179

Invited Paper, Mustererkennung '94
16th Symposium of DAGM and 18th Workshop of ÖAGM, Vienna, Austria

TRADE-OFFS IN THE RECONSTRUCTION AND RENDERING OF 3-D OBJECTS

F. Leberl, M. Gruber, P. Uray, F. Madritsch*)
Institut für computerunterstützte Geometrie und Graphik
Technische Universität Graz
Münzgrabenstr. 11, A-8010 Graz, Austria
Tel (0316) 84 17 66, E-Mail leberl@icg.tu-graz.ac.at

ABSTRACT

We present a progress report of our current work to create geometric models of real world objects. We discuss and illustrate some useful image analysis methods, and the effect of certain modeling parameters on the outcome of an object's visualization. This issue is of increasing importance in the context of computer -aided design, multi-media data bases, and information systems. We show the importance of a systems view as opposed to a view at components for image analysis alone.

1. INTRODUCTION

The demand for computer models of real-world objects increases and results in a closer interaction between the fields of image analysis and computer graphics (Durisch, 1992; Hildebrand, Selzer, 1993, Thibault, 1994). Applications are numerous and range from the factory floor, medical diagnostic tools and multi-media data bases via documentation of objects in museums, architecture and facility management to 3-D data bases for city planning, environmental modeling and many other fields. This has led us to focus our work on the study of methods and their parameters for modeling real world objects, and optimize these methods in the context of computer graphics applications of the object's models. A major source of data for such models is in the form of images and 2D image sequences (e.g. ZPF, 1993 with 6 papers).

The problem domain is characterized by a need to *automate* the task of data extraction from images and to accomplish *synergy* between five component technologies of (1) electronic imaging, (2) image analysis and pattern recognition, (3) computer processing systems, (4) object visualization concepts and (5) interactive data editing.

---

*) The results reported in this paper have been obtained through the efforts of an entire team of University collaborators, doctoral candidates and diploma students at Graz University of Technology, whose enduring enthusiasm is greatly appreciated.

We describe first our ongoing effort to build a laboratory-type end-to-end system to create, edit and visualize 3-D models of various kinds of real-world objects. We illustrate the importance of a system view on the entire process. We are concerned with the electronic imaging subsystem and how it affects the score of an automated shape extraction element; how shape extraction itself can support the definition of the minimum point density to accomplish a desired geometric accuracy in an object's surface description; how the geometric accuracy of an object's 3-D model affects both the required computing resources for rendering and the realism of the visual result; how the use of photo texture relaxes the need for geometric accuracy; how the computing architecture affects the throughput of a given process element.

Relationships exist between imaging, information extraction, rendering methods, geometric detail and computing resources. Some of these are illustrated by various examples from our work, to include the reconstruction of a prototype engine, of an historical monument, of an outdoor city scene and of a human face. Performance of a chosen modeling process is of great interest in terms of geometric accuracy, visual realism, use of computing resources, need for interactive intervention and overall throughput. Initial attempts at such assessments have been made; but significant further development will be needed and defines our research strategy for the near future.

2. SYSTEMS FOR 3-D OBJECT RECONSTRUCTION

The construction of geometric and radiometric models of real world objects needs a component for sensing, for analysis and for geometric image processing. In traditional terms the issue is one of photogrammetric data collection. However, the focus of photogrammetric research has since many decades been the so-called "2.5-dimensional" Earth's surface, not the truly 3-dimensional small or medium-sized objects of the human environment. As a result many of the photogrammetric techniques must be adjusted to the changed focus, as documented by the confluence of machine vision and so-called "close-range" photogrammetry (Grün, Kahmen, 1993).

"Modeling" consists not only of the creation of a geometric description of an object, but also of surface properties, reflective properties and surface texture. In this context, machine vision and photogrammetry need to be augmented by computer graphics and visualization concepts. The description of the interior of a closed space such as a room or industrial plant requires a sensing concept different from that for the collection of urban 3-D city-scapes with their roofs, building shapes and facades. The model of an industrial part may be obtained from source material that is again different from these, and from a system optimized for modeling human faces.

We believe that *sensing* is a central issue for a modeling system. Applications-relevant imaging with optimized illumination to support photometric stereo (Horn, 1989), the use of multiple views of an object for stereopsis or for multi-image shape-from-shading (Thomas et al.,1991) will greatly affect the throughput and accuracy of the modeling system. Sensing for 3-D object reconstruction will therefore need to rely on a tool-box of technologies not only for sensing alone, but augmented by an array of related techniques to calibrate the sensing element under various conditions of operations (for example Horaud et al., 1993).

While object reconstruction in controlled laboratory or commercial environments may be based on an optimized choice for a sensing component, there exist numerous circumstances where sensing methodology has not been selected for object reconstruction. This is typically the case in satellite imaging of the Earth's or other planetary surfaces, or in forensic applications.

Various laboratory-type systems have recently been described to use 2-D images for the creation of models of real objects for computer graphics applications. This is typically driven by computer graphics needs, for example in the work by Thibault (1994), Hildebrand and Selzer (1993), Pitschke and Garny (1993). An interesting approach is that by Koch (1993) who focusses on the *automation* of the process.

3. EXAMPLES OF RECONSTRUCTION TASKS

3.1 Mechanical Parts

Prototypes of industrial parts are frequent examples for the need of a detailed and accurate inspection, as well as for a comparison with design goals. For objects not exceeding certain sizes there exist commercial products for shape and texture modeling, preferably employing a laser-based illumination system (Cyberware, 1992). We demonstrate a strictly photogrammetric method using the reconstruction of a prototype engine.

