OBJECT RECOGNITION USING TACTILE IMAGE ARRAY SENSORS

Ren C. Luo

Dept. of Elec. and Comp. Eng.
North Carolina State University
Raleigh, N.C. 27695-7911

Wen-Hsiang Tsai

Dept. of Inform. Science
National Chiao Tung University
Hsinchu, Taiwan, R.O.C.

Abstract

The objective of this paper is to develop an object recognition system through the combination of 2-D tactile image array and visual sensors. A video camera is used to acquire a top view image of an object and two tactile sensing arrays mounted on a gripper are used to detect the tactile information about the lateral surfaces of the object. 3-D reference object models are established as a decision tree, and recognition of unknown objects is accomplished through measuring and comparing input object features hierarchically with these of the reference objects associated with the decision tree.

The clustering process and recognition procedures are described. The recognition scheme has been implemented. The resulting decision tree is also presented.

I. INTRODUCTION

Conventionally, object recognition usually is performed using visual information. With the advent of tactile array sensors [1-3] also useful for object shape measurement, we propose in this paper a new approach to object recognition which combines the use of visual and tactile information. Non-visual information of object shapes is obtained from a video camera mounted right above the work platform, and tactile information of lateral object shapes is measured by two array sensors mounted onto the robot grippers. Both types of 2-D shape information are utilized for recognition. The recognition scheme is defined in a hierarchical manner, so that an object is recognized first with the 2-D visual information along, followed by, if necessary, the 2-D tactile information measured when the object is grasped by the robot gripper.

The proposed approach will use moment invariants [8,9] of object silhouette shapes as the features for object recognition. The reason for this is twofold. First, the lateral object shapes measured with the tactile array sensor are in low resolution because of the small array size available. Also, the shapes result from touching or taction of object surfaces on array sensor elements, so the number of points included in a shape ranges from one point (e.g., when the object is a sphere), a line (e.g., when the object is a cylinder), a surface patch (e.g., when the object is a polyhedron), to possibly a combination of the former three cases. The boundaries of such shapes, especially of the first two (a point and a line), are not meaningful enough for most boundary-based shape descriptions (such as Fourier descriptors, chain codes, syntactic string representations, etc.) to be applicable here. Instead it is better to base the shape analysis on shape regions, and moment invariants are appropriate for this purpose. Next, since robot manipulation on the objects is necessary, it is often required to find out the position and the orientation of a given object so that proper grasp of the object with the robot gripper can be accomplished. For this, moments turn out to be the best choice. In particular, object centroids and principle axes, which defines object positions and orientations, can be easily derived as functions of low-order moments. The hierarchical recognition scheme is based on a decision tree [10] which can be constructed automatically in the learning phase from a set of given objects. Object shape ambiguity is resolved further as more tree levels are expanded until all shapes are discriminated or until further resolution is impossible. In the recognition phase, the decision tree is traversed when input objects features are compared with reference object features until a decisive tree node is reached or unitl the input object is determined indiscriminable in its current stable state. For the latter case, the gripper is operated to pick up and rotate the object so that a new object state can be obtained. Another phase of object recognition is then started again.

II. SYSTEM CONFIGURATION AND TACTILE INFORMATION MEASUREMENT

A. System Configuration

The system we use for object recognition and manipulation is shown in Fig. 1. Both the TV camera and the array sensors are controlled by a microcomputer. The camera is mounted right on top of a work platform on which objects are laid. It is assumed that the camera is far enough from the platform so the perspective effects on object images can be reduced to a minimum. The camera optical axis (going through the camera lens center) is made perpendicular to the platform plane.

The gripper we use includes two square-shaped 16x16 array sensors. The elements are attached close enough to the array edges so that slant contact of an object surface with any array edge can also be detected. The tactile information measured by touching an identical object from a fixed lateral direction (fixed with respect to the principal axis of inertia of the object, as will be discussed) will always be identical or stable.

