Knowledge Based Modelling of Natural Scenes

Oliver Grau and Ralf Tönjes

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung

Division "Automatic Image Interpretation", Chair Prof. C.–E. Liedtke

Universität Hannover, Appelstrasse 9A, 30167 Hannover, Germany

E–mail: grau@tnt.uni–hannover.de, toenjes@tnt.uni–hannover.de

Abstract

A knowledge based approach for the automatic generation of 3D models from stereo images is presented. Two different application fields are shown: The modelling of houses and the modelling of landscapes from aerial images. The models are used for computer graphic applications. A human observer is in possession of a priori knowledge about the scene. To meet his expectations and thus to improve realism the presented approach employs explicit knowledge. The knowledge base is exploited to derive constraints for surface reconstruction and to complete partially occluded models.

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, architecture and landscape visualization and film industry have a high demand for realistic models. Quantity, precision and the kind of models ask for methods that automate the model generation of real objects and are applicable to outdoor scenes, i.e. do not require active sensors or special illumination.

An approach for generating realistic 3D models of natural scenes based on photogrammetric stereo [1,2,3] is presented. Reconstruction of a 3D model from 2D projections is an inverse and underconstrained mathematical problem. Two kinds of solutions for this problem can be distinguished: 1. the data driven integration of data from multiple sources (multiple sensor, multiple views, multiple methods) and 2. the use of knowledge in a model driven approach.

The data driven 3D modelling attacks the ambiguity by assuming simple, i.e. smooth surfaces [1,2]. 3D reconstruction is achieved bottom–up by consecutive grouping of depth measurements to a fixed kind of surface primitive like a triangular patch [2] or a Bézier or NURBS patch.

The knowledge based or model driven approach incorporates a priori knowledge about the scene in the 3D modelling process. The reconstruction is achieved mainly top–down by generating hypotheses about the scene from the knowledge base and

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

testing the hypotheses with the data. Further on the model driven and the data driven approach can be integrated. In this case the data driven phase of the modelling process is controlled by the knowledge based system. An advantage is that the system can limit the parameter space and select more specific surface primitives, e.g. either polygons or splines.

Another important application of the knowledge based system is the completion of the 3D model to obtain a plausible model for the human viewer even when no image information is available. These cases are caused by occlusions or insufficient image resolution.

In the presented approach the scene objects are modelled by using a polygonal mesh for the geometry and mapped texture for fine sub–structure.

The principal goal, a realistic looking scene, can be reached in two ways: The first approach is to model the objects as precise as possible and with the maximum level of detail. The other approach is to take care of what is important for the impression of a human viewer. The latter approach leads to more compact models, a common requirement for many applications, and can be achieved by the use of knowledge in the modelling process. Chapter 2 explains the general kinds of knowledge applicable for the modelling of 3D objects and how these are processed in the presented scene analysis system AIDA.

The chapter 3 describes the use of knowledge in the application of modelling houses. The chapter 4 describes an other application field, the knowledge based modelling of landscapes. The paper terminates with a conclusion.

2 Using explicit Knowledge for 3D Modelling

Five processing steps can be distinguished in the presented approach for the modelling of 3D objects:

I. Interpretation of the signal
II. Deriving constraints from the knowledge base
III. Reconstruction of the object geometry using the obtained constraints
IV. Completion of object parts not supported by the signal (optional)
V. Texturing of the 3D model

The interpretation of the signal assigns a meaning to the scene content represented in its 2D image projections. This part of the AIDA system is an image understanding system [7,6]. The assignment of meanings in terms of the generic model description is necessary for the selection of the proper constraints (step II) that can be applied for the reconstruction of the object surface (step III) and optional the completion of the object parts which are not visible (step IV). In the last step of the modelling process the images are projected onto the reconstructed object surface and stored in a texture map to obtain a photorealistic model.

Both kinds of knowledge the (generic) description of the model used in step I and the constraints used in step II can be formulated in a single knowledge base as a semantic net. The following section describes the used formalism.

2.1 Representation of the Knowledge

In AIDA the knowledge is mainly represented as semantic nets [4,5] implemented in a frame–like structure. Semantic nets have been applied to a number of image and signal understanding problems [5,6]. Each node of the net is represented by a frame, which can be considered as a collection of attributes. Further the frames constitute the concept of inheritance, a basic feature of object oriented programming. In the part of the semantic net that describes the generic world model, the nodes are called concepts, while the nodes describing the objects found in the scene are called instances. The edges or links of the semantic net form the relations between the nodes.

