial, representation, using geometric and radiometric transform characteristics.

In the semantic net the sensor transformation is modelled by the 'con–of' relation (fig. 2). Methods assigned to the object attributes restrict the expected range of attributes top–down and obtain the measured value bottom–up from the sensor. To model uncertainties the attributes are described by minimum and maximum values.

3.1.4 Strategy: Strategies state how and in which order scene analysis has to proceed. A strategy is represented by a set of rules. These rules exploit the knowledge represented in the semantic net to control interpretation. For each relation exist associated 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. The action part adapts the interpretation state of the focussed node accordingly. The knowledge, for example, that an object n_0 is detected and can be denoted as 'complete instance', if all its parts $n_i \in P$ are detected, is represented by following rule:

Rule–complete–instanciation:

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

3.2 Knowledge for Landscape Modelling

Figure 2 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 semantics. 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. The geometry and material layer is sensor independent and is used for sensor fusion. The application dependent domain knowledge is represented in the scene layer. Each layer uses a common appropriate vocabulary: The geometry layer uses a world coordinate system (e.g. Gauss–Krueger) while the sensor coordinate system is referring to the image raster.

Here, in the context of landscape modelling, vegetation, traffic net, and plants are distinguished. The vegetation is assumed to be composed of grassland and forests. They are detected in the sensor data based on their material dependent texture properties. Roads and rivers appear as homogenous bright and dark stripes in the sensor data respectively. The purification plant consists of at least two water pools and an administration building, which show up as dark circles and bright rectangles in the sensor data. The initial concepts 'textured 2D–region', 'dark circle', 'bright rectangle', and 'homogenous 2D–stripe' possess methods for segmentation.

4. Segmentation

The initial concepts in the sensor layer are detected, i.e. instanciated, by segmentation of images. Presently, two different segmentation methods are implemented: The segmentation of homogenous stripes and textured regions [10].

The extraction of homogenous stripes is based on a gradient filter, parallel line finding and thresholding. The segmentation of textures is achieved by a supervised segmentation. The different texture statistics are modelled by Gibbs random fields, a special kind of Markov random fields [11]. Figure 3 shows the segmentation results.

The segmented image regions and stripes are represented as nodes in the semantic net, becoming instances of 'image region' and 'image stripe' respectively. For the segmented regions various texture and shape features are computed.

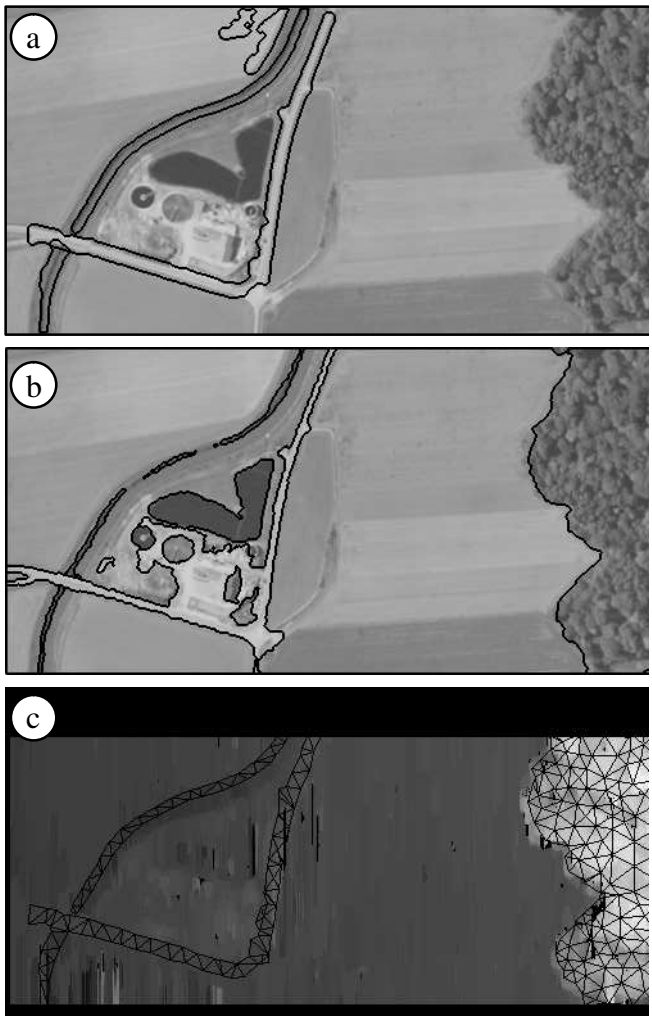

Fig. 3: Aerial image with overlaid segmentation results for a) homogenous stripes and b) textured regions. c) Disparity Map with overlaid 3D−surface mesh of forests, roads, and rivers extracted from a) and b) by interpretation.

