Automatic Object Recognition
within an Office Environment

Michael Wünstel, Reinhard Moratz
Technologie-Zentrum Informatik (TZI)
Universität Bremen
Postfach 330 440, D-28334 Bremen
{wuenstel,moratz}@informatik.uni-bremen.de

Abstract

The visionary goal of an easy to use service robot implies some key features like spatial cognition, speech understanding and object recognition. Therefore such a system needs techniques to identify objects in scenes, i.e. to assign the natural category (e.g. "door", "chair", "table") to new objects based on their prototypical geometry.

Our approach uses $2\frac{1}{2}$ D laser range data to recognize basic objects like chairs or tables within an office environment. It is based on the concept of affordances; established on the work about form and function we identify certain geometries that lead to certain functions and therefore allow their identification. Our approach currently is restricted to basic objects but not limited to a special form. This is achieved by spatial abstraction where we assign the data to three layers. In identifying components in these layers of altitude, we reconstruct the basic form of the object, conclude its function and finally determine the object.

Keywords: Object Recognition, Laser Range Scanner Images, Affordances, Form and Function

1. Introduction

In this section a brief overview about existing object recognition strategies is given.

There are many diverse object/pattern recognition algorithms which can be classified with respect to different characteristics [1], e.g. the data representation, the object representation or the process of comparing an object with a reference. The most obvious characteristic is the given data representation. Here a distinction between light density images (2 D), range or surface images ($2\frac{1}{2}$ D - resulting e.g. from structured light or laser range systems) or object images (3 D - resulting e.g. from a CAD model or multiple laser range scans) is possible. However the type of model can differ from the type of image used during the object recognition process. Some systems like MORAL [3] contain a complete CAD reference model - nevertheless, during the actual recognition step "only" 2 D-density images are used. The next distinguishing mark is the type of object representation for the recognition process, which can be subdivided in the appearance based case where the model is given by a certain number of views, and the model based case where the components of the structure of the objects are known up to a certain point of detail. These two types are also known as recognition-via-localization and recognition-followed-by-localization [9]. In the first case, the image (or the dominant part of it) is used as a whole as input to the object recognition process like in the wavelet base approach in [7]. In the second case, the features first have to be extracted and then compared with the models of the object reference data base. The last overall distinction can be done regarding this classification process and covers methods like numerical classification, alignment, interpretation tree, relaxation or statistical methods and many more.

Many of the existing approaches use precise geometrical prior knowledge of objects and are designed to unify a currently sensed object with an object already known. A more complex task is to categorize unknown, novel objects by interpreting them with respect to semantic concepts (e.g. “chair”, “door”). Solutions of this problem would be very helpful in the context of human-robot interaction. A cognitive theory (“Recognition by Components”) about how humans find natural categories from visual input was proposed by Biederman [2]. It partitions objects into a set of 3D-primitives (geons) that are generalized cylinders. A semantic object category comprises a prototypical relative spatial arrangement of specific geons.

2. Related Work and Own Approach

In this section the motivation for our approach is presented and an introduction to our system is given.

2.1. Affordances, Form and Function

J. J. Gibsons theory on perceptions says that the perception occurs in order to act with the environment. The set of offered possible actions are the affordances. These affordances are properties of the environment relative to the actor. A comprehensive list of affordances can be found in [6]. Fire, for example, affords burns in one case or affords warmth in the other case. Zhang [11] develops a taxonometry of affordances based on Gibson which contains biological (”a healthy mushroom affords nutrition, a toxic mushroom affords dying”), physical (”a flat horizontal panel can only be pushed”), perceptual affordances (”switches of the stove provide controlling the burner”) and others more. In the case of the physical affordance there exist useful properties of objects that are fundamentally understandable to interact with. This has consequences for a functional design. As a simple example Norman [5] mentions a button that for everybody evidently is used by turning it or that slits are used to throw something in them.

These circumstances have some implications for the object recognition. When admitting that some objects have certain forms resulting from their functions, this form can be used to help identifying the objects function and ultimately the object itself. Then the affordance of an object is a visual clue to its function. As an example, stairs are characterized by a ratio of height to width of their steps that allow for climbing them comfortably [4]. These affordances are often closely related to the spatial arrangment of the object's components.

