A KNOWLEDGE-BASED OBJECT RECOGNITION SYSTEM FOR APPLICATIONS IN THE SPACE STATION

NAG 9-203
IN-35-CR
120031
25P.

P.I. Dr. Atam P. Dhawan
Assistant Professor, Dept. of electrical Engineering
University of Houston, Houston, TX 77004

<u>Final report on the project funded by NASA/JSC from Feb. 1, 1987 to Jan. 31, 1988</u>

A knowledge-based three-dimensional (3D) object recognition system is being developed at the University of Houston. The system uses primitive-based hierarchical relational and strucutral matching for the recognition of 3D objects in the two-dimensional (2D) image for interpretation of the 3D scene. The system under developement has several expert systems working in both stand-alone and cooperative modes. They are responsible for multi-level processing and analysis of the acquired information at their respective levels. The modules for mult-level processing have been designed and implemented. They have also been tested on simple images of low-complexity on individual basis. The overall system has been designed to work in a blackboard-oriented fashion in order to provide integrated multi-level processing. The complete integration of the system has not been completed. The funding support has been asked for continuation of the project to complete the integration and evaluation of the system.

The complete 3D object recognition process in the system has six major steps: (1) the entry-level pre-processing to enhance features and remove noise in the input image data; (2) the low-level preliminary segmentation and initial feature detection followed by the rule-based expert segmentation to yield suboptimal meaningful segmented and labelled regions; (3) the intermediate-level specific-feature processing and decomposition of the segmented image data into valid primitives (boxes, cylinders, and spheres) based on the geometric reasoning provided by the "primitive viewing knowledge-base" (PVKB) (4) intermediate-level geometric reasoning based on the "primitive viewing knowledge-base" (PVKB) to identify, hypothesize, and establish the type of primitive and its camera-oriented viewing angle; (5) creation of 3D primitive-based description of the objects seen in the 2D image of the 3D scene; and finally (6) high-level interpretation and recognition by first selecting the candidate models based on the established 3D primitive-based description and then by detail frame-based matching of the image data to the selected model through structural and relational matching for the established viewing angle. In case of a mismatch because of either lack of information or corrupted information, the model-driven top-down feedback are issued by the high-level system. These top-down feedbacks are focused over the selected window area and directed by the expected goal in order to reject or accept the current hypothesis.

At present, the pre-processing, low-level preliminary segmentation, rule-based segmentation, and feature extraction have been completed. The data structure of the "primitive viewing knowledge-base" (PVKB) has also been completed. We have also developed new algorithms and programs based on attribute-trees matching for decomposing the segmented data into valid primitives. We can now hypothesize their viewing angles using PVKB by matching the hierarchical structural and relational attribute-trees. The frame-based structural and relational descriptions of some objects (similar to those seen in the simulated video show at NASA for the space station) have been created and stored in a knowledge-base. This knowledge-base of the frame-based descriptions has been developed on the MICROVAX-AI microcoputer in LISP environment. Other expert systems, related to the segmentation, decomposition and geometric reasoning have also been developed on the MICROVAX-AI station. We have successfully interpreted the simulated 3D scene of simple non-overlapping objects as well as real camera data of images of 3D objects of low-complexity.

The initial results of the knowledge-based low-level analysis system, the intermediate-level primitive decomposition system, and the high-level recognition system have been reported in two research papers [1,2] to be published and presented at the SPIE Digital and Optical Shape Representation and Pattern Recognition Symposium, and Applications of Artificial Intelligence VI conference to be held from 4-8 April at the Orlando Peabody Hotel, Florida. Three students who were supported from this project have already completed their Master of Science theses. They are:

(1). Himanshu Baxi, "A Low-Level Image Analysis System", completed, 1987.

(2). Nilesh Thakkar, "Intermediate-Level Feature Extraction and Object Representation in a Knowledge-Based Vision System", completed, 1987.

(3).Sushma Ghiya, "Intermediate-Level Analysis for Primitive-Based Decomposition of Image Data", completed, 1987.

One student used the low-level and the intermediate-level processing modules for interpreting 3D medical images. This thesis was also completed during this project. This theis is listed as following:

(1) Sridhar Juvvadi,"A Knowledge-Based Approach for Interpreting Computerized Tomography (CT) Images", completed, 1987.

Two other students are just about to complete their theses on the topics related to this research. They are:

(1).Htam Hmam, "Geometric Reasoning from Perspective Distortions of 3D Scenes", to be completed, May 1988.

(2).Chih-Ho Chao,"High-Level Matching for 3D Primitive-Based Object Recognition System", to be completed, May 1988.

PUBLICATIONS:

(1).Dhawan, A. P., Ghiya, S., Thakker, N., and Chao, C. "A primitive-based 3D object recognition system", accepted for presentation and publication, **Digital and Optical Shape Representation and Pattern Recognition, 1988 SPIE Technical Symposium on Optics, Electro-Optics, and Sensors**, April 4-8, Orlando, Florida, 1988.

(2). Dhawan, A. P., Baxi, H, and Ranganath, M.V., "Knowledge-based low-level image analysis for applications in object recognition and scene interpretation systems", accepted for presentation and publication, **Applications of Artificial Intelligence VI Conference**, SPIE and IEEE Computer Society, April 4-8, Orlando, Florida, 1988.

(3). Dhawan, A. P., Baxi, H., and Ranganath, M.V., "A hybrid low-level image analysis system", submitted to the **Computer Vision, Graphics and Image Processing**, 1988.

Technical Description of the System:

A paper describing the technical aspects of the system is enclosed. The paper is entitled "A Primitive-Based 3D Object Recognition System".

A PRIMITIVE-BASED 3D OBJECT RECOGNITION SYSTEM

ABSTRACT

A knowledge-based 3D object recognition system has been developed. The system uses the hierarchical structural, geometrical and relational knowledge in matching the 3D object models to the image data through pre-defined primitives. The primitives, we have selected, to begin with, are 3D boxes, cylinders, and spheres. These primitives as viewed from different angles covering complete 3D rotation range are stored in a "Primitive-Viewing Knowledge-Base" in form of hierarchical structural and relational graphs. The knowledge-based system then hypothesizes about the viewing angle and decomposes the segmented image data into valid primitives. A rough 3D structural and relational description is made on the basis of recognized 3D primitives. This description is now used in the detailed high-level frame-based structural and relational matching. The system has several expert and knowledge-based systems working in both stand-alone and co-operative modes to provide multi-level processing. This multi-level processing utilizes both bottom-up (data-driven) and top-down (model-driven) approaches in order to acquire sufficient knowledge to accept or reject any hypothesis for matching or recognizing the objects in the given image.

INTRODUCTION

The basic problem of recognizing 3D objects from a single perspective 2D image of a 3D scene is not only complex from the geometric reasoning point of view, but is an ill-posed problem with incomplete high-level information. This is primarily due to the processes of image acquisition and non-uniqueness of low-level region extraction. Further, to teach computers to recognize an object and interpret a 3D scene, one needs a very strong representation of the structural, geometrical and relational knowledge of objects. Also, how to provide adequate reasoning for using these sources of knowledge for creating hypotheses for candidate models and then matching image data to the model, is another central issue.

Early machine vision systems worked exclusively in the "block world" domain trying to separate out and identify each polyhedron in a scene (Guzman, 1968; Huffman, Clowes & Waltz, 1971 & 1978; Agin, 1973). The use of constraint analysis was introduced and physical constraints on edges and vertices were applied (Huffman, 1978). The "block world" objects were basically modeled by surface-edge-vertex representations. With such representations it is difficult to define or explain complex objects. The use of relational models and geometrical reasoning was developed later for describing objects in a simpler way (Barrow & Tannenbum, 1976; Hanson, 1978; Brooks, 1981; Parma, 1981). Then, with the advances in computerized processing, emphasis was shifted to advance control mechanism such as pyramid structures and discrete relaxation processes to provide tools for object matching. With the help of knowledge-based sytems and AI techniques, it now seems possible to develop model and hypotheses driven vision systems for object recognition and scene understanding (Shapiro, 1981, 1983, 1985). Of course, the limitations related to data management, storage, data processing speed, and the need for more sophisticated methods to represent knowledge in a more efficient form, etc., still exist but for a specific application and a finite object domain, the research efforts towards development of new systems and techniques should be useful and rewarding and must be encouraged.

VISIONS ("Visual Integration by Sementic Interpretation of Natural Scenes", developed by Hanson & Riseman, 1978); and ACRONYM (developed by Brooks, 1981) are two good examples of complex computer vision systems. Other model-based vision systems include MSYS (Borrow & Tenenbaum, 1976); Kanade's scene analysis system (Kanade & Reddy,1981,), and ARGOS (Rubin, 1978). In the VISIONS system, analysis of a scene is a task of model building and constructing a description of the major objects. There are four components involved in the model building. First, a multi-level representation of the model being built and of the stored world knowledge. Second, data processing between levels of representation. Third, high level control, and finally, a tree search mechanism. All four components are hierarchical in nature.

Other interesting approaches used for developing image understanding systems include inexact graph matching in object recognition(Eshera & Fu, 1986); dynamic programing based topological structure matching in outdoor-scene analysis (Levine, 1978); and rule-based interpretation based on overall spatial and structural consistency for aerial imagery (McKeown et al., 1985).

Lack of a powerful, accurate and efficient low-level analysis and descriptive process, an adequate representation of the high-level knowledge, and the model-driven top-down feedback process to modify and update the knowledge required for high-level recognition have been the common problems of these systems. Because of the inherent problems of of image acquisition including the geometric limitations, digitization and segmentation, the process of interpretating a 3D scene from 2D image becomes so ill-posed that the high-level recognition must not depend much on the quantitative measurements and analysis. Instead, more symbolic representation of the key attributes of structural and relational details defining 3D objects must be used. Also, both bottom-up and top-down analyses must be performed to make better predictions and interpretations. Only one type of approach was used in some systems, e.g., Borrow & Tenenbaum, 1976 used only bottom-up analysis, while Bolls, 1976] and Garvey, 1976 used the top-down analysis. Nagao & Matsuyama,1980 incorporated both types of analyses but used ad hoc rules to determine which type of analysis is to be used at what stage of processing, in the system developed for understanding aerial photographs. Such system requires a large set of domain dependent control knowledge to control the overall system.

In order to recognize objects and interpret the scene in the environment of robotic automation, such as in the space station, a powerful knowledge-based vision system is required. In such applications the object domain is finite and mostly of man-made objects. These objects can be described and decomposed into three basic primitives: 3D rectangular box, cylinder, and sphere. Thus, if the 2D image data can be decomposed into these primitives, by analyzing the combination of these primitives hypothesized in the image, we can create a 3D primtive-based description of the objects present in the scene. This primitive-based description is then utilized in the high-level matching nad interpretation analysis to recognize the objects. First, the types of primitives and their attchments are considered to hypothesize and instantiate the models stored in the data-base, and then detailed matching is performed using the detailed description to verify the hypotheses.

We are developing a knowledge-based 3D object recognition system that uses the structural, geometrical and relational matching of 3D object models to the image data through pre-defined primitives. The primitives, we have selected to begin with, are 3D boxes, cylinders, and spheres. The system has several expert and knowledge-based systems working in both stand-alone and co-operative modes to provide multi-level processing. This multi-level processing utilizes both bottom-up (data-driven) and top-down (model-driven) approaches in order to acquire sufficient knowledge to accept or reject any hypothesis for matching or recognizing the objects in the given image.

The complete 3D object recognition process in the system, we are developing, has six major steps: (1) the entry-level pre-processing to enhance features and obtain the preliminary segmentation; (2) the low-level global feature extraction followed by the rule-based expert segmentation to yield suboptimal meaningful labeled regions; (3) the intermediate-level specific-feature extraction and decomposition of the segmented image data into valid primitives (boxes, cylinders, and spheres) based on the geometric reasoning provided by the "primitive viewing knowledge-base"; (4) intermediate-level geometric reasoning based on the "primitive viewing knowledge-base" (PVKB) to identify, hypothesize, and establish the type of primitive and its camera-oriented viewing angle; (5) creation of a 3D primitive-based description of the objects seen in the 2D image of the 3D scene; and finally (6) high-level interpretation and recognition by first selecting the candidate models based on the established 3D primitive-based description and then by detailed frame-based matching of the image data to the selected model through structural and relational matching for the established viewing angle. In case of a mismatch because of either lack of information (the information that may have been washed out during segmentation, e.g., deletion of a weak edge) or corrupted information, the model-driven top-down feedback are issued by the high-level system. These top-down feedbacks are focused over the selected window area and directed by the expected goal in order to reject or accept the current hypothesis (see Figure 1).

We have discussed the entry-level preprocessing, preliminary segmentation, rule-based segmentation, and window processing elsewhere [Dhawan et al., 1987]. In this paper, we present the overall approach for the decomposition of image data and high-level recognition. The discussion includes the data structure and the development of the Primitive-Viewing Knowledge-Base, the intermediate-level processing to decompose the segmented data into valid primitives, and hypothesizing the viewing angles using PVKB. Also, we present primitive-based detailed structural and relational matching for the high-level recognition. The high-level structural and relational knowledge about the model is stored in frames. The frame-based matching of the data to the model has been implemented using an expert system building tool, KEE version-3, on a SYMBOLIC-3640 computer. The preliminary results are presented.

PROCEDURES AND METHODS

Primitive Viewing Knowledge-Base (PVKB)

To begin with, we have selected only three primitives: box, cylinder, and sphere. The selection of these primitives is largely based on the type of objects the proposed system is being designed to recognize in a space station. Out of these three primitives, the box is the one with the largest structural variations when viewed from different angles. We assume that we have an imaging system that gives us a single 2D perspective view of the 3D scene having 3D objects. The approach can be easily extended to the case of dealing with orthographic views, if the camera is located too far from the objects. We first compute, several views of the box primitive by rotating it by fixed increments in all three directions. Each view is then represented by a graph having the structural and relational attributes in the form of an ordered tree. All of these views are stored in a Primitive Viewing Cube (PVC) which is represented by an octree. Thus, any view can be accessed by accessing a node in the octree and the viewing angle can be found by reading the node position. Figure 2 shows the concept of the PVC. Similar, but less complicated PVC for other primitives: cylinder, etc., are computed and stored in the PVKB. The resolution of each PVC, i.e. the increment in the rotation angles is based on finding a significantly different structural and relational information.

In the structural and relational tree graphs, the complete primitive, as viewed, is placed at the root of the tree. Root node then has closed regions as children. Each region node has segments as children nodes. Each segment is classified as a line, or an arc, or a closed curve. Each segment node is then linked with other segment nodes through attributes and values, as shown in Figure 4(b). The segment-segment links are visualized in two modes: *connecting* and *facing*, e.g., the lines can either be connecting or facing. In case of *connecting* lines, the attribute is defined by the values of line length and the angle by which it joins another line. These measurements of lengths and angles are transformed into appropriate pseudo-symbolic form, such as angles are categorized as less than 45, greater than 45 but less than 90, equal to 90, greater than 90 but less than 135, greater than 135 but less than 180, equal to 180, etc. Other defining attribute combinations of the connected line, arc and closed curve segments are shown in Figure 3. While in case of *facing*, the attribute is defined by length or area, the distance and the values which are parallel, converging, or diverging. For connecting arc with arc or line, the angles are defined as the angle between line joining two end points of the arc with another connecting line or another line joining two end points of the connecting arc. The closed curve can have an attribute touching (same as connecting), or facing, or concentric. (The crossing or overlapping curves are broken into arcs.) The attributes of the closed curve can be defined by the values of the area enclosed, and the distance between the centroids (if another closed curve is touching, facing, or is concentric) and by the length of touching (in case of touching). Figure 3 shows a table of the attributes and the values by which they are defined. At present, we are ignoring the length of the line or arc, and the area of the closed curve. Angle has been taken as the major attribute parameter in the *connectivity* attribute and type of facing (parallel, convex or converging, concave or diverging) is taken as the major parameter in the *facing* attribute.

ATTRIBUTE: CONNECTING OR TOUCHING

	Line	Arc	Closed curve
Line	ll, θ	ll, θ_j	ll, tl
Arc	al, θ_j	al, θ_j	al, tl
Closed Curve	A, tl	A, tl	A, tl

ATTRIBUTE: FACING

	Line	Arc	Closed curve
Line	ll, d, x	ll, d, z	ll, d, z
Arc	al, d, z	al, d, z	al, d, z
Closed Curve	A, d	A, d, z	A, d

Note: ll: line length; al: arc length; tl: touching length; A: area; θ: angle between two lines; θ_j: angle between a line and the line joining two ends of the adjacent arc; d: average distance; x: parallel, or converging, or diverging; z: convex or concave.

Two concentric closed curves is taken as a special case of closed curve facing another closed curve where the d is measured between the two centroids.

Figure 3. The table showing the attributes and their properties used for creating SRG trees for the structural and relational matching.

Decomposition of image data for creating 3D description

The segmented image is scanned region-by-region by a knowledge-based system having the knowledge of primitives viewed from several angles covering the valid viewing range. The structural and relational graphs for image data are created and then used for matching with those stored in the PVKB. For example,

Figure 4(a) shows the primitive "cylinder" viewed by rotating the cylinder about x axis by 45 deg in anticlockwise direction (the y axis is aligned to the axis of the cylinder). Figure 4(b) shows the structural and relational graph of Figure 4(a) presented in the form of a tree. The attribute links shown in Figure 4(b) are corresponding to the attribute "connecting". The values assigned to these links, as per table shown in Figure 3, are not shown in this and subsequent graphs. The attribute "facing" will have different attribute links; only two in this case, between C1 and A1; and between L2 and L1. Similarly, Figure 4(c) shows the primitive "box" viewed by rotating the cylinder about x axis by 45 deg in anticlockwise direction; and Figure 4(d) show its "connecting" structural and relational tree. Figure 5(a) shows the image data restructured and simulated from an input image of a "cylinder placed on a box". Figure 5(b) shows the structural and relational graph (SRG) of the image data. Now the region-by region matching of the image data to the stored primitive models is started. First, based on the intermediate level features such as shape of the region, type of segments forming the region, number of segments in the region, etc., a candidate primitive is hypothesized in a data-driven mode; and then weighted SRG tree matching is performed in a model-driven mode. Weights for each node are assigned on the basis of the area covered by the node. After a primitive's SRG tree has been matched, the possible viewing angles are obtained from the PVKB just by finding out the viewing angles with similar SRG trees from the index of the PVKB for the primitive. Now, if some apriori knowledge is available as the restrictions imposed in the real image world on the rotation of the objects (such as rotation about a particular axis is allowed only) and/or from the camera location, it is used to strike out the angles which give similar views but are not valid. After a primitive view is identified in the image data, the corresponding segments are deleted, and the open nodes are linked to form optimal number of convex regions. This is performed by first identifying the nodes of mistmatch with the candidate SRG tree and then executing the "extend-segments" or "delete-nodes" rules to obtain convex regions. Thus, the image data is decomposed into a set of primitives and their hypothesized valid viewing angles. We now have a 3D description of the image scene in terms of the 3D primitives.

High-Level Matching and Verification of the Candidate Model-Object

The 3D description of the image data is used to find candidate 3D object models stored in the model-knowledge-base. For example, for the image data shown in Figure 5(a), the description obtained after the decomposition will include

the number of primitives: 2;
type of primitives: the cylinder and the box
the common viewing angle(s); around 45 deg
attachment: box and cylinder;
relationship: involving faces, covering full.

This description is used for creating hypotheses of object models and frame-based detailed structural and relational matching and verification for the established viewing angle(s).

Each object model is described in terms of parts (components). Each part is a composition of one or several primitives. For example, the Figure 6(a) shows a toy which has only two real parts: a cylindrical stick and a rectangular box with a hole, but from primitive decomposition point of view, the toy has three parts, a cylinder attached to a brick attached to a cylinder. The model description is developed in a frame based hierarchy. Another model of a small box over a big box is shown in Figure 6(b). This can be decomposed into two primitives only. Figure 6(c) shows a scene having the toy object on the brick structure.

At the highest level, for a general description of the model, there are three slots: (1) Parts slots that contain all type of components, (cylinder and box, for example, for a toy shown in Figure 6(a)); (2) Structure slot containing the exact sequence of the components (cylinder, box, cylinder, for the example); and (3) Coordinate slot describing the relative orientation of stacking of components. The frame of the general description of the toy figure is

(toy(parts(value stack box))

(structure(attach cylinder box cylinder))
(coordinate (((base-frame cylinder) (base-frame box))
((base-frame cylinder) (base-frame box))).

The complete decription of the toy figure is shown, later in this paper, which is KEE version of the description.

In the 3D scene description obtained from the image data, each primitive which is visible from a particular viewing angle has been already identified. In each primitive frame, in the model description, there are properties with structural descriptions. The properties include relative size, orientation, generic class, etc. The structural description include attachment and spatial (positional) relations with respective to other primitives. The procedure of the high-level matching and recognition is as following:

(1) Each primitive description from the image data is first scanned to see if it is a component of the models stored in the knowledge-base. If yes, we put the model on the candidate list on the lowest level.

(2) These candidates are now scanned on the basis of their attachment primitives. The models having those primitives having no attachment evidence similar to the image data will be dropped out. The candidate models having attachment primitive similar to image data are now put on the second level in the most-likely-candidate hierarchy.

(3) The description of the attachment(s) is now analyzed, and most-likely candidate models having similar type of all or most of attachments are put on the top of the hierarchy.

(4) For each model, a focus of attention is created on a primitive having the largest number of the neighbors which are also parts of the model. Starting from the focus of attention primitive, we search in detail all primitives belonging to the selected model through the attachment relationship. In order to reduce the search space, first the type of attached primitives to the focus-of-attention primitive is examined. In case of a match, the fine details of the attachement (such as partially or completely attached) are examined, otherwise, a new focus-of-attention is created. After the matching, a new frame is created to show the

model with parts which are found and matched in the data, and the parts which are not found labeled as missing. A score is assigned to this frame indicating the belief and confidence in overall matching.

(5) If the score of matching is perfect, or near perfect (above certain threshold) and there is no other competing model hypothesis, the image data will be declared "recognized" as per model. If it does not happen, the missing information or primitive(s) are identified as per candidate model (having a reasonable score for matching) spatially and windowed in the image data. The top-down feedback is created for performing the low-level analysis again. The top-down feedback low-level modification analysis is discussed in the accompanying paper [Dhawan et al., 1987]. In case of some new information at the low-level, intermediate-level processing is also modified, and the resulting effect is interpreted at the high-level in the knowledge of the model. If the modification returned by the top-down feedback raises the matching score above the acceptance threshold, the model is accepted and the process is terminated. If this does not happen, the model is rotated for the established viewing angle(s) (hypothesized at the intermediate-level). The description frame is again created to find out whether the "missing information or primitive" is still a part of the model or not. If yes, the model is rejected. If not, the model is accepted.

RESULTS AND DISCUSSIONS

We implemented the intermediate-level decomposition of image data into 3D primitives, and high-level matching and interpretation on a SYMBOLICS-3640 computer using the KEE-3 (an expert system building tool) environment. Frame-based structure was used for representing knowledge in both stages.

For the discussion of high-level matching, we will now use a scene. The scene contains both objects; the toy shown in Figure 6(a) and the brick structure shown in Figure 6(b). The segmented image data of the scene (toy and the brick structure) after the final segmentation and intermediate-level feature processing is shown in Figure 7. Figure 8 shows the KEE version of the primitive-based frame description of the toy. The descriprion contains a unit called a stick which has been further described as a cylinder in the hierarchical frame structure. Thus, a hierarchy of frames is implemented for the complete description of the model objects.

The model objects can be put together to create other model objects of greater complexity for the high-level knowledge-base. The attributes, properties, values, etc. can be inherited to define bigger model objects from the smaller model objects (primitives at the lowest level). Thus we can expand the existing knowledge-base after a part of the scene (or, the complete scene) has been interpreted in terms of the model objects. The complete scene (or, a sequence of the scenes having same objects) can then be interpreted using the expanded knowledge-base.

The instantiated primitive descriptions after going through the process of structural and relational nmatching, as described above, creates the object description. For example, in this case, two primitives P3 and P5 were instantiated and verified as the object "brick-stack". The premitives P1, P2, and P4 were also instantiated and verfied as the "toy" object. The final scene description was creaated as the "toy" completely attached with the "brick-stack".

CONCLUSION

We have developed an intermediate-level knowledge-based system for decomposiing the segmented data into 3D primitives to create an approximate 3D description of the real world scene from a single 2D perspective view. We have also developed a knowledge-based approach for high-level primitive-based matching of 3D objects. The intermediate-level decomposition and the high-level interpretation both are based on the structural and relational matching and are implemented in a frame-based environment. The preliminary results show the successful recognition of the simple objects in a non-ambiguous situation. These results are quite encouraging. We are expanding the knowledge-base to include more complex objects. This is to be noted that the proposed system is being developed for a specific application of recognizing 3D objects in a space station. The objects expected to be present in the space station are the ones which can be described by the combination of the selected primitives: 3D box, cylinder, and sphere. The computer-aided descriptions of these objects are avilable to the high-level interpretation system for detailed matching. The approach used in our system is therefore based on first creating a 3D primitive-based description of the scene from the 2D perspective image data and then matching it to the models of the objects stored in the data-base. Future studies include evaluation and modifications of our present approaches and procedures to analyze and interpret more complex scenes.

ACKNOWWLEDGEMENT

This work was supported, in parts, by the grants from the NASA Johnson Space Center, Houston, and the Texas Advanced Technology and Research Program (TATRP). Authors are very grateful to Dr. J.K. Aggarwal, University of Texas, Austin, for the encouragements and comments.

REFERENCES

Agin, G.J. & T.O. Binford, Computer description of curved objects, 3rd Int. Joint Conf. on AI, Stanford, August, 1973.

Bogdanowicz, J.F., An evolving system for image understanding, SPIE vol. 538, Digital Image Processing, 1985.

Bolles, R.C., Verification vision within a programmable assembly system, Stanford Artificial Intelligence Lab. report AIM-295, Dec. 1976.

Borrow, H.G., & J.M. Tanenbaum, MSYS: A system for reasoning about schemes, SIR AI Tech., Report 121, 1976.

Brooks, R.A., Symbolic reasoning among 3D models and 3D images, Report STAN-CS-81-861, Department of Computer Science, Stanford University, June 1981.

Brooks, R.A., R. Greiner & T.O. Binford, The ACRONYM model-based vision system, Proc. IJCAI-6, Tokyo, PP. 105-113, 1979.

Brooks, R.A., Symbolic reasoning among 3D models and 2D images, Ph.D. Dissertations, Stanford University, 1981.

Clowes, M.B., On seeing things, Artificial Intelligence 2(1), pp/ 79-116, 1971.

Davis, L.S., The SIGMA image understanding system, Proc. 3rd Workshop on Computer Vision, Representation and Control, pp. 19-26, 1985.

Dhawan, A.P., H. Baxi & M.V. Ranganath, A hybrid low-level image analysis systems, submitted for presentation at the IEEE Conference on CVPR, 1988.

Dhawan, A.P., S. Juvvadi, H. Baxi & N. Mullani, Knowledge-based analysis and understanding of 3D anatomical and functional images, to be presented at the IEEE EMBS conference, Nov. 13-16, 1987.

Edwards, G.R., et al., Image understanding application project results, Proc. AIAA Computers in Aerispace Conf., 1983.

Eshara, M.A. & K.S. Fu, An image understanding system using attributed symbolic representation and inexact graph matching, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-8(5), pp. 604-618, 1986.

Garvey, T.D., Perceptual strategies for purposive vision, Tech note 117, SRI International, Sept. 1976.

Guzman, A, Computer recognition of 3D objects in a visual scene, MAC-TR-59, M.I.T. Dec, 1968.

Hanson, A.R. & E.M. Riseman, VISIONS: a computer system for interpreting natural scenes, in Computer Vision Systems, ed. A. R. Hanson & E.M. Riseman, AP, 1978.

Huffman, D.A., Impossible objects as nonsense sentences, in Machine Intelligence ^, pp. 295-323, 1978.

Kanade, T. & R. Reddy, Image understanding at CMU, Proc. ARPA Image Understanding Workshop, McLean, Va, PP. 199-207, 1981.

Shafer, S.A. & T. Kanade, Using shadows in finding surface orientation, DARPA Image Understanding Workshop, 1982.

Shapiro, L.G., A structural model of shape, IEEE Trans, PAMI, PAMI 2(2), pp. 112-126, 1980.

Shapiro, L.G. & R.M. Harlick, Structural description and inexact matching, IEEE Trans, PAMI, PAMI-3(5), pp. 504-519, 1981.

FIGURE CAPTIONS

Figure 1. The schematic block diagram of the proposed knowledge-based 3D object recognition and scene interpretation system.

Figure 2. The concept of the Primitive Viewing Cube stored in the form of an octree. Each node of the octree stores information about corresponding SRG tree.

Figure 3. The table showing the selected attributes and their properties for creating SRG trees.

Figure 4(a). The primitive "cylinder" as viewed by rotating it about the x axis by an angle of 45 deg in anticlockwise direction (the y axis is aligned to the axis of the cylinder).

Figure 4(b). The structural and relational graph (SRG) tree of Figure 4(a) for the attribute "connecting". The attribute "facing" will have only two attribute links: between C1 and A1; and between L2 and L1.

Figure 4(c). The primitive "box" as viewed by rotating the cylinder about x axis by 45 deg in anticlockwise direction.

Figure 4(d). The "connecting" structural and relational tree of Figure 4(c).

Figure 5(a). The simulated data of an image of a 3D scene having a "cylinder placed on a box".

Figure 5(b). The structural and relational graph (SRG) of the image data shown in Figure 5(a).

Figure 6(a). An image of the object "Toy".

Figure 6(b). An image of the object "Brick Structure".

Figure 6(c). An image of the scene having the "toy" and the "brick structure".

Figure 7. The segmented image data of the scene (toy and the brick structure) after the final segmentation and intermediate-level feature processing.

Figure 8. The primitive-based frame description of the toy.

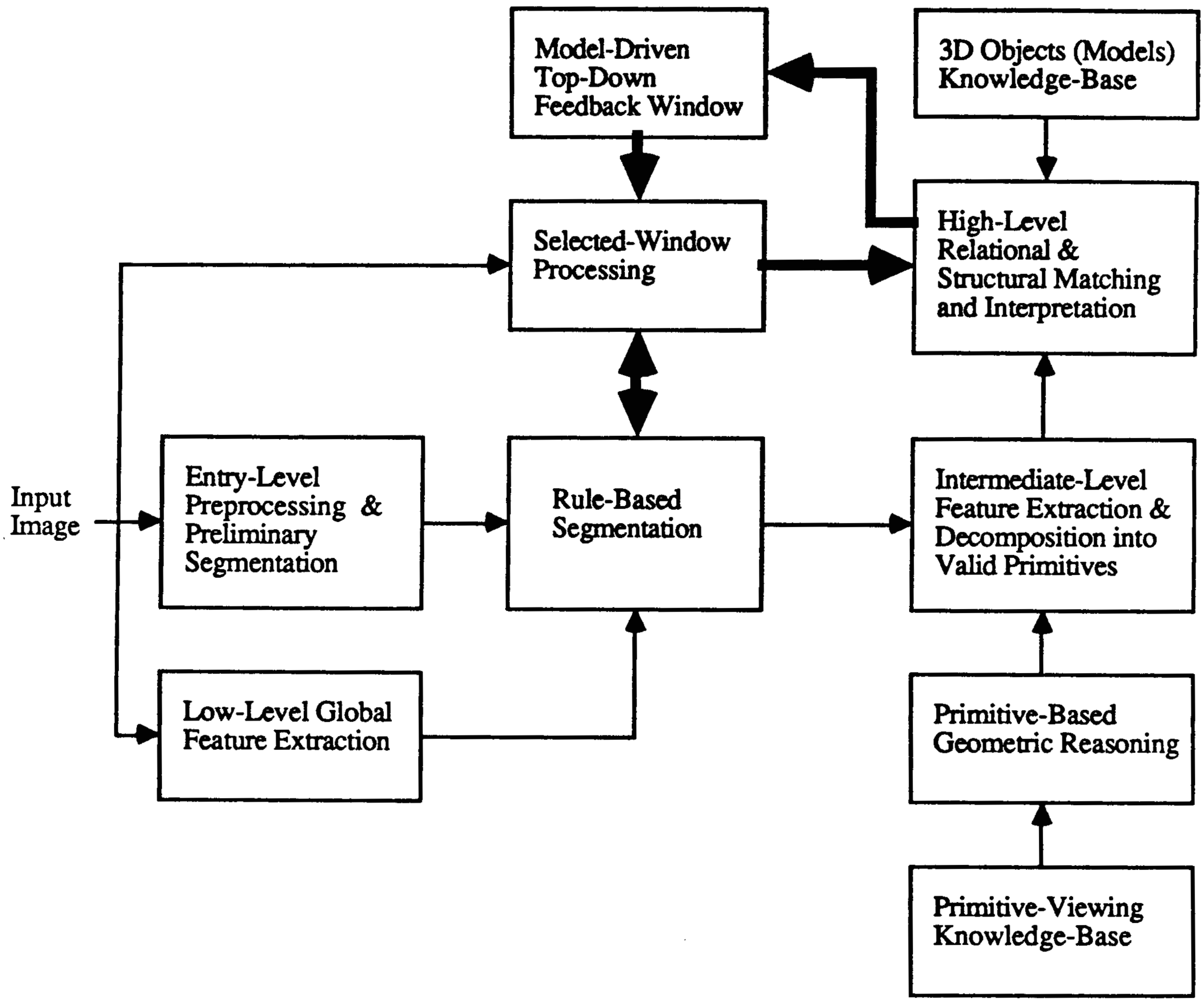


Figure 1: The schematic block diagram of the proposed knowledge-based 3D object recognition system.

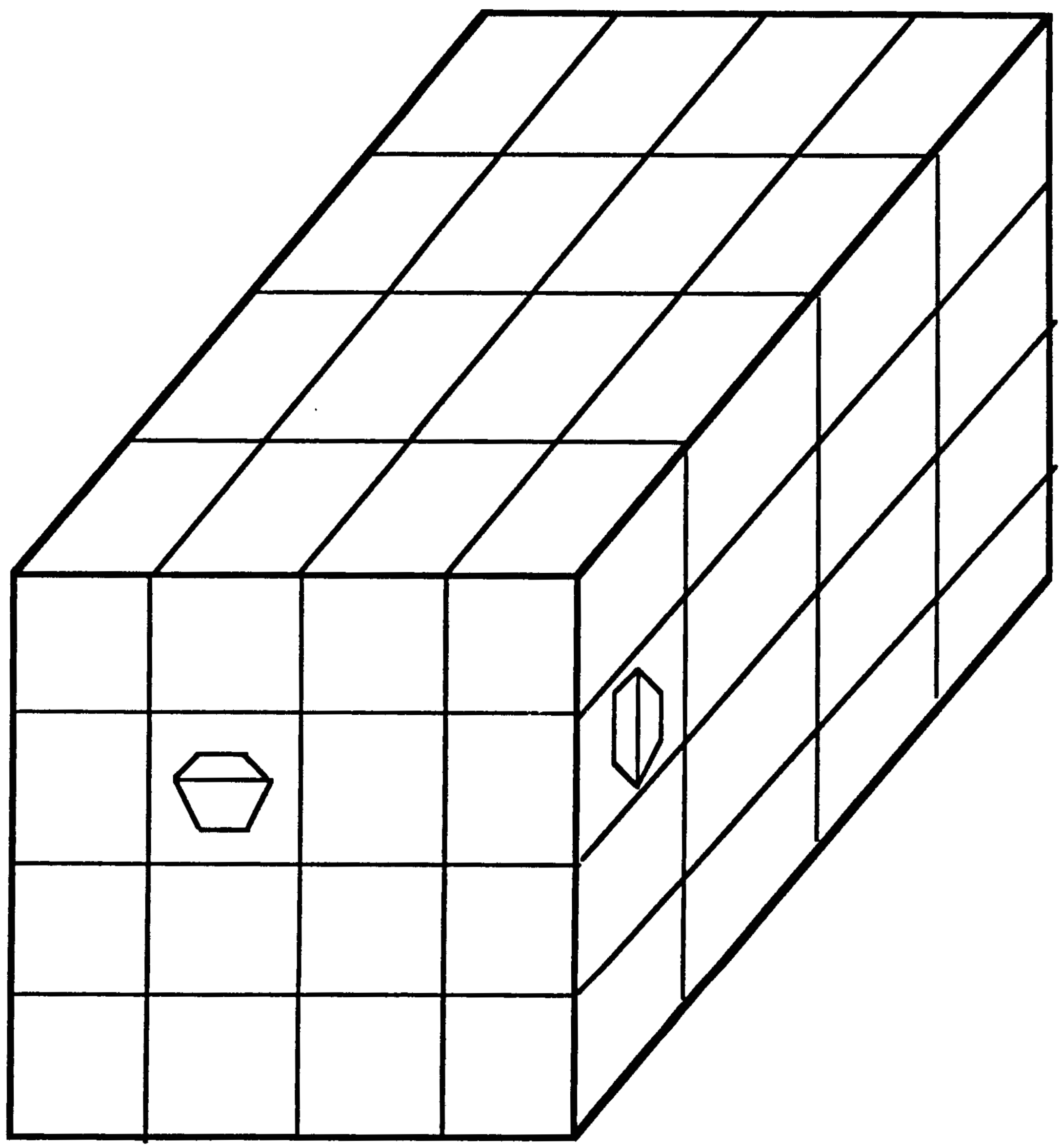


Figure2.The concept of the Primitive Viewing Cube (PVC) for PVKB.

Figure 3 - See text.

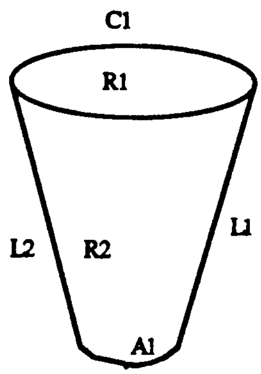


Figure 4(a)

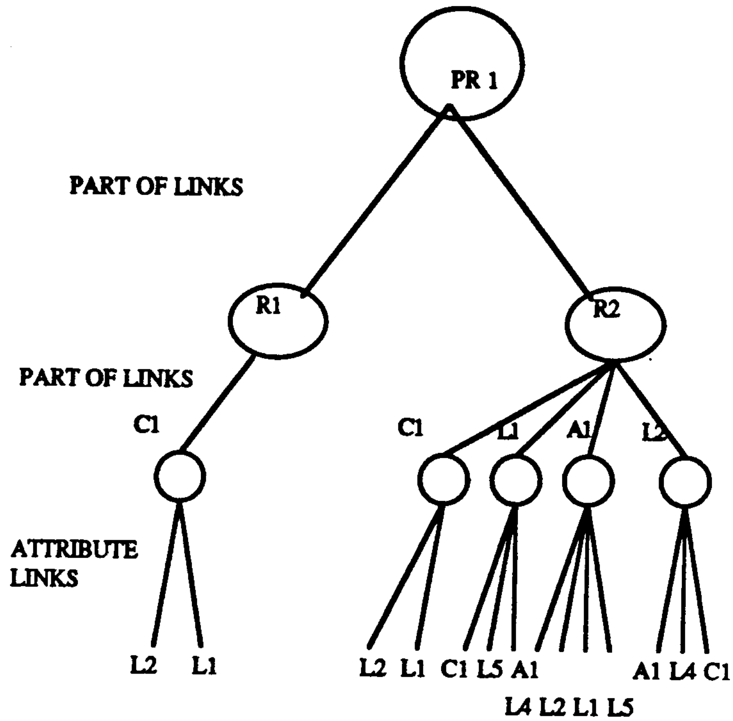


Figure 4(b)

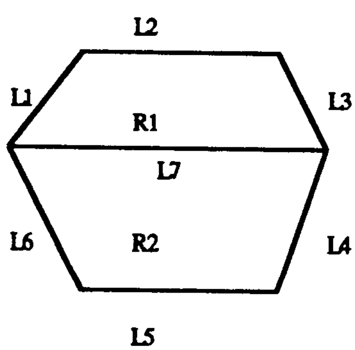


Figure 4(c)

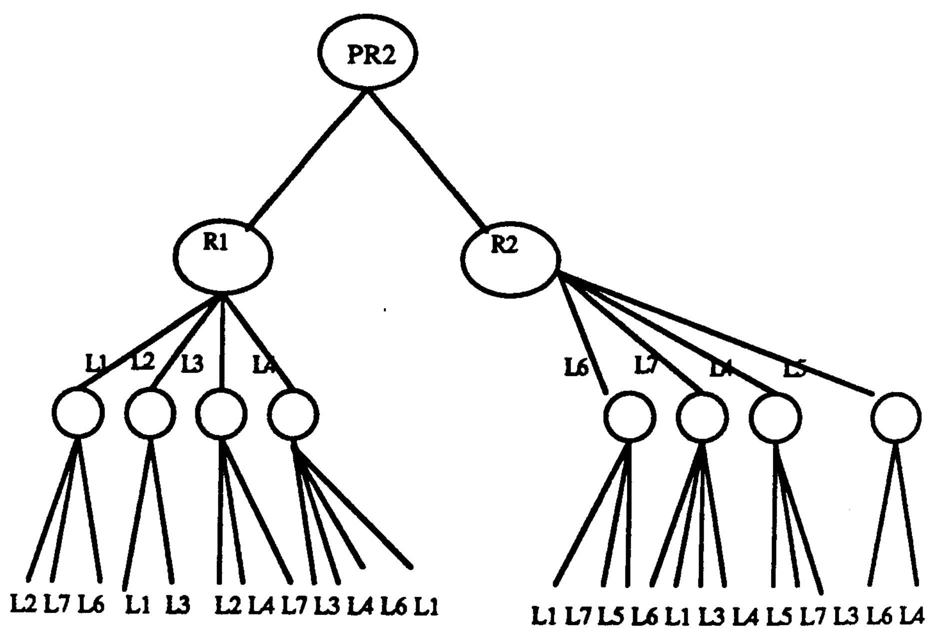


Figure 4(d)

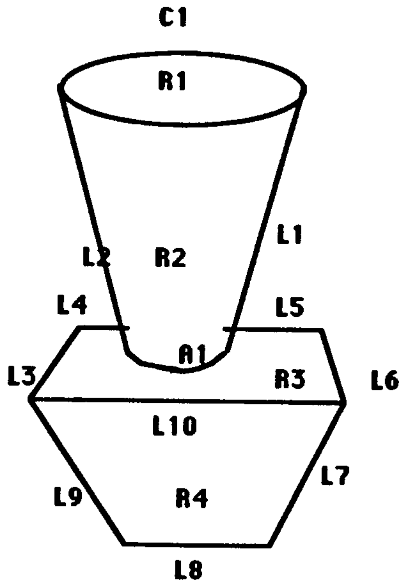


Figure 5(a)

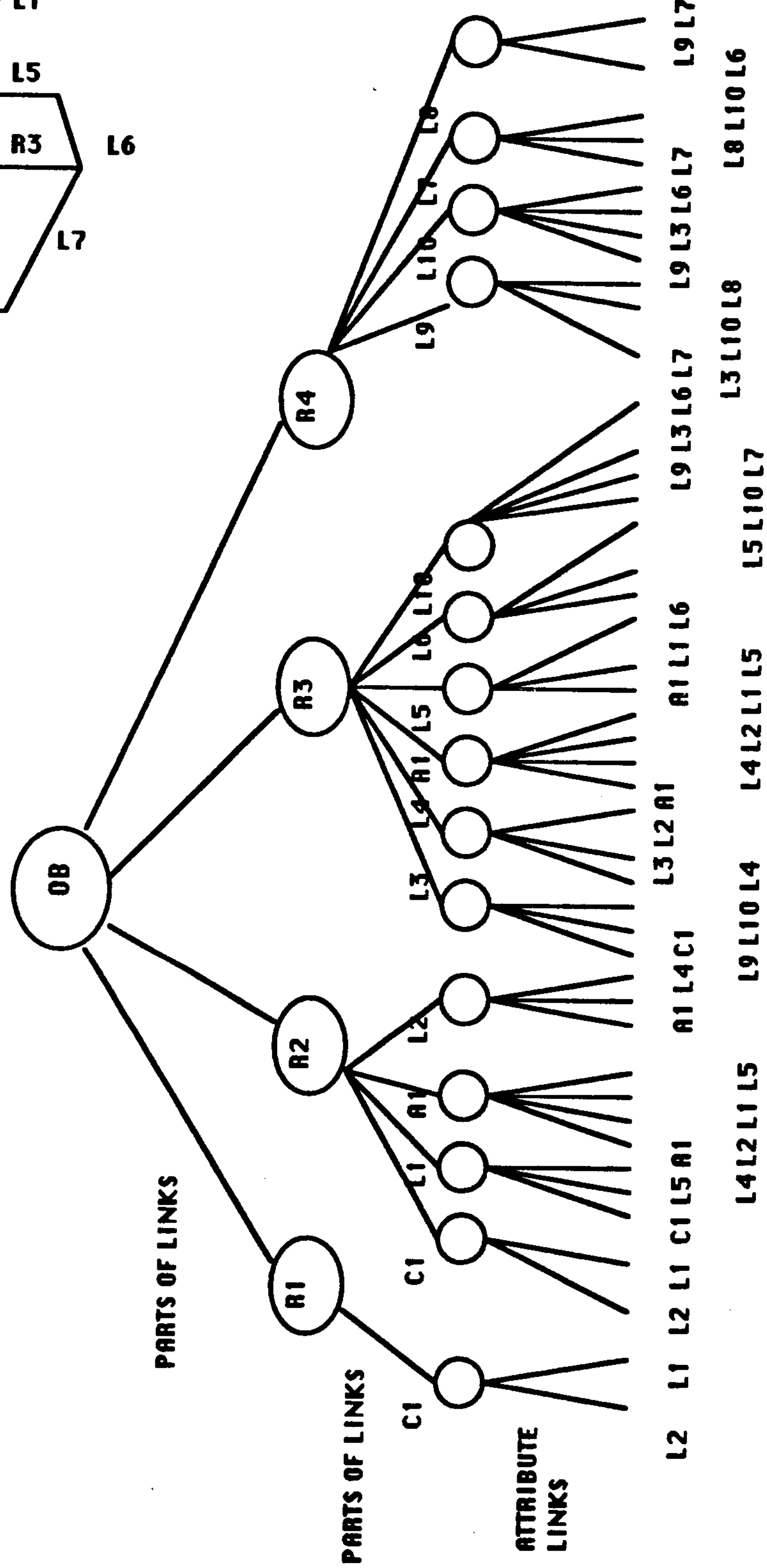


Figure 5(b)

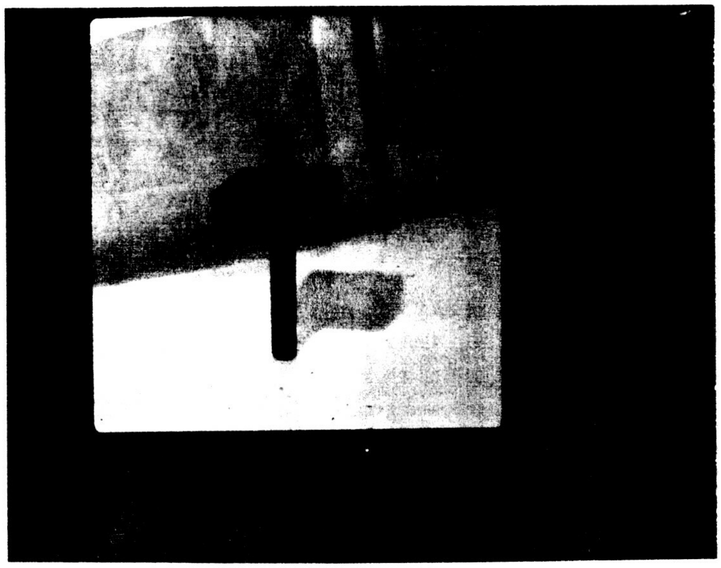


Figure 6(a)

ORIGINAL PAGE IS
OF POOR QUALITY

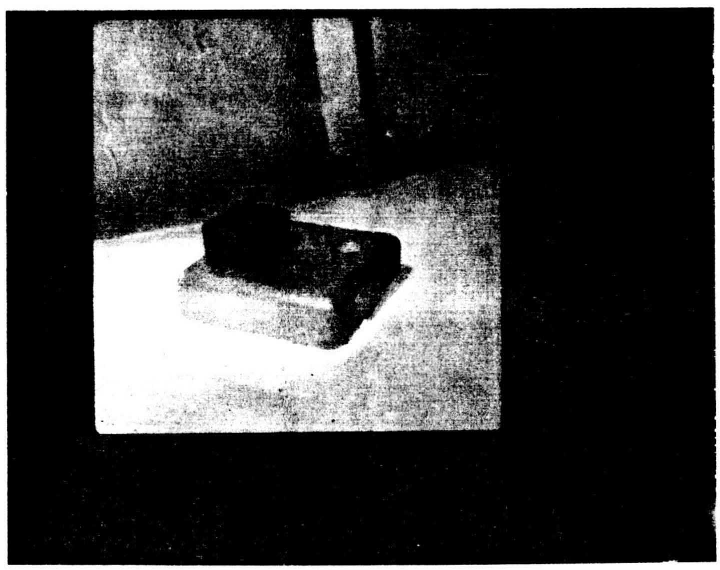


Figure 6(b)

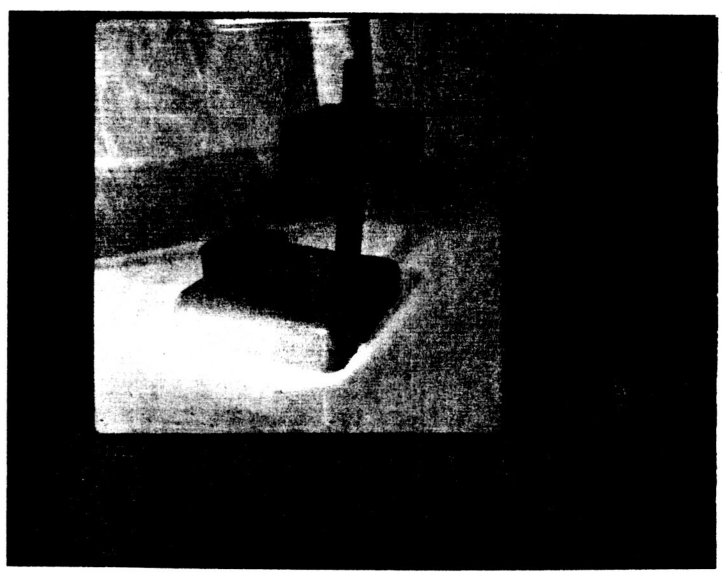


Figure 6(c)

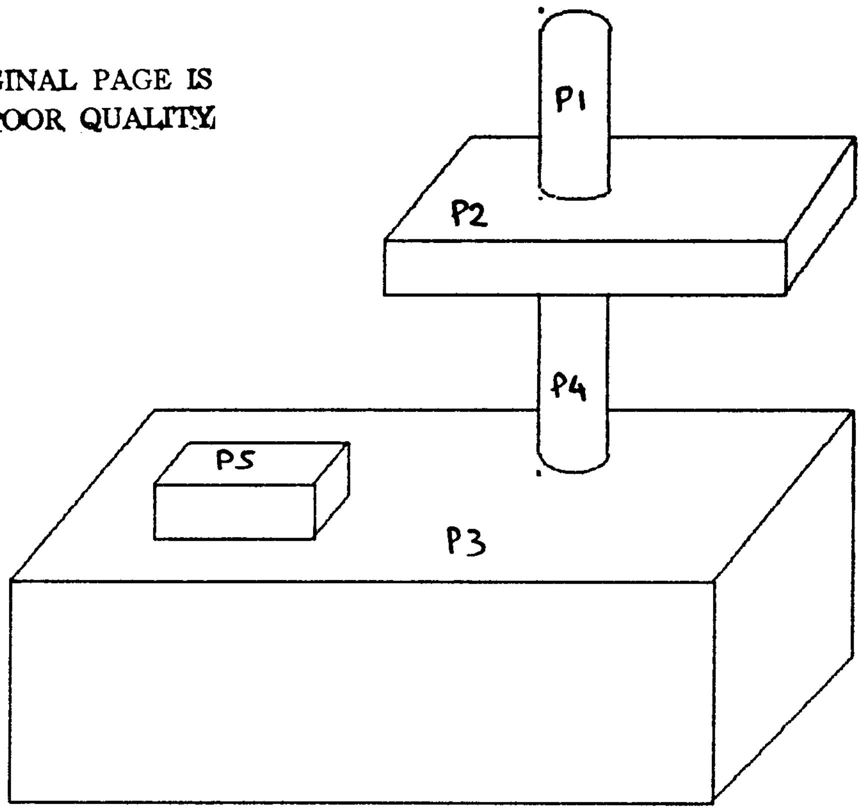
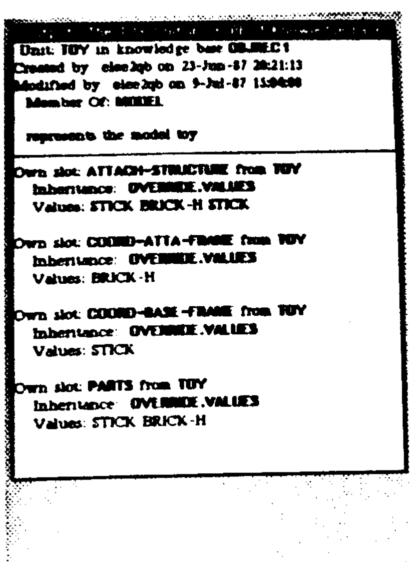


Figure -7



Unit: TOY in knowledge base OBJEC1
Created by elee2qb on 23-Jun-87 20:21:13
Modified by elee2qb on 9-Jul-87 15:04:00
Member Of: MODEL

represents the model toy

Own slot: ATTACH-STRUCTURE from TOY
Inheritance: OVERRIDE.VALUES
Values: STICK BRICK-H STICK

Own slot: COORD-ATTA-FRAME from TOY
Inheritance: OVERRIDE.VALUES
Values: BRICK-H

Own slot: COORD-BASE-FRAME from TOY
Inheritance: OVERRIDE.VALUES
Values: STICK

Own slot: PARTS from TOY
Inheritance: OVERRIDE.VALUES
Values: STICK BRICK-H

Figure-8