The features (e.g. moments, fourier descriptors of contour, etc.) are used to select circles and rectangles from the 'image region' instances.

5. Interpretation

5.1 Control

The aim of image interpretation is to assign a symbolic description, i.e. semantics *S(D)* to the data *D*. Image interpretation exploits the semantic prototype net G_c to instantiate hypotheses of objects expected in the scene. Four different states of the object recognition are distinguished: hypotheses I_H, missing instances I_M, partial instances I_P and complete instances I_c. Hypotheses are not yet verified in the sensor data. Falsification in the sensor data results in missing instances, while verification obtains partial and complete instances successively. Partial instances contain all objects of the sub–layer and all obligatory context independent parts. Additionally, complete instances possess all context dependent parts. Interpretation is complete when the optimal instance I_c^* of the goal concept C_g , here landscape, is found based on the semantic prototype net G_c and the features F which are segmented in the image data. To judge the quality Q of the interpretation for each instance a compatibility measure is computed. Thus the resulting symbolic description is defined as

$$S(D) = I_C^*(C_g) = argmax\ Q(I(C_g)|\ G_c,\ F) \quad (1)$$

Interpretation employs an inference engine which controls the execution of the rules. According to the set of rules and their priority different strategies are possible. To avoid an explosion of the search tree, here, hypotheses are primarily propagated top–down from the scene layer to the sensor layer to verify them in the sensor data. At the sensor layer the propagated hypothesis 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. Each possible interpretation is documented by a search node which contains all concepts and their current interpretation state. If 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. An A^*–algorithm selects the interpretation judged best for further investigation.

5.2 Verification of Hypotheses

In the following the verification of hypotheses about road detection is described (fig. 2). First, a hypothesis for a road is instanciated. The condition of the rule for complete instanciation (see 3.1.4) is not fulfilled because the obligatory parts are not present. Hence the rule for top–down propagation of a hypothesis is executed, instanciating I_{H1}*(road–segment)*. The rule is executed repetively until the initial concept 'homogenous–2D–stripe' is instanciated, which can be tested in the sensor data by calling a segmentation function. On success a complete instance I_C*(homogenous–2D–stripe)* is generated. Now the condition of the rule for complete instanciation is true resulting consecutively in I_c*(asphalt–3D–stripe)*, I_c*(road–segment)*, and I_c*(road)*.

This scheme can also be used to verify (old) map information of a GIS in more recent aerial images. In this case a semantic net of hypotheses is instanciated from the GIS data and consecutively tested in the image data by top–down propagation.

5.3 Completion of Missing Object Parts

In aerial images the edges of forests are usually not visible. Thus they are modelled as context dependent parts (cdpart–of) with no representation in the sensor layer (fig. 2). According to the priority of the rules the control of interpretation looks at first for context independent parts, i.e. forest roofs. When a complete instance I_c*(forest–roof)* exists, forest can be instanciated partially, but not completely since the forest edge is still missing. A request for the forest edge is propagated top down. Because 'wood–3D–surface' possesses no 'data–of' link, the I_H*(wood–3D–surface)* can be converted to a complete instance I_c*(wood–3D–surface)*. This information is propagated bottom–up causing complete instanciation of I_c*(wood–3D–surface)* and I_c*(forest)*. The completion of occluded objects is important for 3D–reconstruction.

5.4 Recognition of Complex Structures

Complex structures like the purification plant in figure 2 are composed of parts. Since plants have a road access and the purified water is usually released into a river, the location is constrained. This spatial relation is represented by an attributed relation 'close–to' that guides the search of neighboured objects. The value of the attribute 'distance' for the relation constrains the search area in the neighbourhood. For these locations the hypothesis I_H*(purification–plant)* is tested by sequentially generating and testing hypotheses for the first water pool, the second pool, and finally the building.

6. Reconstruction

The model–driven reconstruction exploits a priori knowledge about object geometry to restrict the parameters for reconstruction. While data driven reconstruction uses only a few and general geometric constraints, interpretation offers the facility to exploit object specific geometric properties. Interpretation yields the segmentation of aerial images into various regions, such 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. Furthermore the object semantic controls mesh generation. Roads and rivers 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. Figure 3c shows the net for selected objects.

The semantics attached to the model parts allow an object specific post processing: Objects can be artificially refined by adding details which are invisible to the sensor from computer graphic libraries. For close–up views of forest edges, for example, synthetic trees with fine transparent leaf structure are placed in front of the edges (fig. 4).

