OBJECT RECOGNITION USING A TREE-LIKE PROCEDURE GENERATED FROM 3-D MODEL

Yasukazu Okamoto, Yoshinori Kuno, Kazunori Onoguchi,
Mutumi Watanabe and Haruo Asada

Research and Development Center
Toshiba corporation
1, Komukai-Toshiba-cho, Saiwai-ku、Kawasaki 210,Japan

ABSTRACT

This paper presents a vision system which automatically generates an object recognition procedure from a 3-D model, and recognizes the object by executing this procedure. The change in object appearances due to the viewpoint is a main issue in 3-D model vision. In this system, the appearances of an object from various view points are described with visible 2-D figures, such as parallel lines and ellipses. These figures are projections of linear or circular visible elements in the model. Then, the appearance descriptions are compared. Similar figures are extracted and merged into a new description, which represents a common and general appearance of the old ones. This process is iterated and ends in a tree-like procedure.

The system search the procedure tree and recognizes the object. At a visited node of the tree, figure is estimated by the recognition procedure and looked for in the image. If such figures are detected, a child of the current node is selected for the next visit. Detected figures are used for estimation at the next node. The number of object candidates increases as the number of detected figures increases, and propagated candidates construct a tree. The system searches the candidate tree in addition to the procedure tree.

Experimental results show the efficiency of proposed system.

Introduction

Vision is the most important sense for intelligent robots. In practical applications, many of them will be used in manufacturing factories and plants. For such a robot working in an artificial environment, a model-based vision system is a practical solution, because the objects to be recognized are artificial and the models can be easily described. ACRONYM[1] is the most successful model-based vision system. It proved the effectiveness of the model-based approach for a vision system. The problem with ACRONYM, however, was that it took a long time to match features extracted from an image with the 3-D model. Several systems have been proposed to overcome this difficulty. Goad[2] and Lowe[3] proposed a method to generate a procedure from a polyhedron model. Their purpose was to make object recognition efficient by generating a recognition procedure from a 3-D model in advance. The object is recognized by searching for a combination of line segments under the constraint given by the recognition procedure. Grimson[4] used a similar method for depth images.

Though these approaches have improved efficiency, some problems still remains. The features used for recognition are only local ones, such as line segments or plane surfaces. The number of local features increases as the image complexity does. As a result, search space becomes large which prevents efficient recognition.

This paper proposes a new approach to a model-based vision. In this system, the appearance of an object is described with global figures, such as ellipses and parallel lines, instead of local ones. Because the number of global figures is smaller than that of local features, the search space can be kept small.

Approach

The authors' approach is as follows.

1 The relation between the model of an object and its appearance is explicitly expressed. A 3-D model for an object is constructed as indicated in Fig.1. An object model consists of cylindrical primitives. A primitive is a combination of elements whose projections on a 2-D space indicate certain figures. A cylinder primitive consists of two disc elements and a cylinder side element. The projection of a disc element on the 2-D space is an ellipse figure, and the projection of a cylinder side element is a parallel line figure. The appearance of an object from a certain view point is represented by a list of figures.

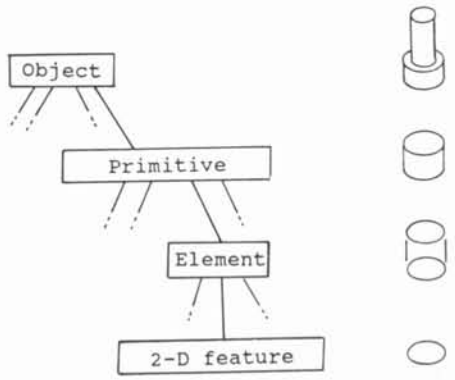


Fig.1 3-D object

2 The efficiency of the recognition procedure which treats a 3-D object is considered. The appearances of a 3-D object changes according to the viewpoints. In the proposed system, appearances are sampled from many viewpoints. The system looks for an object which has any one of the sampled appearances. Because it is tedious to look for each appearance of the object, figures which describe the appearances are looked for in the order of common visibility.

