3D Shape Detection for Mobile Robot Learning

Chapter · January 2009

DOI: 10.1007/978-3-642-01213-6_10

CITATIONS
5

READS
385

2 authors:

Andreas Richtsfeld
TU Wien
22 PUBLICATIONS **469** CITATIONS
SEE PROFILE

Markus Vincze
TU Wien
475 PUBLICATIONS **7,113** CITATIONS
SEE PROFILE

3D Shape Detection for Mobile Robot Learning

Andreas Richtsfeld and Markus Vincze
Technische Universität Wien
Gußhausstraße 27-19 / E376
1040 Wien, Austria
ari@acin.tuwien.ac.at, vincze@acin.tuwien.ac.at

Abstract

If a robot shall learn from visual data the task is greatly simplified if visual data is abstracted from pixel data into basic shapes or Gestalts. This paper introduces a method of processing images to abstract basic features into higher level Gestalts. Grouping is formulated as incremental problem to avoid grouping parameters and to obtain anytime processing characteristics. The proposed system allows shape detection of 3D such as cubes, cones and cylinders for robot affordance learning.

Introduction

Humans can solve even complex tasks such as manipulating or grasping objects by learning and using the relationship of object shape to the intended behaviour - also referred to as affordances. Robotics would profit considerably if a robot could learn these affordances [1] and relate them to visual percepts. However, at present visual perception delivers only image-based features. It is the objective of this work to present a method that extracts the basic 3D shape of objects for enabling more realistic robot learning. Fig. 1 gives an example of a robot that attempts to learn concepts such as moveability of objects.

There are several paradigms of making robots learn affordances. [2, 3] present an approach where a mobile robot observes the environment and autonomously learns a prediction model from its actions and observation data. An overhead camera is used to track colour blobs of the robot and the objects. Another possibility is to discover object affordances [4]. To track objects and estimate the object behaviour, they are also using colour blobs and a set of visual features to describe the 2D blob shape. As an example, a circle can roll but a rectangle cannot. Another possibility is learning high-level ontologies, as proposed in [5, 6], by clustering the point data from a laser range scanner with an occupancy grid map.

All these approaches for autonomous learning are dealing with simple vision data, which are determining relative positions of objects and use image-based point or blob features. However, to understand the usage of different objects the capability to abstract shapes from images into 3D shapes is necessary. This is par-

ticularly true for object grasping but also holds for navigation, where an abstraction of point data into geometric features reduces complexity, thus enables efficient learning and yields an effective description of affordances.

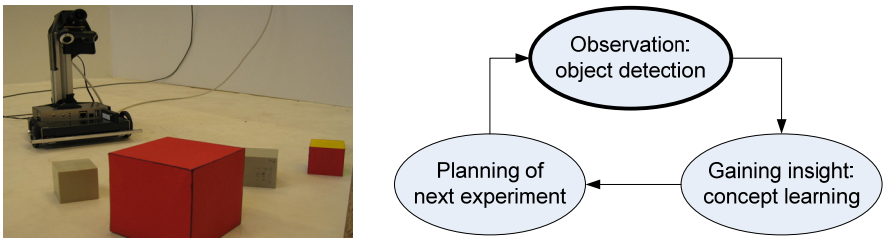

Fig. 1 Robot attempting to learn moveability of objects and the basic learning loop of planning robot motions, detecting objects (the focus of this paper) and gaining insights [2].

Hence we propose a system using methods of perceptual organisation to estimate basic object shapes in a hierarchical manner such that affordances can link to different abstractions as required. This aims at offering a generic tool for autonomous learning of robots in their environments.

Related Work

Perceptual recognition is a well known problem in computer vision and was research topic since the Eighties. Seminal works regarding the theory of perceptual recognition of simple objects are [7] and [8]. Biederman [7] proposed that there are non-accidental differences between simple objects, which can be derived through so called geons. [9] discusses the potential of geons in computer vision systems. Perceptual grouping is a bottom-up method to estimate Gestalts, where new hypotheses will be estimated from Gestalts of lower-level Gestalt principles. It can also be seen as geometric grouping, because most of the Gestalt principles implements geometric restrictions.

The main challenge of perceptual grouping is to limit the combinatorial explosion. Indexing and thresholding of less salient hypotheses are normally used to solve this problem [8, 10]. Indexing divides the search space for relations between elements into bins and each element will be allocated to a bin depending on the grouping relationship. Further all elements in a bin can be analysed whether they fulfil the given relation or not. A method for indexing in image space is proposed in [11], where search lines for a Gestalt feature are used to find intersections. Indexing into the image pixels limits the search space and avoids the problem of comparing all combinatorial possibilities to find connections. Depending on the type of search line, different types of groupings can be found.

A dilemma of using search lines is the length definition. When the length of the search lines is too small, some important Gestalts could not be found and on the other side, if they are too long, enhancement of computation time could be pro-

voked because many combinations may occur. This problem can be avoided with incrementally growing search lines [11], where in every processing circle another search line grows one pixel. When drawing into the indexing image, intersections with existing search lines are detected immediately in the respective bin. The biggest advantage of this method is that any preset parameters or thresholds are now moved into longer and longer processing. The results are no longer dependent to the line length, but to the processing time. The longer the process operates, the more junctions can be found and thus more other Gestalts can be detected. This gives a possibility to evaluate the best possible hypotheses for a limited processing time and on the other side this makes it feasible to obtain results at any time (a quality referred to as anytimeness). This grouping approach is suited for robotics, since it allows to sequential group higher level Gestalts such as cubes or cylinders in short processing time.

System Overview: Perceptual Grouping to Detect Object Shape

We propose an incremental grouping method to enable efficient abstraction of image pixel data into a hierarchy of basic geometric Gestalts shown in Fig. 2. Triggered by search lines in the image each Gestalt creates new hypotheses and delivers them to Gestalts of higher levels for robot learning.

When following the proposal of [11] it is possible to build a perceptual grouping system which processes most of the Gestalt principles incrementally. To extend from this work, we firstly applied the idea of incremental indexing to ellipses. Hence, we insert the Ellipse-Junction as Gestalt element. Boxes in the non-shaded area of Fig. 2 are referring to basic geometric features, whereas the boxes in the shaded area indicate Gestalts obtained by incremental processing. This shows that the essential Gestalts are the Line- and Ellipse-Junctions. They are are providing the system with incremental indexing capabilities.

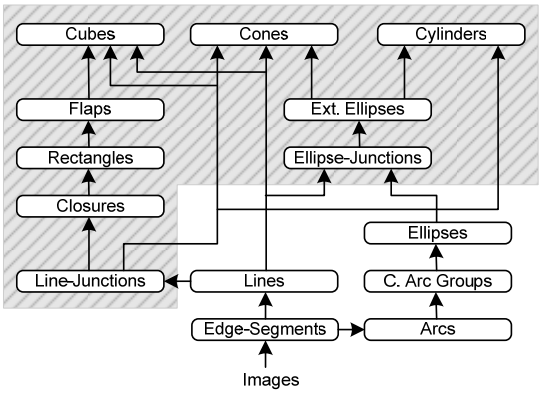

Fig. 2 Gestalt tree: basic geometric features and higher-level Gestalts incrementally built up (shaded).

In a first processing step the basic features are processed all at once and grouping uses neighbourhood constraints. Next, incremental processing of Line- and Ellipse-Junctions starts and the higher level Gestalts receive new input incrementally: each cycle extends one search line and all new hypotheses are processed at once to obtain new Gestalt hypotheses. If one of the Gestalts can hypothesise for a new Gestalt one level up the tree, the Gestalt from the next higher level begins to work, and so on. Hence the next junction will not be processed until all higher Gestalts are processed and no more hypotheses can be produced. This functionality of the processing tree guarantees that the best Gestalt hypotheses for the detected junctions are calculated in every processing circle.

The steps for building new Gestalts are the creation of new hypothesis, the ranking and masking, see Fig. 3. Whenever a Gestalt is informed about a new hypothesis, the processing starts and tries to build new hypotheses within this incoming Gestalt. The creation of new hypotheses exploits Gestalt principles from geometric constraints.

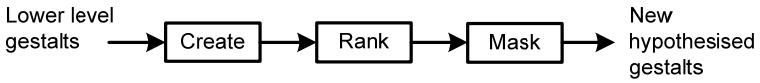

Fig. 3 Processing pipe for a Gestalt in the Gestalt tree.

After creation of new Gestalt hypotheses they will be ranked by quality criteria using the geometric constraints. Ranking is important to firstly calculate the best results and secondly for the masking of hypotheses. Masking is used to prune the poor results, e.g. when different hypotheses disagree about the interpretation of one element. Masking Gestalts is optional and is used to increase the performance and also to avoid combinatorial explosion in the following high-level principles.

Gestalt Principles

In this section we explain which Gestalt principles are used to build up the Gestalt tree of Fig. 2. The grouping of Edge-Segments into Lines, Arcs, Convex Arc Groups and Ellipses has been described in [11, 12]. Here we start with an explanation on how Line- and Ellipse-Junctions are formed to then introduce the approach to obtain the basic 3D shapes Cube, Cone and Cylinder.

Line-Junctions

The fundamental principles for incremental processing are Line and Ellipse- Junctions. Each processing circle starts with incrementing one search line of a line or an ellipse. Depending on which search line was chosen, the next hypothesis can be an Ellipse- or a Line-Junction. Line Junctions are connections between two lines. For each line six different search-lines are defined, as shown in Fig. 4, at each end one in tangential direction and two in normal direction. Different combinations of

intersecting search lines are generating different junctions. We distinguish between Collinearities, L-Junctions and T-Junctions, where T-Junctions could also be interpreted as two L-Junction.

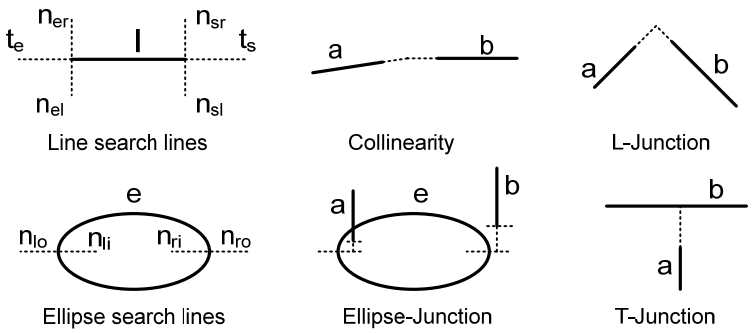

Fig. 4 Search lines for lines and ellipses and different junctions of lines and ellipses.

Ellipse-Junctions
Ellipse-Junctions are connections between a vertex of an Ellipse and a Line, which can also be found with search lines. Ellipse-Junctions have been defined to detect the higher-level Gestalts Cone and Cylinder. The search lines for Ellipses are growing on the main axis of the Ellipses, beginning at the vertices into both directions. For the growing of the search lines, we are using the same growing algorithm as for the search lines of Line-Junctions. Fig. 4 shows the search lines defined for Ellipses on both vertices.

Closures
Closures are closed convex contours, built from the Gestalts Lines and Line-Junctions, as proposed in [11]. Whenever the principle will be informed about a new Line-Junction, a new closed convex contour could be detected. The task is formalised as followed:

- Connect neighbouring lines to form a graph G = (V,E) with Lines as vertices (nodes) and Junctions between Lines as edges E of the graph.
- Perform a shortest path search on G, while making sure this path constitutes a roughly convex polygon.

To find these contours, Dijkstras algorithm for shortest path search is used, where only paths consisting of non-intersecting Lines, Collinearities and L-junctions of the same turning direction are allowed, thus ensuring simple convex polygon contours.

Rectangles
Rectangles can be directly derived from Closures by using geometric restrictions, which are given through the description of Rectangles in a perspective view. When considering these perspective projections as shown in [13] and neglecting

one vanishing point (one-point projection), it is possible to recognize Rectangles, whenever a new Closure appears. We build a new Gestalt hypothesis, when:

- A Closure contains four L-Junctions, and
- Two opposing Lines between the four L-Junctions are approximately parallel.

Flaps

A Flap is a geometric figure consisting of two Rectangles that do not overlap and that have one line in common. Detecting Flaps is the intermediate step for detecting Cubes, because two sides of a Cube are always building a Flap in a perspective view of a camera. When seeing a Cube aligned to one side only a Rectangle or Flap is visible, making the Flap an obvious Gestalt element. We can formalise: Whenever a new Rectangle is hypothesised, a new Flap can be built, if

- It is possible to find another Rectangle, which shares one line with the new hypothesised Rectangle, and
- The two Rectangles do not overlap.

Cubes (Cuboids)

An image taken from a camera may show different perspective views of three-dimensional objects. In the case of a cube there may be one, two or three rectangles visible. The chance that one observed rectangle indicates a cube is small and increases for a flap, but both can occur accidentally. Only when we are able to find three adequate rectangles, we can conclude that the robot observes a cube. Cubes and cuboids can be built up from Flaps, Lines and Junctions in two different ways, shown in Fig. 5. We can formalise the detection of Cubes in the following way: Whenever a new Flap is hypothesised, build a Cube, if

- There are two Flaps (upper line in Fig. 5), which share one Rectangle with the new Flap (area 1 and 2) and the two Flaps are sharing one Rectangle (3), or
- There is one L-Junction and Lines that close the Flap from the two outer corners at the smaller inner angle of the Flap (lower line in Fig. 5).

The second method for finding cubes is more general than the method with three Flaps and leads to more poor results, because connections between corners of one Flap can also occur accidently. On the other side it gives the possibility to detect cubes, where one rectangle could not be detected.

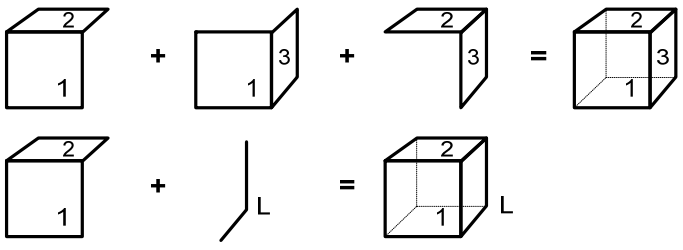

Fig. 5 Composing a cube from three Flaps or from a Flap with two Lines and an L-junction.

Extended Ellipses
Extended Ellipses are Ellipses with Lines attached trough Ellipse-Junctions. They are needed to build later the higher-level Gestalts Cone and Cylinder. We can describe the building of Extended Ellipses in the following way: Whenever a new Ellipse-Junction could be hypothesised:

- Build a new Extended Ellipse, if there is not one with the delivered Ellipse, or
- Assign the new Ellipse-Junction to the existing Extended Ellipse.

Cones
Cones are geometric figures consisting of ellipses (circles), lines and junctions between these components. We are using Extended Ellipses, Lines and Line-Junctions to find the object shape as given in Fig. 6. Building cones can be described in the following way:

- Whenever an Extended Ellipse (with at least one E-Junction at both vertices) could be hypothesised, build a Cone, if there is one L-Junction and Lines, which can close the Extended Ellipse.
- Whenever a new L-Junction can be hypothesised, build a new cone, if it is possible to close an Extended Ellipse with one Junction at both vertices.

Fig. 6: Composing a cone and a cylinder from lower level Gestalts.

Cylinders
The object shape of a cylinder consists of an ellipse and lines, as shown in Fig. 6. Therefore Cylinders can be grouped from Extended Ellipses, Lines and Junctions. Building new Cylinders can be formalised to:

- Whenever an Extended Ellipse (with at least one E-Junction at both vertices) could be hypothesised, build a Cylinder, if there is another Extended Ellipse, whose lines are connected with the lines of the new Extended Ellipse.

Experiments and Results

The incremental grouping method has been evaluated with a mobile robot moving among simple geometric 3D objects. Fig. 7 shows an example image and the edge image containing several three-dimensional objects. The picture indicates the typical problem of grouping, namely that shadows or image noise creates spurios features such as lines or arcs. A grouping into higher level Gestalts sometimes accidentially includes a wrong feature though, more often the Grouping principle constrain the search to actual higher level Gestalts.

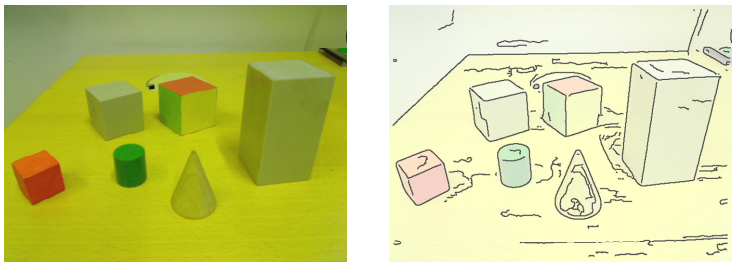

Fig. 7 Original image (left) and edge image with underlying half-transparent image (right).

With the incremental approach object detection depends on processing time. Fig. 8 shows in every row the results for different processing times. The first image (a) presents the resulting search lines, the second (b) the detected closures, the third (c) the best results of lower level Gestalts and the right column (d) the detected basic object shapes cube, cone and cylinder.

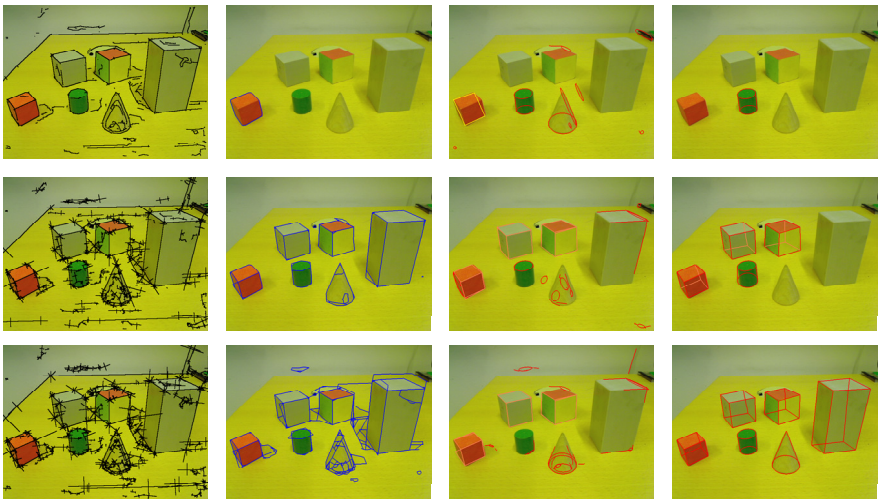

Fig. 8 Search lines (a), closures (b), the best ten lower level results (c) and simple object shapes (d) after 184ms (first row), 328ms (second row) and 468ms (third row).

Fig. 9 shows eight images from a sequence of a robot to a playground-scene with cubes (see Fig. 1). The whole sequence consists of 148 images, where 404 cubes could be observed in all images. For the detection of some cubes the processing time was too short or the view to the object too bad to detect the cube. In these cases the highest level Gestalt detected is shown. In nearly every image is at least one rectangle or a flap of a cube visible. Hence it is fair to assume that tracking of rectangles would be sufficient to follow cubes over degenerate views. For this sequence we evaluated the cube detection rate depending on processing time. The detection rate is 46% at 220ms processing time per image, 71% at 280ms (ex-

amples shown in Fig. 9) and 82% at 350ms, calculated with an Intel Core Quad Q6600 with 2.40GHz. We can see that the detection rate grows for increased processing time as expected, but we note that increasing processing time to more than 500ms does not lead to further improvements: search lines are long enough to detect what is possible.

This clearly indicates the potential of the approach. Fig. 10 shows therefore results of real world images in an office and living room scene. While previous robot learning methods had to cope with pixel data [5] or 2D blobs [4] this could be exploited to learn from object to robot relationships in 3D in [3].

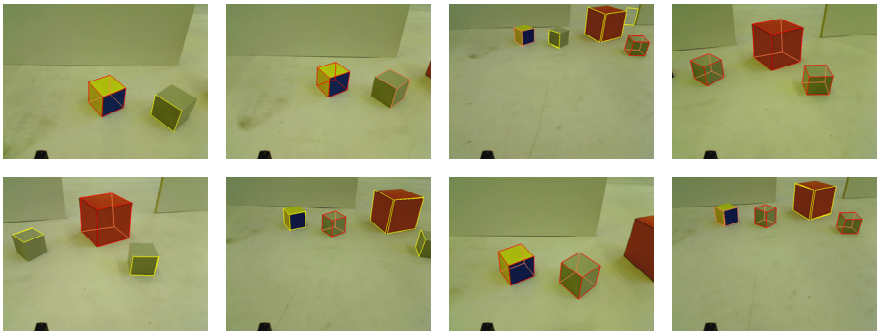

Fig. 9 Results for every 21 image from a 148 image sequence (from left to right) in a playground scene with 280ms runtime per image.

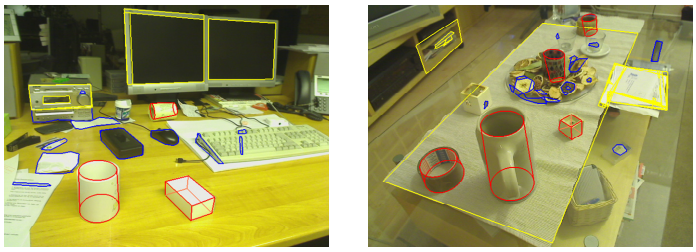

Fig. 10 Detected high-level Gestalts in an office and living room scene.

Conclusion and further work

We presented a hierarchical visual grouping system for the detection of basic geometric Gestalts such as cones, cylinders and cubes. With the incremental processing approach the problem of having parameters for each Gestalt principle is reduced to the single parameter processing time. The evaluation shows that basic shapes are detected with more than 80% detection rate. This makes it possible to follow the object motion over sequences. In [3] this has been used to learn affordances such as moveability and relate it to object size. And Fig. 10 indicates that

this method is capable of extracting the outlines of objects in everyday scenes for future use in service and home robotics.

In future work it is interesting to investigate how the approach in [4] can be extended from learning affordance relations to 2D image blobs to different object shapes and their behaviour when pushed or grasped. This would also exploit the ability to calculate the relative 3D positions of the objects and also of size and orientation under the condition of knowing the mounting point of the camera to the ground plane when using a triangulation.

Presently added is shape tracking using lower-level Gestalts. With such a tracking algorithm there will be no need to observe the objects from a viewpoint where several or all sides are visible. Once a three-dimensional object is detected, it is possible to follow it, e.g., when only a rectangle or a flap of a cube is observable.

Acknowledgments
The work described in this article has been funded by the European Commission's Sixth Framework Programme under contract no. 029427 as part of the Specific Targeted Research Project XPERO ("Robotic Learning by Experimentation").

References

1 Gibson J (1979) The Ecological Approch to Visual Perception. Boston, Hougton Mifflin
2 Zabkar J, Bratko I, Mohan A (2008) Learning Qualitative Models by an Autonomous Robot. 22nd International Workshop on Qualitative Reasoning, 150-157
3 Zabkar J, Bratko I, Jerse G, Prankl J, Schlemmer M (2008) Learning Qualitative Models from Image Sequences. 22th International Workshop on Qualitative Reasoning, 146-149
4 Montesano M, Lopes A, Bernardino J, Santos-Victor J (2008) Learning Object Affordances: From Sensory-Motor Coordination to Imitation. Robotics, IEEE Transactions on [see also Robotics and Automation, IEEE Transactions on]. Vol 24(1):15-26
5 Kuipers B, Beeson P, Modayil J, Provost J. (2006) Boostrap learning of foundational representations. Connection Science 18.2; special issue on Developmental Robotics
6 Modayil J, Kuipers B (2004) Bootstrap Learing for Object Discovery. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems; 742-747
7 Biederman I (1987) Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review, Vol 94(2):115-147
8 Lowe D G (1987) Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence, Vol 31(3):355-395
9 Dickinson S, Bergevin R, Biederman I, Eklund J et al (1997) Panel report: The potential of geons for generic 3-d object recognition. In Image and Vision Computing, 15(4):277–292
10 Sarkar S, Boyer K L (1994) A computational structure for preattentive perceptual organization: Graphical enumeration and voting methods. IEEE Transactions on System, Man and Cybernetics, Vol 24(2):246–266.
11 Zillich M (2007) Making Sense of Images: Parameter-Free Perceptual Grouping. Ph.D. Dissertation, Technical University of Vienna
12 Zillich M, Matas J (2003) Ellipse detection using efficient grouping of arc segments. In 27th Workshop of the Austrian Association for Pattern Recognition AGM/AAPR, 143-148
13 Carlbom I, Paciorek J (1978) Planar Geometric Projections and Viewing Transformations. Computing Surveys, Vol.10(4):465-502