Recognition and Reconstruction of 3-D Objects using Model-based Perceptual Grouping*

In Kyu Park
School of Electrical Eng.
Seoul National University
Seoul, 151-742, KOREA
pik@sting.snu.ac.kr

Kyoung Mu Lee
Dept. of Electrical and Electronics Eng.
Hong-Ik University
Seoul, 121-791, KOREA
kmlee@wow.hongik.ac.kr

Sang Uk Lee
School of Electrical Eng.
Seoul National University
Seoul, 151-742, KOREA
sanguk@sting.snu.ac.kr

Abstract

In this paper, we address a new algorithm for recognition and reconstruction of 3-D polyhedral objects, based on perceptual grouping and graph search technique. Perceptual grouping is performed in a model-based framework, in which decision tree classifier is employed for learning and retrieving geometric information of the 3-D model object. On the other hand, in order to extract the polygonal patch structure, initial grouping result is represented by a Gestalt graph. Polygonal patch hypotheses are then generated by graph search and verified by the consistency test with the model. In the experiments, it is shown that the model-based grouping reduces the number of the generated hypotheses efficiently, and furthermore, robust recognition and reconstruction are achieved by means of the graph search technique.

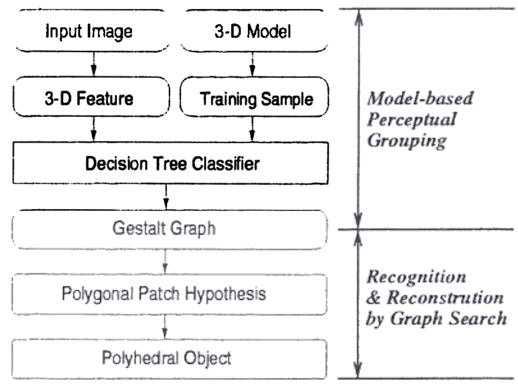


Figure 1. Block diagram of the proposed algorithm

1 Introduction

Over the past several years, computer vision community has noticed various applications of perceptual grouping in many computer vision tasks including stereo matching, model indexing, contour completion, figure-ground segmentation, change detection, and more [1][2][3][7][8]. The behind notion of perceptual grouping is that image features are not distributed at random, but exhibiting regular pattern, known as the Gestalt principles [2]. According to the Gestalt psychology, elements in an image are grouped from part into whole during the recognition process, based on proximity, parallelism, closure, symmetry, and continuation.

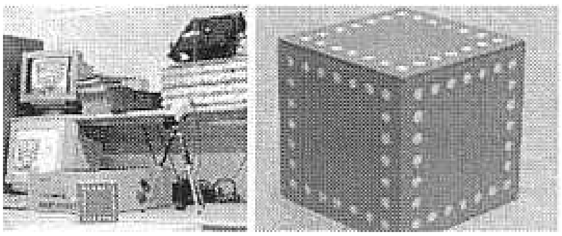
Although perceptual grouping for generic object is not a simple task, owing to the structural regularity of the underlying primitive features, the perceptual grouping approach is considered as a powerful methodology for the problem of extraction and recognition of polyhedral objects and man-made structures. Furthermore, the bottom-up grouping hierarchy is also useful for the reconstructing 3-D structure.

In this paper, a robust recognition and reconstruction algorithm is proposed, based on the model-based perceptual grouping algorithm. In order to achieve 'model-based', the proposed algorithm employs decision tree classifier [6], which is learned from the training samples extracted from the model. The input instances are fed into the decision tree, yielding Gestalt relationships between them. Gestalt graph is then established to model the network of 3-D features. Based on this representation, graph search is performed to generate polygonal patch hypotheses, in which each of them consists of the consistent chain of features. In this procedure, subgraph decomposition is used to reduce the search space. The overall block diagram of the proposed algorithm is shown in Figure 1.

