Generic Modeling of 3D Objects from Single 2D Images

Guillaume-Alexandre Bilodeau and Robert Bergevin
Laboratoire de vision et systèmes numériques, Pavillon Adrien-Pouliot
Université Laval, Sainte-Foy (QC), Canada, G1K 7P4
bilodeau@gel.ulaval.ca, bergevin@gel.ulaval.ca

Abstract

This paper addresses the problem of building generic 3D models of structured objects on the basis of single 2D intensity images. In the context of this paper, generic modeling refers to the situation where analysis of the image information is performed on the sole basis of generic knowledge. That is, no a priori knowledge about the specific quantitative shape properties of the objects of interest is ever assumed. Moreover, images of interest are realistic. For instance, they may contain complex foreground 3D objects with textures and shadows, and a cluttered background. Objects are modeled by their constituent parts and connections. Therefore, a partly occluded object could be recognized from its model. Part models are based on geons, which are a set of qualitative generalized cylinders. An overview of the architecture of the modeling system is presented, along with the functionality of each subsystem and processing results.

1. Introduction

With the constant evolution of complex high-level control systems and on-line image and video database, there is a growing need for efficient and robust 3D object recognition systems. In this context, this paper presents results from an ongoing project aimed at building such a system.

The intended application of the system in development is flexible, efficient, and robust content-based access to image and video databases. Our main assumption is that qualitative 3D geometric models of object shapes are needed for such a task.

In the past, a variety of object modeling and recognition systems have been proposed. Many systems restrict themselves to planar shapes [1]. A number of systems model and recognize 3D objects by their constituent parts [2],[3],[4],[5],[6]. They differ in the way they model the parts and in the type of input data. Some systems use intensity images and generalized cylinders (GC) [2], or superquadrics [3] models. Range images are also used with such volumetric models [4], [5]. Other systems use contour data and qualitative 3D models [6]. Finally, a number of systems recognize objects based on their 2D appearances [7].

While most of these systems perform well for the type of images and objects they were designed for, no one is appropriate for our application. 2D models are too restrictive. Range images are not as common as intensity images in existing databases. GCs and superquadrics models are not qualitative enough. Finally appearance or viewer-based models are too cumbersome.

The system presented here builds a qualitative 3D model of a structured object from a single realistic 2D intensity image.

The paper is divided as follows. First, an overview of the system is presented in Section 2. This is followed by a description of each of its modules: feature extraction, part segmentation, and object modeling in Sections 3, 4, and 5, respectively. Finally, Section 6 concludes the paper.

2. System overview

The approach followed is based on works by Lowe [8], which restated that similarity, colinearity, proximity and symmetry are important clues used by humans to recognize objects. Works by Biederman [9] have also shown that humans recognize objects by their constituent parts. The main idea is thus to group image contour primitives (arcs and segments) based on generic models of projected volumetric parts. Each group is then associated to one member of a finite set of qualitative parts, the geons. The models obtained are well-suited for our application. They are qualitative and they allow recognition despite missing or spurious parts.

As shown in figure 1, the system is divided into three main subsystems: feature extraction, part segmentation and object modeling. The goal of the feature extraction subsystem is to extract arcs and segments from a 2D intensity image, and then extract the outline of the object from them. The intensity image may contain shadows, a complex background, and textured objects. From the features extracted, the part segmentation subsystem then groups the arcs and segments into parts structures using geometric relationships. An additional grouping clue is

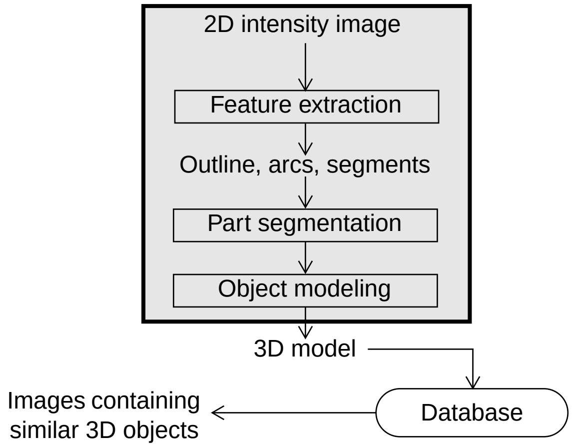

Figure 1. System overview.

needed to group lines with less ambiguity. Often, junctions are used for this purpose. However, it is still quite difficult to obtain reliable junctions from realistic intensity images. This is why the outline has been chosen as the additional clue. It is easier to extract and it also conveys information about the object structure. The part segmentation subsystem also computes the relationships between these parts. The object modeling subsystem then approximates the parts into a representation suitable for the description as geons. The spatial relationships between the parts are also modeled qualitatively by this subsystem.