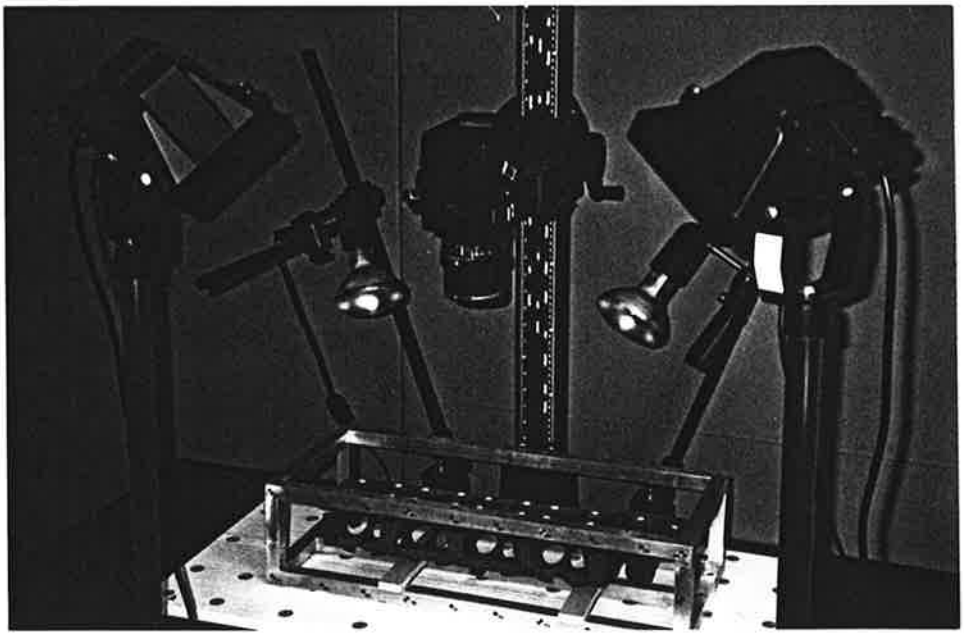


Figure 1: The water-cooling jacket of an engine of a length of about 500 mm placed inside a calibration rig, and the imaging scenario. Images were taken with a film camera and then scanned.

Figure 1 presents the imaging scenario currently being used to obtain source data from which the engine's geometry can be modeled.

"Sensing" initially consists of taking a total of about 50 centrally perspective photographs. An elaborate photogrammetric triangulation system (Knopp, 1990) permits one to reconstruct the so-called "exterior" orientation of the individual camera images. Then a cloud of surface points is obtained to describe the engine's geometry. The cloud consists of a total of 50,000 surface points at an accuracy of ± 0.2 mm.

Figure 2 presents part of the resulting surface points for the object in Figure 1. While the imaging system offers full flexibility in collecting about any kind of object, it is not optimized for automation. The multitude of centrally perspective images creates a need for managing multiple data files, cope with complicated calibration issues, and merge disjoint individual object surface patches obtained from selected pairs of overlapping stereo photographs.

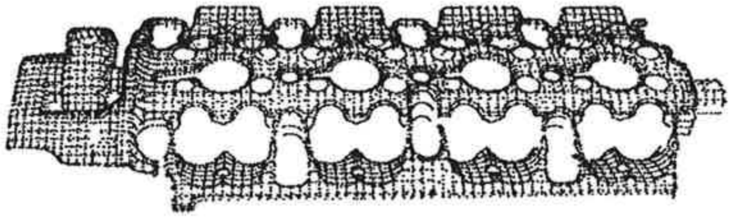


Figure 2: Point cloud to represent the lower half of the engine of Figure 1, showing 29,000 of the total number of points which is about 50,000. Each point has a positional error of ± 0.2 mm.

3.2 A Cultural Monument

Figure 3 illustrates the sensing scenario to collect source images for a 3 meter statue of Emperor Karl VI in the Austrian National Library. The surface of the statue is represented by more than 9,000 points (Figure 4). The data source again consisted of 20 centrally perspective photographs at scale 1:35, so that the issues in the reconstruction of the engine in Figure 1 reocurred here as well: multiple images, assembling a large object from individual patches, and a concern for collecting a sufficient number of points to describe the surface with adequate accuracy. Individual points have errors of about ± 1.5 mm. The mean distance of the points on the surface amounts to about 20 mm. The accuracy of the surface itself was not a prevailing concern, instead the "specification" merely states that a "realistic" model be created. We estimate that the surface deviates from its model by about ± 1.5 mm, thus by the amount of uncertainty in individual points.

In contrast to the point cloud in Figure 2, this surface now needs to support a system for computer graphics rendering of the statue. Therefore the point cloud needs to be triangulated, resulting in a total of 12,377 triangles (Figure 5).