The is–a link connects a general node with a specialization, like flat roof as special kind of roof. The concept of inheritance in semantic nets is realized by this type of link. For example a flat roof has all the attributes of the general concept roof except these that are defined different in the flat roof node. Further a flat roof inherits all links from roof. If for example roof has a part–of link to chimney, then flat roof has this link too.

The right side of figure 1 shows a part of a semantic net describing a small scene. The scene is composed of parts, indicated by the part–of links from the sub objects to the parent node. The links can have attributes like the condition that there are more or equal than four parts or that a link is optional. So a ”car” is optional in figure 1 and must not be present in the scene.

Another important link is the concrete–of link introduced by the ERNEST system [6]. It links two conceptional layers or levels of abstraction. In figure 1 the node ”wall” is connected by a concrete–of link with the node ”polygon”. The link assigns a concrete geometric realization to ”wall”. Currently three conceptional layers are implemented in AIDA: The semantic scene layer, the geometric 3D layer and the 2D layer.

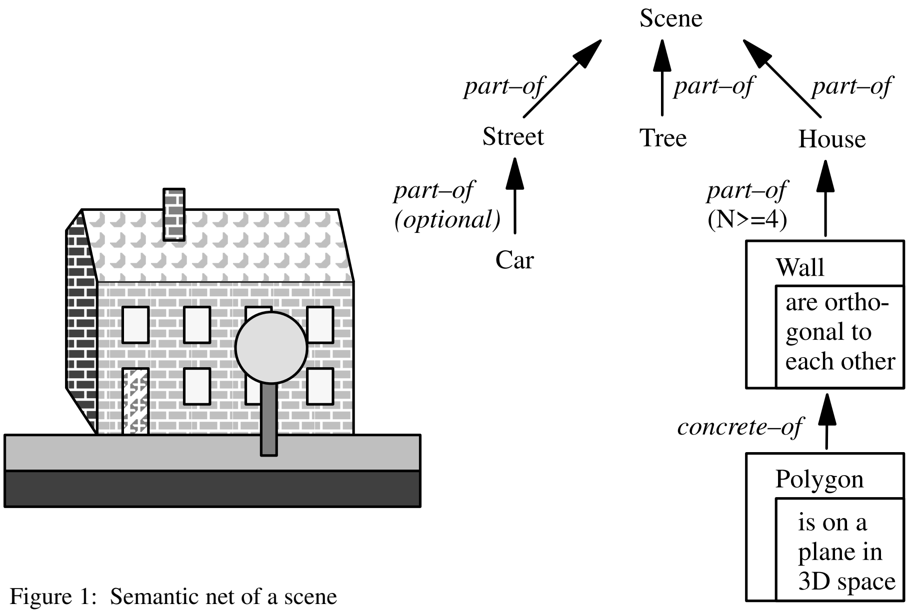

Figure 1: Semantic net of a scene

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

Following the concrete–of links from the top level nodes, the scene specific semantic of an object is transformed to a geometric or photometric concrete realization in the image data.

The last principal link type is the instance–of link. This link is established at the interpretation phase of the analysis and connects an instance supported by the signal with a concept in the generic model description. The next section gives a short outline of the algorithm that performs the interpretation.

2.2 Interpretation of the Signal

The AIDA system employs a graph processing algorithm to instantiate nodes and links, i.e. to assign a meaning to the signal. The algorithm is based on the speech and image analysis system ERNEST [6] extended by some features needed for the 3D scene analysis. The problem independent control strategy is implemented as a small set of rules for instantiation guided by an modified A* graph search algorithm to select the optimal instance.

In a model driven phase hypotheses for the appearance of a wall in the image are generated, judged, and the most probable is selected. The algorithm is also applicable for a bottom–up analysis. In this case the initial goal is not a node of the top conceptional layer like ”house”, but a node from a lower layer like ”2D polygon”. In both cases the algorithm stops if the scene is fully interpreted.

An important feature of the interpretation system is that it limits the parameter space of segmentation and low level search algorithms in the model driven phase. If a new view of the scene is analyzed, the position of already found objects can be predicted thus reducing the search area.

The assignment of the semantic nodes in the lowest conceptional layer to the image data is achieved by image segmentation or provided by a human operator. In the first case the system represents an image understanding system. The presented approach supports both ways. The goal is to create a system that works fully or nearly automatic, but in cases the system fails a human supervisor should have full control of all system components. Another point of view of the system is, that it offers the flexibility of an interactive system, but assists the operator in laborious and difficult tasks and minimizes the interactions neccessary.