CH2282-2/86/0000/1248$01.00 © 1986 IEEE

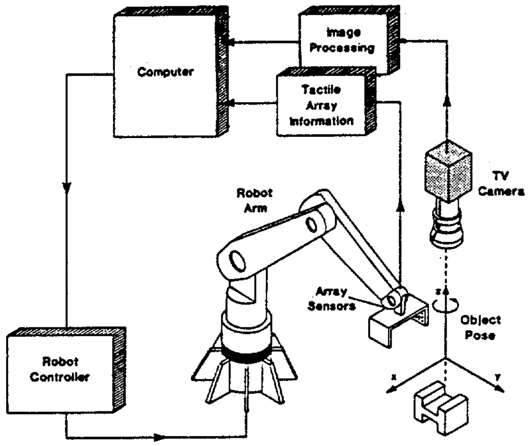

Fig. 1. System Configuration for Object Recognition and Manipulation

Before the tactile information measurement process can be described, we define some notations to facilitate geometric descriptions of the system configuration. Let A_R denote the right-hand side array sensor as viewed from the gripper wrist, and A_L denote the left-hand side one. In some cases, A_R and A_L will also be used to specify the planes going through the sensing-element surfaces. Each array, after touching any object surface, will provide an array image of tactile information. Let I_R denote the image provided by A_R and I_L provided by A_L. Let O_R be the origin of the image coordinate system for I_R. O_R is chosen to be the center of A_R. The origin O_L for I_L is similarly chosen. The line going through O_R and O_L will be called the gripper lateral axis and denoted by L_L. Also, when A_R and A_L are opened apart to the maximum, the middle point of the line segment joing O_R and O_L will be called the gripper center and denoted by G. The plane

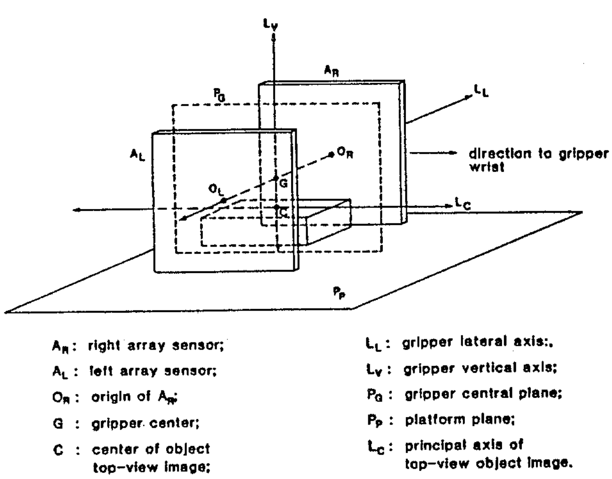

Fig. 2 Spatial relation among various system planes, axes, and centers. "+" denotes positive directions of axes L_L and L_C. The Object is being measured for its tactile information from direction 90°.

defined by the surface of the platform will be denoted by P_P. And the line going through G and perpendicular to P_P will be called the gripper vertical axis and denoted by L_V. The direction defined by the vector from G to O_R will be called the positive direction of L_L. Another useful structure is the plane going through G and L_V, and perpendicular to L_L (and so parallel to A_R and A_L), which will be called the gripper central plane denoted by P_G. Fig. 2 shows the spatial relation among all the geometric structures defined above.

III. Overview on Object Learning and Recognition Schemes

A. Object Shape Learning

Accordingly, a block diagram showing the major steps of the learning phase is included in Fig. 3 which needs some further explanation. After all the reference objects are discriminated according to their top-view silhouette boundaries, there may still exist some groups of objects, each containing several objects which are still visually indiscriminable. Then, for each such group, the gripper is operated to measure tactile information for further discrimination. This precedes by first selecting a set of preselected lateral directions and then measuring a pair of tactile images for each of these directions. The lateral direction most effective for discriminating each group of visually-ambiguous objects is selected. Stop if all groups of ambiguous objects are discriminable; repeat the process by selecting another most effective lateral direction from the remaining directions for each subgroup of still ambiguous objects until all lateral directions are tried. The above learning procedure also involves selection of shape features from visual and tactile object images for object discrimination. The result of learning will be a hierarchical decision tree with each tree node represented by a group of ambiguous or indiscri-minable objects and each tree link associated with a most effective lateral direction θ. The emphasis here is that the whole learning procedure can be made fully automatic.