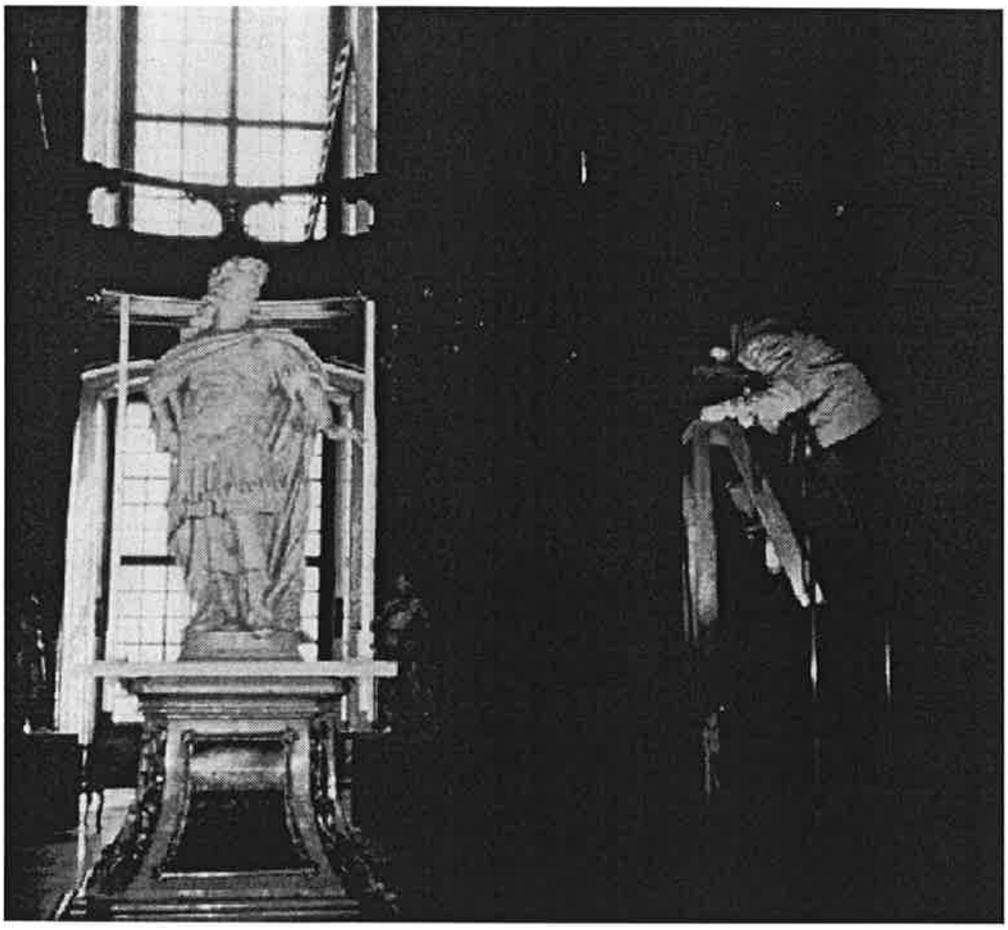


Figure 3: Sensing scenario for the statue of Emperor Karl VI at the National Library in Vienna, Austria. The statue is about 3 m tall. About 20 photographs were used to develop the geometric model.

While triangulation is a well developed tool in two dimensions (for example Correc and Chapuis, 1987; or Ho-Le, 1988), we have a requirement here to obtain optimum surfaces patches in three dimensions. The example in Figure 5 was obtained from a projection of the point cloud into various projection planes so that then the triangulation would simplify into a 2-D problem. This was an unsatisfactory solution which required considerable manual interventions. Therefore we need to improve this method by a transition to fully 3-D triangulation methods. We performed some experiments with the so-called x-shapes (Edelsbrunner, Mücke, 1992) and weighted x-shapes (Edelsbrunner, 1992), but were not able to obtain a more complete result than that from 2-D projections.

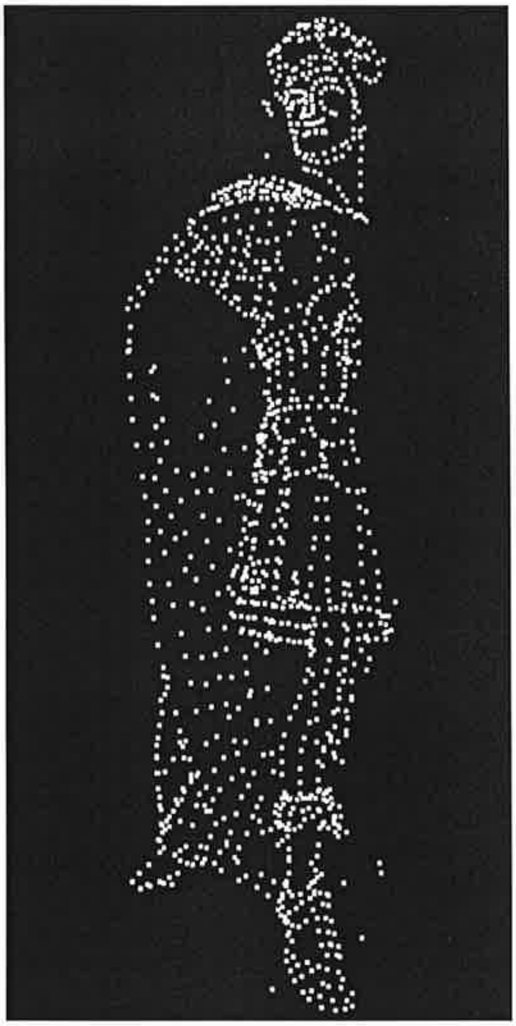


Figure 4: Point cloud to describe part of the statue of Figure 3. A total of 9,100 points was measured, of which 7,195 were used, each point with a positional accuracy of ± 1.5 mm.

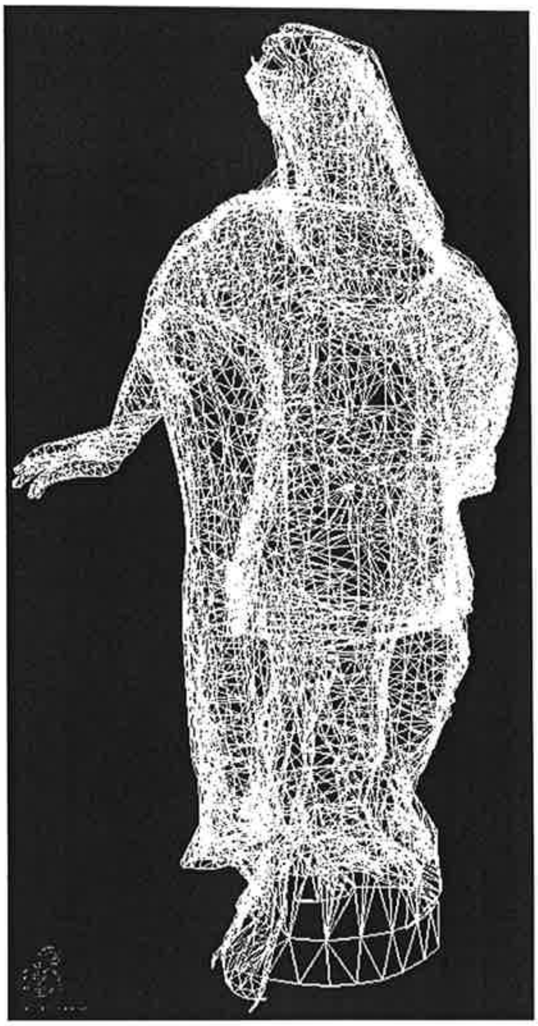


Figure 5: Triangulated surface using the point cloud of which part is shown in Figure 4. Triangulation is accomplished in 2-D by projecting the points into various projection planes.

