Use of Explicit Knowledge for the Reconstruction of 3–D Object Geometry

C.–E. Liedtke, O. Grau, S. Growe

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Division "Automatic Image Interpretation", Universität Hannover, Appelstr. 9a,
D–30167 Hannover, Germany, E–Mail: liedtke@tnt.uni–hannover.de

Abstract

The automated generation of 3D CAD models of real objects from different camera views poses frequently problems in regard to man made objects. Models do not match the expectations of a human observer, because house walls are not perpendicular, streets are not planar, windows and doors are not rectangular, etc.. The new knowledge based modeling system AIDA handles these problems by using an explicit knowledge base about the semantics of the scene to be modeled including knowledge about the visual appearance of scene objects. During the analysis of the scene constraints for the modeling are derived automatically and are applied during model generation.

Keywords: Image Processing, Scene Analysis, Knowledge based System, 3–D Modeling, CAD Models, Virtual Reality

1 Introduction

Presently a great demand for highly realistic looking 3–D models of real objects can be observed. These models are used in developing flight and driving simulators, in the movie and TV production, in advertising, education, documentation, in city and landscape planning, etc.. The manual construction employing CAD tools is usually too time consuming and expensive. In addition there is very often a lack of naturalness to be observed. For this reason methods have been developed which derive 3–D models automatically from multiple camera views of natural objects by using binocular stereo [1]. The coarse geometry of the object surface is approximated by a mesh of polygons. The geometric fine structure and the photometric surface properties are modeled by projecting a photo texture onto the surface polygons.

The problem, which arises especially in modeling of man–made objects, like town buildings, is that the automatically modeled surface geometry does frequently not match with the experience and understanding of the viewer. House walls do not become planar, window– and door frames are not perpendicular, walls are not connected, streets are not planar, etc.. The reason is, that due to noise in the data or unsuitable surface properties of objects, the surfaces cannot be modeled with the required accuracy. The human observer becomes confused because of his prior knowledge about the expected surface geometry. There are different approaches to solve the problem. One approach is to increase the measurement accuracy by using more sophisticated methods, by increasing the number of views or by modifying unsuitable surface properties, like surfaces without structure,

Accepted Paper for: 6th International Conference CAIP'95
Computer Analysis of Images and Patterns. Sep. 6–8 1995 Prague.

semitransparent, transparent or highly reflective surfaces for instance by spraying with color spray. In any case, the effort in time, equipment and computational power has to be increased considerably.

The proposed knowledge based system AIDA (Automatic Image Data Analyzer) uses a different approach employing explicitly formulated prior knowledge about the scene [2]. During the analysis of the sequence of camera views the scene is partially interpreted. The modeling process is then made dependent on the interpretation of the respective scene detail, i.e. the context. The goal is to derive 3–D models from camera views which match in their visual appearance with the subjective expectations of the viewer.

Various knowledge based systems for scene analysis have been mentioned in the literature. They differ in the field of application, the paradigms of knowledge representation, the structure of data and the control strategy. For modeling 3–D objects (a) the inaccuracy and unreliability of sensor based features like contours, regions, depth maps must be considered, (b) occlusions must be handled, and (c) large data sets must be managed.

Nearly all systems try to cope with the unreliable results of low level vision modules for instance by permitting combinations of bottom up and top down image analysis strategies like SIGMA [5] or ERNEST [8]. Many have been developed for the interpretation of single images, frequently aerial photos ([4],[5]). None of these can to our knowledge handle the occlusion problem. This special aspect gets more attention in the literature related to close range photogrammetry.

Various paradigms for the representation of domain knowledge and control knowledge and for the structuring of knowledge have been reported. MESSIE [7] uses a blackboard system and production rules. Semantic nets or frames are found in SCIN [3], ERNEST [8] and MESSIE [7]. In SPAM [4] and Foresti [6] different layers of abstraction are described.

Since none of those systems alone could fulfill all of the requirements stated above the new system AIDA has been developed. It permits the automatic modeling of 3–D scenes from various camera views under exploitation of human prior knowledge about the scene content and its geometry.

2 System Overview

Fig.1 gives an overview over the system components. The knowledge base contains prior knowledge about the objects which are expected to be observed in the scene, the relations between the objects, the geometrical properties in 3–D and in 2–D as they may appear in a camera image. The knowledge is represented explicitly in a semantic net.