3. Feature extraction

To obtain arcs and segments from an intensity image, the SE2D system [10] is used. SE2D is based on the Canny edge detector and custom contour segmentation algorithms. The outline is then extracted.

The outline extraction method consists of making a cycle with the arcs and segments of the image corresponding to the outline of the object. To do that, a first primitive line (arc or segment) is found. Then, from this line, a clockwise cycle is made using a proximity and orientation criterion to extend the cycle path. Each time the path reaches a dead-end, it is resumed at the last line where there were multiple possibilities for the next line (called a multi-possibilities point or MPP). The proximity and orientation criterion is used as followed:

$$l_n = \max_{l_i \in R} (\angle l_i - \angle l_c) \qquad \text{(EQ. 1)}$$

where l_n is the new line to add, l_c is the current line, l_i is a line in the search area R around l_c. This algorithm is simple and robust to image textures and noise. For instance, the next line to be added to the path is always the one making the largest counterclockwise angle (the line that is potentially the more outside of the object) relative to the current line. This way, interior textures and noise do not affect the path. However, as a result, the algorithm has more difficulty on images with complex backgrounds. For this reason, a number of starting lines are found by casting rays from various positions and directions around the image border. The best outline defined by length and regularity criteria is the one retained. Figure 2 shows the results of feature extraction on the image of a lamp. 168 contour primitives were extracted from the intensity image and 57 MPPs were encountered. The best outline was extracted in 0.22 sec. on a UltraSPARC II, 248 Mhz.

4. Part segmentation

The goal of this subsystem is to group lines such that they correspond to a substructure of the object. To do so, a similarity, proximity, and symmetry (SPS) criterion is used. Colinearity is also used to merge lines before the grouping process. The lines outside the outline are initially removed from the image lines.

Line grouping is done as followed. Two lines are paired if they are mostly symmetrical and nearby. The more two lines are parallel, close, and of about of the same length, the best the pair is considered. Therefore, the best pair is:

$$BP = \min_{l_i, l_j \in I} \left(\alpha \frac{\left\| \|l_i\| - \|l_j\| \right\|}{max(\|l_i\|, \|l_j\|)} + \beta d \right) \qquad \text{(EQ. 2)}$$

where l_i and l_j are two given lines of image I, d is the separation between the lines, and α and β are normalization and weighting factors. $\beta > \alpha$ and both factors are weighted by the level of parallelism of the pair of segments or arcs. Once two lines have been paired, they are considered as the two main *sides* of a part. The part boundary is completed for this pair by finding two paths joining their extremities. Lines may be grouped either with the help of the outline or simply using the geometric relationships.

4.1. Outline part extraction

Parts touching the outline should be the most reliable, since more information (geometric relationships and outline) are used to decide the line grouping. These parts are extracted by first selecting the longest available outline line. Then, the other outline lines are tested against the selected line to find the best pair using the SPS criterion. Once the best pair is determined, the boundary of the part is

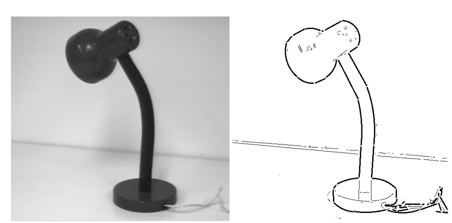

Figure 2. Feature extraction.

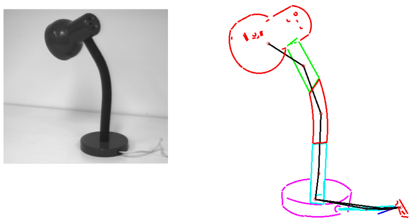

Figure 3. Parts of the lamp and their connectivity.

completed. The object lines lying inside the part boundaries are added to the part as its interior lines. All lines assigned to the part are removed from the image lines in order to speedup following processing, and a new outline line is selected to be paired. This goes on until it is no more possible to make a new valid pair. A variation of this method is then used to extract the internal parts, as described next.

