OBJECT RECOGNITION USING MULTI-SENSOR FUSION AND ACTIVE EXPLORATION

Claus Brenner and Michael Hahn
Institute of Photogrammetry
Stuttgart University
P.O.B. 106037
70049 Stuttgart / Germany

ISPRS Commission V, Working Group I

KEY WORDS: Active exploration, object recognition, multi-sensor fusion, free-form surfaces

ABSTRACT
This paper presents a concept for the recognition and localization of objects which relies on multi-sensor-fusion and active exploration. Today research in photogrammetry generally agrees that the use of complementary sensors, e.g. ranging and imaging cameras, is important for simplifying interpretation related tasks. But no notice has been taken so far on the role of active exploration. Our work is part of a research program where five institutes of Stuttgart University cooperate to develop an experimental measuring system for flexible inspection and gauging. The system will be capable of determining automatically the shape, form and class attributes of an industrial object. It then solves in a self-acting manner the measuring task associated with that object. The paper briefly describes the experimental measuring system and the used sensors. It then focuses on the object recognition concept. With first results a number of subsequent processing steps of the whole procedure is illustrated.

1 INTRODUCTION

For years, the production of many plastic and metal parts, for example in the automobile industry, has been automated. While at the same time the complexity of the parts increases, often the production lots get smaller. This has led in many areas to a more flexible assembly of the products. Quality assurance in this case is often performed with individually prepared gauges or specialized measuring systems, which is a very unflexible solution. Additionally, the need for a 100% quality control has grown in areas such as automobile production, since more and more complicated parts are assembled by suppliers.

Optical measurement techniques have several properties that make them interesting for flexible gauging and inspection tasks: they are able to carry out quickly measurements and are applicable to a wide range of materials including deformable objects. When used in conjunction with photogrammetric techniques, optical measurement can yield very accurate results.

Despite those advantages, optical measurement techniques are not well accepted in the industry (Grün 1994). One reason for this is that traditional measurement techniques, such as coordinate measuring machines (CMM), are very well established whilst optical systems with comparable performance have not been commercially available until recently. This may change now with new very high resolution sensors and projectors. Another drawback is that optical techniques are often considered to be too complicated to be operated under factory conditions. Testing of geometric specifications is simply solved by an unskilled worker by placing a part on a part-specific gauge. Further, optical measurements often give accuracies which depend on the specific object. In unfavourable cases, e.g. if an object's surface is soiled, measurement may become impossible using fixed sensor and lighting positions. But changing this conditions (e.g. by changing the sensor, lighting or object positions) usually requires some skilled person familiar with that particular measuring system.

To overcome these difficulties, two key components of inspection and measurement systems will be

1. multi-sensor-fusion and
2. active exploration.

Multi-sensor fusion gives a system the opportunity to select another measurement technique if the measurement task can not be performed with the currently used technique. Thus, sensor-specific flaws can be avoided. Moreover, combining sensors with different resolutions allows for precise measurements with the accuracy of the fine-resolution sensor whilst the field of view is that of the coarse-resolution sensor.

Active exploration is the process of choosing sensors and sensor positions automatically when capturing the object under consideration. Ideally, this is some hierarchical process, much like humans proceed when they examine some object: first an overall view of the object is taken which reveals some general properties and hints on interesting features of this object. Then those features are inspected in detail. For a measurement system this means that different sensors are used for different levels of detail. Active exploration also allows to change sensor and lighting positions

