

MAPPING HAPTIC EXPLORATORY PROCEDURES TO MULTIPLE SHAPE REPRESENTATIONS

Peter K. Allen

Department of Computer Science
Columbia University
New York, New York 10027

Abstract

Research in human haptics has revealed a number of Exploratory Procedures (EP's) that are used in determining attributes of an object, particularly shape. We have used this research as a paradigm for building an intelligent robotic system that can perform shape recognition from touch sensing. In particular, we have found a number of mappings between EP's and shape modeling primitives. In this paper, we discuss the choice of shape primitive for each EP, and present results from experiments with a Utah-MIT dextrous hand system. We also present a vision algorithm to complement active touch sensing for the task of autonomous shape recovery.

1. INTRODUCTION

A number of interesting properties of the human haptic system have been investigated by Lederman and Klatzky and their colleagues [16-18]. This work has shown that an important component of the haptic system is its ability to recognize attributes of three-dimensional objects quickly and accurately. Among these attributes are global shape, hardness, temperature, weight, size, articulation and function. An outcome of this research is the identification of hand movement strategies that are used by humans in discovering different attributes of three-dimensional objects. They have labeled these EP's, or Exploratory Procedures, and have reported success rates of 96-99% in identifying different object properties using two-handed, haptic exploration. Given these well-developed and useful human skills, we have found it natural to create analogs of these human capabilities in our robotic domain. Future robotic systems will need to use dextrous robotic hands for tasks such as grasping, manipulation, assembly, inspection and object recognition.

2. MAPPING HAPTIC EXPLORATORY PROCEDURES TO SHAPE REPRESENTATIONS

Our research is focused on recovering shape attributes of rigid 3-D objects using active touch sensing. The question of what shape representation to use is an important open question in robotics. While we do not have a definitive answer to propose here, it seems that a key component of any working system is to use *multiple* shape representations. Different shape representations seem to be intuitively better based upon both the task and object domain. Our effort in implementing robotic EP's has led us to find "natural" mappings between some well-known shape representations and EP's. In the sections below, we will describe the 3 EP's we have implemented on our robotic hand system, and motivate our choice of shape representation. Experiments have been carried out using our hand system to recover shape using the EP's discussed below. Our hand system is described in detail in [2]. It consists of a Utah-MIT hand mounted on a PUMA 560 arm and Interlink tactile sensors mounted on the distal links of each finger (figure 1).

2.1. Mapping 1: Grasping by Containment → Superquadrics

Consider the EP of grasping by containment. In humans, this EP is based on encompassing an object's volume with the hands, deriving sparse but global shape information. We have chosen superquadrics as a representation for recovering shape with this EP. They have been used by a number of researchers [4, 5, 19] to recover shape. What makes superquadrics particularly relevant for this EP is the following:

- The representation is volumetric in nature, which maps directly into the psychophysical perception processes suggested by grasping by containment.
- The models can be constrained by the volumetric constraint implied by the joint positions on each finger.
- The representation can be recovered with sparse amounts of point contact data since only a limited number of parameters need to be recovered. There are 5 parameters related to shape and 6 related to position and orientation in space. Global deformations (tapering, bending) add a few more.
- In addition to the use of contact points of fingers on a surface, the surface normals from contacts can be used to describe a dual superquadric which has the same analytical properties as the model itself.
- The recovery process uses a non-linear least-squares estimate of a fit function. This approach is especially relevant with touch sensing, in which there is evidence that the human tactile system serves essentially as a low-pass filter[18].

Our implementation of this EP, described in detail in [3], is to obtain a number (typically 30-100) of finger contact points by encompassing the fingers of the hand around the object. The data is from all sides of an object, giving justification to the least-square fit of the data from the entire object. Figure 2 shows a number of objects whose shape has been recovered using the Grasping by Containment EP.

The problem of generating a good initial hypothesis is central to robust object recognition. If we can generate a good initial shape estimate, then we will be much more successful as we try to discover and refine further object structure. The requirements for an initial shape estimator are that it be efficient, stable in the presence of noise and uncertainty, and able to use sparse, partial data. Our recovery of global shape using superquadrics for an initial estimator of shape meets these criteria.