This paper is organized as follows. In Section 2, a method for extracting linear feature and range image is pre-

*This work was supported by the Agency for Defence Development (ADD), Taejon, Korea, and the Automatic Control Research Center (ACRC) in Seoul National University, Seoul, Korea.

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE



(a) (b)

Figure 2. Space encoding range finder (a) System configuration (b) Calibration box with known reference points

sented. Learning and evaluating decision tree is described in Section 3. In Section 4, algorithms for constructing polygonal patch and polyhedral shape are given, based on the proposed graph representation. The experimental results are provided in Section 5. Finally, we conclude in Section 6.

2 Feature Extraction and Imaging System

Our main concern is to group image features perceptually. In our approach, it is assumed that the model object has polyhedral representation, which is common situation for the most man-made objects and aerial scenes. Since the line edge drawing in 2-D characterizes 3-D polyhedral model efficiently, line segment could be useful feature in our work.

2.1 2-D Line Feature Extraction

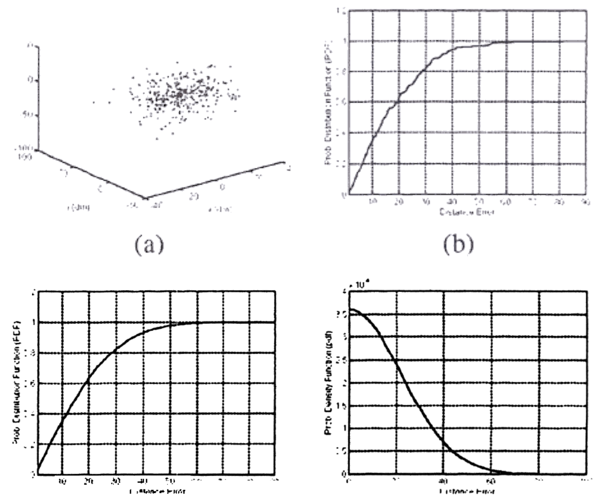
In order to obtain line segments, we first extract edgels using the Nevatia-Babu edge detector. Note that the performance of the Nevatia-Babu edge detector shows comparable performance to the Canny's edge detector, especially when we deal with objects of line features as in our case. Besides, it is much faster than the Canny's. The detected edgels are linked together by searching 8-connectivity neighborhood. Then, each linked chain is fitted by several straight line segments, taking into account the approximation error and minimum length.

2.2 3-D Range Image Acquisition

The grouping process is performed in 3-D in our approach. In our work, the space encoding technique is used for 3-D range imaging. The system configuration is shown in Figure 2 (a). Note that other techniques, such as laser scanner and stereo matching, can also be used if available.

2.3 Noise Modeling of the Range Imaging System

In order to make the further process be more reliable and robust, noise modeling is performed by measuring and analyzing the noise distribution of some reference points with known 3-D coordinates. Figure 2 (b), shows a box with 72 reference points on it. Several hundreds of range data are measured for those points, while rotating the box. The error distribution of measured data is shown in Figure 3 (a), in which the measurement error is scattered in omnidirection, forming a cluster.



(a) (b) (c) (d)

Figure 3. Error modeling of the range finder (a) Error distribution at the reference points (b) Measured probability distribution function (c) Estimated Gaussian distribution function (d) Estimated Gaussian density function

Since the direction of the error is observed to be almost random, the distance from the true position to the measured can be regarded as a random variable $\hat{X}$. In order to derive the *a posteriori* probability density of $\hat{X}$, the cumulative distribution function $\hat{F}(x)$ is first considered as follows.

$$\hat{F}(x) = Prob(\hat{X} \leq x) = \frac{\#\text{ (Samples with error less than } x)}{\#\text{ (Total error samples)}} = \frac{1}{N}\int_0^x \sum_{k=0}^{N-1} \delta(s - s_k)ds, \quad (1)$$

where s_k is the position of the distance error. In Figure 3 (b), obtained $\hat{F}(x)$ is shown. Since $\hat{F}(x)$ is not differentiable in this case, functional fitting is performed by assuming Gaussian density to find the variation σ_m^2, with which the resultant cumulative distribution is best fit to the measured $\hat{F}(x)$. The obtained fitted distribution function, $F(x)$, actually approximates the $\hat{F}(x)$ very well as in Figure 3 (c). The resultant density function, $f(x)$, is also shown in Figure 3 (d).