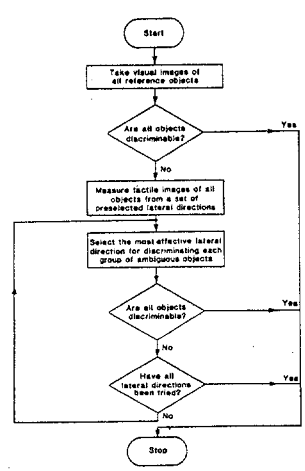

Fig. 3 Flowchart of learning procedure

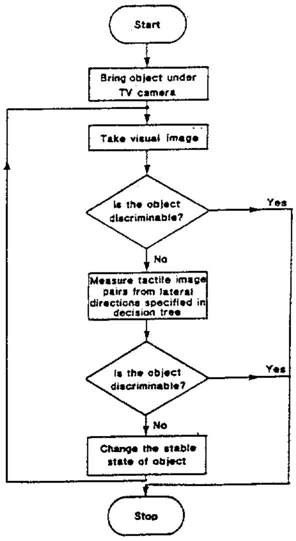

Fig. 4 Flowchart of recognition procedure

B. Object Shape Recognition

As shown in Fig. 4, the recognition procedure begins with taking the top-view visual image of a given unknown object after it is brought right under the TV camera on the platform. The object is then discriminated according to features extracted from the visual image. If the object is not discriminable with its visual image alone, a pair of tactile images are then measured from lateral direction specified in the decision tree. After object features are extracted, the object is discriminated further. This step may be repeated more than once if the number of tree levels is more than two. Most objects can be recognized after this step.

But as mentioned previously, there still exist objects which are indiscriminable if they are in certain stable states on the platform. One way to solve the problem is to change the stable state of the input object so that the originally "invisible" and "untouchable" object portion can become "visible" and "touchable."

IV. TACTILE AND VISUAL FEATURE SELECTION

A. Shape Features for Visual Recognition

Since the TV camera can take images with a rather high resolution and since object silhouette boundaries reveal, in most cases, enough shape information for object recognition, the point set used for defining moments is chosen to include just the shape boundary instead of all silhouette points. This set will be denoted as B_V. This also makes the features, to be defined next, more informative about minute details of the shape boundaries. Furthermore, moment computation can speed up significantly because much less points are involved. All features for recognizing visual images will be denoted as v_i.

The first feature v_1 we use is $m_{00} = \bar{m}_{00}$ which is the area of B_V. Since B_V is the silhouette boundary, v_1 actually is the perimeter of the boundary [11]. another useful shape feature is shape elongateness or eccentricity. An eccentricity measure in terms of central moments is described in [12] as:

$$e = [\bar{m}_{02} - \bar{m}_{20})^2 + 4\bar{m}_{11}^2] / \bar{m}_{00}$$

The third feature V_3 we use is the normalized moment of inertia around the centroid defined as follows:

$$V_3 = \sum_{i=1}^{N} [(x-\bar{x})^2 + (y-\bar{y})^2] / \bar{m}_{00}$$

$$= (\bar{m}_{20} + \bar{m}_{02}) / \bar{m}_{00}.$$

In the following sections, we use V to denote the feature vector composed of the three features V_1 through V_3 described above, i. e., $V = [v_1, v_2, v_3]$. V will be called the visual feature vector.