2.2. Mapping 2: Lateral Extent → Face-Edge-Vertex

The second EP we have implemented with our robotic hand system is a Lateral Extent EP. This EP is used to explore a continuous, homogeneous surface such as a planar face, and to determine its extents. Research with human subjects suggests that this strategy is used until a discontinuity is found, which can cause a change in haptic sensing strategy.

We have found that this EP maps into a winged-edge type of Face-Edge-Vertex model suggested by Baumgart [6]. This EP is capable of determining the extents of a planar surface, and by using multiple fingers, different planar faces can be explored. The intersections of these planes form the edges and vertices of the object in question, and are easily computed.

Our implementation of this EP uses the hand's index finger. While the index finger is held in an extended position, the PUMA arm is moved until the tactile sensors on the index finger contact a surface (if no contact is detected, the procedure terminates). After the initial contact, the Cartesian position of the contact point is noted. The hand and arm then begin an iterative search for the boundaries of the surface by performing the following sequence: (a) lift the finger off the surface until tactile contact is lost; (b) move the arm in a direction parallel to the surface; (c) if the finger is in contact after the movement, note the new contact location, otherwise lower the index finger until it makes contact with the surface again; (d) repeat steps (a)-(c) until the finger fails to make contact in step (c). In step (d), if the finger does not contact the surface, then either the finger has moved beyond the edge of the surface, or the surface is too far away from the finger to be detected. To check for the latter case, the arm must be moved toward the surface. After completing the first collection of data points and finding the edge of the surface, the index finger is moved back to the position of initial contact, and a second mapping of the surface is undertaken in a direction 180° opposite. This procedure continues until a second surface edge is detected. The search now continues as before but in a direction perpendicular to the first two traces. This procedure then is able to map out a set of contact points on the surface, describing its extent. Each time the fingertip contacts the surface, the Cartesian coordinates of the contact are retained. Figure 3 shows a pattern of traces on 2 adjacent planar faces of a rectangular block using this EP. Least-square planes were fit to each of the traces and the computed angle between the recovered planes is 96° (the actual angle is unknown but assumed to be 90°). The intersection of the planes forms an edge which can be used to build the winged-edge model of the object.

2.3. Mapping 3: Contour Follower → Generalized Cylinders

The Contour Follower EP in humans is described by Klatzky and Lederman as "a dynamic procedure in which the hand maintains contact with a contour of the object. Typically, the movement is smooth and non-repetitive within a segment of object contour, stops or shifts direction when a contour segment ends, and does not occur on a homogeneous surface [15]." It seems natural that this EP will report information that can be used to recover a shape that can be represented as a class of generalized cylinders. Generalized cylinders have been proposed by many researchers beginning with Binford [7] as a shape modeling primitive. Other researchers have expanded on this idea of a swept volume by creating classes of generalized cylinders or cones, depending upon the nature of the axis curve, sweeping rule and cross sectional curve. These primitives have special appeal in the recognition of elongated objects and objects that provide strong visual contours.

Fearing [10] has attempted to recover the shape of a class of these generalized cones (RLSHGC - Right Linear Straight Homogeneous Generalized Cones) using extremely sparse amounts of data. He has characterized the necessary and sufficient conditions for being able to recover the axis and orientation of these cones, given limited, multi-fingered tactile sensor data that includes point contacts, surface normal direction, and surface curvature information.

We have also chosen to use a class of these primitives for shape recovery. The class we are using is surfaces of revolution, which are RSHGC with a circular cross section function (no linear scaling of the cross section is required). These surfaces may be completely described by the rotation of a plane curve about the axis of symmetry (see figure 4). If we take this axis to be the Z axis, the surface will intersect the O-XZ plane in the plane curve

$$r_c(u) = p(u) \mathbf{i} + z(u) \mathbf{k} \quad (1)$$

and the surface of revolution has the equation

$$\mathbf{r} = p(u) \cos \theta \mathbf{i} + p(u) \sin \theta \mathbf{j} + z(u) \mathbf{k} \quad (2)$$

From the definition, we can see that the cross-section curve is circular, but the contour generator curve can be quite arbitrary.