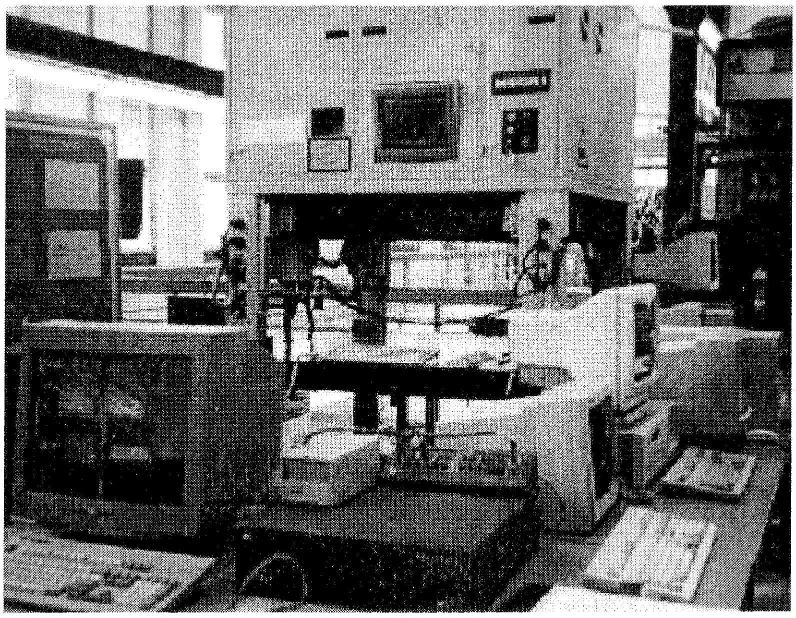

Figure 1: View of the measuring cell with actor modules (background) and general purpose computing hardware used for sensor data processing and system control (foreground).

in order to obtain favourable capturing conditions. Thus, in connection with sensor fusion, active exploration promises to increase the operability of optical measuring systems under factory conditions.

2 THE EXPERIMENTAL MEASURING SYSTEM

2.1 The currently used sensors

To investigate concepts on active exploration and multisensor fusion, an experimental measuring system is under development (see Fig. 1). The system currently uses

- a laser projector which can be used with the coded light approach for range image measurement and as a point triangulation sensor. With the ability to generate variable light patterns the laser can also be used for texture projection to support stereo image analysis.
- a multi-parameter color CCD-camera which has the ability to change parameters like focus, focal length and aperture and several electronic parameters based on the interpretation of the current image.
- a stereo camera.
- several light source arrays consisting of regularly spaced individually controllable light spots located at the edge of the measuring volume.

The sensors are mounted on three actor modules with a total of 13 axes. The size of the measuring volume is about $1000 \times 1000 \times 700\,mm^3$.

2.2 On the role of recognition and pose estimation

Object recognition and location is an important part of the overall system concept since it increases the flexibility of the measuring system. Of course, there are inspection tasks in which the object is known a priori and there is no need for recognition or identification of the specific object. In this case the task is to determine the object's pose which allows to avoid object specific conveyors needed to position the object accurately in the measuring volume. But may be already in the near future the ability to recognize objects will be the conspicious feature of such systems. Object recognition aims at finding out the object type among a library of possible types, but can also be used to identify parts of objects.

In the next paragraph we discuss how sensor fusion and active exploration can be exploited for the purposes of object recognition and point out our basic recognition concept.

3 OBJECT RECOGNITION

3.1 General aspects of object recognition

Object recognition can be defined as the problem of assigning the correct label to an object present in a scene. For that, an object has to be detected and may be localized and then identified by comparison with a given model of the object. In simple cases, global approaches can be used. Usually, global properties like volume, roundness or higher order moments are computed, forming a vector of parameters. Matching objects to models then reduces to a comparison of parameter vectors, i.e. matching is done in parameter space. In general, however, global methods are not considered to be robust enough, particularly in the presence of occlusion and clutter (Grimson 1990). Thus, many

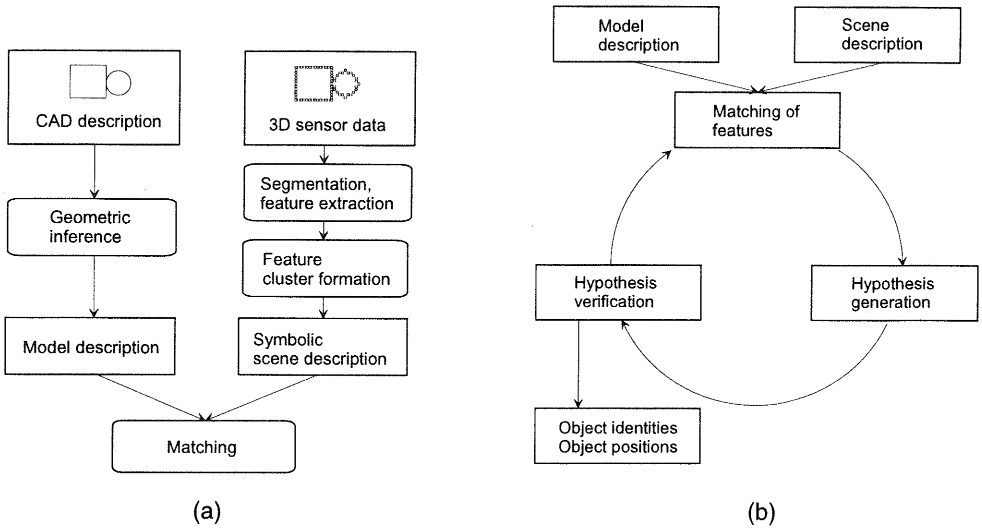

Figure 2: (a) Block diagram of a typical object recognition system for recognizing industrial parts. (b) Standard matching scheme.

object recognition systems rely on the matching of geometric primitives such as points, lines and planes. In this case, matching the object to the model is solved by establishing a number of feature correspondences which determine the object and its pose uniquely.