3 For the purpose of ensuring efficiency and flexibility, This system is divided into two stages to ensure efficiency and flexibility, as shown in Fig.2. One is the analysis stage, while the other is the recognition stage. In the analysis stage, an object recognition procedure is generated from the 3-D model of the object. This stage is finished prior to recognition, and the procedure for each object to be recognized can be prepared. In the recognition stage, the object is recognized using only the generated recognition procedure. The system need not match the image and the 3-D model.

To accomplish these approaches, the object recognition procedure is generated by the analysis of the appearance descriptions. Commonly visible figures are extracted from the appearance description from various viewpoints. The recognition procedure generation is accomplished in the opposite direction from that of using it.

Analysis stage

Appearance descriptions from various viewpoints are compared with each other, and nodes of the recognition procedure which represent commonly visible part of the appearances are generated from them. The generated nodes and the original appearance descriptions are connected with an appropriate parent-child relation to construct a tree structure. The comparison is repeated for the generated nodes, and new nodes are generated from previously generated ones.

Appearance description

Appearances are sampled and described from 60 viewpoints. These viewpoints are placed at the vertices of a polyhedron, like a soccer ball, and the object is put at the center of the polyhedron.

Figures which describe the appearance have 3 attributes as follows:

1) Size

A figure has size attributes for each kind of it. The size of an ellipse is represented by its major axis length and minor axis length. The size of parallel lines is represented by the length of the lines and the width between the lines.

2) Positional relation

The positional relation of figures is represented by the relation that of its center on a certain coordinate system. The direction of a figure is determined by its kind, the major axis for an ellipse and either line for parallel lines show the figure direction. The coordinate system is determined by the position and direction of the first figure in the list. If the figure has a direction ambiguity caused by its symmetry, the direction of the coordinate system is determined uniquely by the position of the second figure on the list.

3) Visibility

There are two kinds of values which determine the visibility of the figure: perspective and occlusion. Perspective is the ratio of the figure's size to the element's size. Occlusion is the ratio of the occluded part to the whole of the figure.

The viewpoints for the appearance description are sampled to represent some viewing direction area. The figure attributes change a little when seen from various viewpoints in the area. The attributes are described with a range to represent the whole area.

There may be figures which are invisible from a certain view point because of self-occlusion and perspective. The appearance of the object is described with figures which have larger visibility than the threshold value.

Recognition procedure node generation.

A recognition procedure node is generated from the similar figure list in the appearance descriptions. Two figure lists are judged to be similar when all figures, which are at the same position in the figure list, have similar attributes. One-dimensional attributes are judged to be similar when their ranges overlap. Two-dimensional attributes are judged to be similar when their region overlap. As shown before, there is ambiguity about the direction of a figure. The direction attributes are judged to be similar when any range which the direction attributes represents, overlap.

A figure is generated from similar figures, and a new figure list is generated from similar figure lists. For one-dimensional attributes, the generated figure range is determined by uniting the original figure ranges. For two-dimensional attributes, the region for the generated figure is determined by the circumcircle of the original figure regions. So, the attributes for the generated figure must contain the original attributes. The new node is made from generated figure list description.

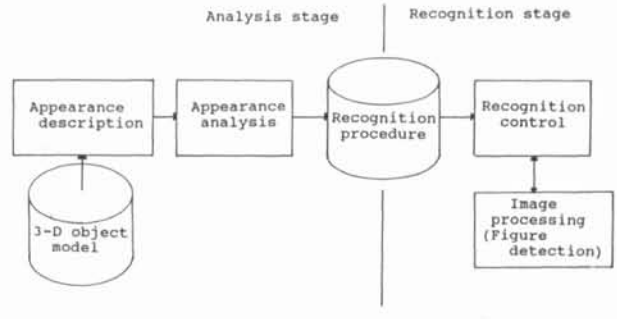


Fig.2 System overview

Recognition procedure

Figure.3 shows the result of appearance analysis. In this figure, the nodes in the extreme right end actually show the original appearances from each viewpoint. The other nodes show the recognition procedure nodes. The relations between the nodes and appearances are shown by arcs. A node is represented by two circles. The right-hand circle shows a rough sketch of the figures described at the node. The left-hand circle shows the ratio of the number of viewpoints for which the figures are visible.