3.3 City Models

Three-dimensional models of cities are currently a widely discussed "upgrade" from strictly 2-dimensional geographic information systems (GIS). A recent review is by Förstner and Pallaske (1993). Applications will span the entire range of GIS which have firmly taken root in city administrations world wide.

Data sources for a 3-D city model of course include the existing GIS-files. Roofs are obtained from aerial photography. Building shapes are either reconstructed from aerial imagery as well (Lang and Schickler, 1993), or from

GIS-files in combination with aerially assessed roof data. Details of building exteriors finally derive from photographs taken at street level.

Figure 6 illustrates a demonstration area in the 7th district of Vienna, Austria. This city block consists of 28 buildings and 570 meters of street. All of Vienna has 220,000 buildings. In our initial work we have chosen to produce a model of this city block that is "frugal", encompassing merely the crude building shapes (see Figure 7). Detail about the building's facades will have to follow.

The current work is based on a preliminary set of methods which are in need of significant improvement. Most important is the ability to obtain efficient images of the building's sides. Instead of currently used perspective cameras it would be useful to employ strip imagery in a form which supports the stereo analysis of the facade's geometric details.

Second is the ability to automate the extraction of shape information, for example by implementing automated methods for the description of the building's shape, using the GIS data as an input jointly with the aerial images.

Third is the extraction of details about the buildings geometry from street level imagery, using single image shape reconstruction by means of shape-from-shading and by using the knowledge of horizontal and vertical lines as well as the fact that walls are vertical and plane.

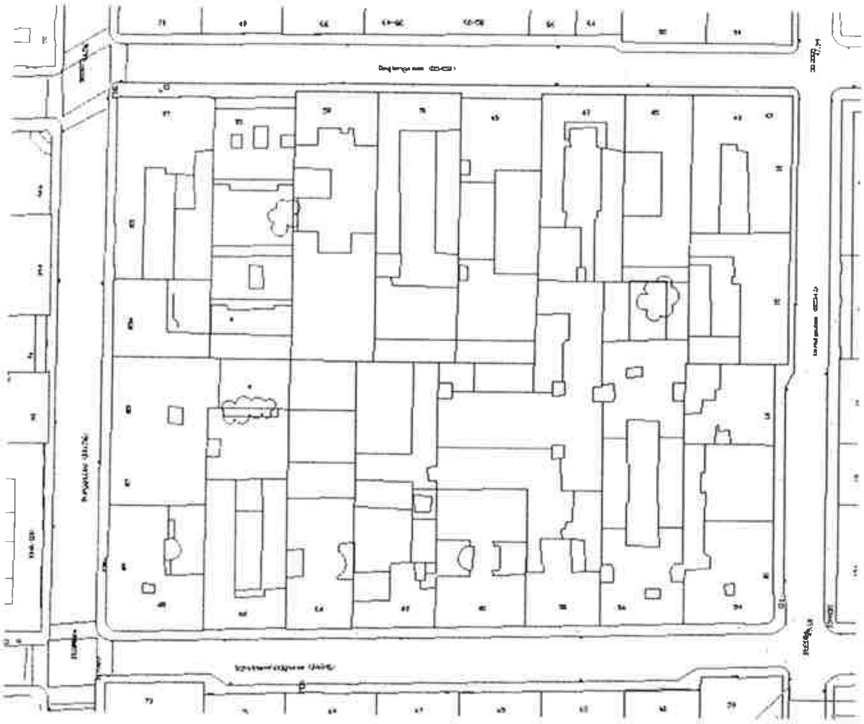


Figure 6: Geographic information from a segment of the 7th district of Vienna (Austria). This information, straight from AutoCAD-data in the city's repositories, is being augmented into a fully three-dimensional data file. Area comprises 570 m of streets and 28 buildings.

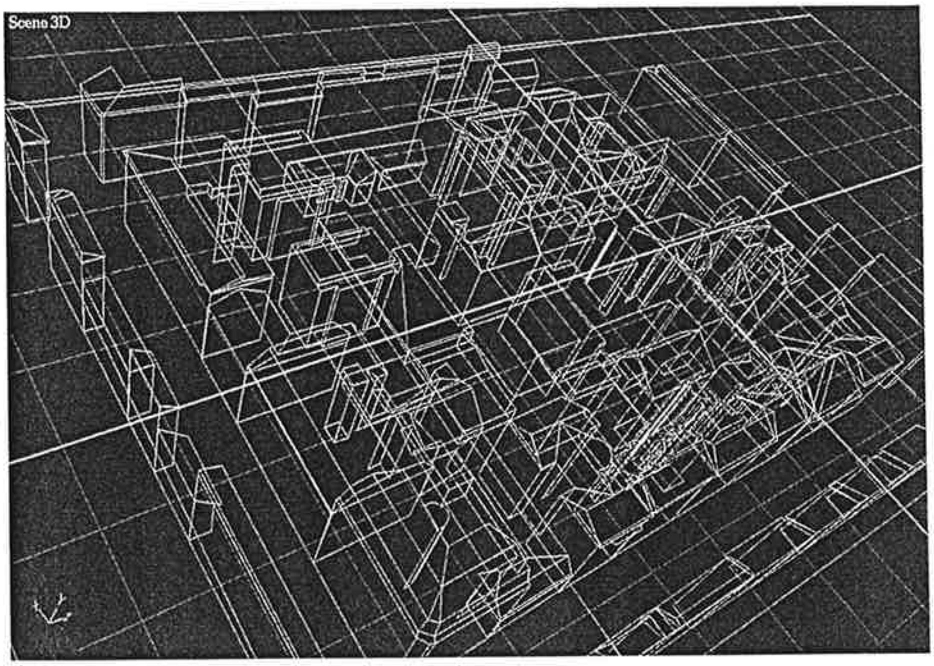


Figure 7: Perspective view of the buildings of a segment of the 7th district in Vienna (Austria). This data set is extracted from digitized aerial photographs and the existing AutoCAD 2-D file of the city (see Figure 6).