B. Shape Features for Tactile Recognition

Since tactile images I_L and I_R are measured with a much lower resolution, all non-zero points in I_L and I_R (resulting from contact of array sensing elements on the surfaces) will be used to compute the moment values. In the following, the non-zero points in I_L and I_R will be denoted as B_L and B_R, respectively. Features extracted from B_L and B_R will be denoted as t_i^L and t_j^R respectively, and called tactile features. Also, the moments m_{pq} (or $\bar{m}_{pq}$) computed from B_R and B_L will be superscripted as m_{pq}^R and m_{pq}^L (or $\bar{m}_{pq}^R$ and $\bar{m}_{pq}^L$), respectively.

Again, the areas of B_L and B_R can be used as tactile features.

So we select t_1^L as $m_{00}^L = \bar{m}_{00}^L$ and t_1^R as $m_{00}^R = \bar{m}_{00}^R$. Next, since B_L and B_R are mesured with array sensors A_L and A_R fixed spatially in position with respect to the platform plane P_P. the locations of the centroids and the directions of the principle axes of B_L and B_R reveal a certain amount of 3-D structural information about the object. Therefore, we choose t_2^L and t_3^L to be $\bar{x}_L$ and $\bar{y}_L$ t_2^R and t_3^R to be $\bar{x}_R$ and $\bar{y}_R$, respectively, and t_4^L

and t_4^R to be:

$$t_4^L = 1/2 \tan^{-1} \left(\frac{2\bar{m}_{11}^L}{\bar{m}_{20}^L - \bar{m}_{02}^L} \right),$$

$$t_4^R = 1/2 \tan^{-1} \left(\frac{2\bar{m}_{11}^R}{\bar{m}_{20}^R - \bar{m}_{02}^R} \right).$$

Finally, t_5^L and t_5^R are selected similarly to v_2 as follows:

$$t_5^L = (m_{02}^L - m_{20}^L)^2 + 4(m_{11}^L)^2 / m_{00},$$

$$t_5^R = (m_{02}^R - m_{20}^R)^2 + m_{11}^R)^2 / m_{00}.$$

The two vectors $T^L = [t_1^L, t_2^L, t_3^L, t_4^L, t_5^L]'$ and $T^R = [t_1^R, t_2^R, t_3^R, t_4^R, t_5^R]'$ together will be called the tactile feature vector and will be collected vector $T = [T^L, T^R]'$.

V. REFERENCE OBJECT LEARNING

A. Decision Tree Construction Overview

Let the set of reference objects be denoted by R_1, R_2, ..., R_m. For each object R_i, there may exist more than one stable state. Each stable state actually should be, from the view point of object recognition, regarded as distinct 3-D object shape. In the following the 3-D shape of the jth state of the ith object R_i will be denoted as S_{ij}. Once an unknown input shape is recognized to be a particular S_{ij}, we can conclude that it is just the jth stable state of object R_i. Let S denote the set of all the 3-D shapes, i.e., $S=\{S_{ij} \mid i = 1, 2, ..., m, \ j = 1, 2, ..., n_i\}$. Then the decision tree construction may be regarded as a consecutive partitioning of S into smaller subsets, with each level of partitioning based on either the visual or the tactile shape features, until S is partitioned into all non-partitionable subsets. Each subset of S is partitioned into smaller subsets according to a clustering procedure. Thus, the decision tree has S as its root and all non-partitionable sets as its leaf nodes.

A non-partitionable set as a leaf node is either a single-element set or a multiple-element set. The former case means that the single element, say S_{ij}, in the set can be uniquely discriminated from other 3-D shapes in S. And the latter case means that any two 3-D shapes in the multiple-element set are indiscriminable from each other using the visual feature vector and the tactile feature vector pair measured from any lateral direction. On the other hand, all non-leaf tree nodes are multiple-element subsets of S. In the following, we use S_1, S_2, ..., S_n to denote the n_1 first-level child nodes of the decision tree, $S_{1.1}$, $S_{1.2}$, ..., $S_{1.n_{1.2}}$ to denote the $n_{1.2}$ second-level child nodes S_1, and so on. In general, $S_{i_1, i_2, ..., i_{m-1}, i_m}$ denote one of the $n_{1.2...m}$ mth-level child nodes of $S_{i_1, i_2, ... i_{m-1}}$.