In the recognition stage, the recognition procedure is searched for from the left side to the right side of Fig.3, and a figure is looked for at each node.

In the figures at each node, one of them shows the new figure looked for and the others show the figures detected at the ancestor nodes. The latter are used for estimating the former. At each node, the order used to select a child node is determined. The number of viewpoints for which the figures are visible, as well as the sizes and kinds of the figures that are looked for at the node, are used to determine the order.

In Fig.3, a long parallel line is looked for at first. If it is detected, a big ellipse beside the parallel line is looked for. If an ellipse is not detected, a short parallel line beside the long parallel line is looked for.

Strictly speaking, the recognition procedure is not a tree structure. It is permitted that a node has more than one parents. However, the recognition procedure may be treated as a tree for searching. There can be several search paths to reach a node. Even if the system fails to detect a figure at a certain node, other path may be left to reach the descendants of a failed node. This structure increases the possibility to recognize an occluded object.

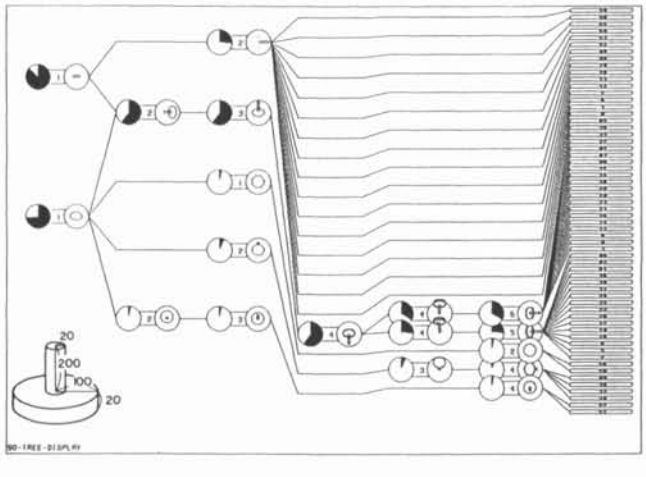


Fig.3 Appearance analysis

Recognition stage

A block diagram of the recognition stage is shown in Fig.4. First, an input image is processed and edges are extracted. Then, line segments are taken from the edges. The recognition procedure is carried out for these line segments.

Edge extraction.

An edge image is extracted by the algorithm as shown in Fig.5. First, the input image is smoothed with a 3X3 filter. Next, 3X3 edge operators for 4 directions are applied to the image. Then, the edge images for 4 directions are thresholded and thinned, respectively. Finally, the 4 binary images are added and small edges are eliminated.

