OBJECT RECOGNITION USING THREE-DIMENSIONAL INFORMATION

Masaki Oshima and Yoshiaki Shirai

Electrotechnical Laboratory
1-1-4, Umezono, Sakura-mura, Niihari-gun
Ibaraki, 305 Japan

ABSTRACT

This paper describes an approach to the recognition of stacked objects with planar and curved surfaces. The range data of a scene are obtained by a range finder. The system works In two phases. In a learning phase, a scene containing a single object Is described In terms of properties of regions and relations between them. This description Is stored as an object model. In a recognition phase, an unknown scene Is described In the same way as In the learning phase. And then the description is matched to the object models so that stacked objects are recognized one by one. Efficient matching is achieved by a combination of data-driven and model-driven search process. Experimental results for blocks and machine parts are shown.

I INTRODUCTION

This paper describes a method for object recognition using three-dimensional data of a scene. Our aim is to deal with scenes which have the following features:

(1) Objects are placed in any 3-D position with any orientation.
(2) Objects may be stacked in a Bcene.
(3) Objects have planar and/or smoothly curved surfaces.

Recognition of real objects has been studied aiming at realization of automatic inspection, assembly and so on. For this purpose, the recognition method should be flexible and efficient enough to analyze scenes with stacked objects and recognize them. It Is well known that even recognition of a simple object Is not easy if the object is allowed to rotate In 3-D space. The shape identification methods using a monocular gray picture [l, 2] assumed many constraints on Input scenes. Recognition of scenes with stacked objects was studied for simple scenes [3].

If range data la available, more flexible recognition can be achieved because 3-D shapes of objects are directly obtained. In order to analyze a scene with multiple objects occluding one another, it is often necessary to make a scene description which Includes useful information for recognition. There have been several studies on scene description using range data [4-12]. Some methods assumed planarity on objects [4, 10, 12]. Sugihara [9] proposed a way for trihedral objects. Agin [5], and Nevatia and Binford [6] described a scene including curved objects with generalized cylinders. Although the method is advantageous in describing body which possesses considerable elongation, there may arise some difficulties for other kind of objects. We have developed a method which describes a scene with planar and smoothly curved surfaces [8]. The method Is applicable to general scenes with real objects. Recognition of objects usually requires pattern matching process: matching a part of a scene description to a part of an object model. The control strategy for matching is important in processing a complex scene. The blind search is inefficient under circumstances where objects are stacked and the position and orientation of them are not known a priori. Our system, at first, tries to get reliable and useful features of a scene and to match them to models so that the result may guide the further processing. Thus the system realizes flexible and efficient recognition.

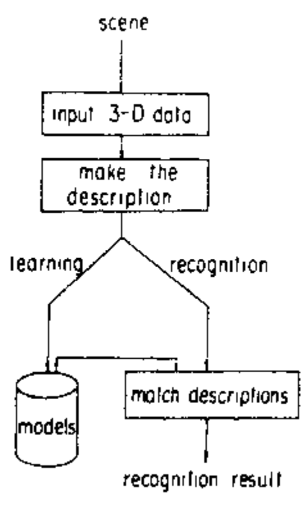

Fig. 1. Main control of the system.

II OUTLINE

The system works In two phases: learning and recognition as shown in Fig. 1. In learning phase, known objects are shown to the system one by one. The system makes a description of a scene in terms of properties of regions(surfaces) and their relations. The description is stored as a model of an object. If one view is not enough to build a model of an object, several typical views are shown. In recognition phase, the system makes a description of unknown scenes in the same way as In learning phase.
The system selects among unknown regions those which seam to be most reliable and useful for recognition. A part of a scene consisting of these regions is called a kernel(see Fig. 2). Then a model is selected which includes regions corresponding to the kernel. Once a candidate model is chosen, regions neighboring the kernel are searched for by a model-driven matching process. When this process terminates, the system decides if the candidate model Is really found or not. This process Is repeated until all regions In a scene are processed. Further details of description and matching are described in the following sections.

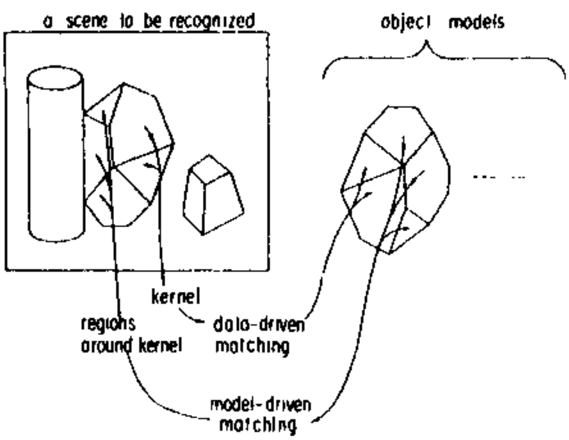

Fig. 2. Matching process.