Figure 8: Perspective view of a segment of the 7th district in Vienna, obtained from aerial photography, Vienna's current GIS data, and augmented by phototextures extracted from aerial and terrestrial photographs. Total data set over one city block (4 street segments) contains 28 buildings along 560 meters of streets.

3.4 Human Faces

Quick and inexpensive modeling of objects is feasible by using a pointer and a pair of cameras. Detecting a pointer's trace on the surface of an object is far easier than developing a surface point from automated image matching. As a result a simple object reconstruction method is illustrated in Figure 9. The point is of a distinct color, in this case a bright red that contrasts well with the remainder of the environment. Two cameras follow the pen with an accuracy of ± 0.5 pixels. Individual image points define a perspective ray in 3-D space; the 2 rays intersect to define a surface point.

Figure 10 is the point cloud obtained for a particular human face within 600 seconds from 2000 points. It is to be noted that the method uses a stereo intersection of two rays, but does not collect the image points in a stereo mode. Instead two independent and "monocular" measurements are being taken in the two images of a so-called "signalized" point.

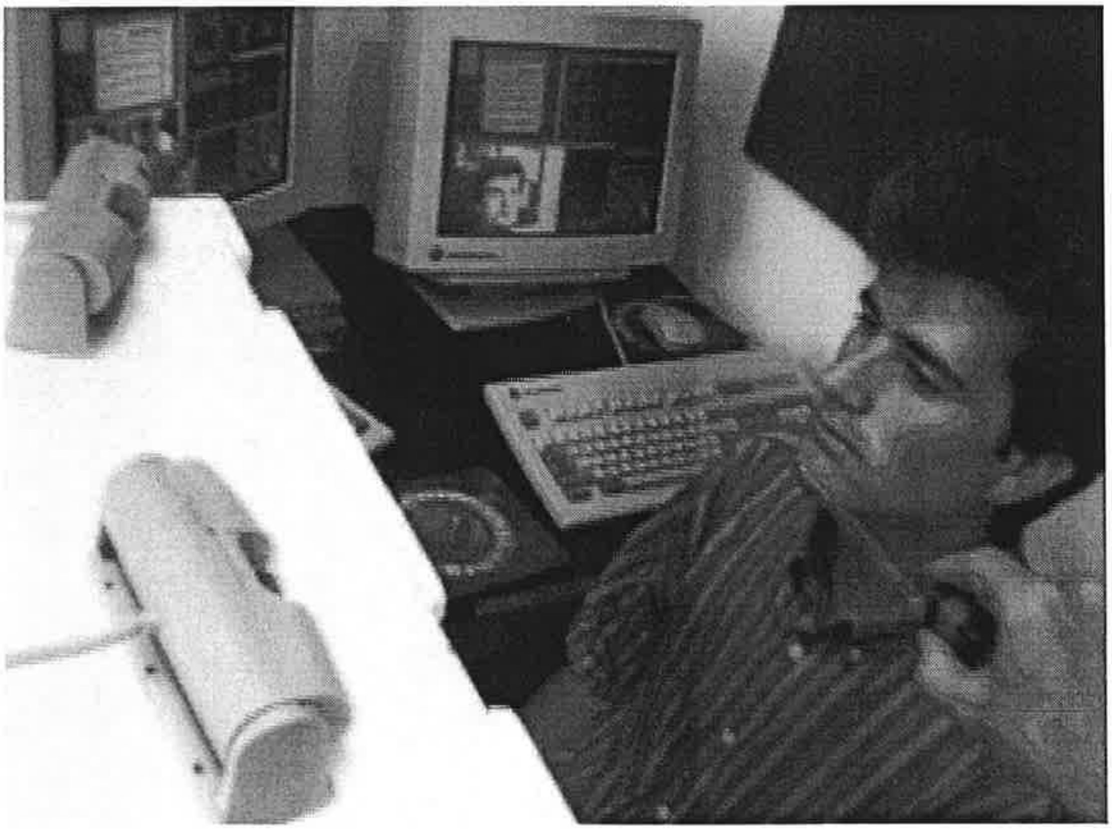


Figure 9: System to model a face by means of tracking a colored pointer moving over the surface. The pointer is either an LED physically straddling the surface, or a projected red point obtained by a small projector pen used for slide presentations. Two cameras feed images into a Silicon Graphics Indy.

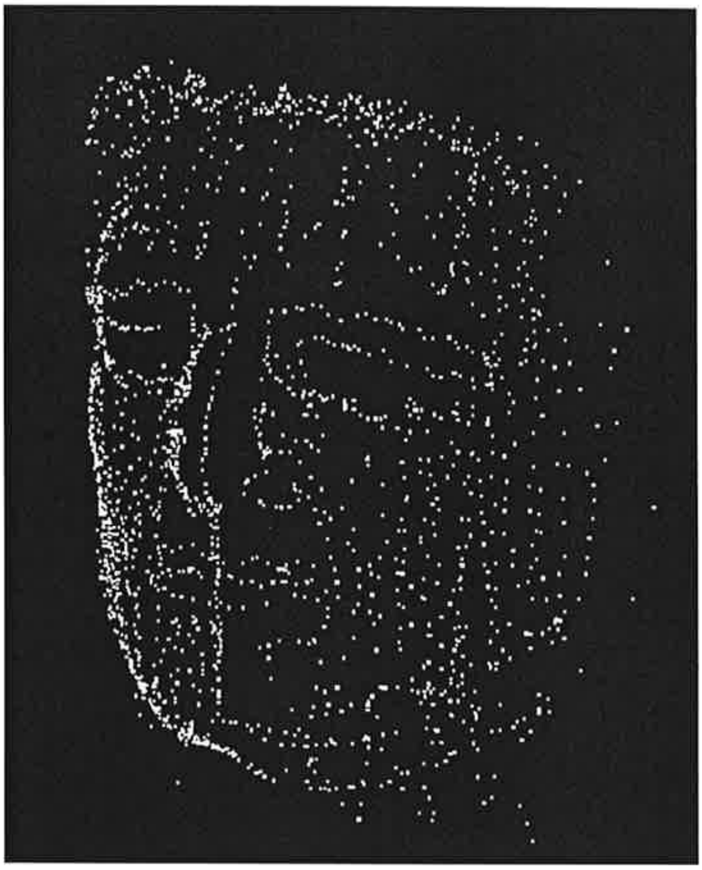


Figure 10: Point cloud developed by the method describe in Figure 9 and obtained from a human face. Point accuracy is about ± 0.5 mm. A set of 2000 points has been use to describe the face. Data collection is an interactive process taking about 600 seconds for this face.