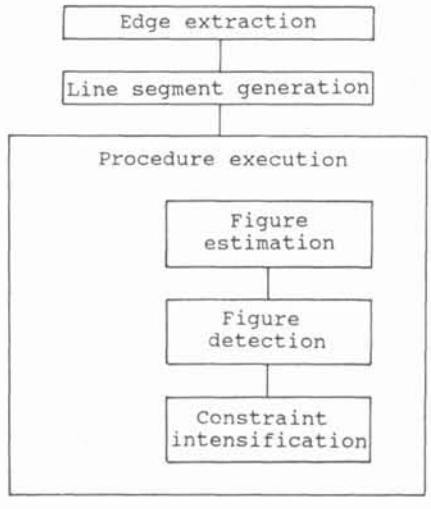


Fig.4 Recognition stage

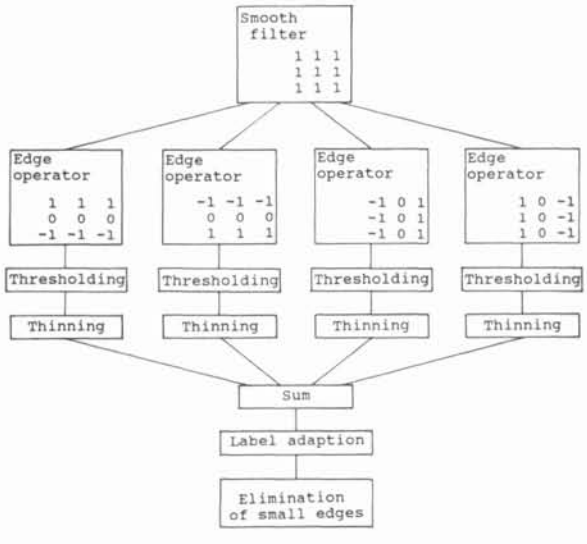


Fig.5 Edge extraction

Line segmentation

Line segments are extracted by vectorizing the edge image. Each edge is split into segments with a curvature. A line segment has information regarding connected segments and break points. A straight line edge is described with one line segment. A curved edge is approximated by the connected line segments.

Execution of recognition procedure

The system starts with the root node of the procedure tree and visit a node according to a search algorithm. At each node, the figure that matches the description in the node is looked for from the line segments. If a figure is detected, then one of the child nodes is selected for the next visit.

Search algorithm.

There are two kinds of trees to search for: One is the recognition procedure tree that is generated in the analysis stage. The other is a candidate tree that is dynamically generated during recognition.

A candidate for the object is a list of figures detected at the nodes from the root node to the current node. A figure is looked for at each node, but more than one figure may be detected in the estimated region, and a candidate increased to the the number of detected figures. The number of candidates increases as the system visit a descendant node and figures are detected. The estimated region for the next figure becomes different, because the position and direction of the detected figures are different. The system must search for a candidate independently from the others.

All the candidates are scored and ordered by their similarity with the figure description in the node. The estimated region is determined to include all the variations of the figure attributes for certain viewpoints. Therefore, the figures detected in an estimated region are equally possible. However, the score of the whole appearance description would be different from the sum of each figure's score. The candidate score value is a weighted sum of each figure's score plus the standard deviations of the attributes. The score for each figure is determined from such reliabilities of the line segments and the differences between the attributes for a detected one and an estimated one.

The number of candidates increases explosively as the detected figures increase. Good candidates are selected by searching the candidate tree, and the procedure tree is searched for each selected candidate. In the candidate tree, the increase in the candidates is shown as branches from the parent candidate to the child candidates. Actually, these trees are duplicated, as shown in Fig.6. A circle represents a procedure node and a square represents a candidate. The candidate tree branches in a vertical direction, and each descendant procedure tree that each candidate searches is a copy of the original nodes.

For searching this duplicated tree, the depth-first-search algorithm was chosen for the procedure tree, and the beam-search algorithm was chosen for the candidate tree. The system searches this duplicated tree as follows. First, the system selects a fixed number of candidates according to the scores. The number is defined as the beam width beforehand. Then, the system selects a procedure node to be executed for each candidate. The search on the procedure tree for each candidate is carried out in a depth-first manner. An estimated region is determined from the appearance description of the current node, and a figure is looked for in this estimated region. If some figures are detected in the estimated region, a child of the current node is selected for the next execution. If no figure is detected, the search backtracks to select another node.

If the position and the direction of the object can be calculated by comparing the detected figures with the 3-D model, no further search in the recognition procedure is necessary. When sufficient figures to calculate the position and the direction of the object are detected, the recognition procedure is stopped. If there is an ambiguity in the detected figures, further search using the recognition procedure may be used for verification. When to stop the execution is given in the recognition procedure.

The search in the recognition procedure will be described, assuming that some figures have been detected. The initial conditions for starting the search on the procedure tree will be presented later.

Process at the node

In each node, figures are described by their positional relations. Some of them correspond to figures that were detected from the image in the search path to reach the node. The remaining one corresponds to the figure to look for at the node. An estimation is made concerning the region where the figure to look for exists. This region is determined by the positional relations of the figure descriptions, the position of the detected figures, and the approximate distance from the camera to the object. Not only the position of the figure, but the attribute ranges, such as size and direction, are estimated. The distance constraint will be described later.