The measurement error is observed to be increasing in proportion to the distance D from the focal plane to the object. The effect of the different system configuration can be compensated by considering the random variable X, which

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE

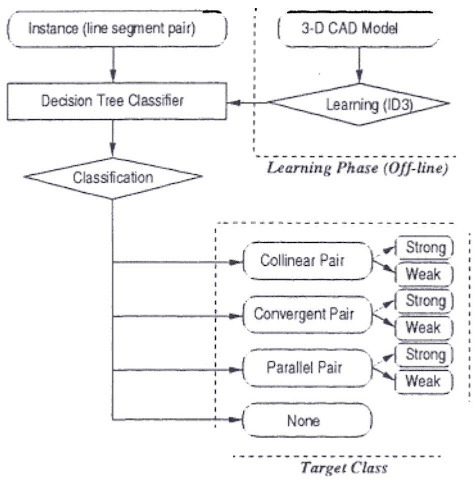


Figure 4. Block diagram of the classification using decision tree classifier

is given by

$$X = (\frac{D}{D_m})\hat{X}, \tag{2}$$

where D_m represents the mean distance from the focal plane to the reference points. Then, the noise variance is obtained as follows.

$$\sigma^2 = (\frac{D}{D_m})^2\sigma_m^2 \tag{3}$$

3 Learning Gestalt Principles Using Decision Tree Classifier

In the first-level grouping procedure, collinear, parallel, and convergent pairs of 3-D lines are extracted by employing decision tree learning technique [6]. Each of them is then classified as strong or weak group according to the combinational completeness. In our approach, decision tree is constructed by learning training examples, which are actually obtained from a given 3-D CAD model. Learning task is performed by using the ID3 algorithm [6]. The overall block diagram of the classification is shown in Figure 4.

3.1 3-D Reference Model and Training Samples

As a usual model-based approach, we have 3-D CAD model as shown in Figure 5 (a). It consists of several line segments and polygonal faces. The training samples are basically the set of each pair of 3-D line segments of the model, which is represented in a vector form.

3.2 Target Class

The target class consists of four categories including collinear (**CL**), parallel (**PP**), convergent (**CV**), and none (**NO**), and each class is further classified into strong (**S**) or weak (**W**) class according to the confidence level. The confidence level is measured by evaluating the supportive evidence of each instance. The supportive evidence of collinear pair is the distance, δ, between the end points. The distance is compared with the prespecified threshold, yielding the decision rule as follows.

$$\begin{aligned} \mathbf{CL}_S &= \{\mathbf{CL} \mid \tfrac{\delta}{min(l_1,l_2)} \leq \alpha\} \\ \mathbf{CL}_W &= \{\mathbf{CL} \mid \tfrac{\delta}{min(l_1,l_2)} > \alpha\}, \end{aligned} \tag{4}$$

where l_1 and l_2 denote the length of two line segments, respectively. The supportive evidence of the parallel pair is the overlapping ratio after the normal projection of one on another. The test is formalized as follows.