The interest in this approach is driven by one's need to edit three-dimensional point clouds, particularly when relating them to the input images from which the clouds derive. In this case a stereo viewing person will be presented with the points , with a marker in 3-D space, and with the images from where the points were derived.

This is then an environment for interactive 3-D editing of 3-D data sets. Of course such editing is simultaneously also a method for original data collection as illustrated here.

4. TEXTURE MAPPING AND RENDERING

In applications to computer graphics, modeling may also include the surface's thematic or qualitative properties. This may be accomplished by parameters of the surface material, such as in architectural design of objects not yet in existence. These properties are used to "render" the object in the visualization of the data.

However, for real world objects the option exists to employ photographic texture to describe the surface properties. The use of such texture requires that surface patches be associated with a segment of a photograph. In essence this is the same process by which photogrammetrists produce so-called "ortho-photos" or "geocoded" images. The surface patch itself is a triangle. The triangulated points need therefore to be projected into the source imagery, or new images are being used and must be related to the surface's point cloud.

Figure 11 illustrates the example of using photo texture. The complexity of the data structure is a result of the number of triangles for the surface and the associated triangles of photo segments to be superimposed over the surface. A data volume of 13,6 MB is being used.

Figure 12 illustrates the added value of phototexture. To visualize the surface of the Emperor's face one needs to make an assumption about the surface reflective properties. This assumption becomes irrelevant when photo texture is being used. In the example of the emperor's face, the difference is not as dramatic as with the buildings of a city. Texture may include the windows, doors and many geometric details which are being replaced by the texture (see for example Figure 13).

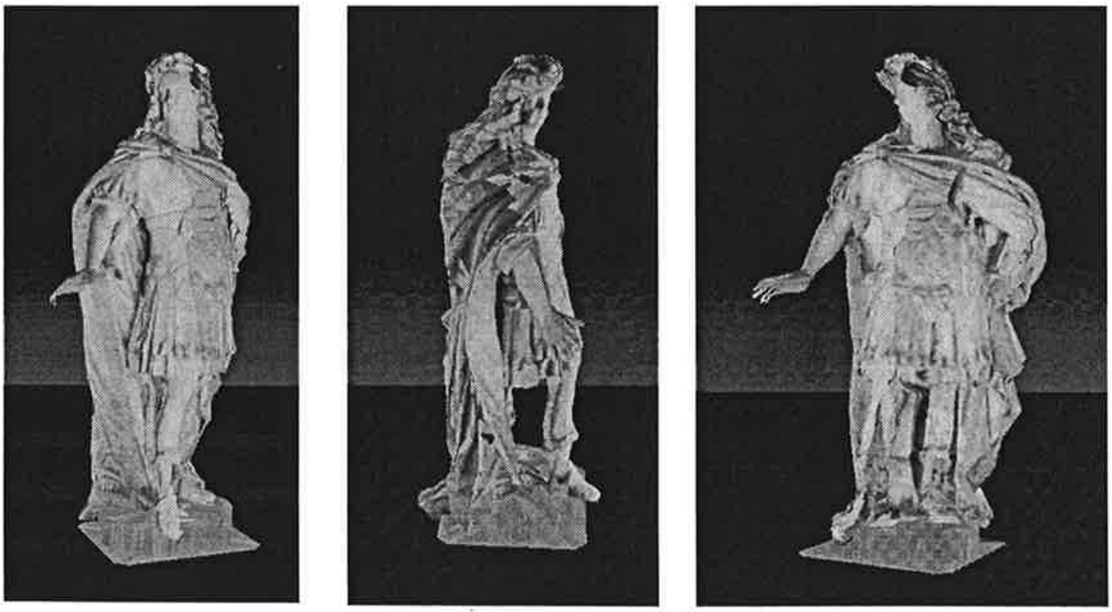
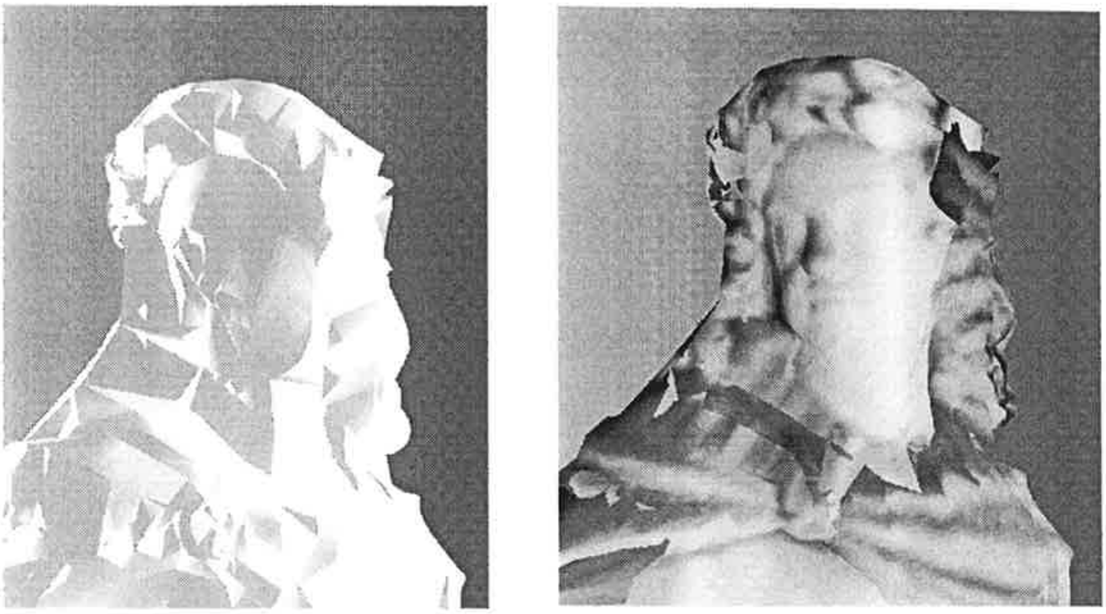


Figure 11: Three perspective views of the 3-D model of emperor Karl VI, using more than 7,000 points, 12,000 triangles and 13,6 MB of phototextures.