Fig. 2(a) shows the block diagram of a typical system for recognizing industrial objects (Bhanu & Ho 1987, Flynn & Jain 1991). For matching, sensor data and model data are ultimately mapped into a common domain. The type of this domain can be chosen either sensor-specific or model-specific or somewhere in between. The sensor-specific representation reduces the complexity of the right branch in Fig. 2(a) and increases the amount of preprocessing necessary in the left branch. One extreme example would be to sample the "important" views of the object (either by acquiring sensor data or by rendering). However, besides the fact that the "important" views are usually very large in number and are — especially for curved objects — not easy to compute (Eggert, Bowyer & Dyer 1992), the matching module has to deal with the sensor-specific representation and thus each new sensor requires to develop a new matching module. On the other hand, choosing the domain very model-specific leads to the well-known problems of extracting a useful symbolic description from the data. E.g., it is so far not possible to use a CAD representation directly for matching, because automatic extraction of the high-level CAD objects from sensor data is not feasible.

Since it is desired to obtain object models automatically and many industrial parts today are manufactured using CAD/CAM technology, several investigations have used CAD descriptions as a basis for the object models (Bolles, Horaud & Hannah 1983, Hansen & Henderson 1989, Flynn & Jain 1991). Typically, CAD data is attributed with local feature data, binary feature relations and global feature lists. Examples for adding local features are the computation of line length (based on the starting and ending points given in the CAD data) or of the parameters of a plane in which a circle (given by CAD data) lies. Binary relations between geometric primitives include the relative orientation of primitives or the distance between two primitives. Global feature lists consist e.g. of all planes with common surface normals or of all circles whose radii lie in a given range. All those approaches so far concentrated on geometric properties, as they are based on CAD data, which does not contain other properties.

Now we consider the box labeled "Matching" in Fig. 2(a). A typical block diagram of its contents is depicted in Fig. 2(b). Matching usually operates in search space, which contains all possible combinations of model features and scene features. Since an exhaustive search of this space is not feasible in all but the most simple cases, several methods are proposed to cut down search complexity. One standard method is to prune the search tree by using unary and binary constraints (Grimson 1990). E.g. a object circle is only matched to a circle present in the scene if it has approximately the same radius. This prunes complete branches of the search tree. Another approach is to limit the depth of the tree by using a combination of search space and pose space techniques. It is called alignment or hypothesize and test method. The idea is to match just enough features to hypothesize the transformation needed to map the object model into the scene. Since this is a 3D–3D transformation, very few feature correspondences (e.g. three point pairs) are needed. Then, in a second step, this hypothesis must be verified or refuted, which is usually done by predicting and verifying the location of additional object features in the scene. As shown in Fig. 2(b), this process can be viewed as being circular, since after the hypothesis verification step, either the hypothesis is refuted and the matching of features restarts, or additional evidence must be gathered, in which case feature matching proceeds and leads to a refined hypothesis.

3.2 The proposed recognition concept using multi-sensor fusion and active exploration

Our object recognition scheme using active exploration relies on the possibility of the measuring system to capture new sensor data whenever we want. Therefore, we propose to extend the circular matching process (Fig. 2(b)) as