2.3 Knowledge based Surface Reconstruction

The most important application of using knowledge in the modelling of 3D objects is to obtain a better conditioning of the surface reconstruction. The goal is to reconstruct the depth information, which is irreversible lost in the projection of a surface and can be reconstructed only noisy and unreliable with local depth measurement techniques like shape from X, e.g. shape from stereo.

A model driven approach offers several possibilities for the use of geometric constraints which can be formulated as explicit knowledge and in general the demand of independent depth measurements can be reduced significantly. The following constraints are used in the presented approach:

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

a) The most specialized surface primitive is selected. E.g. a plane polygon is used to model a wall.

b) Geometry is restricted by spatial relations. E.g. the walls of a house are forced to be orthogonal to each other.

c) Knowledge about the object structure like connections of sub–parts is exploited. E.g. two roof parts should be closed at the peak.

The used geometric constraints are application dependent. In the chapters 3 and 4 some examples for the two presented application fields will be explained in detail.

2.4 Knowledge based Completion of Models

The model reconstructed from the image data may be incomplete due to lack of signal caused by occlusions or insufficient image resolution. The knowledge base provides the missing parts. Selection of the right model from the knowledge base is achieved by image interpretation. Model completion employs different kinds of knowledge:

a) Knowledge about the object structure is used to add invisible parts like occluded backwalls of houses or in aerial images invisible edges of forests.

b) The geometric and photometric features of the parts without signal are obtained from the generic model and further restricted by the detected parts of the object. E.g. the texture of the wood edge depends on the visible adjacent part of the wood.

3 Knowledge based 3D Modelling of Houses

Houses are composed of very regular geometric structures with several spatial relations, like plane polygons which are orthogonal to each other. These facts are common knowledge for a human observer. Hence he will not accept a model that has rounded edges or non planar walls. In a knowledge based approach the geometric relations are used to improve the realism of the 3D models.

This application takes full advantage of the AIDA knowledge representation. A semantic net contains a prototype for houses, their components and the applicable geometric constraints.

Figure 2 shows the use of the knowledge that the wall geometry represents a polygon in 3D space. The local depth measurement performed by a stereo disparity estimator is noisy (Fig. 2 b), because of quantization and failure in locations where no texture is present. It is known that the regarded wall lies on a plane. Hence a regression for all disparity measurements within the marked wall region is performed. Given the disparity each image point p_i can be reprojected to a 3D point P_i according to the used camera model. The regression weighted with the certainty yields a more accurate estimation of the plane wall. The regression integrates the depth measurements of several stereo image pairs assuming known camera position.

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

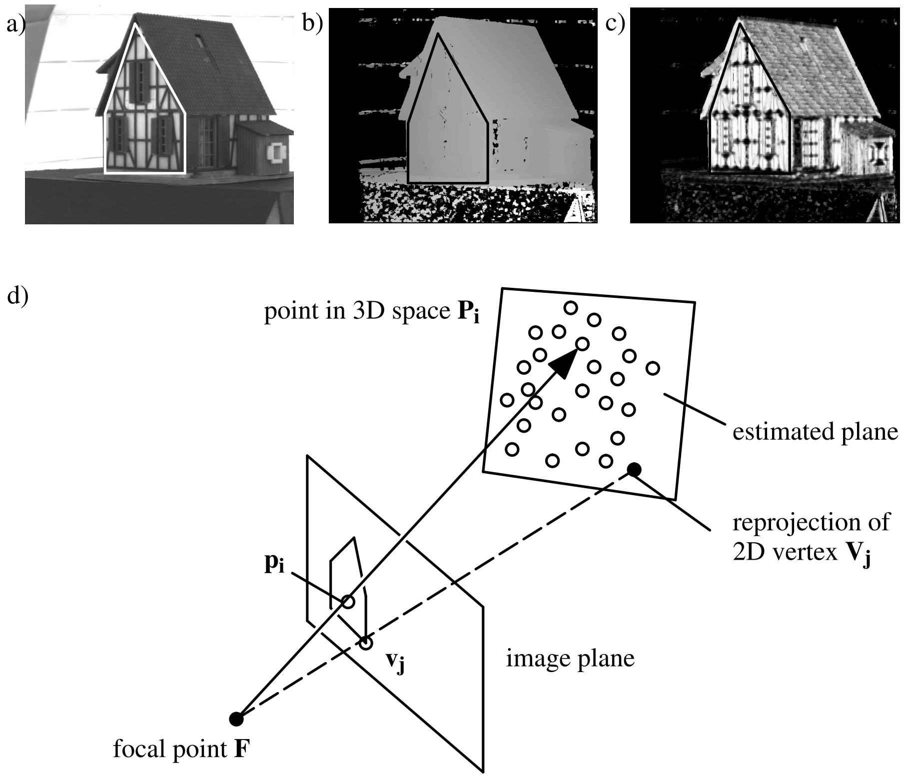

Figure 2: (a) Luminance image with marked region, (b) noisy disparity map, (c) certainty map of the disparity measurements, (d) and regression over all region points

To get a polygon in 3D as a first step the vertices of the 2D polygon are reprojected onto the 3D plane and form a 3D polygon. The reprojection is found as the point of intersection of the line of sight from the focal point **F** through the image plane at the vertex $\mathbf{v_j}$ and the estimated 3D plane.