III DESCRIPTION OF SCENES

Our range finder employs a vertical slit projector and a TV camera to pick up the reflected light. By rotating a projector from the left to the right, many points in a field of view are obtained.

Three-dimensional co-ordinates of the points are calculated by triangulation (Fig. 3(a)). The processing [8] proceeds as follows:

(1) Group the points into small surface elements and assuming each element to be a plane, get the equations of the surface elements(Fig. 3(b)),
(2) Merge the surface elements together into regions(elementary regions, Fig. 3(C)).
(3) Classify the elementary regions into planar and curved ones(Fig. 3(d)).
(4) Try to extend the curved regions by merging adjacent curved regions to produce larger regions(global regions) and fit the quadratic surfaces to them(Fig. 3(e)).
(5) Describe the scene in terms of properties of regions and relations between regions(Fig. 3(f)).

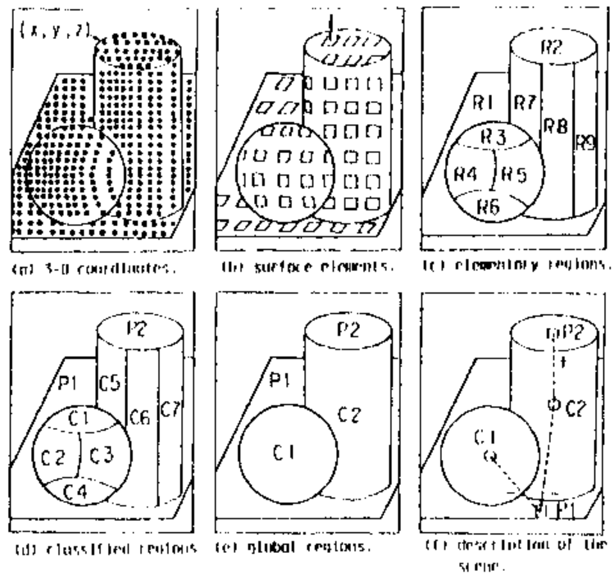

Fig. 3. Conceptual scheme of the scene description process.

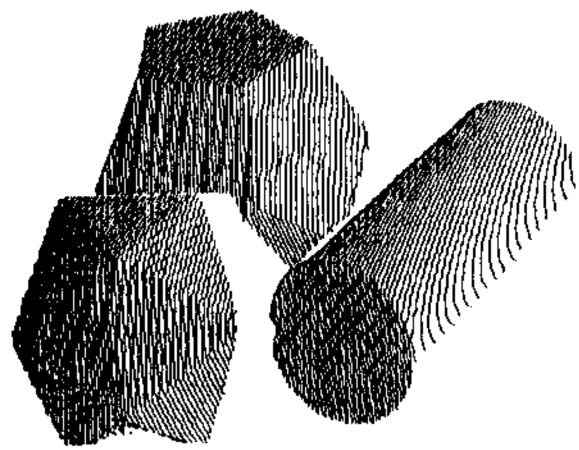

Fig. 4. Slit image of a scene.

Figure A shows an example of a slit Image of a scene to be processed, which includes a dodecahedron, an icosahedron, and a cylinder. In Fig. 5 elementary regions of the scene are shown.

Fig. 5. Elementary regions for the description.

The properties of a region consist of:

(a) type of a surface fitted to the region(planar or curved, and type of quadratic curve for curved region)

(b) equation of the surface in a 3-D space,

(c) 2-D properties of the region(area, perimeter, compactness (4 π area/perimeter^1), mean radius, standard deviation of radii, minimum radius, and maximum radius, etc.)

(d) three-dimensional centroid of the region

(e) number of adjacent regions.

The relations between regions consist of:

(a) adjacency

(b) type of intersectlon(convex or concave)

(c) angle between regions(For non-planar regions, planes are fitted to them.)

(d) relative positions of the centroids.

The properties and relations are used in matching process described in the following section.