Fearing was able to recover the shape of generalized cylinders from a minimal set of tactile contacts that were static in nature, not dynamic. Our attempt to recover the shape of these objects from touch is based upon receiving less accurate tactile information than the system used by Fearing. In particular, our tactile sensors are not capable of reporting accurate surface normals or surface curvatures; but they are capable of producing localized 3-D positional contacts of fingers to the object. Our experience is that information such as surface curvature is very difficult to accurately sense, particularly with sensors that have low dynamic range and are used in an active, exploratory manner. Our approach is to use the Contour Follower to recover the contour curve described above. If we obtain two such contour curves that are on either side of the object, we can estimate the axis of the surface of revolution, and recover the shape. This procedure maps naturally into a two-fingered Contour Follower EP, in which an object's contour on either side are sensed using the thumb and index finger.

The problem of using an active tactile device to trace a surface on an object is a complicated one. Our previous work [1] using a one-fingered tactile sensor mounted on a PUMA traced along a curved surface by calculating a weighted vector of constraint directions that tried to follow the surface curvature while preserving smoothness of the trace and a constraint having to do with creating regions bounded by traces that were equivalent in size. Hor [13] traced contours of planar objects using a planar four-fingered "chopstick" like manipulator. Strain gauge sensors on the fingers of this device would calculate surface normals and move tangentially along a surface, recording the contour. Stansfield [22] used a planar LORD tactile sensor mounted on a PUMA to trace edges and other features on objects.

Our Contour Follower EP is now described. First, the PUMA is moved to a location near one end of the explored object, and the thumb and index finger are opened enough to allow them to encompass the object without making contact with it. Then the thumb is slowly moved toward the object until the sensors detect contact between the thumb and the object. Next, the index finger follows the same movement. After detecting contact, the positions of the two contact locations are noted, and the fingers are backed off the object so that they are no longer in contact. The arm and hand are moved a small amount along the axis of the explored object, and the process is repeated. This exploratory procedure ends when one of the fingers moves toward the object and fails to make contact. (The location of the object and its axis are not currently determined automatically, but with human aid.)

The detection of contact and conversion to Cartesian coordinates is a process that requires several steps. The fingers are moved toward the object in a number of discrete intervals. After each movement, two checks are performed. First, did the tactile sensor detect contact? And second, did the finger move the entire distance that was commanded? If the tactile sensor detects contact, then the location of the center of the contact region is found. To find the center of the contact, the first moments of the tactile array are taken. Then a transformation is performed from the fingertip coordinate frame to the hand coordinate frame, and finally, from the hand coordinate frame to world coordinates. The second check is to see if the finger has moved the entire distance commanded (and there is no tactile contact). This event would signal that something is impeding a finger from moving. In this case, no centroid of the contact region is found and the data point is thrown out. After detecting contact that does not involve the tactile sensor, the exploratory procedure continues looking for valid contact points along the original search axis.

We have performed a series of experiments that try to recover the shape of a number of different surfaces of revolution including a wine bottle, a beer bottle, a coke bottle and an Orangina bottle (a flask like object). The procedure begins with exploring the object along an exploration axis that is assumed to be perpendicular to the support table (but can be inferred from vision sensing described below). The points generated from these contour traces are then linked into a set of linear contour segments. Circular cross section curves are then fit perpendicular to the exploration axis that include trace points from each of the contours. The recovered shapes are shown in figure 0. The shapes are clearly distinguishable from this sparse data. An additional discriminating characteristic is actual 3-D size and volume which are calculable from these representations.

3. REQUIREMENTS FOR AN AUTONOMOUS HAPTIC SYSTEM

The EP's discussed above can be thought of as a set of primitive haptic functions that can be used as the building blocks for an active, autonomous haptic recognition system. The requirements of such a system are more complex than a similar vision based recognition system, primarily due to the active control needed for the hand. We now discuss how more autonomy may be built into such a system using vision as an intelligent front end to the haptic EP's.