The input data from the scene to be modeled consists of sequences of stereo image pairs. In a first pre–processing step the stereoscopic camera is calibrated and each image is rectified. The calibration estimates the radial distortion of the lenses and the relative external orientation of both cameras. From the input stereo images a disparity map is calculated based on standard photogrammetic correlation techniques. The camera parameters are used for the calculation of depth values from the disparity map. Interpolation

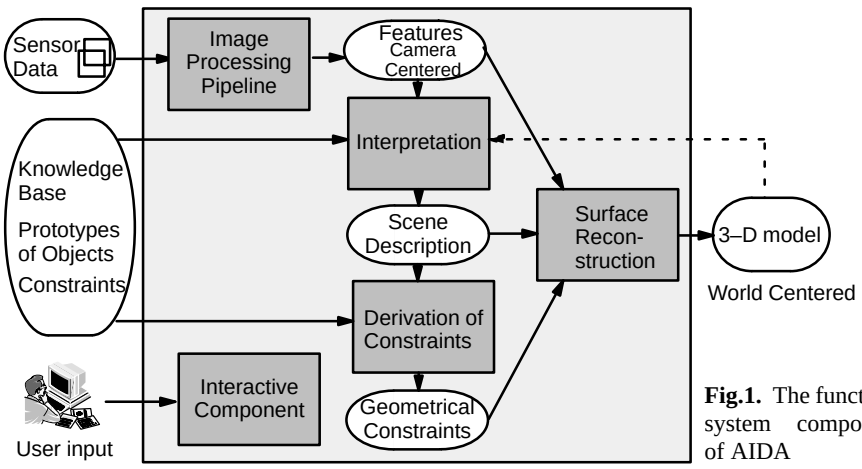

Fig.1. The functional system components of AIDA

techniques are used to arrive at a dense depth map. The methods have been described in more detail elsewhere [1]. The images are segmented using photometric and depth information. An example for the segmentation of the toy–house is presented in Figure 2.

The knowledge base is used for the interpretation of the segmented image resulting in a symbolic scene description. From the symbolic scene description geometrical constraints for modeling are derived from the knowledge base. These geometrical constraints are used on one hand to support the interpretation of the individual scene objects. On the other hand the constraints are used to improve the modeling during the surface reconstruction phase. In this connection wall surfaces are remodeled to become planar, edges become straight and perpendicular, etc..

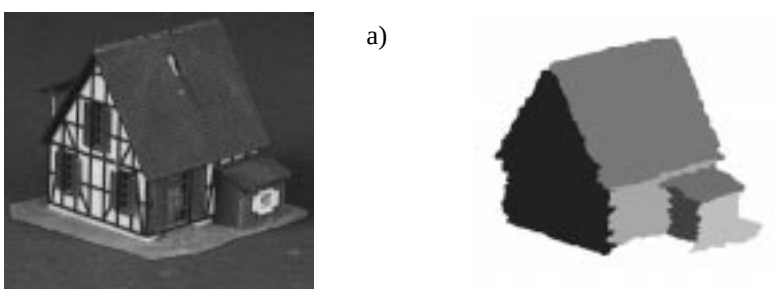

Fig. 2. a) Input image b) Segmented image based on photometric and depth information [1]

3 Knowledge Base

The knowledge about the scene to be modeled is represented explicitly in a semantic net. For the description of objects and their relations a problem independent net language was defined which resembles the net syntax of ERNEST [8]. Extensions have been made to meet the special requirements of 3–D modeling, like the handling of 3–D occlusions.

The knowledge base of AIDA is subdivided into three conceptual layers. The lowest is the *camera centered layer*. It contains all information about the sensor data and the processing results of the image processing pipeline like the regions and contours of the segmented image. As an example, an image region may be described by a 2–D polygon. Images are projections of the 3–D world onto the 2–D camera target. Therefore, the 2–D polygon is interpreted as the projection of a 3–D polygon onto the image plane. The 3–D geometrical objects form the second conceptual layer, the *world centered layer*. The highest layer, the *scene layer*, represents the interpretation of the 3–D geometric objects from the layer below. In this connection the 3–D polygon may be interpreted to be a wall or a window in the scene.