B. First-Level Partitioning Based on Visual Feature Vector

The first-level partitioning is always based on the visual feature vector. The reason for this is to obtain the most effective first-level partitioning because the visual shape is measured with a much higher resolution than the tactile shape. Note that the faster the set S is decomposed into single-element subsets (i.e., the fewer the tree levels are), the faster the recognition based on the resulting decision tree will be. Let S_i $(i = 1, 2, ..., n_1)$ be a resulting first-level node. Then, S_i includes all those 3-D shapes in S which have a specific similar visual feature vector. this common feature vector for all shapes in S which have a specific similar visual associated with the tree link from S (the tree root) to S_i. V_i will be used in the recognition phase for visual feature matching.

C. Multiple-Level Partitioning Based on Tactile Feature Vectors

The remaining levels of partitioning are based on the tactile feature vector pairs measured from a predetermined set H of lateral directions. H is determined in this study to include angles θ which are multiples of 30°, with $\theta_\circ = 0°$ coincident with the direction of the direction of the principal axis of the top-view shape. Therefore, $H = \{\theta \mid \theta = i \times 30°, i = 0, 1, ..., 5\}$. Note that only θ_i up to 180° needs to be included because each time two lateral sides are measured. Practical choice of H depend on the surface properties of the objects to be recognized. In general, we assume that H include d lateral directions. In the following, we describe how to partition each first-level node S_i $(i = 1, 2, ..., n_1)$ into multiple levels of child nodes. Each level of partitioning will be made most effective according to a certain effectiveness measure defined subsequently. First, from each lateral direction θ_ℓ in H, $\ell = 1$, ..., d, we measure the tactile feature vector pair for each 3-D shape in S_i.

Based on the resulting feature vector pairs, S_i can be partitioned into a set of second-level child nodes denoted as $S^{\ell}_{i.1}$, $S^{\ell}_{i.2}$, ..., $S^{\ell}_{i.n_{1.2}}$. $S^{\ell}_{i.j}$ are different lateral directions θ_ℓ in H set P_i, defined as

$$P^{\ell}_i = \{ S^{\ell}_{i.j} \mid j = 1, 2, \ldots, n^{\ell}_{1.2} \}$$
$$i = 1, 2, \ldots, n_1, \ \ell = 1, 2, \ldots, d,$$

be used to denote such a partitioning. All P^{ℓ}_i will be called the basic partitioning of S_i. Each P^{ℓ}_i is a candidate for the final second-level partitioning. Let $\#P^{\ell}_i$ denote the number of nodes $S^{\ell}_{i.j}$ in P^{ℓ}_i.

Obviously, the larger the number $\#P^{\ell}_i$ is, the more effective the measured tactile features from direction θ_ℓ are for shape discrimination (because the shapes in S_i are separated into more groups). Therefore, we define $\#P^{\ell}_i$ as the effectiveness measure lateral direction θ_ℓ. Then, we can now choose the basic partitioning $P^{\ell_i}_i$ with the most effective lateral direction θ_{ℓ_i} as the desired second-level partitioning of S_i for use in the final decision tree. We add a subscript i to the index ℓ_i of the most effective direction θ_{ℓ_i} for S_i to emphasize the dependency of ℓ_i on S_i. Therefore, for each first-level node S_i in the decision tree, we now have $S^{\ell_i}_{i.1}$, $S^{\ell_i}_{i.2}$, ..., $S^{\ell_i}_{i.n_{1.2}}$ as its child nodes which are in the second level of the tree. Also, the lateral direction θ_{ℓ_i}, and the corresponding tactile feature vector pair, denoted as $T_{i.j}$, will be said to be associated with the tree link from S_i to $S^{\ell_i}_{i.j}$, $j = 1, 2, \ldots, n^{\ell_i}_{1.2}$. They will be used in the recognition phase for tactile information measurement and feature matching. This completes the second-level tree node generation.