3.1. Using Multiple Sensing Modalities for Autonomy

While the focus of this paper is on the acquisition and interpretation of touch sensor data, our overall approach to the problem of robotic object recognition lies in a multi-sensor approach; we believe no single sensing modality is currently powerful enough to robustly perceive and recognize its environment. Just as humans exploit a multitude of sensor systems, robotic systems need to use multiple sensors for perception as outlined in Allen [1] and Kak and Chen [14]. A central idea in using multi-sensor data is that over-reliance on one sensor can cause error. It has been empirically observed that trying to extract too much information from a single sensing modality results in a degradation of results; however, using only the most reliable and highest confidence sensor data allows one to proceed along a path that is known to be correct. We call this principle "less is more," in that reduced amounts of reliable data from a single sensor are more useful than large amounts of data which may be spurious. By combining the data that is most reliable from each of a number of sensors, more accurate results may be computed. An interesting analogy to our work is in the ability of blind people to perceive the world. While blind people can and do function in complicated environments, small amounts of other sensory data (such as verbal visual guidance cues from sighted people) can extend their ability to perceive.

3.2. Example: Finding an Exploration Axis for an EP

Determining the exploration axis is a key part of the Contour Follower EP. Knowing in which direction to trace the object is important to higher level recovery procedures which need to use this information in the recognition process. Once the hand makes contact with the object, it explores the contour along a known axis which we calculate a priori. We have implemented a vision based technique to determine this axis. Our method of visual recovery of the exploration axis exploits the recent work of Wolff [23] in stereo line matching. Point-based stereo techniques tend to be unreliable in that multiple correspondences between images can cause mismatches and error. *More stable matching can occur using larger primitives such as lines [12].* Even using line-based matching, problems can still occur. Matching the endpoints of lines can be prone to errors in the output of the line finder which may break a single line into multiple segments due to differing edge strengths along the line. The problem here is that 3-D depth is being computed, which requires an absolute correspondence of points (whether from point-based or line-based methods).

Our method alleviates this dependence on absolute matching of unstable primitives to generate 3-D depth. All we require of the algorithm is an *orientation* vector in 3-D. We do not need to have its absolute depth, but need to generate a match between a family of parallel lines sharing the same orientation. This orientation can then be used by the active hand as the exploration axis. The 3-D depth has already been determined from the contact of the hand with the object. Given this 3-D depth from tactile contact, we can follow the 3-D axis determined by the line based stereo matcher to continue our exploration.

It is important to note that this method is less sensitive to matching errors and baseline measurement, another common cause of stereo error. In addition, it is also less prone to the effects of physical point mismatches as the baseline increases, since we are still matching a larger entity, the line itself. Intuitively, the method creates a 3-D plane in space from the camera center and any two points on the line. This plane and a similar plane from the other camera are all that are needed to create a 3-D intersection line which we can use as the exploration axis.

3.3. Vision Module

The vision module consists of a pair of cameras that image the object in the scene. To extract linear features from the images, an algorithm developed by Singh and Shneier is used [21]. This is a two stage process that uses the real-time edge detection capability of the PIPE image processing system to apply smoothing, gradient, threshold and thinning operators to the image in real-time. The PIPE allows images to have local operators work on multiple images in time as well as pipelining images from processing stage to processing stage for sequential processing of a multi-step image algorithm. Each processing step takes one field time (1/60 sec). The initial operator is a first derivative of a Gaussian smoothed image, which is used as an edge strength measure. This image is then thresholded using a local averaging technique over a 5x5 window to isolate edges. This procedure, while able to be realized in real-time, has the unwanted side effect of isolated "spot" edges. These can be eliminated by sending the edge image to another processing stage in the PIPE where a morphological thinning operator can be applied, again in one image field time. The thinning is accomplished by two operators, one for removing spots and another restoring template to merge multiple parallel edges.

To create full linear features, a non-local edge following algorithm is needed. This is done by shipping the edge marked binary image to a host over a high speed interface. The edges are linked into linear features by using a raster scan connected component analysis. The connected edge pixels are broken into linear pieces using a recursive subdivision algorithm. These linear-connected edge pixels are then fit to a least-square fit line that minimizes the sum of the pixels distances from the line. The lines end-points are calculated by projecting the ends of the connected pixel chains onto the fitted line. Attributes of each line (orientation, extent, length) are also calculated.

The vision module performs the algorithm for each of the two images. The cameras are calibrated but no attempt is made at scan line registration. The intent is that the linear features are much more stable as candidates for stereo matching than pixel level data.