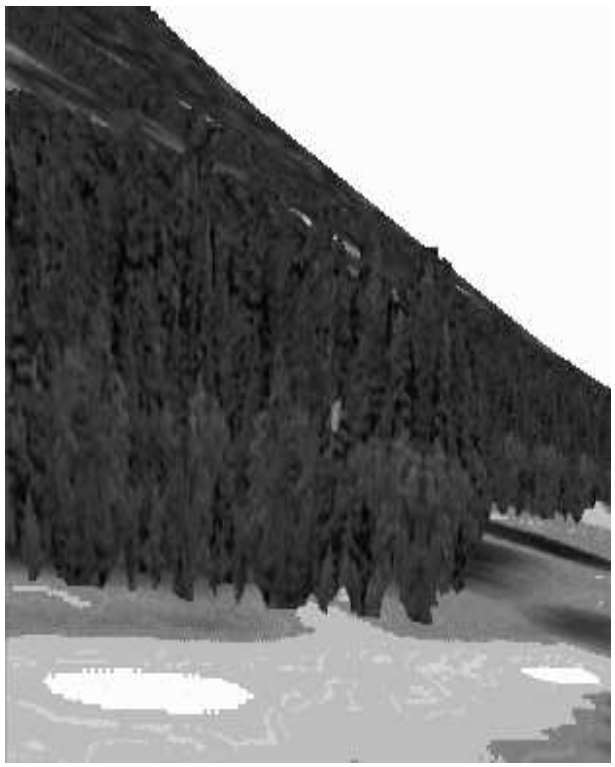

Fig. 4: Close up view of a forest edge

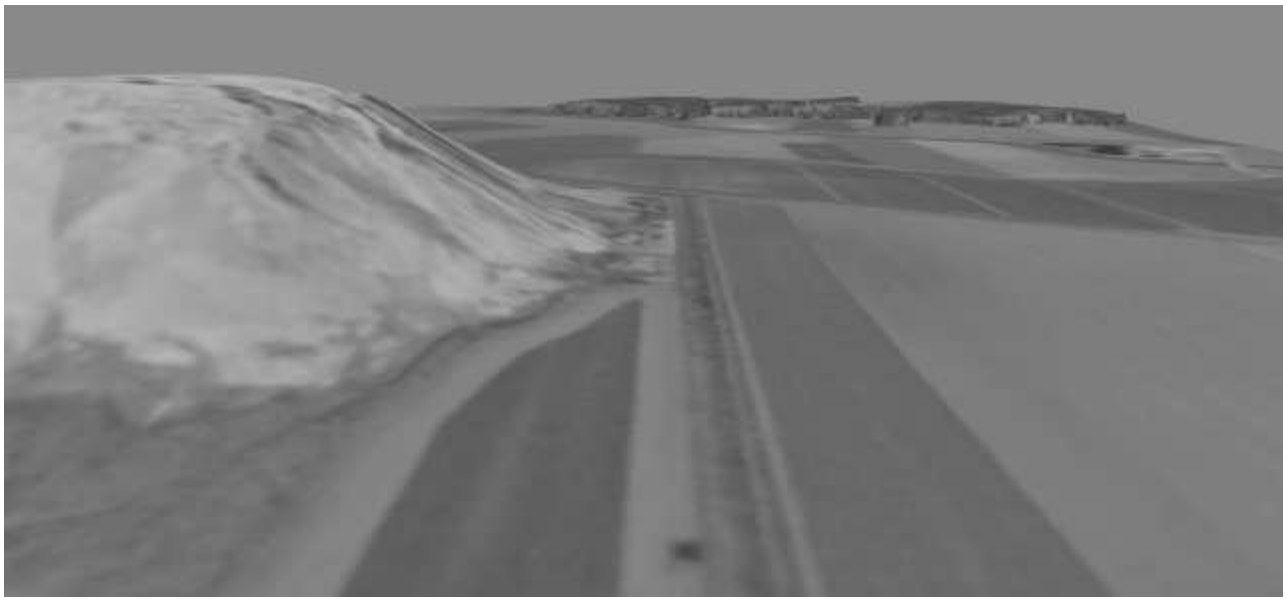

Fig. 5: Synthesized view of a landscape model with continuous roads and rivers and elevated forests

7. Results

Figure 5 shows the synthesized view of a reconstructed landscape model with continuous roads and rivers and height steps at the edges of forests. The above mentioned section with the purification plant can be seen on the right in the background. On the left a hill of a potassium plant can been seen. The landscape (1km x 1.2km) consists of approximately 10.000 Polygons and a texture map of 1024x1280 pixel. For interactive exploration of the scene in real–time the model can be visualized on a graphic computer.

The control algorithms of the system AIDA are implemented in C++ with a Tcl/Tk interpreter interface. The low level algorithms for segmentation, stereo matching, and mesh generation are realized in C and C++.

8. Conclusion

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

Further work will focus on the integration of GIS data (the german topographic and cartographic information system ATKIS) for interpretation.

References

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

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

[3] **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.

[4] **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.

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

[6] **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.

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

[8] **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.

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

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

[11] **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.