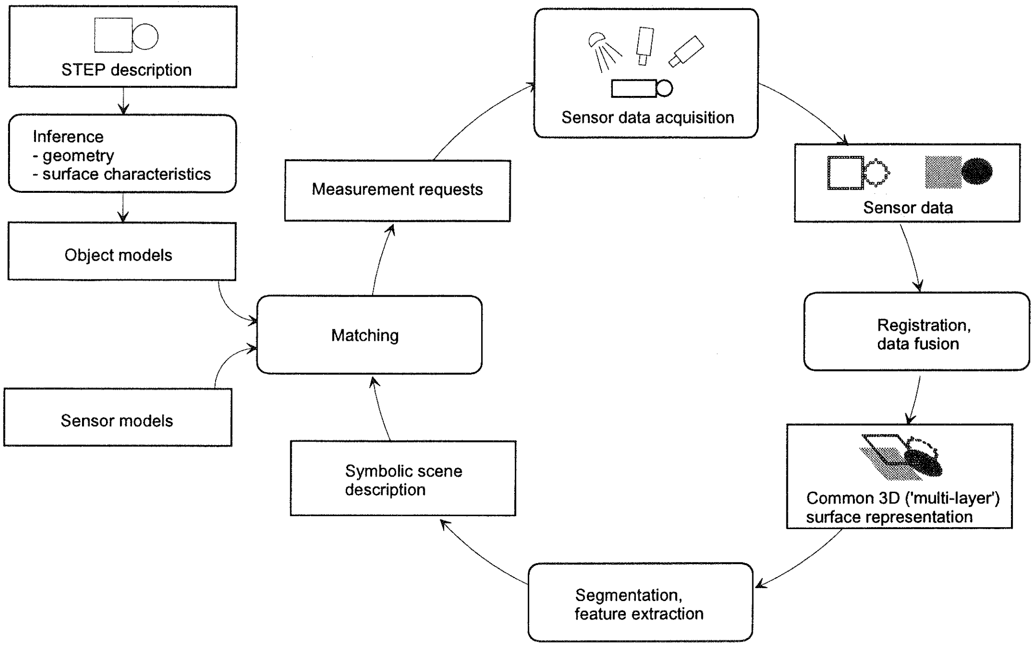

Figure 3: Object recognition approach using sensor fusion and active exploration.

shown in Fig. 3. The main idea behind this concept is that the complexity within the recognition process must be kept small. We start recognition with a small number of captured sensor data. In this case the search space is small but it is to be expected, of course, that matching will not come up with the recognized object. The hypothesis generation and verification scheme is now used to call for new sensor data. Next goal-driven new measurements are carried out which over several refinement steps may lead to a recognition in steps: at the beginning the object class is identified and at the end of the analysis the unknown object is recognized.

In the circular process (Fig. 3) the fusion of multi-sensor data is the other important characteristic. Clearly, the use of information from different sensors can be used to improve the quality of the segmentation result. E.g. range images contain information about the 3D shape of the imaged object more explicitly than intensity images. Therefore, segmentation of range images in physically meaningful parts is often much easier than the segmentation of intensity images. However, considering the spectrum of available sensors and the variable lighting, non-geometric properties can be captured as well. Surface roughness can be obtained either by high resolution distance imagery or by the use of high resolution intensity images (in connection with dedicated lighting). Surface color is captured by the color CCD camera. Using different light incidence angles, a general surface classification can be obtained from image sequences.

Having this data there are two tasks to be done: it must be incorporated into the segmentation and into the modelling of the object. Concerning the modelling we chose the ISO 10303 standard (STEP, (ISO 1994)), and particularly the application protocol "Core Data for Automotive Mechanical Design Processes" (10303-214) as a basis for deriving object models. This protocol allows surface properties like surface coating or surface roughness to be specified.

A prerequisite for the segmentation is that all sensor data is transformed to a common representation. In our case, all data is projected onto the reconstructed object surface, forming several layers of information. This requires the registration of the data. Often this step is done using the orientation of the sensor given by the measuring system, which makes a very precise and thus expensive positioning necessary. Another possibility is to use given sensor orientations just as an approximation and to fit the data according to positions of points which can be identified automatically in both datasets. We demonstrate this approach in the next paragraph.

4 FIRST RESULTS

To investigate our object recognition and location concept, we have carried out some experiments. Fig. 4(a) shows an industrial object as seen from one camera of the stereo camera. The object is made of free-form shaped sheet metal. At the dark areas in Fig. 4(a), the metal has been cut out by a laser cutter. The images have a resolution of 512×512 pixels.

By image matching, the relatively coarse height model shown in Fig. 4(b) is obtained. As expected, this coarse model cannot deal properly with the breaklines of the cut-out regions of the object. Nevertheless, since the cut-out regions show up very well in the intensity imagery, we can extract them using standard image processing. As shown in Fig. 4(c), however, this usually yields some spurious data as well. Thus, to improve our results, we use the larger of the detected features to form areas of interest which are then captured using the range sensor.