(a) Without Photo Texture (b) With Photo Texture

Figure 12: Example of a surface modeled without and with photo texture . The statue in Figure 3 is being used. The detail of the face illustrates the improvement in the photo-realism when employing photo texture. Note that change in contrast in the texture is due to the use of different source photographs. A radiometric adjustment is yet to be accomplished. The trade-off between required geometric detail and photographic texture is evident.

5. TRADE-OFFS

5.1 Automation versus Accuracy

There exist several reasons to automate the modeling of 3-D objects. One is throughput which would increase, say from one point collected every 3 seconds by hand, to 100 points collected automatically by machine. Another is to remove the need to understand complicated procedures to be performed by hand. Instead a "black-box" approach can be made available to a "naive" user. Automated object shape from stereo matching is, of course, a widely studied topic. Its specific variation for computer graphics is particularly applicable in implementations which include surface texture maps (Koch, 1993).

Typically, however, automated surface modeling produces results that are less accurate than those obtained from an experienced human operator (see for example Leberl et al., 1994). This is overcome in practice by combining the machine process with a manual editing and augmentation step: surface points collected automatically by stereo-matching are then inspected visually, accepted or changed manually. When accuracy requirements are high then the manual editing may totally invalidate the throughput advantage otherwise enjoyed.

5.2 Geometric Detail versus Photo-Texture

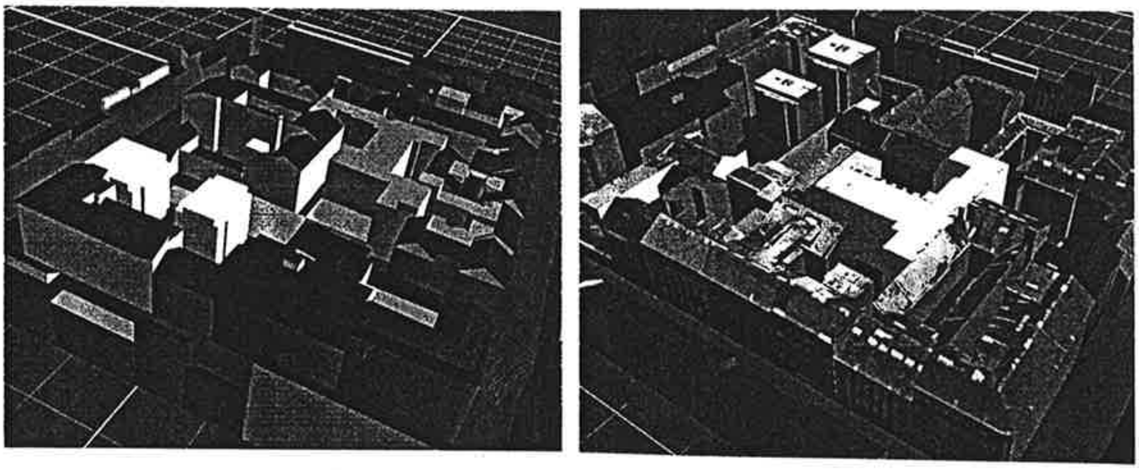
Obviously the surface detail of a 3-D real world object is needed to create a visually pleasing or photorealistic rendering of that object. However, when supplemented by photo texture, the need for detail is no longer justified by realistic rendering. This trade-off is not well understood at this time, although it is easily perceived as in the Emperor's face (Figure 12).

The example of the city model is based on very simple building shapes without any geometric details. Yet for the purpose of rendering, the combination of simple boxes for the buildings and the photo texture seems to suffice (Figure 13).

There would seem to be no reason to simplify the geometric shapes, since data quantities in 3-D models are entirely dominated by the photo texture, not by the detail of the geometric shapes. Yet there is a need to keep the shape description simple due to the desire to provide rendered surfaces at a rate of 20 to 30 per second for real-time interaction with the data. Real time, at this stage of computer technology, cannot be accomplished with arbitrarily large data sets. It is therefore important to trade off the detail in geometry with the resulting number of polygons, against the ability to interact with the data sets in real-time.

5.3 Sensing versus Image Analysis

The opportunity exists in most vision scenarios to affect the feasibility and efficiency of a solution so that it largely is determined by the sensing method. In the examples presented in this paper, shape-from-shading and photometric stereo are expected to greatly increase the detail and morphological accuracy of the geometric shapes. Therefore the illumination needs to be integrated into the sensing and subsequent reconstruction processes.



(a) Geometric shape

(b) Adding photo texture

Figure 13: Segments of a city-scape using simple geometric shapes and photo texture (detail of Vienna's 7th district).

In the example of city-scapes the use of centrally perspective photographs unnecessarily complicates the work: slit camera technology is expected to simplify the task.

5.4 Detail versus Throughput

An automated process may create a large number of surface points. The problem is then to discard the outliers, either again automatically or by interactive editing. "Detail" is a function of both the number of surface points as well as the location of those points along significant edges of the object.

In the procedure by Ackermann and Schneider (1992) denoted as Match-T the strategy is to collect a very large number of surface points and then to extract from those raw points a finite element description of the surface. About 10 x 10 raw points may serve to determine one point in the finite elements.

However, in the context of computer graphics applications, detail, as previously discussed, is not only an issue of creating the object's model, but also an issue of rendering throughput. Increased detail slows down the creation of an image of the object. Therefore the detail to obtain a model of an object needs to be assessed, preferably by automated means. A promising approach may the concept of "progressive sampling" (Makarovic, 1973). some rules need to be established which relate the type and characteristics of an object as well as the rendering requirements to a measure of the resulting detail in the object's model.