After the reprojection of all polygons the knowledge about connections of sub–parts is exploited. This is explicitly formulated for the walls and the roof of the house. A spatial relation introduces the constraint that the neighboring polygons meet at the line of intersection of their 3D planes. This constraint yields closed models.

3.1 Results

The figure 3 shows the result of the presented modelling approach. Only three stereo pairs from different views were used to model the house from a laboratory scene. The model is only composed of a small number of polygons. The walls are typically modelled as one plane polygon and are connected at the line of intersection as outlined in the previous section. The knowledge base used consists of a semantic net with 41 nodes and 64 links. The net is able to represent simple houses like the one depicted in figure 3.

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

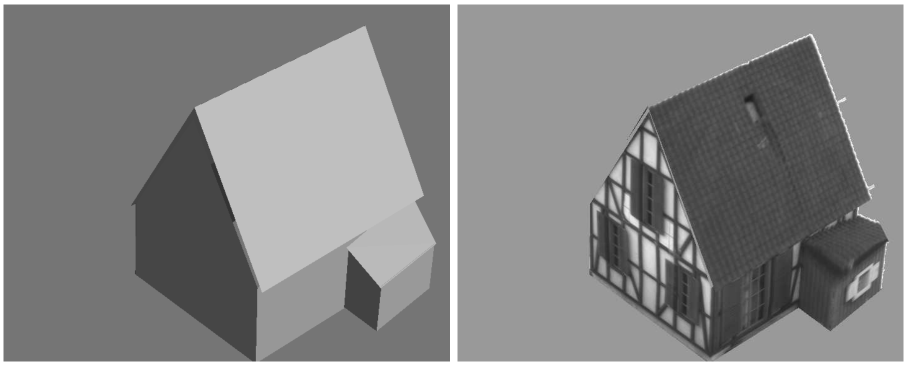

Figure 3: Synthesized view of the model in (a) shaded and (b) textured presentation

Since currently no automatic segmentation methods of the 2D polygons in the image are implemented, the region must be marked manualy. However, no high accuary is required because of the regression of the 3D plane and the calculation of the intersecting polygon edges.

4 Knowledge based 3D Modelling of Landscapes

Conventional landscape modelling ignores the object meaning and applies only stereo matching techniques to construct a terrain model. However this approach hardly works for complex scenes and fails when a high level of detail is required. Natural culture (wood, trees) and man made objects (buildings, roads) are not modelled or appear faulty. A human observer is in possession of a priori knowledge about the scene and expects height steps at the edges of forests and roads to run continuously. To obtain more realistic models this a priori knowledge must be integrated in the modelling process using image segmentation and interpretation.

a) b) c)

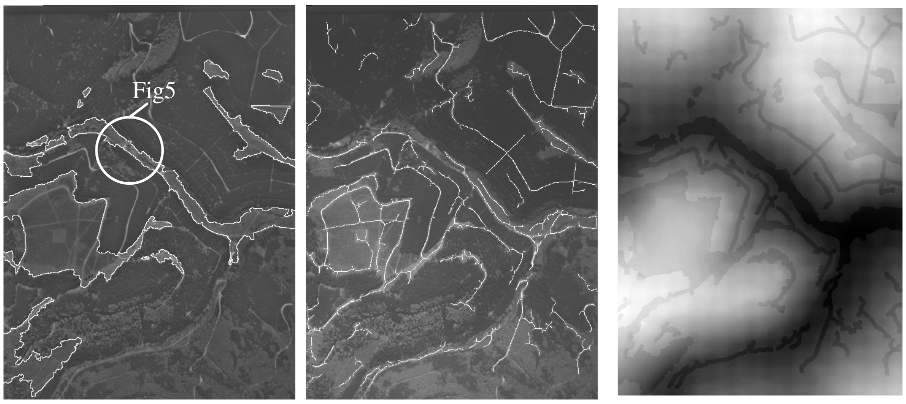

Figure 4: Aerial image with (a) overlayed segmented woods, (b) roads and (c) height map with elevated woods