A related approach for object recognition using the term *Form and Function* [10, 8] will now be presented in more detail. This approach recognizes objects by analyzing how well they support the functionality of the possible object categories. Such an object category is defined by one or several functional properties. The *straight_back_chair* has the properties: *provides_sittable_surface*, *provides_stability* and *provides_back_support*. These functional properties in turn correspond to certain knowledge primitives which represent certain physical properties of shapes or orientation between shapes. The knowledge primitive *relative orientation* for example determines if the orientation between two surfaces falls below a certain threshold. Other primitives are dimensions, proximity, clearance or stability. The goodness of congruence between analyzed and reference object is measured using fuzzy logic. The system has been successfully used in a laser range finder environment. For that purpose 14 different shapes have been scanned 216 times. In a first step 211 images could successfully be segmented and afterwards topological relations could automatically be extracted. Afterwards, these descriptions were analyzed by a function-based reasoning module. With only few exceptions the images could successfully be assigned (partly correctly not definite) to a furniture.

2.2. Our Approach: Object Recognition system ORCC

We developed the object recognition system *ORCC*[1] which can handle the input data, can render it and perform our object recognition algorithms.

The fundamental processing chain of our object recognition algorithm is shown in Figure 1. The chain consists of three main parts: Data Acquisition, Object Modelling and Object Classification. Stark's [8] original form and function algorithm (as described in section 2.1) contains a segmentation work step within the $2\frac{1}{2}$ D data, which is finally based on geometric limitations of the allowed objects. They only used objects composed of cuboids with sharp corners. In our environment scenario freeform objects (preliminarily restricted to tables and chairs) have to be identified. A three-dimensional segmentation based approach is not suitable or necessary as we neither have nor need fully defined three dimensional object model descriptions.

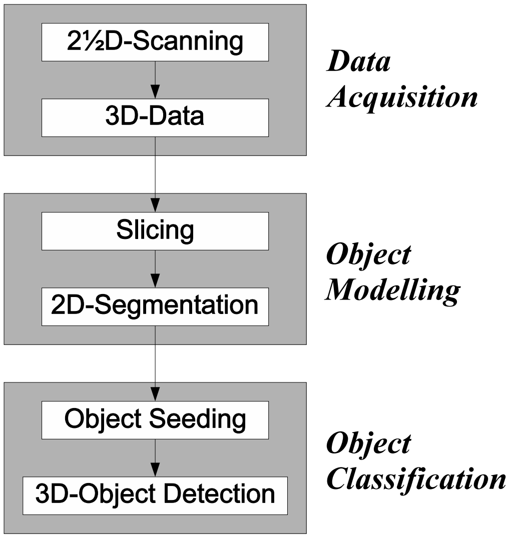


Figure 1. *ORCC* **Object Recognition Processing Chain**

The three-dimensional data resulting from the laser range image (see Figure 6) is sliced into several layers. We then project this data into a two-dimensional

[1] The acronym *ORCC* stands for *O*bject *R*ecognition using *C*ognitive *C*omputing

plane. Within this plane we now can robustly segment object parts by using standard methods. These segments, representing object parts in certain heights, are then used to identify the whole three-dimensional object. Finally, during this slicing process a spatial simplification of the scanned object is done. This approach consequently is only feasible if the resolution of the object during this simplification step does not completely disappear. Nevertheless, it is appropriate to detect certain objects that afford certain functions or like in our task certain furniture with sufficient precision.

3. Experiments and Results

In this section a more detailed view of our approach is given and results within a scanned office environment are presented.

3.1. Data Acquisition

The data used here has been received using a SICK[2] laser range scanner mounted on a pan-tilt unit. The scanner has a scanning angel of 180 ° and an angel resolution of 0.5 ° which corresponds to 361 points per scanning plane. The scanner is tilted about 180 times within a scanning angel of 90 ° (Scene A). The result is a $2\frac{1}{2}$ D picture of the scene. In a first step these data are converted into 3 D data using the position and resolution information of the scanner.

3.2. Object Modelling

During the object modelling work step, the different components which build up an object are extracted. As our models consist of segments in three different layers we first divide the three-dimensional data into three intervals concerning their z-axes value. The boundary values are found by training a gaussian Bayes classifier on a small sample. It turned out that the lowest interval is about 30-60 cm from the bottom (see Figure 2), the middle slice to an interval of 60-80 cm and the highest to an interval of 80-120 cm.

These slices are then projected to a plane with projection direction of the z-axis. Corresponding object segments are then obtained using standard morphological opening and closing operations. This leads to several segments per slice. Smaller segments which mostly result from noise in the data are detected and deleted.

Figure 2 shows the projected scan points in the lowest slice. The upper cloud corresponds to a sitting surface of a chair, the lower corresponds to a beverage box. The Figure 3 shows the resulting segments.

The different layers are used to build a slice model of the real object which is then compared with the given model objects.