To partition any second-level multiple-element node S_{ij} further, we have to select the most effective lateral direction $\theta_{\ell_{i.j}}$ from all those in H except the one θ_{ℓ_i} already used in partitioning S_i. To obtain the partitioning $P^{\ell}_{i.j}$ defined similarly to P^{ℓ}_i as

$$P^{\ell}_{i.j} = \{ S^{\ell}_{i.j.k} \mid k = 1, 2, \ldots, n^{\ell}_{1.2.3} \},$$

for $S_{i.j}$ for a fixed θ_ℓ, the basic partitioning P_i can be utilized to speed up the process. Actually two shapes in $S_{i.j}$ should be separated into two distinct $S^{\ell}_{i.j.k}$ if the two shapes appear in two different $S^{\ell}_{i.j}$ in P^{ℓ}_i (i.e., the two shapes are dissimilar according to their tactile feature vector pairs measured form lateral direction θ_ℓ). Once $P^{\ell}_{i.j}$ for all θ_ℓ (except θ_{ℓ_i}) are obtained, the remaining steps to generate the third-level tree nodes are all the same as those for generating the second-level ones.

Finally the above process for generating the third-level nodes is repeated to generate all the nodes of lower levels, if necessary, until every tree node is found to include just a single shape (or several shapes but all from a single object), or until no more lateral direction from H is available for partitioning any multiple-element tree node. This completes the construction of the decision tree.

VI. EXPERIMENTAL RESULTS

A 16x16 simulated tactile image is used to conduct the experiment at this stage. Visual images are acquired with a TV camera. Fig. 5 shows the sketches of the ten objects chosen as the test sets. they include most angular and curved surfaces encountered in common industrial applications.

Each stable state of an object was considered to be a separate entity by itself. Therefore, the recognition problem consisted of being given an unknown object, taking its visual image and tactile images as required, and then determining what object it is and in which stable state it lies. A cross reference table (Table 1) is included which transposes between the actual object names and stable states and the object names indicated in the decision tree. The top views (visual shapes) of the stable states of some objects do not possess principal axis owing to their rotational symmetry. Such object states are not included in the experiment. Treatment of rotationally symmetric shapes has been studied in another research [13]. The resulting decision tree is shown in Fig. 6.

VII CONCLUSIONS

We have demonstrated the feasibility of a system that utilizes a combination of visual information and taction in the identifcation and discrimination of three dimensional objects.

A recognition scheme has been implemented which succeeds in identifying any of the ten reference objects placed in any of their permissable stable states. Although the decision tree was constructed using these objects in four specified orientations as the "Training Set", the "Test Set" consisted of objects placed in random orientations and this scheme successfully identified all these stable states with satisfactory accuracy.

ACKNOWLEDGEMENT

This work has been supported in part by NSF grant No. DMC 8505166.

REFERENCES

[1] R. C. Luo, F. Wang, Y. Liu, "An Imaging Tactile Sensor with Magnetostrictive Transdustion," Proceedings of the International Conf. of Intelligent Robots and Computer Vision, Cambridge, MA., Nov. 1984.

[2] M. Briot, "The Utilization of an 'Artificial Skin' Sensor for the Indentification of Solid Objects," Proceedings of the 9th International Symposium on Industrial Robots, Washington, D.C. 1979.

[3] P. Dario, R. Bardelli, D. De Rossi, L. R. Wang, P. C. Pinotti, "Touch-Sensitive Polymer Skin Uses Piezoelectric Properties to Recognize Orientation of Objects," Sensor Review, Vol. 2, 1982.

[4] M. Oshima and Y. Shirai, "Object Recognition Using Three-Dimensional Information," IEEE Trans. Patt. Anal. Mach. Intell., Vol. PAMI-5, No. 4, pp. 353-361, July 1983.

[5] Y. Sato and I. Honda, "Pseudodistance Measurement for Recognition of Curved Objects," Ibid. pp. 362-372, July 1983.