5.6 Cost versus Photo-Realism in Rendering

"Cost" is mostly for computing power to accomplish real-time photo realism. Rendering such a scene with 10,000 polygons and complex texture may take several minutes. "Realism" requires the computation of shadows, for example via ray-tracing, and needs the inclusion of so-called "participating media" like fog, clouds, dust, etc. If we also require that two eye positions be computed for stereopsis, then we will have to conclude that we have use for graphics work stations which are 10,000 times faster than what we have available today.

It is therefore important at this time to be conscious of rendering performance, even in the context of the extraction of 3-D information about real-world objects.

6. OUTLOOK AND EXPECTATIONS

We have begun to develop a work program that is directed at the creation of 3-D models of real world objects, in part for the application in computer graphics and for photo-realistic rendering in real time. Several examples have shown that numerous trade offs need to be considered. The geometric accuracy of a model is affected by the amount of manual work invested to correct results obtained by automated methods. Detail of a geometric model is determined by the extent to which photo texture can be used, and the extent to which photorealism is adversely affected by too coarse a geometric model.

Throughput is an issue of photo realistic rendering. This limits the detail in the geometric models of objects and the resolution of photo texture. Throughput also creates costs for computing resources, in particular costs for photo realism at real time and for stereo-viewing.

Generally we find that the application of information extracted from sensor data determines the sensing methods, the amount of detail and the accuracy of the information. Currently, however, these requirements are only poorly understood. Significant work is needed to improve this knowledge and to set up some rules to relate the information requirements for 3-D models of real world objects to the desired use in computer graphics systems.

We will continue on this path and contribute to a better understanding of these issues.

REFERENCES

Ackermann F., W. Schneider (1992) *Experience with Automatic DEM Generation.* International Archives of Photogrammetry and Remote Sensing, Washington D.C., Vol. XXIX, Part B4, pp. 986-989.

Hildebrand A., H. Selzer (1993) *Erfassung und Simulation realer Objekte.* Computer Graphic Topics, Mitteilungen aus dem Haus der Graphischen Datenverarbeitung, Darmstadt, 4/93, pp. 11-12.

Correc Y., E. Chapuis (1987) *Fast Computation of Delaunay Triangulations.* Advances in Engineering Software, Vol. 9-2, pp. 77-83.

Cyberware (1992) *Product Literature on Rapid 3-D Digitizer 30303.* Monterey, California.

Durisch P. (1992) *Photogrammetry and Computer Graphics for Visual Impact Analysis in Architecture.* International Archives of Photogrammetry and Remote Sensing, Vol. 29, B5, pp. 434-445.

Edelsbrunner H., E. Mücke (1992) *Three-Dimensional Alpha-Shapes.* Proceedings of the Boston Volume Visualization Workshop, Computer Graphics.

Edelsbrunner H. (1992) *Weighted Alpha-Shapes.* Techn. Report UIUCDCS-R-92-1760, Dept. Computer Science, Univ. of Illinois at Urbana-Champaign, Urbana, Ill.

Förstner W., Pallaske (1993) *Mustererkennung und 3-D Geoinformationssysteme.* Zeitschrift für Photogrammetrie und Fernerkundung, Vol.61, No.5, pp.167-177.

Grün A., H. Kahmen (eds.) (1993) *Optical 3-D Measurement Techniques II.* Proceedings, Wichmann-Verlag, Karlsruhe, Germany, 624 pages.

Ho-Le K. (1988) *Finite Element Mesh Generation Methods, Delaunay Triangulation, Mesh Topology, Voronoi Diagram.* Computer Aided Design, Vol. 20-1, pp. 27-38.

Horaud R., R. Mohr, B. Logecki (1993) *On Single-Scanline Camera Calibration.* IEEE Transactions on Robotics and Automation, Vol. 9, No. 1, pp. 71-75.

Horn B. (ed.) (1989) *Shape-from-Shading.* The MIT Press, Cambridge, Mass.

Knopp D. (1990) *Interactive Realisation and Calibration for Mapping 3-D Objects.* Int. Archives of Photogrammetry and Remote Sensing, Vol. 28/B5, Zürich, pp. 728-735.

Koch R. (1993) *Automatic Reconstruction of Buildings from Stereoscopic Image Sequences.* Eurographics '93, Blackwell Publishers, Vol. 12, No. 3, pp. C339-C350.

Lang F., W. Schickler (1993) *Semi-Automatische 3D-Gebäudeerfassung aus digitalen Bildern.* Zeitschrift für Photogrammetrie und Fernerkundung, Vol. 61, No. 5, pp. 193-200.

Leberl F., W. Maurice, J. Thomas, M. Millot (1994) *Automated Radar Image Matching Experiment.* International Journal of Photogrammetry and Remote Sensing, Vol. 49, No. 3, pp. 19-33.

Makarovic B. (1973) *Progressive Sampling for Digital Terrain Models.* ITC Journal, Vol. 3, Enschede, The Netherlands, pp. 397-416.

Neugebauer P.J. (1994) *Objektlokalisation in Grauwertbildern unter Verwendung interner 3D-Objektmodelle.* Proceedings, 1st Workshop on Visual Computing, Fraunhofer-Gesellschaft Darmstadt, 16-17 March 1994.

Pitschke I., P. Gorny (1993) *Stereo-View - Ein System zur interaktiven 3D-Geometrie-Rekonstruktion.* Zeitschrift für Photogrammetrie und Fernerkundung, Vol. 61, No. 5, pp. 185-192.

Thibault G. (1994) *Numérisation 3D an plus près du réel.* Proceedings Imagina 94, pp. 101-105.

Thomas J., W. Kober, F. Leberl (1991) *Multiple Image SAR Shape-from-Shading.* Photogrammetric Engineering and Remote Sensing, Vol. 57, No. 1, pp. 51-59.

ZPF (1993) *Industriephotogrammetrie, Special Volume on Close Range Photogrammetry.* Zeitschrift für Photogrammetrie und Fernerkundung, Vol. 61, No. 2, 6 papers.