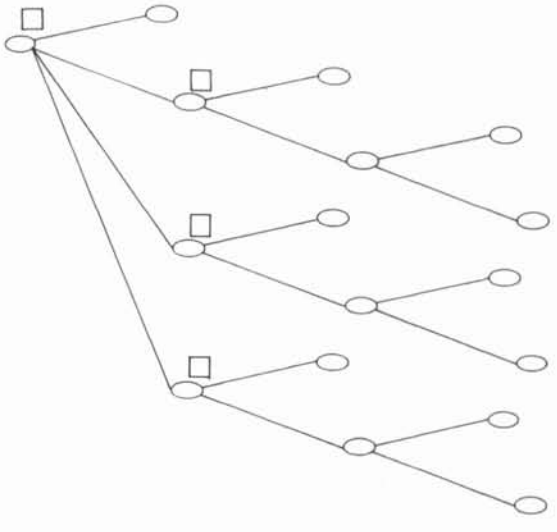


Fig.6 Duplicated tree

The figure that matches the description is looked for in the estimated region. The system uses a top-down hypothesis verification method, for detecting a figure. First, long line segments in the estimated region are tested and hypothetical figures that match the description are detected by them. Next, the hypothetical figures are verified by testing how much they match the other line segments, and figures are detected. Using this hypothesis verification method, it is possible to detect partially hidden figures.

At the start of the procedure, a rough distance from the camera to the object is given. This constraint is narrowed by newly detected figures. Because the size of the figure taken by the camera varies according to the distance between the object and the camera, the distance can be determined by the size of the figure. In a figure, there are several attributes invariant for viewpoint changes and vary only by the distance changes. For example, in an ellipse, which is a projection of a disc, the major axis is invariant for a viewpoint change. In the case of parallel line, which are projections of cylinder sides, the parallel line width is invariant for viewpoint changes. From these attributes regarding a detected figure, the distance constraint is narrowed and used to determine the next estimated area.

Experimental results

Experiments were carried out to confirm the efficiency of the proposal system. Three criteria were adopted to evaluate the efficiency. They were speed, robustness for the complexity of a scene, and robustness for change in the viewpoint. A valve was selected as a target object as an example of a complex 3-D object.

The input image is shown in Fig.7(a). This image was taken by a monochrome CCD camera under unstructured lightings. The distance from the camera to the valve was about 1.8 meters. Figure.7(b) shows the edge image. Then, line segments were generated from the edge image. The square is the estimated region and ellipses in the square are the detected figures. Six was used for the beam width to select the candidates. Figure.7(c) shows the candidates detected at the start of recognition. As there were no detected figures at this time, the estimated region can't be limited, and it involved the whole scene. Figure.7(d) shows the candidates detected after 3 figures were detected. The estimated area was narrowed by the detected figures. Figure.7(e) shows the candidate which was recognized as the valve, that is, the one which had the highest score. The recognized figures for the recognized candidate were matched with the 3-D model, and the position and attitude for the object was calculated. In Fig.7(f), the model is overlaid on the input image. Figure.7(g) shows the recognized object in another attitude.

Discussion

Speed

It took about 2 minutes by using a 16.7 MHZ 68020 CPU to recognize the object in Fig.7(f). This was still a long time for practical use in robots.

Robustness for complexity

The valve had curved surfaces and a metal gloss. Under an unstructured lighting condition, the edges were lacking, distorted and broken into small pieces, as shown in Fig.7(b). Moreover, the side flange was partly occluded by the object itself. However, the object could be recognized because the figures were detected in a top-down method.

Robustness for viewpoint changes

As shown in Fig.7(f) and Fig.7(g), an object in different attitudes could be recognized with a single procedure. Though their attitudes did not accurately match that for the 60 sampled appearances, they could be recognized by matching the detected figures with the 3-D model. However, if the object were in a special attitude, such as seen from just beside, the system may fail to recognize the object. In such an attitude, greatly distorted figures will be eliminated even if the object were big. For example, flat ellipses can't be treated as lines in the current system. To overcome this difficulty, the system must change the kind for the distorted figure and use it.