[6] R. Nevatia, "Prescription and Recogniton of Curved Objects," Artificial Intelligence, Vol. 8, 1977.

[7] R. Bajcsy, "Three-Dimensional Object Representation," in Patt.'Recog. Theory and Appl. (J. Kittler, K. S. Fu, and L. F. Pau, Eds.) D. Reidel Publishing Co., U. K., 1982.

[8] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," IRE Trans. Inform. Theory, Vol IT-8, pp. 178-187, Feb. 1962.

[9] S. A. Dudani, K. J. Breeding and R. B. McGhee," Aircraft Identifcation by Moment Invariants," IEEE Trans. Computers, Vol C-26, No. 1, pp. 39-45, Jan. 1977.

[10] J. K. Mui and K. S. Fu, "Automated Classification of Nucleated Blood Cells Using a Binary Tree Classifier," IEEE Trans. Patt. Anal. Mach. Intell., Vol. PAMI-2, No. 5. Sept. 1980.

[11] A. Rosenfeld and A. C. Kak, Digital Picture Processing, Academic Press, New York, 1982.

[12] D. H. Ballard and C. M. Brown, Computer Vision, Prentice-hall, New Jersey, 1982.

[13] W. H. Tsai, "Automatic Detection of Inherent Shape Orientation by Moment Functions," Submitted for publication.

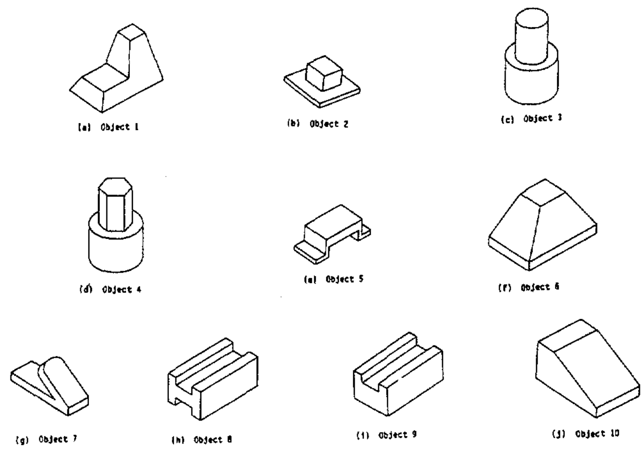

Fig. 5 Sketches of reference objects used in experiments

OBJECT NAME IN TREE	CORRESPONDING OBJECT NUMBER	STABLE STATE
A	7	a
B	8	a
C	8	b
D	8	c
E	9	a
F	9	b
G	9	c
H	2	a
J	6	a
K	6	b
L	6	c
M	10	a
N	10	b
O	10	c
P	10	d
Q	3	a
R	4	a
S	1	a
T	1	b
U	1	c
V	1	d
W	1	e
X	5	a
Y	5	b
Z	5	c
CC	9	d
DD	7	c
EE	7	d
FF	7	e

Table 1 Cross reference table for object names and stable states.

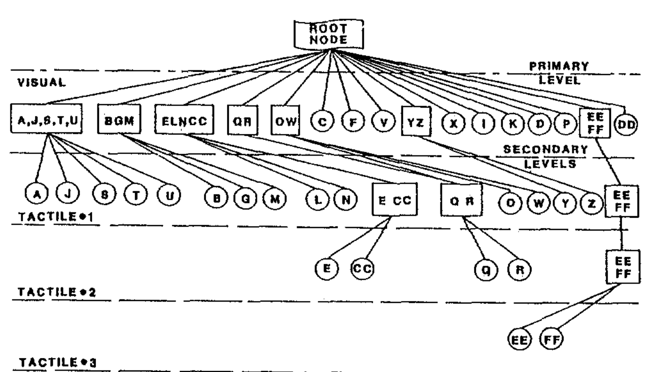

Fig. 6 Decision tree constructed in the experiment