4.2. Internal part extraction

An internal part, is a part which has either no line or a single line on the outline. This implies that this type of part may only be extracted with the help of the geometric relationships. Because of that, the internal texture lines can influence the internal part extraction process. This was not the case for the outline parts (texture lines are never on the outline). However, the SPS criterion and other tests ensure that texture lines are grouped as rarely as possible with part boundary lines. For example, other tests address special cases such as close-by parallel parts and self-intersecting part boundaries.

To extract the parts, the longest available line is selected first. It is tested against any other line (outline or interior line) of the image to find the best pair based on the SPS criterion (EQ. 2). The boundary is completed and the interior part lines are extracted as for outline parts. However, the lines are not removed for the image lines list, in case an erroneous pair is formed. This way, a bad pairing will not influence the other possible pairs.

By reviewing the complete line grouping process, it is clear that part segmentation rests on the quality of the outline. Also, it is clear that if the internal outlines (the outlines of the holes in the object) of an object were also found, the outline part extraction step could extract all object parts. This might improve the future performance of the system.

4.3. Parts configuration

After all parts are extracted, their spatial relationships are computed. At this time, the only relationship computed is connectivity. For instance, "part X is connected to part Y". This relationship will be refined at the next step.

To obtain these connections, the outline is used again. Parts that are consecutive on the outline are connected. However, this does not permit the computation of connections of internal parts. Proximity is used in this case. Since proximity is less reliable, the internal part connections are validated at the modeling step. Figure 3 shows results of part segmentation. To extract these eight parts, 949 pairs (29.3% of all possible pairs) were tested. Processing time was 5.5 sec.

5. Object modeling

To obtain an object-centered representation, the parts must be modeled as 3D shapes. For our purpose, we have chosen the geons. Those used here have either straight or curved cross-section edges, a straight or curved axis, are truncated or not, and have constant, expanding or expanding-contracting sections. In all, this makes 15 different geons. For more generality, we have added the sphere and the half-sphere to this set.

Because of our choice of part models, we are not concerned with the exact cross-section shape of the part. This is convenient since noise and textures make it very hard for some parts to determine the exact cross-section. Knowing whether a cross-section has arcs or segments is the only generic information to be determined from the image. Hence, 3 or 4 lines (arcs or segments) can represent the projection of any geon on a plane with the cross-section approximated by a single arc or a single segment without any loss of information. Taking these facts into account, the extracted image parts are first approximated by 3 or 4 lines and then associated to one or more geons according to the characteristics of the approximation.

5.1. Approximation of parts

Parts are approximated by searching on their boundary, points where there is a fast orientation change. The set of points obtained (T) is then reduced by keeping only the best set of 3 or 4 points. The optimizing criterion asks for a surface that covers as well as possible the area of the original part while forming corner angles that are as close as possible to 90 degrees. That is:

$$A = \min_{P_i \subseteq T} \left(\eta \left| Area(P_i) - AO \right| + \tau \sum_j \left| \angle P_i(j) - 90^o \right| \right) \quad \text{(EQ. 3)}$$

where P_i is a subset of 3 or 4 points in T, AO is the area of the original part, $P_i(j)$ is one of the points in P_i and η and τ are weighting and normalization factors. In practice, the second term is minimized first. The result is then compared according to the first term with an approximation built with the main sides of the part.

5.2. Merging the approximations

Before interpreting the approximations as geons, they are merged, if needed. Sometimes, because of texture and

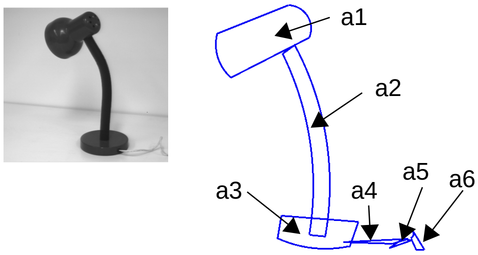

Figure 4. Parts approximation of a lamp.

noise, a substructure of an object is divided into two or more parts. Since the approximations are simple (3 or 4 points linked), it is easy to match their points to identify the parts that possibly need to be merged. A verification of the axes of the approximations and their area helps decide if the parts are indeed to be merged. Figure 4 shows the lamp after approximation. The approximation a2 results from merging three parts. The approximated model retains only information essential to geons modeling. This ensures that simple rules can be used for this purpose, as described next.

5.3. From approximations to geons

To model the approximations as geons, a set of rules is used. For example, if an approximation is made of two sets of parallel segments, it can be modeled as a prism with x% of certainty or a cylinder with y% certainty. In this example x is greater than y, because a cylinder can be approximated as a prism in very few viewpoints. These probabilities will be useful to adjust the quality of a match. Table 1 shows the results for the parts of the lamp.

Table 1: Geons obtained from the approximation

	Geon (%probability)	Geon (%probability)
a1	cylinder (89%)	lemon shape (11%)
a2	curved prism(70%)	curved cylinder (30%)
a3	cylinder (80%)	lemon shape (20%)
a4	pyramid (80%)	prism (20%)
a5	pyramid (80%)	prism (20%)
a6	cylinder (89%)	lemon shape (11%)

5.4. Connections modeling

The spatial relationships between geons are computed using the axes of the geons as referential. A relationships is described, if applicable, by its type (top-to-bottom, top-to-side), its position on the axis (top, bottom or middle) and by which side of the axis the connection occurs.

6. Conclusions and future works

This paper has presented a new object modeling system. It uses qualitative 3D volumetric shapes to model an object from a real 2D intensity image. Results obtained from a number of different images seem to be suited for the purpose it was designed for. Among the parts obtained (see Fig. 5), very few parts are missing or spurious and they can be modeled successfully as geons in most cases.

More work needs to be done to test and complete the system. Then, an efficient database architecture has to be developed to use the obtained models in the intended flexible, efficient, and robust content-based access to image and video databases application.

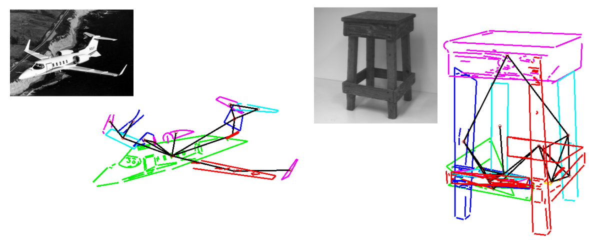

Figure 5. Parts of an airplane and a stool.

7. Acknowledgments

This work is supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC).

8. References

[1] F. Stein and G. Medioni, Structural Indexing: Efficient 2-D Object Recognition, IEEE Trans. Patt. Anal. Mach. Intell., Vol. 14, No. 12, December 1992, pp. 1198-1204.
[2] M. Zerroug and R. Nevatia, Three-Dimensional Descriptions Based on the Analysis of the Invariant and Quasi-Invariant Properties of Some Curved-Axis Generalized Cylinders, IEEE Trans. Patt. Anal. Mach. Intell., Vol. 18, No. 3, March 1996, pp. 237-253
[3] M. Pilu and R.B. Fisher, Model-driven grouping and recognition of generic object parts from single images, Robotics and Autonomous Systems, 21, 1997, pp. 107-122.
[4] F. Solina and R. Bajcsy, Recovery of Parametric Models from Range Images: The Case for Superquadrics with Global Deformation, IEEE Trans. Patt. Anal. Mach. Intell., Vol. 12, Feb. 1990, pp. 131-147
[5] D. Jr Dion, D. Laurendeau, and R. Bergevin. Generalized cylinders extraction in a range image. In Proc. Int. Conf. on 3D Digital Imaging and Modeling, Ottawa, May 1997, pp. 141-147.
[6] R. Bergevin and M.D. Levine, Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings, IEEE Trans. Patt. Anal. Mach. Intell., Vol. 15, January 1993, pp. 19-36
[7] C. Huang, O.I. Camps and T. Kanungo, Object Recognition Using Appearance-Based Parts and Relations, in Proceedings of IEEE Comp. Vision and Patt. Recog., Puerto Rico, June 1997, pp. 877-883.
[8] D.G. Lowe, Perceptual Organization and Visual Recognition, Kluwer Academic Publishers, 1985.
[9] T. Biederman, Human Image Understanding: Recent Research and a Theory, Comp. Vis. Graph. Image Proc., Vol. 32, 1985, pp. 29-73.
[10] J.L. Arsenault, R. Bergevin and D. Laurendeau, Extraction of 2D groupings for 3D Object Recognition, SPIE Orlando '94, Visual Information Processing III, Orlando, Florida, April 1994.