Fig. 4(d) shows a range image of the lower left part of the object. The image consists of 256×256 3D data points. Clearly, besides capturing the breaklines very well, the data

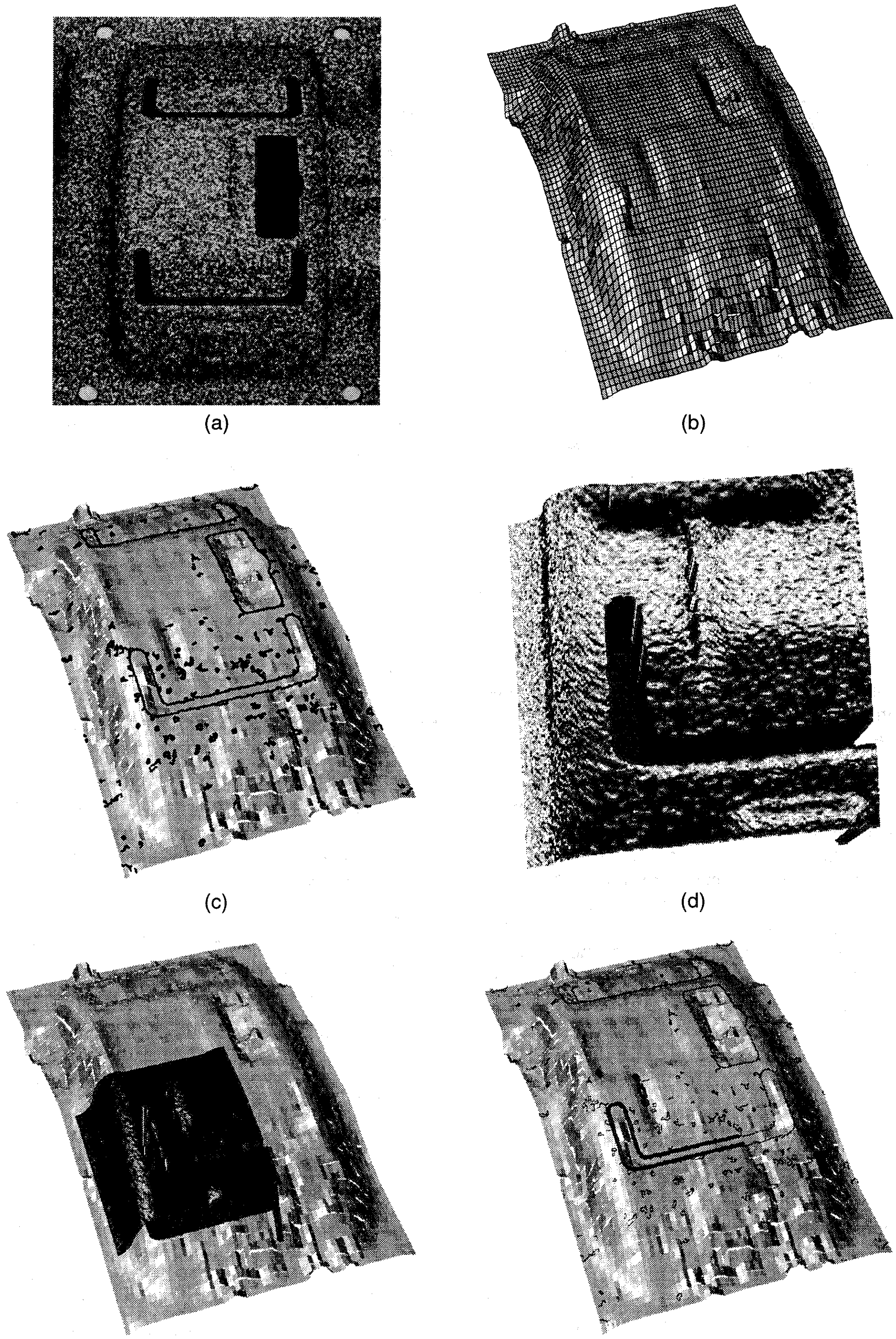

Figure 4: (a) original image of the object. (b) coarse resolution height model obtained by image matching. (c) result of intensity image feature extraction, transformed into 3D space. (d) range data set. (e) result of fitting range data set to the height model in (b). (f) overlay of features extracted from intensity and range images.

is also suited to assess the surface roughness. Since we have for both the photogrammetric and the range data height model synchronous intensity information, we used image matching to register both datasets and obtain the result depicted in Fig. 4(e).