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

Knowledge is used both, for scene interpretation and model improvement. In the first case the knowledge indicates what objects are to be searched for and which are the eminent features for detection. In the second case knowledge is used to correct erroneous and complete missing signals. The landscape modelling employs knowledge about the photometric object features for object detection and geometric knowledge for model improvement. The height discontinuity of wood edges is missed by depth measurement due to image resolution. However, in aerial images woods can be segmented based on their texture [8]. Figure 4a shows the segmented woods represented by white contours. The segmented wood region is used as a mask to elevate the woods in the height map, where bright areas correspond to higher terrain (figure 4c). Further, to consider roads passing woods the road regions are detected (figure 4b) and not elevated.

For efficient visualization the model geometry is approximated by a polygon mesh. The polygon representation must consider height steps and breaklines. Continuous surfaces such as wood and roads are modelled by a separate polygon mesh.

In the aerial image the edges of woods are invisible. This lack of signal is overcome by a priori knowledge about wood models to photo texture the inserted wood edges.

Currently only a bottom–up analysis is implemented, allowing no iterating refinement. However, even integration of solely data driven image interpretation in the modelling process improves model realism significantly.

4.1 Results

Figure 5 shows a synthesized view of a landscape modelled from overlapping aerial images. Realism is improved by approximating wood and roads by a separate polygon mesh and inserting a height step at the edge of the forest.

The symbolic model description eases the postprocessing with computer graphics techniques. Thus the inserted edge of the forest is textured and the segmented but flat building is replaced by a CAD model.

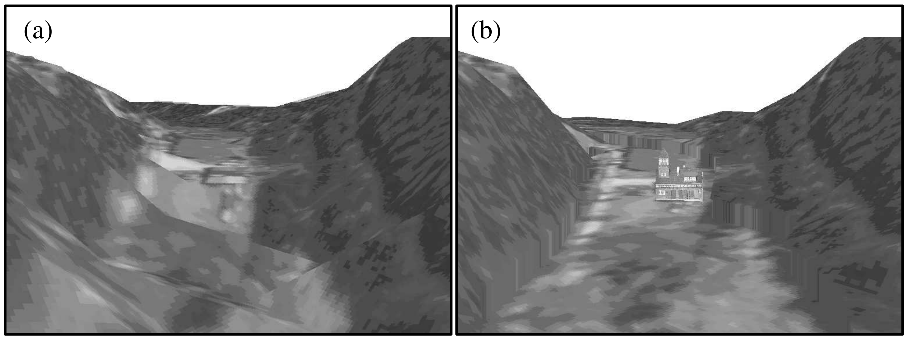

Figure 5: Synthesized close up view of landscape section marked in figure 4 (a) without and (b) with usage of a priori knowledge by image interpretation.

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994

5 Conclusions

The paper shows how a priori knowledge about scene objects in 3D modelling can be used to improve the realism of the visualized models. Knowledge based modelling focuses on model features that are sensitive for faults a human observer would detect.

The explicit formulated knowledge is used to derive geometric constraints, that lead to a better conditioning for the solution of the inverse projection and reduce the demand for independant depth measurements significantly. So the house model presented in section 3.1 could be modeled only from three camera views.

Further the knowledge base tackles the problem of incomplete objects due to lack of sensor signal as shown for the modelling of wood edges.

3D scene analysis results in an iconic 3D model and a symbolic scene description implemented as semantic net. The symbolic scene description offers the possibility to answer inquiries. For example the net can be asked for the number of houses with flat roofs in the scene. The query can also be extended with an action directive like: 'Set the color of roofs to red if the house has more than two floors'. The symbolic scene description establishes an interface between computer vision and computer graphic techniques.

6 References

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

[2] Koch, R., "Model–Based 3D Scene Analysis from Stereoscopic Image Sequences", *ISPRS Conference,* Washington, USA, Aug. 1992.

[3] Tönjes, R., Grau, O., Koch, R.: "Analyse durch Synthese Modellierung von 3D Objekten in Stereobildfolgen", 1. Workshop visual computing, Darmstadt, 1994.

[4] Winston, P.H., "Artificial Intelligence", Addison–Wesley Publ. Comp. 1992.

[5] Niemann, H., "Pattern Analysis and Understanding", Springer Verlag, 1989.

[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] Tönjes, R., "Realistic Landscape Modelling with High Level of Detail", IEEE International Conference on Image Processing, Austin, USA, Nov. 1994.

Published in: Image Processing for Broadcast and Video Production
Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, 23–24 November 1994