III MATCHING

A. Selecting a kernel and assuming the most probable model

The system firstly selects the most promising region. The criterion is based on the type of the region, the area of the region, and the number of adjacent regions In the following manner. We consider that planar regions are more reliable and useful than curved ones because the relative directions of planar surfaces of an object are consistent under rotation of partial occlusion. Regions with larger area are also better because properties of larger regions are less sensitive to noise. Regions with many neighbors are also more useful because many relations can be used in matching. For each region S_t, the following function Is calculated. The region S_4 which maximizes the function is selected as a part of the kernel:

$$f_1(S_i) = \lambda_1 U(S_i) + \lambda_2 A(S_i) + \lambda_3 N_a(S_i) \qquad (1)$$

where

$$U(S_i) = \begin{cases} 1 & \text{if } S_i \text{ is curved} \\ 2 & \text{if } S_i \text{ is planar} \end{cases}$$

$A(S_i)$ and $N_a(S_i)$ are the area of a region S_i and the number of adjacent regions respectively. λ, ,λ_2 and λ_2 are weights.

The model to be selected must have a region that corresponds to S*. Suppose the region S4 has no neighboring regions. Then the kernel contains only one region. Let My denote i-th region of J-th model. The dissimilarity between S_i and M_i is evaluated by the differences of their properties defined in section 3. Let a denote I-th property of a region and D denote an operator to evaluate the difference between property values. The system calculates the following function:

$$f_2(S_k, M_{ij}) = \sum_l \mu_l D_l(p_l(S_k), p_l(M_{ij})). \qquad (2)$$

where μ_l denotes a weight.

The system selects a model which minimizes f_2 if the value Is small enough. If the value is not small, the assumption falls.

If the region S_1 has neighboring regions, the evaluation function should contain the relations between regions. Among neighboring regions a region S(is picked out which maximizes Eq.(I). Now the kernel consists of S_A and S_A. Let M denote a neighbor of M_i, , and q_m denote m-th relation between regions. (the relations are defined In section 3.) The system calculates the following function:

$$\begin{aligned} & f_2'(S_k, S_k', M_{ij}, M_{ij}') \\ &= (f_2(S_k, M_{ij}) + f_2(S_k', M_{ij}')) \\ &\quad + \sum_m \nu_m D_m(q_m(S_k, S_k'), q_m(M_{ij}, M_{ij}')). \end{aligned} \qquad (3)$$

where ν_m denotes a weight.

The system selects a model which minimizes f_2 if the value is small enough.

B. Verification of assumption

The system verifies an assumed model by matching the regions around the kernel to those of the model. Since some regions of the model may not be seen in the scene, regions in the scene are picked up one by one, and are matched to those of the model object.

Starting from a kernel, the system tries to establish a correspondence between scene regions and model regions and to find all the regions of the object in the scene. At each step, the system selects a new region S_n among those which are adjacent to known regions (that is, the regions which have been selected and matched in the earlier steps). Again,Eq.(l) is used for the selection of a new region.

Whenever S_n is selected from the scene, the corresponding region is searched for in the model. The search is based on the dissimilarity function which evaluates dissimilarity of properties and that of relations between S_n and supposed model region M_{ui}. (see Fig. 6).

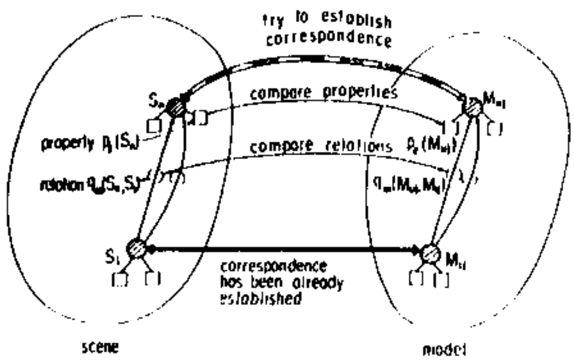

Fig. 6. Schema of comparison.

The function compares properties of the region $p_l(S_n)$ with that of the region in the model $P_p(M_{ui}*)$^ The function also compares relations q $(S_nS_i;)$ between S_n and the known regions S_i of the same object with those $q_m(M_{ui}.M_i)$ of the corresponding regions M_{ui} and M_i. In the model. Suppose the number of already known regions is K. The dissimilarity of region S_n and M_{ui} is defined as follows:

$$f_3(S_n, M_{u_j}) = \sum_l \beta_l D_l(p_l(S_n), p_l(M_{u_j})) + \sum_m \gamma_m \left(\sum_{i=1}^{k} A(M_{i_j}) D_m(q_m(S_n, S_i), q_m(M_{u_j}, M_{i_j}))\right) / \sum_{i=1}^{k} A(M_{i_j}). \quad (4)$$

where P_1 and T_m denote weights. Note that the areas are used as weights so that more reliable regions contribute more to f_3.