Fig.3 illustrates the structure of the AIDA knowledge base. It contains all the relevant information from the sensor related 2–D domain up to the abstract scene interpretation. This is necessary to support both, data driven and model driven types of processing strategies. In Fig.3 besides the house a camera is modeled in the semantic net. Its parameters describe the position and orientation of the sensor in 3–D space and permit a transformation between world coordinates and image coordinates.

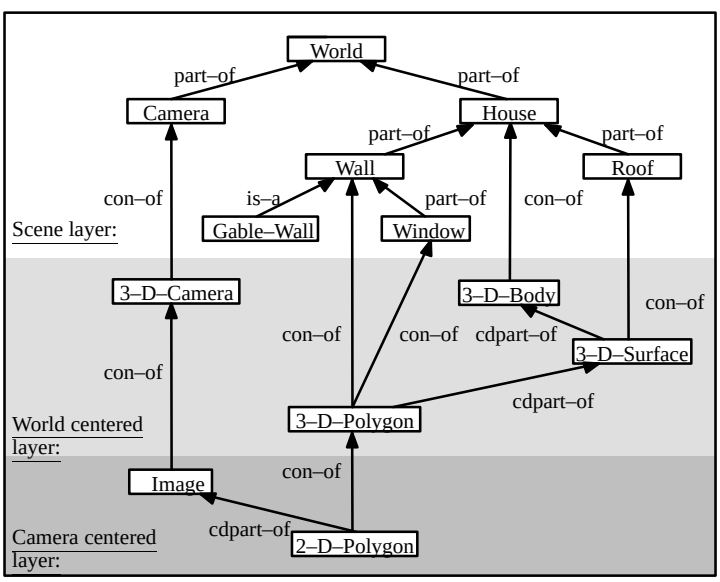

Fig. 3. Structure of knowledge base in AIDA

Between the nodes, which can be *concepts* or *instances*, i.e. copies of concepts, certain types of links are permitted. The *part–of* link describes the decomposition of an object in its components. A special form of the *part–of* link is the *context–dependent–part–of* link abbreviated *cdpart–of*. It connects a concept with a related obligatory context. A 2–D polygon describing an image region cannot be established until the context, i.e. the related image, is instantiated. Thus the 2–D polygon is a context dependent part of the image. Specializations of objects are described by the *is–a* link. In Fig.3 the concept *gable–wall* is a specialization of the concept *wall*. If no exceptions are defined, the

specialized concept inherits automatically all parts, concretes and concept properties from its superior concept.

The *concrete–of* link abbreviated *con–of* connects nodes in different conceptual layers. The concept *wall* belongs to the *scene layer* and is connected to the (for image analysis purposes more concrete) concept *3–D–Polygon* in the *world centered* geometry layer via a con–of link. In addition AIDA provides *constraint–links* in order to represent explicitly constraints for modeling purposes. These links are only established during the process of scene interpretation. So a link *perpendicular–to* is created between two wall instances to express their particular geometrical relation. Information about these links is used for the selection of appropriate processing methods during the surface reconstruction phase.

So far objects are described by their membership to object classes, their parts and their geometric appearance in other conceptual layers. In addition the concepts are described by properties which may be divided up in attributes and constraints. Attributes are measurable properties of an object like the height of a wall. For each attribute provisions are made to state the permitted range of attribute values and the procedural knowledge how this particular attribute value can be obtained from the sensor data or from other instances. During image interpretation all attributes are calculated and valued by comparison with predefined or expected attribute values.

In contrast to attributes constraints represent restrictions in properties of or relations between different objects. The fact that house edges connecting neighboring walls form usually a perpendicular line can for example be modeled by using constraints. They are used to create the constraint–links which have been mentioned before. The constraints can be used in two ways. On one hand the position and orientation of a second wall can be predicted if the first wall is known. Thus the search space for the second wall can be restricted. On the other hand the constraint is used during surface reconstruction of the 3–D model and will enforce a perpendicular connecting edge despite of errors during segmentation and depth measurement which might lead to inaccurate wall positions.

4 Scene Interpretation

The strategy of scene interpretation will be explained on the example of finding two house walls in the scene of Fig.4. It is assumed that there exists a knowledge base as has been described above, which contains the relevant prior knowledge about the parts of a house in a scene. In the beginning the goal concept *House* is established. The instantiation rules of the system permit instantiation of a concept only if all obligatory parts and concretes (concepts connected via a *con–of* link) of an object have been found. Therefore, search starts for one of at least four obligatory house walls. According to Fig.3 each *wall* is represented by a *3–D–Polygon* in 3–D space or a *2–D–Polygon* in the camera plane. The system establishes in a model driven strategy top–down instances of these concepts as hypotheses (*GableWall–1, 3–D–Polygon–1, 2–D–Polygon–1)*. Since the concepts on the bottom layer have neither further parts nor concretes they represent *initializing concepts* because they can directly be instantiated from the data, here image data.

During instantiation of the initializing concept *2–D–Polygon* several regions from the segmented image are returned for further investigation. They are considered to be individual candidates competing for the interpretation of being a house *Wall* and are processed independently. Each interpretation is valued with respect to the prior expectations of the system. For example the attributes of the instances are compared against the permissible range values in the knowledge base. Following the example of finding a house wall a region is only then considered to represent a wall if it is large enough and if its orientation is perpendicular to the ground. During verification the strategy works bottom up through the prototype net from the data to the symbolic description level.

After instantiation of the first wall instances for a second house wall are hypothesized. In this connection prior knowledge about the geometric relation between walls is exploited, for instance the knowledge, that walls are in general perpendicular to each other. This knowledge is represented explicitly by constraint links in the net of instances. Fig.4 illustrates, how the position and orientation of the second wall can be predicted from the first wall. Four competing hypotheses *Wall–1,...,Wall–4* are established based on the constraint of being perpendicular to the first wall. Each of these hypotheses is investigated with respect to its visibility or occlusion in 3–D space.

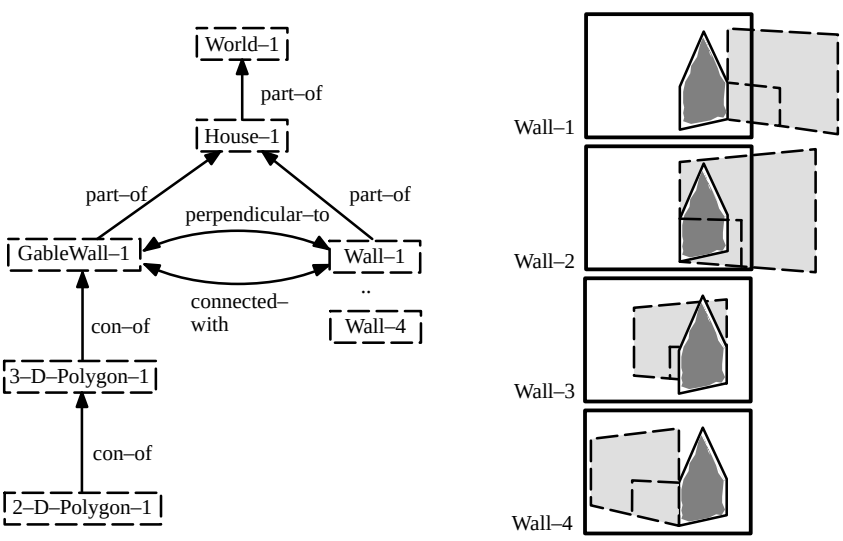

Fig. 4. Four competing hypotheses for the second wall (Wall–1, .., –4) projected onto the image plane and their corresponding representation in the semantic net

The hypothesis *Wall–1* in Fig.4 is rejected because the wall would have to appear in its main part outside of the image plane. The second hypothesis *Wall–2* is rejected because it would occlude the first wall which has already been detected in the first step. The third hypothesis *Wall–3* is valued low, because major parts of the house wall would have been occluded by the first house wall. For these reasons the fourth hypothesis *Wall–4* gets the highest credit. Subsequently that image segment is instantiated which matches to the

largest extent the expected criteria, i.e. which is positioned in the search area for *Wall–4* depicted in Fig.4. The new knowledge, which has been obtained from instantiation, penetrates the semantic net bottom up and by doing this triggers a comparison of the size and the orientation of the second wall with the expectations in the model.

The process of alternative generation of hypotheses top down and instantiations and verifications bottom–up continues until a predefined level of scene interpretation has been obtained. In the example at least four walls and a roof have to be instantiated in order to instantiate the object *House*. If desired the interpretation can be further refined by looking for optional parts of a house like the built on shed (see Fig.2). The interpretation process results in a net of instances which represents a precise description of the scene and attaches symbolic interpretation labels to all scene objects. Part of the interpretation results is the postulation of geometric constraints which are required for a subsequent realistic surface reconstruction of the 3–D model aimed at.

As has been mentioned it may happen that several competing hypotheses are established or several competing instances are found. All these interpretation steps have to be evaluated and processed separately. The control algorithm regards each alternative as a separate state of interpretation in a global search tree. Each processing step corresponds then to a new node in the search tree. The nodes of the search tree are evaluated by calculating a quality value from the individual quality values of all instances obtained so far. The quality of an instance is in turn calculated from the quality values of all attributes and constraints within an instantiated frame. When the search tree is split due to a multitude of possible interpretations for an object or possible new objects for a particular interpretation the decision for the next node in the search tree to be investigated, i.e. the next processing step, is based on the best evaluation result from an A*–algorithm.

5 Surface Reconstruction

As a result of image interpretation geometrical constraints for improved image modeling can be extracted from the knowledge base. Let us assume, that a *wall* is represented by a *3–D–Polygon* in the world–centered layer and a *2–D–Polygon* in the camera–centered layer. By definition the polygon is planar, which corresponds to the prior knowledge that the wall is plane. All points in the depth map which belong to the *2–D–Polygon* and therefore to the *3–D–Polygon* with the interpretation *wall* are assessed by inverted use of the camera model. In order to arrive at a planar wall in the model the set of spatial points is approximated by a planar surface using a linear regression.

6 Results

Fig.5a shows a result from a purely data driven modeling strategy. Fig.5b gives an example for modeling results which have been obtained using the approach presented in this paper. The model consists of 10 polygons only. The semantic net which we used as knowledge base for this example had a size of 41 nodes and 64 links. Fig.3 shows only a small part of the net. During the analysis 311 instances and 1378 links were created.

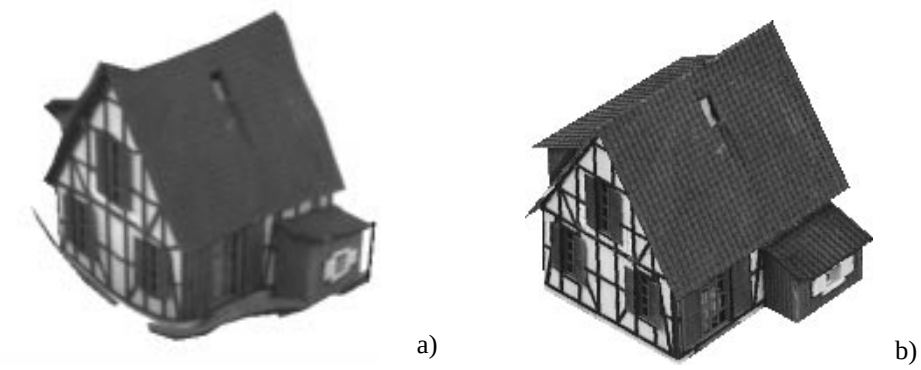

Fig.5. Synthesized view of the textured model a) without use of knowledge b) with use of a priori knowledge

The control algorithm for the reasoning process and the handling of the semantic net was implemented in CLOS (Common Lisp Object System). The algorithms for low level processing in the the image processing pipeline, i.e. the calculation of the depth map from stereo images, image segmentation, etc. and the algorithms for surface reconstruction were implemented in C and C++.

7 References

[1] Koch, R., "3–D Surface Reconstruction from Stereoscopic Image Sequences", International Conference on Computer Vision ICCV'95, Boston, June 1995.

[2] Grau, O., Tönjes, R., "Knowledge Based modelling of Natural Scenes". Proc. of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production. 23.–24. Nov. 1994 Hamburg, Germany

[3] Röning, J., Taipale, T., "Stereo–Based 3–D Scene Interpretation Using Semantic Nets", SPIE Vol. 1608 Intelligent Robots and Computer Vision X (1991)

[4] 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.

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

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

[7] Clément, V., Giraudon, G., Houzelle, S., Sandakly, 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

[8] 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.

8 Acknowledgements

This project has been funded by a grant of the Deutsche Forschungsgemeinschaft.