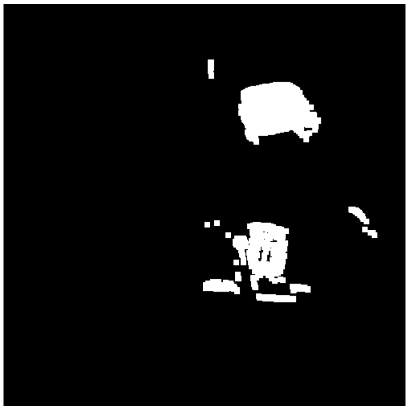


Figure 2. Data points of the lowest level (Scene A)

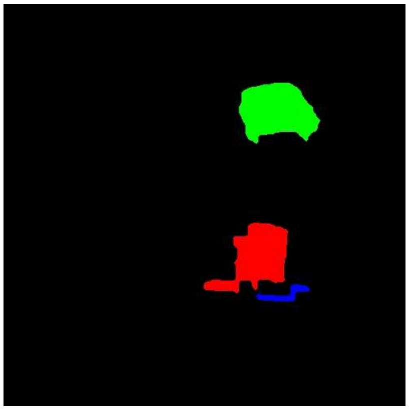


Figure 3. Resulting segments of the lowest level (Scene A)

3.3. Object Classification

The model objects consist of multiple segments with the *properties*:

- height (corresponds to the interval 1 to 3)
- minimal size (in pixels)
- maximal size (in pixels)

[2] see http://www.sick.com

Based on these properties, further ones like the center of gravity can be deducted. The segments are connected via *relation descriptions*. A relation contains the two involved segments, the relation type and when necessary a floating value. The actual integrated relation descriptions are:

- `isLower`
- `isHigher`
- `minDistCenter`
- `maxDistCenter`

`isLower` means that the height of the first segment is lower than the second segment, correspondingly `isHigher`. `minDistCenter` means that the center (center of gravity) of the first segment has a minimum distance as specified to the center of gravity of the second segment, correspondingly `maxDistCenter`. Due to the slicing process the space values have two dimensions.

Mathematically these models can be seen as a directed acyclic graph (DAG) where the nodes correspond to the segments (*properties*) and the weighted edges are the relations between the segments (*relation descriptions*). Figure 4 shows the graph representation of a chair model. It consists of a sitting surface in the lowest level, two arms in the middle level and a backrest in the highest level. The relations are not only given by explicit declaration of edges but also by the property values of the nodes themselves. As shown in Figure 4, the relation `Sitting-Surface isLower than Backrest` can be conducted by the segments level values.

The graph contains a specific node, namely the root. When detecting objects first these base segments are identified, dealing in a way as a seed crystal. Functionally this segment corresponds to the basic function of the object (for a chair it is the sitting surface). Starting from this identified root segment spatially neighboring segments are detected and the validity of attributes (height and size) and their relations to the root segment are verified. This can be iteratively repeated until no further successor exists. The final result is a set of more or less complete objects.

Figure 5 shows the result produced by the ORCC system. The system has recognized nine different segments in all three layers like listed in the upper part. Besides the segments ids their height, their size and their position is listed. For experimental reasons individual segments can be removed from the following processing steps. In the lower edit box the identified objects (in red) and their segment parts (in green) are listed. Simultaneously the segments referring to an identified object are colored with an object specific color in the upper list. These colors can be varied and are used in the rendering of the 3D scene.

Figure 6 shows the original image where the identified object are colored. The table surface is colored blue, the

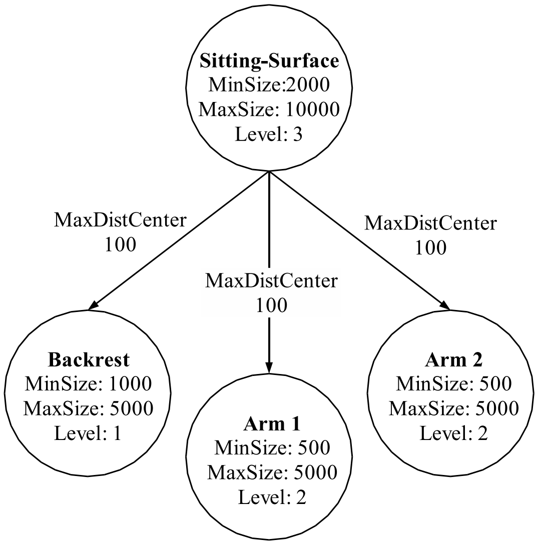
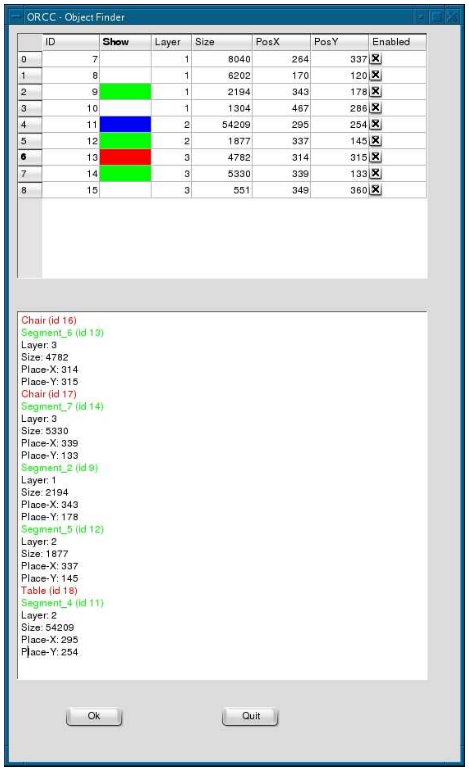


Figure 4. Graph representation of chair model



	ID	Show	Layer	Size	PosX	PosY	Enabled
0	7		1	8040	264	337	☒
1	8		1	6202	170	120	☒
2	9	🟩	1	2194	343	178	☒
3	10		1	1304	467	286	☒
4	11	🟦	2	54209	295	254	☒
5	12	🟩	2	1877	337	145	☒
6	13	🟥	3	4782	314	315	☒
7	14	🟩	3	5330	339	133	☒
8	15		3	551	349	360	☒

```
Chair (id 16)
Segment_6 (id 13)
Layer: 3
Size: 4782
Place-X: 314
Place-Y: 315
Chair (id 17)
Segment_7 (id 14)
Layer: 3
Size: 5330
Place-X: 339
Place-Y: 133
Segment_2 (id 9)
Layer: 1
Size: 2194
Place-X: 343
Place-Y: 178
Segment_5 (id 12)
Layer: 2
Size: 1877
Place-X: 337
Place-Y: 145
Table (id 18)
Segment_4 (id 11)
Layer: 2
Size: 54209
Place-X: 295
Place-Y: 254
```

Figure 5. Detected segments with deducted objects (Scene A)

chair green. Interestingly the box under the table is also recognized as chair (which is correct concerning the defined concept of Form and Function - you can also just simply sit on it).

Another Scene (Scene B) is shown in figures 7 and 8. Figure 7 shows the scene together with the scanning equipment, figure 8 the corresponding rendered range data again with the identified objects. Besides the table on the right hand side the three different chairs have been detected.

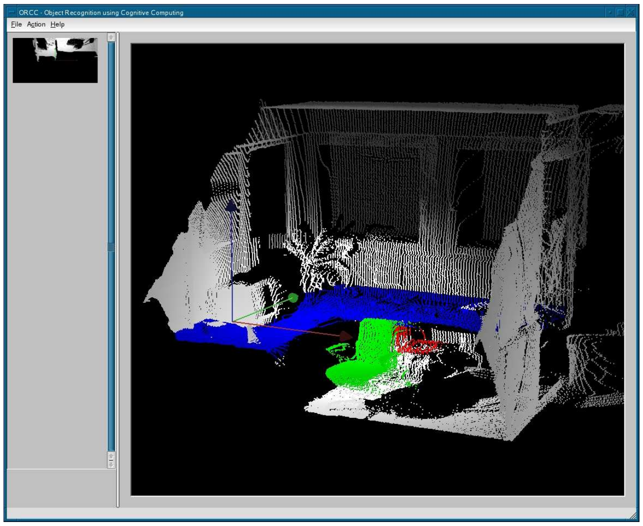


Figure 6. Rendered range data - the identified objects are colored (Scene A)

4. Conclusion and Outlook

The presented approach has been applied successfully within an office environment to detect basic furniture.

It has the advantage that it is not limited to particular features like the form of the surface as it is often the case when using a feature based approach. As the segmentation could be reduced to a two-dimensional case it has clear speed advantages. Our affordance based approach has its technical limits when the object boundaries are in the range of the scanner resolution. Consequently objects like closed doors, information signs or light switches may not be recognized. But as two-dimensional shapes these objects have the advantage that they are suitable for feature based two-dimensional object recognition.

The next steps will be to expand the knowledge base by relations between the objects themselves. In the future the equipment will be expanded by a light camera to combine the two- and three-dimensional methods.