The region with minimum f_3 is chosen. If the value Is small enough, the match Is considered to be acceptable. Generally If a surface of an object is not fully seen, the region may not pass the test and It may remain unknown. However, if a plane Is partially occluded by other surfaces, the type and the direction of the region are not affected by the occlusion. Such a planar region, therefore, is determined to match a model region if the dlssimilarlty of corresponding relative direction is sufficiently small, 3-D position correspondence is good enough, and the area of the region is less than that of the model.

When a matching process for a model terminates, the system checks if the assumption is acceptable or not. If enough portion of the model is found, the system concludes that it has found an object that is identical to the assumed model. The position and orientation of the object are calculated from the correspondence of the regions.

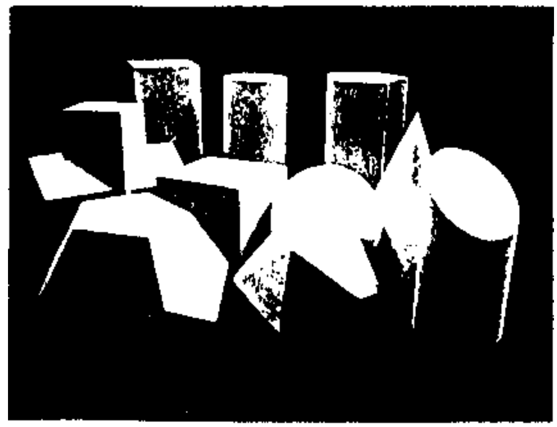

Fig. 7. Model objects.

V EXPERIMENTAL RESULTS

Experiments are made to recognize two kinds of scenes: one with blocks and one with machine parts. In the first experiment, ten kinds of objects with planar and/or quadratic curved surfaces(shown in Fig. 7) are used for object models.
Figure 8 illustrates an example of a description of a model. The result of recognition(which corresponds to Fig. A) is shown in Fig. 9. The first two letters in a region indicate the model object and the letter in the parentheses indicates the corresponding region In the model.

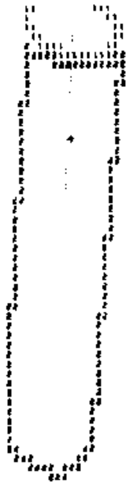

Fig. 8. Example of a description of a model.

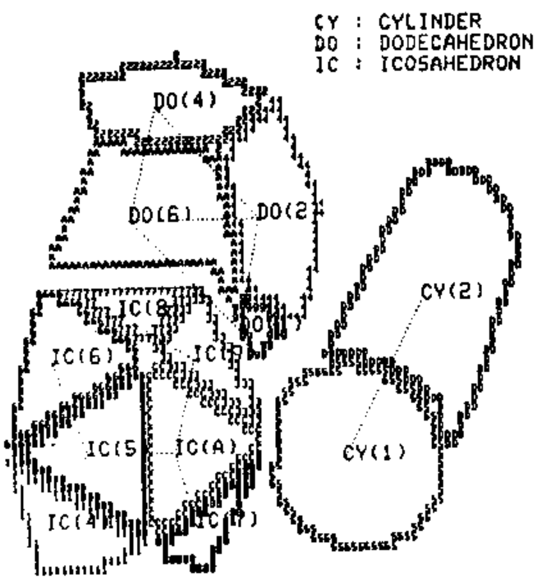

Fig. 9. The result of recognition.

In the second experiment, machine parts(pulley, liner, piston, and conrod of a car) are used for object models. Figure 10 shows an example of an input image. Figure 11 shows elementary regions. In Fig. 12, the result of recognition is shown. Experiments for several similar scenes were so far satisfactory. The processing time for a typical scene such as shown in Fig. 10 is about 3 min. for description and 1 min. for matching.

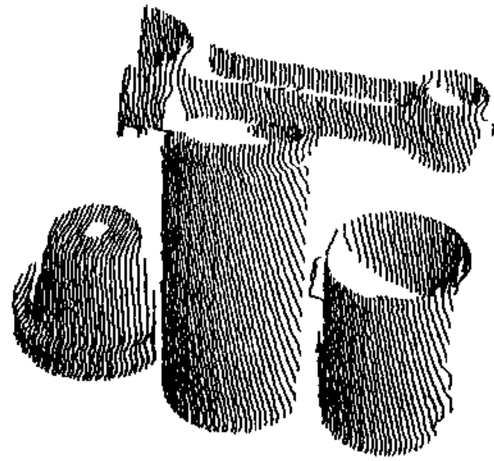

Fig. 10. Slit image of a scene.