In a further step, we extracted features from the range

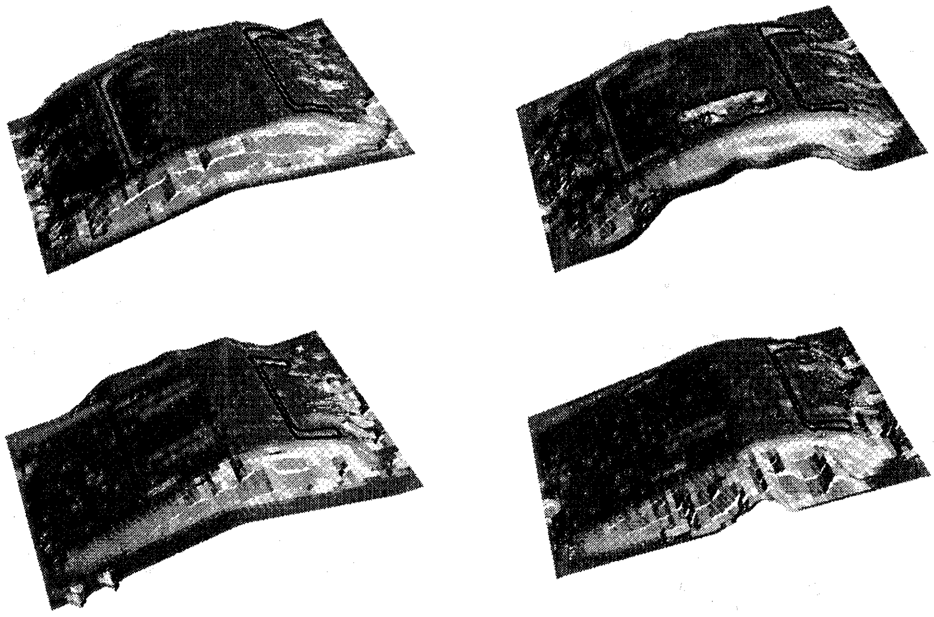

Figure 5: Four different objects with extracted breaklines.

dataset and mapped them to 3D using the registration found in the previous step. As can be seen from the overlay in Fig. 4(f), combining the intensity image extraction and the range sensor extraction gives the breaklines, as desired.

Fig. 5 shows the result of this extraction for four different objects. It can be seen that the upper right object can be distinguished immediately from the others. Thus, we have performed one cycle in our object recognition system of Fig. 3. To distinguish the remaining parts, e.g. the shape can be used. From Fig. 5 we see that the upper two parts are shaped cylindrically, while the lower two parts contain a large plain region. Capturing these differences can be done e.g. by using a camera in connection with dedicated light incidence angles.

5 OUTLOOK

We have proposed a new object recognition concept that incorporates multi-sensor fusion and active exploration into the recognition process. From multi-sensor fusion, we expect to obtain a significant quality improvement of the segmentation result. Active exploration is a key feature to avoid the problem of combinatorial explosion of the interpretation tree. We have shown some first examples which illustrate our proposed concept.

ACKNOWLEDGEMENT
We thank Mr. C. Voland of the Institute for Technical Optics for providing the range images.

References

Bhanu, B. & Ho, C.-C. (1987), 'CAD-based 3D object representation for robot vision', *IEEE Computer* **20**(8), 19–35.

Bolles, R. C., Horaud, P. & Hannah, M. J. (1983), 3DPO: A three-dimensional part orientation system, *in* 'Proc. of the Ninth International Joint Conf. on Artificial Intelligence, Karlsruhe, Germany', pp. 1116–1120.

Eggert, D. W., Bowyer, K. W. & Dyer, C. R. (1992), Aspect graphs: State-of-the-art and applications in digital photogrammetry, *in* 'Proc. ISPRS Congress 1992, IAPRS Vol. 29 Part B5 Comm. V', pp. 633–645.

Flynn, P. J. & Jain, A. K. (1991), 'CAD-based computer vision: From CAD models to relational graphs', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13**(2), 114–132.

Grimson, W. E. L. (1990), *Object recognition by computer*, Series in Artificial Intelligence, MIT Press, Cambridge, Massachusetts.

Grün, A. (1994), Digital close-range photogrammetry — progress through automation, *in* 'Proc. Comm. V Symp. Close Range Techniques and Machine Vision, IAPRS, vol. 30, part 5', pp. 122–135.

Hansen, C. & Henderson, T. C. (1989), 'CAGD-based computer vision', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11**(11), 1181–1193.

ISO (1994), *ISO 103031 Product Data Representation and Exchange Part 1: Overview and Fundamental Principles*.