$$\begin{aligned} \mathbf{PP}_S &= \{\mathbf{PP} \mid \tfrac{L_O}{min(l_1,l_2)} > \beta\} \\ \mathbf{PP}_W &= \{\mathbf{PP} \mid \tfrac{L_O}{min(l_1,l_2)} \leq \beta\}, \end{aligned} \tag{5}$$

where L_O represents the overlapped length. On the other hand, the supportive evidence of the convergent pair is the ratio of the segment length to the virtual perfect one. It measures the completeness of forming corner. The decision rule is given by

$$\begin{aligned} \mathbf{CV}_S &= \{\mathbf{CV} \mid min(\tfrac{l_1}{L_1}, \tfrac{l_2}{L_2}) > \gamma\} \\ \mathbf{CV}_W &= \{\mathbf{CV} \mid min(\tfrac{l_1}{L_1}, \tfrac{l_2}{L_2}) \leq \gamma\}. \end{aligned} \tag{6}$$

where L_1 and L_2 mean the length of the complete line segments, which is extended to the corner point. Note that α, β, and γ in (4), (5), and (6), respectively, denote the decision parameters determined empirically.

3.3 Node Attributes

The node attributes should provide proper measures to distinguish geometric differences among different target classes. In order to distinguish convergent from collinear and parallel, the between-angle property (θ) is first considered. Next, d_{min} is also used, which is actually defined as the minimum distance between infinitely extended line of each line segment, to encode relative position. In addition to them, since parallel segments must be overlap each other, projected overlapping ratio, R, is also taken into account. Note that in 3-D, coplanarity, C, is a necessary condition for a pair of line segments to be grouped as one of the non-null target classes discussed before, and this can be evaluated by calculating the deviation of points from the fitted plane. In summary, an instance is represented by an attribute vector **I**, which is given by

$$\mathbf{I} = \{\theta, d_{min}, R, C\}. \tag{7}$$

3.4 Discretization of the Continuous Attributes

Since ID3 assumes that all attributes are categorical, the numerical attributes should be discretized prior to tree construction. In obtaining training samples from 3-D polyhedral model, since the given model is a perfect CAD model,

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE

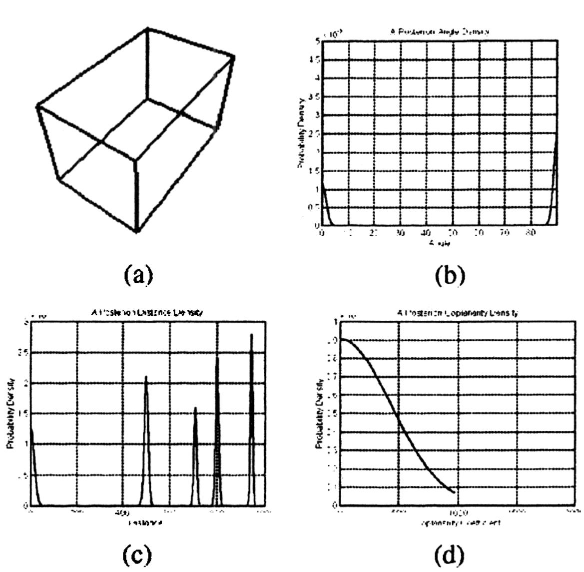


Figure 5. Attribute noise modeling for simple box model (a) Simple box model (b) Probability density of the angle attribute (c) Probability density of the distance attribute (d) Probability density of the coplanarity attribute

θ, d_{min}, and C have a few dominant values. Due to the inherent noise added in range finding, the observed attribute should be distorted by the positional error of the line segments. Since it is not a simple task to compute the probabilistic distribution of the attributes, in our approach, instead, an instance-based learning technique is adopted. That is, lots of random instances are generated using the known positional error distribution. Then, the attribute values are collected, which can be thought that the samples are randomly generated according to the unknown probability density. In this context, the probability distribution is inversely estimated from the samples. Note that it can be also assumed that the density is Gaussian $N(m, \sigma_A^2)$.

In Figure 5, a simple model is shown and the generated probability density is plotted for the angle, minimum distance, and the coplanarity attributes. Using 99 % of confidence level, each density function $N(m, \sigma_A^2)$ produces a discretization interval $[m - 2.58\sigma_A, m + 2.58\sigma_A]$. When the interval is overlapped, the maximum *a posteriori* decision rule is applied. Each interval has the classification class of original model instance. The uncovered interval corresponds to the class **NO**.