3.4. Matching

The linear feature extraction process serves as input to the matcher. The matching process is simplified by filtering the candidate lines on a number of criteria. The initial filtering criteria is edge length. All edges are rejected below a minimum length, which are taken from the statistics of the line feature extraction process. The edges remaining in each image are the longest and most stable lines in the image. The preference for line orientation is again taken from the distribution of the filtered lines, with the maximum of the orientation histogram chosen (an appropriate bucket size of 30° is used). The lines are then matched according to orientation and extent, with a horizontal disparity window imposed to prevent gross mismatches. The procedure works well since it is matching a very small subset of lines, which, because of their length, are quite stable. Figure 6 shows the left and right stereo images of the beer bottle with the extracted linear features overlaid on the bottle, and the matched filtered lines.

As mentioned previously, once the lines are matched, it is still not possible to fully recover depth of the object points, since the matches are not on a pixel by pixel basis. There is still a degree of freedom (translation along the stereo lines matched) inherent in the recovered line match. However, given our problem of finding a 3-D exploration axis on the object, this degree of freedom is acceptable. In fact, it is exactly the axis of exploration, along the matched line. The hand system can locate any point along the line where the object is imaged, move until it contacts this point, and begin exploring along the axis.

4. DISCUSSION: REASONING WITH MULTIPLE SHAPE REPRESENTATIONS

We have described a number of EP's that we have implemented, and the accompanying shape representation that each EP maps into. The next step in the object recognition process is to use these EP's in a coordinated manner to fully explore a complex, multi-part object. Each representation is essentially a constraint system that can be used to interpret the scene. In addition, the active nature of the sensing can allow one set of constraints

developed by an EP to drive further sensing of the object using a different EP. If we begin with the Grasping by Containment EP, which fits a superquadric to the initial grasp data, we have a strong hypothesis about an object's shape. Of particular importance are the two shape parameters in the superquadric representation. The shape of an object can be inferred from these parameters and used to direct further exploration. For example, if the shape parameters appear to be rectangular then the Lateral Extent EP can trace out the planar surfaces and perform a least square fit of the trace data to test the surface's planarity. If the shape parameters appear more cylindrical then the planar faces of the cylinder can be explored with this primitive, and the cylinder's contour can be explored and verified with the Contour Follower EP. Data from the Lateral Extent EP can be used in a matching algorithm that prunes possible interpretations of model object from this sensor data that has been developed by Roberts [20], using constraints similar to those used by Grimson and Lozano-Perez [11] and Ellis et al. [8]. These constraints have proven to be sufficient for recognition of polyhedral objects. Finally, the large body of work with generalized cones can be used for interpretation with the Contour Follower EP, providing a rich and varied spatial reasoning base.

5. ACKNOWLEDGEMENTS

This work was supported in part by DARPA contract N00039-84-C-0165, NSF grants DMC-86-05065, DCI-86-08845, CCR-86-12709, IRI-86-57151, and IRI-88-13119, North American Philips Laboratories, Siemens Corporation and Rockwell Inc. Thanks to Ken Roberts and Paul Michelman for helping to implement the EP's on our hand system.