Conclusion

A new approach to model-based vision has been proposed in this paper. An object recognition procedure is generated in advance of the object being recognized by executing this procedure. The experimental results showed that this system can be used to recognize a 3-D object located in a complex scene.

ACKNOWLEDGMENT

The authors are indebted to F.Umibe for reviewing and revising the original English manuscript. This research and development was conducted under contract with the Agency of Industrial Science and Technology, the Ministry of International Trade and Industry, on the Large-Scale Project, "Advanced Robot Technology ".

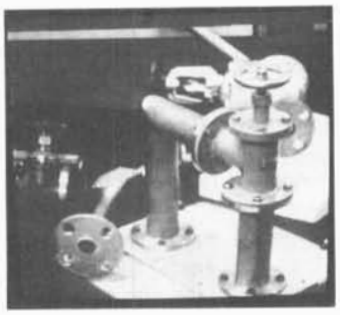
REFERENCES

[1]R.A.Brooks, "Model-based Three-dimensional Interpretations of Two-dimensional Images," IEEE,Trans.PAMI,vol.5-2,pp.140-150,1983.

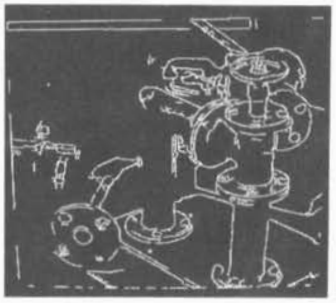
[2]C.Goad,"Special Purpose Automatic Programming for 3D Model-based Vision," Proc.Image understanding workshop,pp.94-104,June,1983.

[3]D.G.Lowe,"The Viewpoint Consistency Constraint," IJCV,vol. 1-1,pp.57-72,1987.

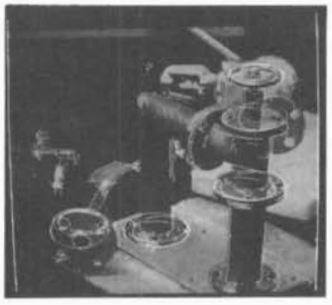
[4]W.E.L.Grimson and T.Lozano-Perez,"Localizing Overlapping Parts by Searching the Interpretation Tree," IEEE,Trans.PAMI,vol. 9-4,pp.469-482,1987.



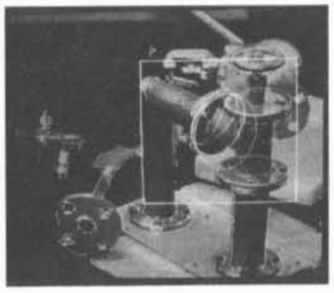
(a)



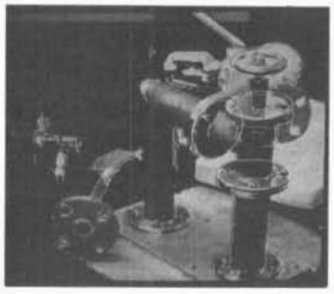
(b)



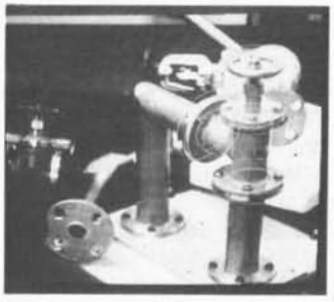
(c)



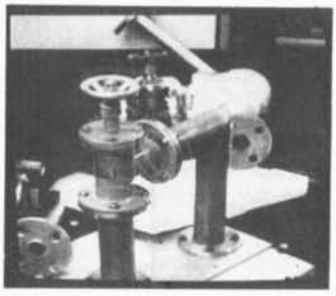
(d)



(e)



(f)



(g)

Fig.7 Experimental results
(a) Input image
(b) Edge image
(c) Candidates at first
(d) Candidates with 4 figures
(e) Recognized figures
(f) Model overlay 1
(g) Model overlay 2