4 Reconstruction of Polyhedral Shape

As a result of the first-level grouping, useful Gestalt relationships between each pair of line segments are obtained. In order to recognize and build the 3-D polyhedral structure, polygonal patches are then first extracted based on the Gestalt graph, which is constructed based on the grouping results.

4.1 Graph Representation and Subgraph Decomposition

In our approach, the network of the Gestalt relationships is generated. Our graph representation puts the derived Gestalt principles together into a graph simultaneously as in [4]. The Gestalt graph, $\mathbf{G}$, is defined as a 3-tuple, given by $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{DT})$, in which $\mathbf{V}$, $\mathbf{E}$, and $\mathbf{DT}$ denote the set of vertices, edges, and the decision tree classifier, respectively. Note that, in case that the output of **DT** is **CV** or **PP**, the edge attribute is not only the Gestalt principle but also the coefficient of the plane on which the attributes lie. Some of the major normal directions are collected. Then, according to the information, the nodes can be clustered into a few subgraphs. Therefore, a subgraph consists of the line segments, which can be inside of a thin plate of space. Using the subgraph constraint, the computational complexity in the patch hypothesis generation procedure can be much reduced.

4.2 Synthesis of Polygonal Patch Hypothesis

In order to extract the 3-D polygonal structure, 3-D alignment method is employed in our approach. First, a super-node is considered, which is defined as a pair of connected nodes, of which edge attribute is either **CV** or **PP**. A pair of super node is selected to determine whether they can be aligned to the geometric structure of the 3-D model. During the alignment, relative angle, corner distance, and projective distance are tested. After the alignment procedure, only 3-D polygonal structures which is formed from the model object are remaining. Note that alignment method is actually second level model-based grouping, while decision tree classifier is first level model-based grouping procedure.

5 Experimental Result

Experiments are carried out on the block test images shown in Figure 6 and Figure 7, in which the input is both intensity and range image. The model object to be recognized is shown in Figure 6 (c) and Figure 7 (c). After the first level and second level grouping process, polygonal patch hypotheses are generated as shown in Figure 6 (d) and Figure 7 (d), respectively, in which all the hypotheses are shown together. Note that in Figure 7 (d), another group is obtained which has very similar geometric properties to the given model. Although the grouping results may yield two or more groups of 3-D polyhedral object, *i.e.*, connected 3-D polygonal patches, we can select the desirable group easily by simple alignment to the model. The whole process takes less than 1 minute on Pentium II 400 MHz processor.

In Figure 8, synthesized photo-realistic model is shown in virtual environment. Note that the texture is extracted

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE

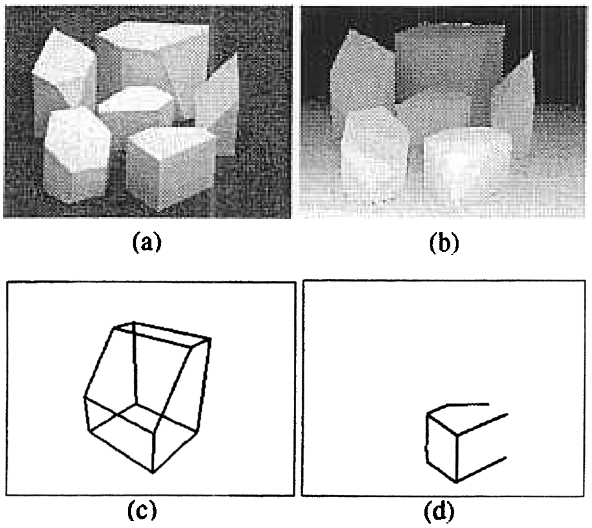


Figure 6. Experimental results (a) Input intensity image (b) Input range image (c) Model object (d) Grouping results