References

- Allen, Peter, "Integrating vision and touch for object recognition tasks," *International Journal of Robotics Research*, vol. 7, no. 6, pp. 15-32, 1988.
- Allen, Peter, Paul Michelman, and Kenneth S. Roberts, "An integrated system for dextrous manipulation," *IEEE Conference on Robotics and Automation*, pp. 612-617, Scottsdale, AZ, May 15-19, 1989.
- Allen, Peter and Kenneth S. Roberts, "Haptic object recognition using a multi-fingered dextrous hand," *IEEE Conference on Robotics and Automation*, pp. 342-347, Scottsdale, AZ, May 15-19, 1989.
- Bajcsy, Ruzena and Franc Solina, "Three dimensional object representation revisited," *Proceedings International Conference on Computer Vision*, London, June 1987.
- Barr, Alan, "Superquadrics and angle preserving transformations," *IEEE Computer Graphics and Applications*, vol. 1, pp. 11-23, 1981.
- Baumgart, Bruce, "Winged edge polyhedron representation," *Stanford A.I. memo 179*, p. Stanford University, Stanford, 1972.
- Binford, T., "Visual perception by computer," *Proc. IEEE Conference on Systems and Control*, Miami, 1971.
- Ellis, R., Edward Riseman, and A. R. Hanson, "Tactile recognition by probing: Identifying a polygon on a plane," *Proceedings of AAAI-86*, pp. 632-637, Philadelphia, August 11-15, 1986.
- Faux, I. D. and M. J. Pratt, *Computational geometry for design and manufacture*, John Wiley, New York, 1979.
- Fearing, Ronald, "Tactile sensing for dextrous manipulation," *Proc. IEEE Workshop on Dextrous Robotic Hands*, Philadelphia, April 1988.
- Grimson, W. E. L. and Tomas Lozano-Perez, "Model based recognition and localization from sparse three dimensional sensory data," A.I. memo 738, M.I.T. A.I. Laboratory, Cambridge, August 1983.
- Henriksen, Knud, "Line-based stereo matching," MS-CIS-87-52, Grasp Lab 109, Department of Computer and Information Science, University of Pennsylvania, Philadelphia.
- Hor, Maw-Kae, "Control and task planning for a four finger dextrous manipulator," Ph.D Thesis, Courant Institute, New York University, October 1987.
- Kak, A. and S. Chen editors, *Proceedings of the AAAI Workshop on Spatial-Reasoning and Multisensor Integration*, Morgan Kauffman, Los Altos, CA, Oct. 1987.
- Klatzky, Roberta and Susan Lederman, "The intelligent hand," in *The psychology of learning and motivation*, ed. G. Bower, vol. 21, Academic Press.
- Klatzky, Roberta, Susan Lederman, and Victoria Metzger, "Identifying objects by touch: An "expert system"," *Perception and Psychophysics*, vol. 37, pp. 299-302.
- Lederman, Susan and Roberta Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, pp. 342-368, 1987.
- Lederman, S. J. and R. A. Browse, "The physiology and psychophysics of touch," *Sensors and sensory systems for advanced robots*, NATO ASI series, Springer-Verlag.
- Pentland, Alex P., "Recognition by parts," Technical Report 406, SRI International, December 16, 1986.
- Roberts, Kenneth, "3D object recognition by active touch sensing," Department of Computer Science Technical Report, Columbia University, New York, August 1989.
- Singh, Ajit and M. O. Shneier, E. W. Kent, M. O. Shneier, and R. Lumia, "PIPE: Pipelined image processing engine," *Journal of Parallel and Distributed Computing*, no. 2, pp. 50-78, North American Philips Laboratories, Briarcliff Manor, NY, 1985.
- Stansfield, Sharon, "Visually-guided haptic object recognition," Ph.D. dissertation, Department of Computer and Information Science, University of Pennsylvania, October 1987.
- Wolff, Lawrence B., "Measuring the orientation of lines and surfaces using translation invariant stereo," *SPIE Conference on Sensor Fusion*, vol. 1003, Cambridge, MA, November 1988.

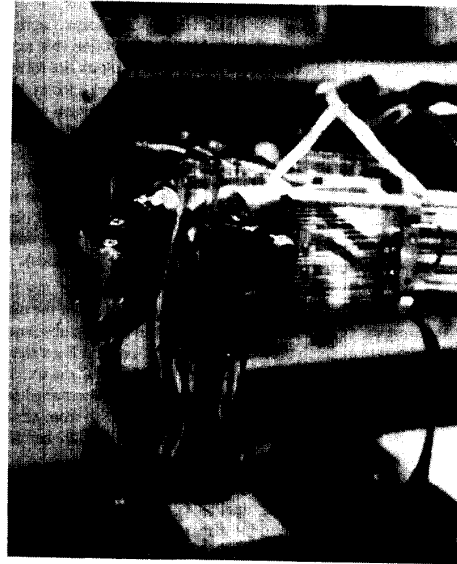


Figure 1: Utah-MIT Hand system with tactile sensors.