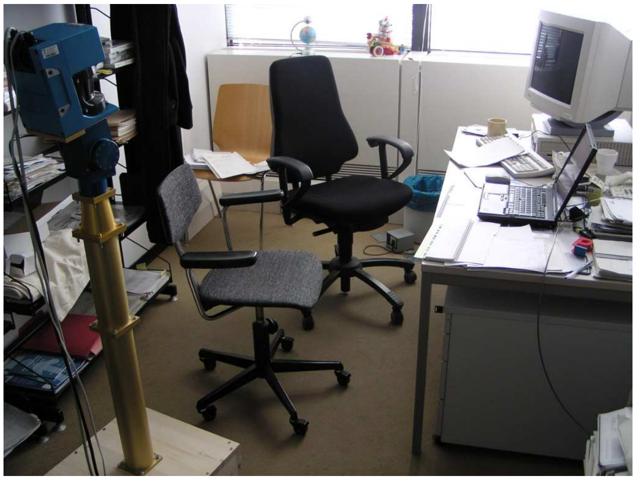


Figure 7. Scene B together with the scanning equipment on the left

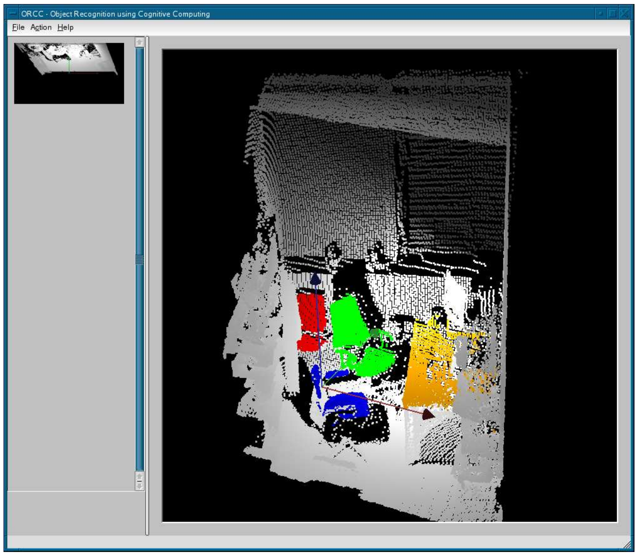


Figure 8. Rendered range data - the identified objects are colored (Scene B)

Acknowledgements: This project is supported by the Deutsche Forschungsgemeinschaft, DFG, through the project Transregio 08, Raumkognition - Schließen, Handeln, Interagieren.

The laser range scans (Scene A) have been generated by our project partners at Freiburg University - many thanks to Rudolph Triebel and Wolfram Burgard.

References

[1] J. Andrade-Cetto and A.C. Kak. *Wiley Encyclopedia of Electrical and Electronics Engineering*, chapter Object Recognition, pages 449–470. Wiley & Sons, New York, 2000.

[2] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

[3] Stefan Lanser, Olaf Munkelt, and Christoph Zierl. Robust video-based object recognition using CAD models. In U. Rembold, R. Dillmann, L. O. Hertzberger, and T. Kanade, editors, *Intelligent Autonomous Systems IAS-4*, pages 529–536. IOS Press, 1995.

[4] L.S. Mark. Eye height-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 13:361–370, 1987.

[5] Donald A. Norman. *Dinge des Alltags, Gutes Design und Psychologie für Gebrauchsgegenstände*. Campus Verlag, 1989.

[6] Edward Reed and Rebecca Jones, editors. *Reasons for Realism: Selected Essays of James J. Gibson*. Lawrence Erlbaum, 1982.

[7] M. Reinhold, D. Paulus, and H. Niemann. Improved appearance-based 3-D object recognition using wavelets. In T. Ertl, B. Girod, G. Greiner H. Niemann, and H.-P. Seidel, editors, *Proceedings of the Vision Modeling and Visualization Conference 2001 (VMV-01)*, pages 473–480, Berlin, November 21–23 2001. Aka GmbH.

[8] Louise Stark and Kevin Bowyer. *Generic Object Recognition using Form and Function*. World Scientific, 1996.

[9] Minsoo Suk and Suchendra M. Bhandarkar. *Three-Dimensional Object Recognition from Range Images (Computer Science Workbench)*. Springer-Verlag, 1993.

[10] Kevin Woods, Diane Cook, Lawrence Hall, Kevin W. Bowyer, and Louise Stark. Learning membership functions in a function-based object recognition system. *Journal of Artificial Intelligence Research*, 3:187–222, 1995.

[11] Jiajie Zhang. Categorization of affordances. Technical report, Department of Health Informatics, University of Texas at Houston, http://acad88.sahs.uth.tmc.edu/courses/hi6301/affordance.html (18 Dec. 2003).