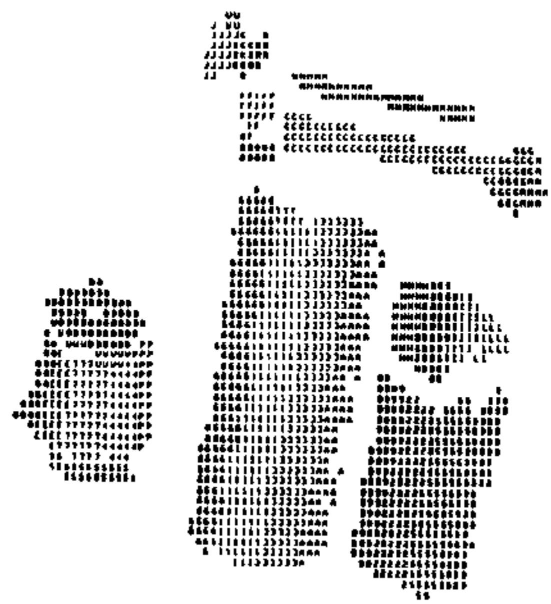

Fig. 11. Elementary regions.

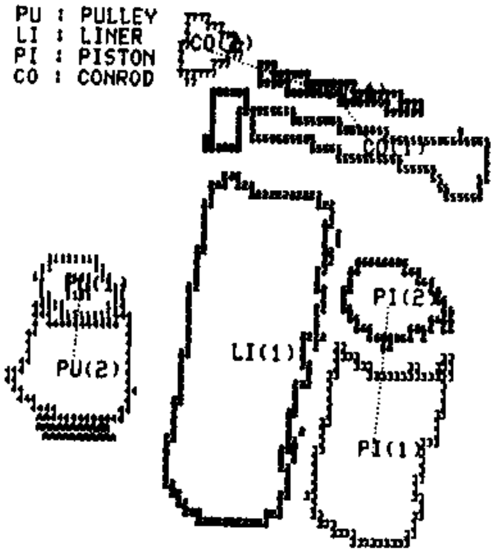

Fig. 12. The result of recognition.

VI CONCLUSION

We have proposed a system to recognize stacked objects using range data. The system describes a scene in terms of planes and smoothly curved surfaces. Models of objects are built in the system by showing them one by one. Objects are recognized by matching the description of an input scene to those of models. The matching program picks up regions which are most reliable and useful for recognition, and matches them to those of a model object. Once a candidate model is chosen, then by a guidance of the model, the rest of the scene regions are searched for. Thus the system has realized flexible and efficient recognition. The result of experiments shows that this scheme is promising.

ACKNOWLEDGEMENTS.

We would like to thank the members of the computer vision section at the Electrotechnical Laboratory for their helpful discussions.

REFERENCES

[I] Barrow, H. and Popplestone, R. "Relational descriptions in picture processing."Machine intelligence 6 (1971) 377-396.
[2] Yachida, M. and Tsuji, S. "A versatile machine vision system for complex industrial parts." Trans. IEEE C-26:9 (1977) 882-894.
[3] Tsuji, S. and Nakamura, A. "Recognition of an object in a stack of industrial parts." In Proc. IJCAI-75. 1975, 811-818.
[4] Shirai, Y. "Recognition of polyhedrons with a range finder." Pattern Recognition 4 (1972) 243-250.
[5] Agin, G. "Representation and description of curved objects." AIM-173 Stanford University, 1972.
[6] Nevatia, R. and Binford, T. 0. "Structured descriptions of complex objects." In Proc. IJCAI-73.. 1973, 641-647.
[7] Popplestone, R. J. et al. "Forming models of plane-and-cylinder faceted bodies from light stripes." In Proc. IJCAI-75. 1975, 664-668.
[8] Oshima, M. and Shirai, Y. "A scene description method using three-dimensional information." Pattern Recognition 11 (1979) 9-17.
[9] Suglhara, K. "Range-data analysis guided by a junction dictionary." Artificial Intelligence 12 (1979) 41-69.
[10] Duda, R. 0. et al. "Use of range and reflectance data to find planar surface regions." Trans. IEEE PAMI-1:3 (1979) 259-271.
(II] Gennery, D. B. "Object detection and measurement using stereo vision." In Proc. IJCAI-79. 1979, 320-327.
[12] Milgram, D. L. and Bjorklund, C. M. "Range image processing: planar surface extraction." in proc IJCPR-80. 1980, 912-919.