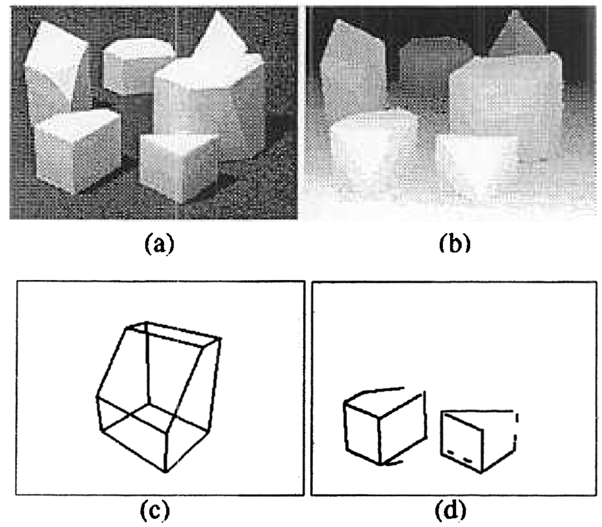


Figure 7. Experimental results (a) Input intensity image (b) Input range image (c) Model object (d) Grouping results

from the input intensity image and mapped onto the generated polygonal hypotheses, yielding reconstructed 3-D polyhedral objects. In Figure 8 (b) shows the back side of the reconstructed object, which does not have texture data in some facet. Note that the complete photo-realistic model could be obtained by using multiple views.

6 Conclusion

In this paper, we have proposed a model-based perceptual grouping for recognition and reconstruction of 3-D polyhedral object. The main idea of the proposed algorithm is to utilize not only the Gestalt principles but also the information provided by the given 3-D CAD model. The noise model is considered to obtain the discrete intervals for the continuous attributes in the decision tree classifier.

Our future work includes grouping-based object recognition and tracking in image sequences and use of other 3-D sensing methods, such as stereo matching. Note that we are also considering using the proposed method in image-based modeling, in which simplified 3-D world is constructed by fitting the objects in the scene to a few simple geometric primitives such as cube, cylinder, cone, and sphere.

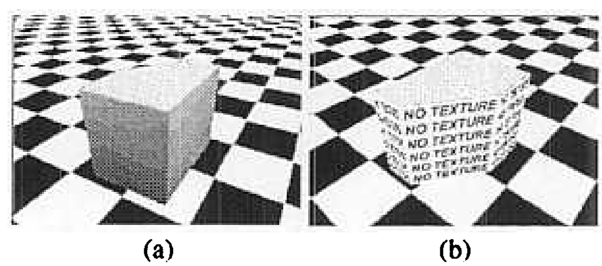


Figure 8. Reconstructed 3-D polyhedral object in virtual environment (a) Photo-realistic object (b) Back side

References

[1] R. Mohan and R. Nevatia, "Using perceptual organization to extract 3-D structures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 11, pp. 1121-1139, Nov. 1989.

[2] S. Sarkar and K. L. Boyer, "A computational structure for preattentive perceptual organization: Graphical enumeration and voting methods," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 24, no. 2, pp. 246-267, February 1993.

[3] P. Havaldar, G. Medioni, and F. Stein, "Perceptual grouping for generic recognition," *International Journal of Computer Vision*, vol. 20, no 1/2, pp. 59-80, 1996.

[4] V. Murino, C. S. Regazzoni, and G. L. Foresti, "Grouping as a searching process for minimum-energy configuration of labelled random fields," *Computer Vision and Image Understanding*, vol. 64, no. 1, pp. 157-174, July 1996.

[5] K. L. Boyer and S. Sarkar, "Perceptual organization in computer vision: Status, challenges, and potential," *Computer Vision and Image Understanding*, vol. 76, no. 1, pp. 1-4, October 1999.

[6] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.

[7] *Proceedings of IEEE Workshop on Perceptual organization in Computer Vision*, Santa Barbara, CA, June 1998.

[8] *Proceedings of IEEE Workshop on Perceptual organization in Computer Vision*, Corfu, Greece, September 1999.

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE