

OBJECT RECOGNITION USING VISION AND TOUCH

Peter Allen and Ruzena Bajcsy

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

ABSTRACT

A system is described that integrates vision and tactile sensing in a robotics environment to perform object recognition tasks. It uses multiple sensor systems (active touch and passive stereo vision) to compute three dimensional primitives that can be matched against a model data base of complex curved surface objects containing holes and cavities. The low level sensing elements provide local surface and feature matches which are constrained by relational criteria embedded in the models. Once a model has been invoked, a verification procedure establishes confidence measures for a correct recognition. The three dimensional nature of the sensed data makes the matching process more robust as does the system's ability to sense visually occluded areas with touch. The model is hierarchic in nature and allows matching at different levels to provide support or inhibition for recognition.

1. INTRODUCTION

Robotic systems are being designed and built to perform complex tasks such as object recognition, grasping, parts manipulation, inspection and measurement. In the case of object recognition, many systems have been designed that have tried to exploit a single sensing modality [1,2,3,4,5,6]. Single sensor systems are necessarily limited in their power. The approach described here to overcome the inherent limitations of a single sensing modality is to integrate multiple sensing modalities (passive stereo vision and active tactile sensing) for object recognition. The advantages of multiple sensory systems in a task like this are many. Multiple sensor systems supply redundant and complementary kinds of data that can be integrated to create a more coherent understanding of a scene. The inclusion of multiple sensing systems is becoming more apparent as research continues in distributed systems and parallel approaches to problem solving. The redundancy and support for a hypothesis that comes from more than one sensing subsystem is important in establishing confidence measures during a recognition process, just as the disagreement between two sensors will inhibit a hypothesis and point to possible sensing or reasoning error. The complementary nature of these sensors allows more powerful matching primitives to be used. The primitives that are the outcome of sensing with these complementary sensors are three dimensional in nature, providing stronger invariants and a more natural way to recognize objects which are also three dimensional in nature [7].

Most object recognition systems are model based discrimination systems that attempt to find evidence consistent with a hypothesized model and for which there is no contradictory evidence [4]. Systems that contain large amounts of information about object structure and relationships potentially reduce the

number of false recognitions. However, the model primitives must be computable from the sensed data. More complex object models are being built [8] but they are of limited power unless the sensing systems can uncover the structural primitives and relationships they contain. The approach used here allows complex and rich models of objects that extends the kinds of generic objects that can be recognized by the system. This is due to the rich nature of the surface and feature primitives the sensors compute. Surfaces are the actual parts of an object that we see; the primitive computed is exactly this. Holes and cavities are important visual and tactile features; these are also computed by integrating touch and vision. Further, the system described here is viewpoint independent. The problems caused by visual occlusion are overcome by the ability to use active touch sensing in visually occluded areas.

The domain that the system works in is one of common kitchen items; pots, pans, cups, dishes, utensils and the like. This is a rich domain and in fact contains objects representative of many other domains as well. The objects are planar as well as volumetric, contain holes and have concave and convex surfaces. They are also decomposable into separate components that have functional semantic meaning; handles are distinct geometric parts that are used for grasping, a cup's central cavity is used to hold liquids, a spout allows one to pour a liquid, a lid covers a cavity. By basing the models of these objects on geometry and topology the system is extensible beyond this domain. The objects are modeled in a hierarchical manner which allows the matching process to proceed at different levels with support or inhibition from higher or lower levels of model matching.

Figure 1 is an overview of the system. The vision system consists of a pair of stereo mounted CCD cameras. They are mounted on a 4 DOF camera frame (X, Y, pan, tilt) under computer control. The tactile system consists of a one fingered tactile sensor attached to the wrist of a PUMA 560 robot. The control module is the overall supervisor of the system. It is responsible for guiding and directing the vision and tactile sensing modules. It also communicates with the model data base during the recognition cycle as it tries to interpret the scene. It is able to use both low level reasoning about sensory data and high level reasoning about object structure to accomplish this task. Both kinds of reasoning are needed and the system's ability to toggle between the two kinds of reasoning makes it powerful. The high level reasoning allows us to use the object model as a guide for further active sensing which is accomplished by the low level sensing modules.

The recognition cycle consists of initial low level sensing which limits the number of object models consistent with the sensed data. The low level sensing elements provide data for local surface and feature matches which are constrained by relational criteria embedded in the models. The system is able to eliminate models that lack the structure uncovered by the low level sensing elements. The system then invokes an object model that is globally consistent with the sensed data and proceeds to verify this model by further active sensing. Verification is done at different levels (component, feature, surface, patch) and according to different

confidence measures at each level. The remaining sections of this paper are a detailed explanation of the various parts of the system.

2. STRUCTURE OF THE OBJECT MODELS

Objects are modeled as collections of surfaces, features and relations, organized into four distinct hierarchic levels. A hierarchic model allows us to do matching on many different levels, allowing support or inhibition for a match from lower and higher levels. It also allows us to nicely separate the low level or bottom up kinds of sensing from the top down or knowledge driven sensing.

The four levels of the model are the object level, the component/feature level, the surface level, and the patch level. Figure 2 is a partial description of the model of a coffee mug. The details of the model are described below.

2.1. OBJECT LEVEL

The top level of the hierarchy is composed of a list of all object nodes in the data base. An object node corresponds to an instance of a single rigid object. Associated with this node is a list containing a bounding box description of the object and a list of all the components (subparts) and features of this object which make up the next level of the hierarchy. For gross shape classification, a bounding box volumetric description of the object is included. The bounding box is a rectangular parallelepiped whose size is determined by the maximum extents of the object in the X, Y and Z directions of the model coordinate system. A complexity attribute is also included for each object which is a measure of the number of features and components that comprise an object.

2.2. COMPONENT/FEATURE LEVEL

The second level of the model is the component/feature level. Each object consists of a number of component (subpart) nodes that are the result of a functional *and* geometric decomposition of an object. The components of a coffee mug are the body of the mug, the bottom of the mug, and the handle. A teapot consists of a body, bottom, spout, handle and lid. They are the major subdivisions of an object, able to be recognized both geometrically and functionally. Each component also has an attribute list consisting of its bounding box, surface area, and priority. The priority field is an aid for recognition in which the components are ordered as to their likelihood of being sensed. High priorities are assigned large components or isolated components in space that protrude (handles, spouts). The protruding parts may show up as outliers from the vision analysis. Obscured components, such as a coffee mug bottom when in a normal pose, are assigned lower priorities. The priority is used to aid the matching probabilistically. If the object is in a regular pose, then certain parts of the object are more prominent which can aid the matching process. Each component node contains a list of one or more surfaces that make up this functional component and that constitute the next level of the hierarchy.

Features are entities that are useful in the recognition process. The features modeled are holes and cavities. These features are important in discrimination tasks for humans and are able to be sensed by the low level sensing. Holes are modeled as right cylinders with a defined axis, centroid and regular cross section. Each hole node contains the hole's axis vector, centroid vector, and a boundary curve that contains the cross section. This curve also encloses a two dimensional area (a "slice" through the hole). The hole node contains the inertial axes of this 2D slice computed from its central moments. By defining holes in this manner, we are treating them as a negative volumetric entity, which has implications in matching. Volumetric elements have an object centered coordinate system that contains an invariant set of orthogonal axes (inertial axes). By discovering such entities and computing these axes, transformations between model and world coordinates can be effected which is a requirement of viewpoint independent matching.

Cavities are features that are similar to holes but are not completely surrounded by surfaces. Cavities may only be entered from one direction while holes can be entered from either end along their axis. An example is the well of the coffee mug where the liquid is poured. These features are modeled as containing an axis vector, a depth, a bottom point and a boundary curve. The boundary curve is closed as in a hole, allowing for a computation of inertial axes for the cavity opening.

2.3. SURFACE LEVEL

The surface level consists of surface nodes that embody the constituent surfaces of a component of the object. The objects are modeled as collections of surfaces. The sensing elements that are used are vision and touch both of which sense surface information. Each surface contains attributes such as bounding box, surface area, a flag indicating whether the surface is closed or not and a symbolic description of the surface such as planar, cylindrical or curved. The surfaces are decomposed according to continuity constraints. Each surface is a smooth entity containing no surface discontinuities. The surfaces contain a list of the actual bicubic surface patches that comprise this surface.

2.4. PATCH LEVEL

Each surface is a smooth entity represented by a grid of bicubic spline surfaces that retain C2 continuity on the composite surface [9]. Each patch contains its parametric description as well as an attribute list for the patch. Patch attributes include, surface area, mean normal vector [10], symbolic form (planar, cylindrical, curved) and bounding box. Patches constitute the lowest local matching level in the system. The patches themselves are represented in matrix form as a matrix of coefficients for a Coons' patch.

2.5. RELATIONAL CONSTRAINTS

It is not enough to model an object as a collection of geometric attributes. One of the more powerful approaches to recognition is the ability to model relationships between object components and to successfully sense them. The relational constraints between geometric entities place strong bounds on potential matches. The matching process is in many ways a search for consistency between the sensed data and the model data. Relational consistency enforces a firm criteria that allows incorrect matches to be rejected. This is especially true when the relational criteria is based on three dimensional entities which exist in the physical scene as opposed to two dimensional projective relationships which vary with viewpoint.

In keeping with the hierarchical nature of the model, relationships exist on many levels of the model. The first level at which relational information is included is the component level. Each component contains a list of adjacent components, where adjacency is simple physical adjacency between components. The features (holes and cavities) also contain a list of the components that comprise their cross sectional boundary curves. Thus, a surface sensed near a hole will be related to it from low level sensing, and in a search for model consistency, this relationship should also hold in the model. At the surface level, again each surface contains a list of physically adjacent surfaces that can be used to constrain surface matching. The patch relations are implicit in the structure of the composite surface patch decomposition being used. Each patch is part of an ordered larger grid of knots that contains relational adjacency automatically. Thus, each patches neighbors are directly available from an inspection of the composite surfaces defining knot grid.

3. LOW LEVEL SENSING

The sensing modalities the system uses are stereo vision and tactile sensing. There are tradeoffs in speed, accuracy and noise in

using each of these sensors. The limitations of the low level sensors will constrain the accuracy of our recognition process.

The sensory data received from the two sensors needs to be integrated. In designing rules for integrating data from these two sensors, there are a few general observations. Vision is global, has high bandwidth, and is noisy. Touch is a low bandwidth, local, sequential process with better noise properties than vision. Vision gives a sometimes confusing view of an object due to the coupling of geometry, surface reflectance and lighting. Touch is better able to measure directly the properties of objects that are desired: their shape and surface properties. It also retains more degrees of freedom in sensing objects than a static camera which is limited by its viewpoint.

The most important difference between these sensors though is the *active*, controlled nature of touch versus the *passive* nature of visual sensing. To use touch it needs to be guided and supplied with high level knowledge about its task. Blind groping with a finger is an inefficient and slow way to recognize an object.

3.1. VISION SENSING

The object to be sensed is placed on a known support surface in an arbitrary position and orientation. It is assumed to be a single rigid object. The cameras have previously been calibrated with the world coordinate system of the robot arm. The stereo algorithm first uses a Marr-Hildreth operator on the images and then matches zero crossings with a stereo matcher developed by Smitley [11]. The output of this is a sparse set of 3D points that form closed contour regions.

These closed contour regions can be analyzed from a connectivity standpoint to form a region adjacency graph. This graph establishes constraints on local surface matches. Figure 2 (upper left) contains the closed contour region analysis that results after the edge finding and stereo matching processes. Stereo cannot provide information about the interiors of these regions, and the tactile system will provide this information.

3.2. TACTILE SENSING

The tactile sensor being used is a rigid finger like device that is made up of 133 pressure sensitive sites. The sites are covered by a conductive elastomer that senses pressure at each site with an eight bit gray scale. The geometry of the finger also allows limited amounts of surface normal information. The sensor is mounted on the wrist of the PUMA 560 robot and is continuously monitored by a microprocessor that is capable of thresholding and ordering the tactile responses at each of the sites. The arm is controlled by VAL-II programs that receive feedback from the tactile sensor's microprocessor. The arm can be commanded to move in an arbitrary path until the sensor reports an over threshold contact or contacts. These contact points, along with their normals are then reported back to the control module.

3.2.1. SURFACE TRACING ALGORITHMS

Algorithms have been developed to have the tactile system do surface tracing. Given a starting and ending point on a surface, the sensor traces along the surface reporting its contact positions and normals as it moves along. There are many potential paths between these two points on the surface of the object. The movement cycle of the sensor begins with contact at the starting point on the surface. The surface orientation can be determined by the location of the contact point on the sensor. The arm then moves then sensor a small distance off the surface in the direction of the surface normal. At this point a movement vector is calculated that is the weighted average of the vector to the ending point of the trace and the vector formed from previous contact points on the surface. The sensor is then moved a small distance along this vector, monitoring contact continuously. If contact occurs, the cycle repeats until the

ending point is reached. If no contact occurs, then contact with the surface is re-established by moving in the surface normal direction towards the surface. This allows the sensor to make progress towards the goal and to stay in contact with a smoothly changing surface. Using the straight line vector to the goal alone will cause cycles and no progress towards the end point of the trace.

3.2.2. HOLE TRACING

Holes are a useful recognition feature that are difficult to find using machine vision alone. A hole explorer algorithm has been developed to find and quantify a hole with the tactile sensor. Holes are modeled as cylindrical, negative volume elements with an arbitrary cross section. They have an axis (the cylindrical axis) from which they can be entered and a centroid which is defined on the cross sectional slice of the hole.

The hole tracing algorithm starts with the tactile sensor probing the region to determine if it is in fact a hole. If the tip does not come in contact with a surface, then the region is determined to be a hole. The arm is then moved in conjunction with tactile feedback around the contour of the hole, reporting the contact points of the hole's cross section.

4. INTEGRATION OF SENSORY DATA

The stereo matching from vision yields a sparse set of 3D points that form closed contours. The interior surfaces of these contours cannot be determined from vision alone. To find out the nature of these closed contour regions, the tactile sensor is used to explore and quantify the interior of these regions. Each region has a contour of 3D points obtained from stereo vision. A least squares plane can be fit to these points and the tactile sensor aligned normal to this plane, forming an approach angle for the sensor. The sensor then moves towards the plane, seeking to establish contact with a surface. If contact occurs, then the surface interpolation process described below proceeds. If no contact is found within a distance threshold of the least squares plane, a hole is hypothesized and a trace of the hole's contour as described above is begun. If surface contact is made after the distance threshold, then a cavity has been discovered, and its cross section is traced similarly to the hole.

The surface interpolation process assumes that the region inside the 3D contour discovered by vision is curvature continuous. If the region was not curvature continuous, then a zero crossing would have been seen inside the closed contour region [12]. Determining the true nature of these surfaces is the heart of the integration process. These contours can be analyzed and points of high curvature chosen as knot points for a bicubic spline interpolation process described in [13]. Once these knot points are chosen, the tactile system is actively guided in tracing the interior of the closed contour region using the surface tracing algorithms described earlier. The knots create 4 boundary curves that comprise the closed contour region. Each surface is traced from the midpoint of a boundary curve to the midpoint of the boundary curve opposite. These surface traces are then combined with the contour data to create a composite bicubic spline surface which preserves the smooth nature of the surface and interpolates the sensed data points. This interpolation can be done to arbitrary precision by tracing each surface at finer and finer resolutions. Typically, one set of traces across the surface is sufficient to obtain a reasonable interpolation of the surface. These surface patches are powerful primitives. They are described by a set of parametric equations that allow easy and efficient calculation of surface areas, normals and curvature properties of the patches which are useful for matching.

The integration of sensory data shows its complementary nature. Visual data can be used to guide active tactile sensing for determination of a visual regions properties. Further, this visual region can be extended into a three dimensional surface by the addition of small amounts of actively guided touch sensing.

5. MODEL INVOCATION

The recognition process begins with the low level vision modules in a bottom up fashion. These modules create a list of closed contour regions and a region adjacency graph. This is input to the tactile system which will either interpolate a surface patch or determine that the region is a hole or cavity. This low level sensing defines the primitives that are sent to the higher level matching system to try to instantiate a model consistent with this sensory data-

5.1. LOCAL MATCHING

Local matching is controlled by a set of rules that drive the process. The rules establish local surface matches for a sensed surface patch by finding surfaces in the model surface list that satisfy all the constraints below:

- The area of the sensed surface patch must be less than or equal to the model surface area.
- The bounding box of the sensed surface patch must be contained within the bounding box of the model surface.
- If the sensed surface patch is planar, the model surface must be planar.
- If the sensed surface patch is cylindrical, the model surface must be cylindrical and the difference of the radii of the cylinders must be small.
- If the sensed surface patch is curved, the model surface must be curved

There is also a set of local matching rules for features:

- If the feature is a hole or cavity, the diameter of the cross section must be within a threshold of the model diameter.
- If the feature is a cavity, its depth must be within a threshold of the model cavity depth.

If the local surface/feature matching is also relationally consistent, then the model is considered feasible. All local surface/feature matches must be consistent with the model relations. If a relationship exists in the scene between two matched surfaces that are not adjacent in the model, the local surface matches are rejected. If a feature is found, its related surfaces must also be consistent. By applying these relational constraints to the local matches, incorrect matches are rejected.

5.2. CHOOSING A MODEL

Once the low level initial sensing has been done, a model must be chosen to drive the recognition process to the next stage. There are three possible outcomes of the initial sensing and local matching process:

- No model is consistent with the data.
- One model is consistent with the data.
- More than one model is consistent with the data.

The first case reflects a possibility of sensory error. The most effective strategy here is to redo the sensing or change the view point to reflect new sensory input. The second case is a desirable state of affairs; this one model is then a candidate for further verification sensing. The third case points to the need for further discrimination among the models. Typically, the local matching process with its geometric and relational constraints will prune the number of consistent interpretation* to a small number of models (two or three). The ambiguity between models can be resolved at the next level of sensing. In this case, one of the two or three models left will be chosen as the invoked model to be verified. Choosing this model will rely on the object priorities discussed above and the model complexity measure. Each consistent model has a matching measure computed based upon the normalized probability of the model components and features matched and the complexity of the object's structure.

$$M = \frac{\sum_k P_k}{\sum_{i=1}^N P_i} \times N$$

where M is the matching probability measure, P_k is the priority of each of the matched components and N is the total number of component/ features for this object.

The choice is weighted in favor of complexity of the objects, and with equal complexity, on finding higher probability matches. The sensory subsystems are more likely to sense the high priority components and features. This is then reflected in the choice of model that is invoked. An incorrect choice will be found during verification, at which time this model is rejected and the remaining ones are again chosen by this rule.

6. VERIFICATION

Once a model has been chosen, the recognition process becomes a verification process, trying to substantiate the hypothesis that the chosen model is the object to be recognized. Verification proceeds initially on a component/ feature level. Some of the component/features of this model have been partially verified by the local surface/feature matching. Partial verification means that a component, possibly made up of a number of surfaces, has had at least one of these surfaces matched at the lower surface level. Any component/feature that has not been partially verified is then put on a list for verification. An important aspect of this verification is using the model to drive the recognition process. To reflect the model in the imaged scene, a transformation matrix relating model coordinates to imaged world coordinates is necessary. The calculation of this matrix will allow accurate scene prediction for features and components.

6.1. SCENE TO MODEL TRANSFORMATION

The transformation between model and scene consists of three translational parameters and three rotational parameters. This transformation can be computed from matched features or surfaces. Holes have an axis vector and inertial axes of their cross section in the model. If hole is matched, these axes can be used to align the model with the sensed hole, thus satisfying the rotational component of the transformation. The centroid of the hole can then be translated to the sensed centroid to calculate the translational parameters. Similarly, cavities are modeled as having an axis and bottom point. These can be used in the same manner as the hole axis and centroid to compute the transformation matrix.

Sensed surface data can also be used to calculate the transformation matrix. The approach is to match points on sensed surfaces with model surface points. Candidates for these matching points are points of maximum or minimum curvature or at surface border vertices. By using several match points a least square fit of the transformation matrix can be calculated [10]. Planar surfaces can also be used to define part of the transformation by aligning sensed surface normals with model normals. Once this matrix is calculated, it can be used to verify that the initial set of local matches of surfaces and features is consistent with the calculated transformation matrix. If they are not, the system returns to the local matching level and determines a new relationally consistent set of matches, recalculating the transformation matrix. If there is only one relationally consistent set of local matches, then those local matches that are inconsistent with the transformation are rejected and the others accepted.

6.2. VERIFICATION SENSING

The verification of the object model takes place at many levels. The top level is at the component/feature level. Verification at this level is done by requiring every component to be at least

partially verified or an occlusion computed from the transformation matrix which accounts for its inability to be sensed. In verifying a component, the surfaces that comprise that component must be sensed. This is done by examining the patches that comprise each surface. The center point of each patch (calculated as the parametric center from the parametric patch equations) is transformed by the matrix and projected into the camera space. If these projected points lie inside a closed contour region that has been locally matched, they are occluded. If they lie outside a closed contour region or inside an unmatched region, then they can be sensed by integrating vision and touch as described earlier. If the points are determined to be occluded, then they must be sensed by the tactile system alone. This is done by actively guiding the tactile sensor to these points and sensing the surface normal at this point, which should be consistent with the transformed mean normal for the patch. By verifying patch center points and normals, partial verification of a component is accomplished.

Holes are verified by applying the transformation matrix to the hole axis and center id and actively using the tactile system to verify the holes existence and cross sectional curve. Cavities similarly are verified by applying the transformation to the cavities axis and bottom point and again using the active tactile system to sense the cavity.

Some components will not be able to be sensed due to support surface occlusion. An example of this is the coffee mug where the bottom component of the mug cannot be sensed since it is on the support surface. The application of the computed transformation matrix to this surface will reveal that the surface is coincident with the support surface.

6.3. CONFIDENCE MEASURES

Verification can be a time consuming and lengthy process as it involves many levels of sensing. Once the component/feature level is verified, a further verification can proceed at the surface level and finally at the patch level. Rather than continue sensing all parts of the object, a confidence measure for each level is established. In the limit, by sensing all modeled surfaces and features and patches, verification will be complete. However, it will not be physically possible to verify all of these parts of the model due to occlusion and inherent limitations in the sensors themselves. Therefore, a measure needs to be established that computes the confidence of the match. One can then predetermine what confidence measures are necessary for acceptance or rejection of a model.

The models used are hierarchic in nature. Confidence measures can be set up at each level of the model allowing acceptance or rejection based on different requirements for each level. At the component/feature level, a measure of confidence is the fraction of total component/features partially matched. We can extend this idea of a partial match at each level by including threshold criteria for accepting a partial match. A partial match at the component level means that some fraction of the surface's that comprise that component are matched. A partial match at the surface level means that some fraction of the patches that comprise a surface are matched. By specifying V_i , a verification fraction at each level i of the model, hierarchical acceptance criteria will determine the amount of sensing to be done. These verification criteria can be global or used on a per object basis, implying different amounts of active sensing for different hypotheses.

One advantage to this approach is that partial matching can be carried out. Since the matching is local with global constraints, partial matches can be made and reported even though a global match is rejected. A further extension would be to articulated parts, where local transformations between components and surfaces would have to be accounted for in the global constraints.

7. CONCLUSION

The system described is currently being tested. The experimental hardware systems, model data base and surface and feature

matching routines are built. Work is presently continuing on the higher level reasoning modules. As the number of objects in the database grows, more sophisticated access mechanisms for indexing into the models will be needed.

8. ACKNOWLEDGEMENTS

This work was supported in part by the following grants: ARO DAA6-29-84-k-0061, AFOSR 82-NM-299, NSF MCS-t219196-CER, NSF MCS 82-07294, AVRO DAAB07 84-K F077, and NIH 1-RO1-HL-29985-01.

References

1. Nevatia, R. and Binford, T., "Description and recognition of curved objects," *Artificial Intelligence*, vol. 8, pp. 77-98, 1977.
2. Brooks, Rodney, "Symbolic reasoning among 3-D models and 2-D images," *Artificial Intelligence*, vol. 17, pp. 285-349, 1981.
3. Shapiro, Linda, Moriarty, J.D., Haralick, R., and Mulgaonkar, P., "Matching three dimensional models," *Proc. of IEEE conference on pattern recognition and image processing*, pp. 534-541, Dallas, TX, August 1981.
4. Fisher, R.B., "Using surfaces and object models to recognize partially obscured objects," *Proc. IJCAI 83*, pp. 989-995, Karlsruhe, August 1983.
5. Grimson, W.E.L. and Lozano-Perez, Tomas, "Model based recognition and localization from sparse three dimensional sensory data," *A.I. memo 738*, MIT AI Laboratory, Cambridge, MA, August 1983.
6. Tomita, Fumiaki and Kanade, Takeo, "A 3D vision system: Generating and matching shape descriptions in range images," *IEEE conference on Artificial Intelligence Applications*, pp. 186-191, Denver, CO, December 5-7, 1984.
7. Binford, T., "Survey of model based image analysis," *Int. Journal of Robotics Research*, vol. 1, no. 1, pp. 18-64, Spring 1982.
8. Shapiro, Linda and Haralick, Robert, "A hierarchical relational model for automated inspection tasks," *Proc. 1st International Conference on Robotics Research*, pp. 70-77, Atlanta, GA, March 13-15, 1984.
9. Faux, I.D. and Pratt, M.J., *Computational geometry for design and manufacture*, John Wiley, New York, 1979.
10. Potmesil, Michael, "Generating three dimensional surface models of solid objects from multiple projections," IPL technical report 033, Image Processing Laboratory, RPI, Rensselaer, NY, October 1982.
11. Smitley, David and Bajcsy, Ruzena, "Stereo Processing of Aerial Images," *International Conference on Pattern Recognition*, Montreal, August 1984.
12. Grimson, W.E.L., *From images to surfaces: A computational study of the human early visual system*, MIT Press, Cambridge, MA, 1981.
13. Allen, Peter, "Surface descriptions from vision and touch," *Proc. 1st International Conference on Robotics Research*, pp. 394-397, Atlanta, March 1984.

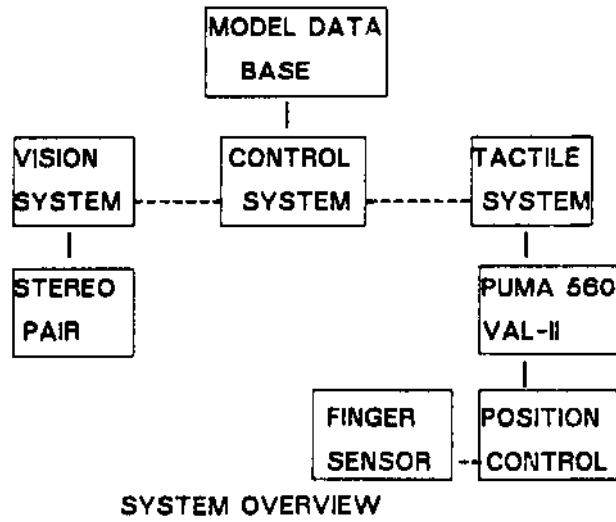
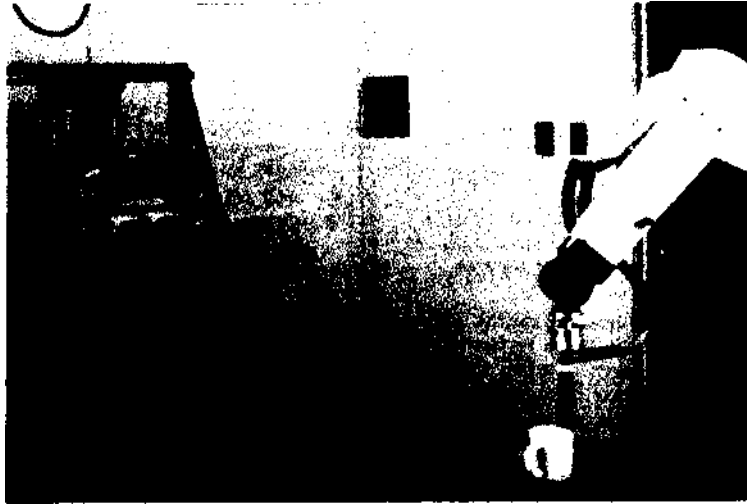


FIGURE 1: SYSTEM OVERVIEW

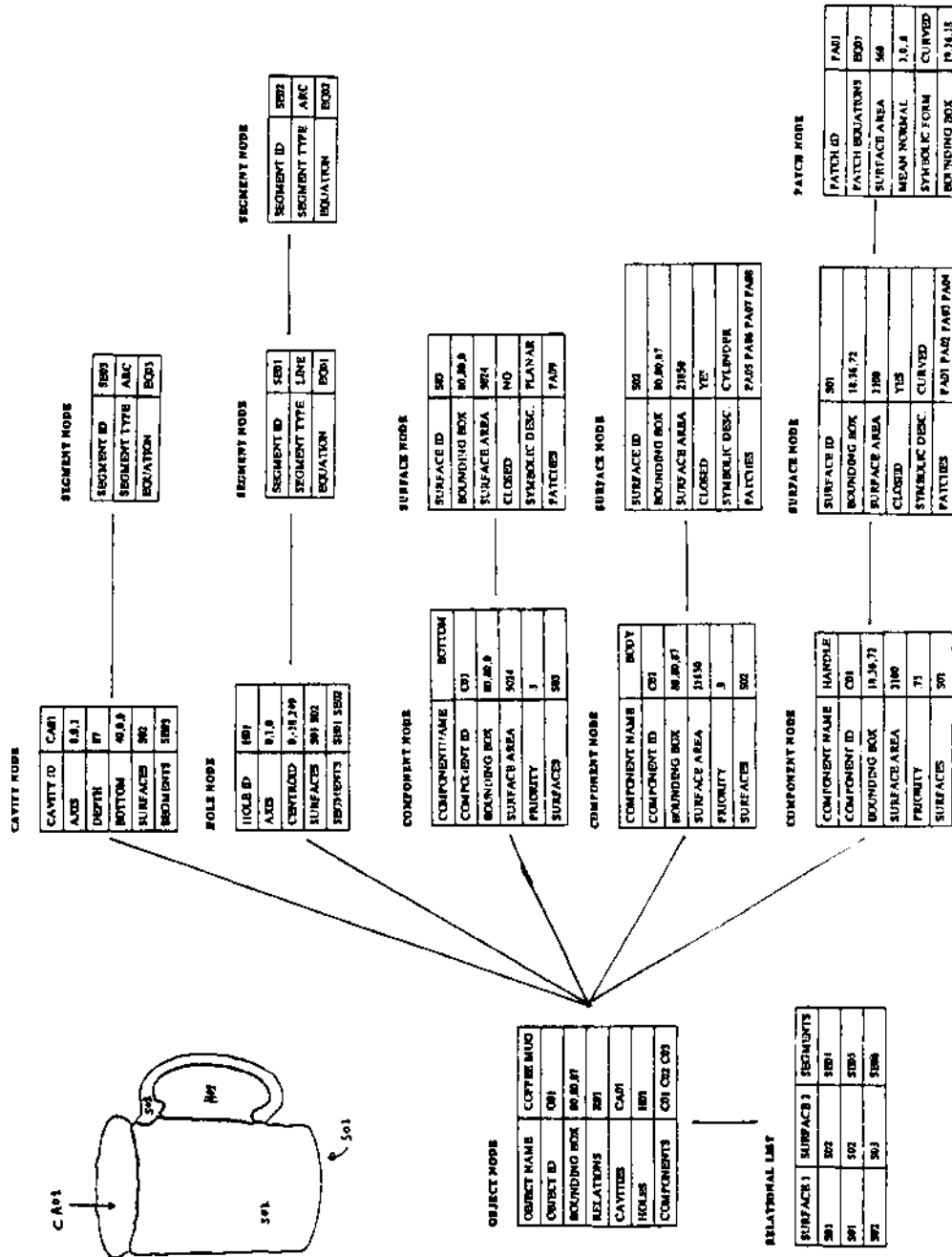


FIGURE 2: OBJECT MODEL OF COFFEE MUG