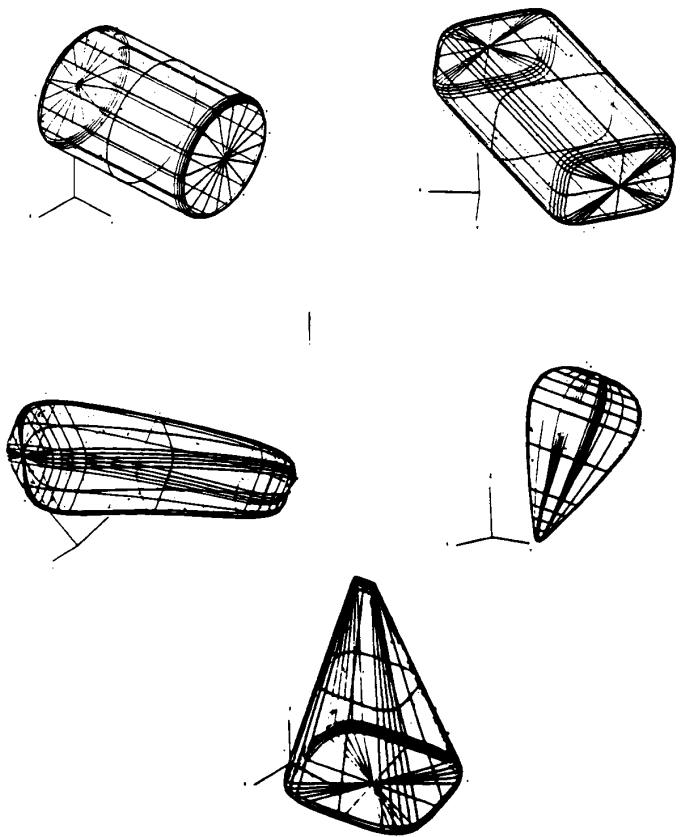


Figure 2: Recovered shape of cylinder, block, wedge, lightbulb and funnel from Grasping by Containment EP.

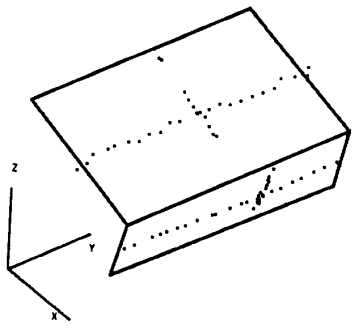


Figure 3: Lateral Extent EP contact points and derived Face-Edge model.

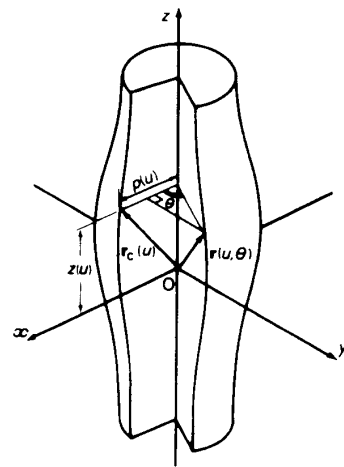


Figure 4: Surface of revolution (from Faux and Pratt [9]).

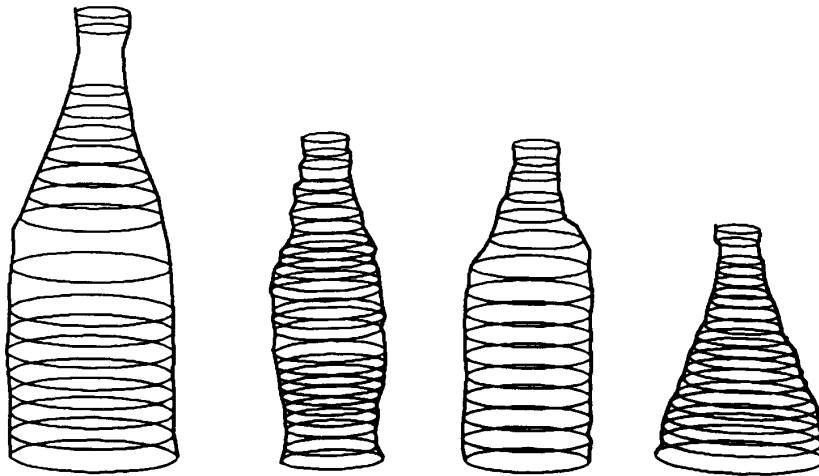


Figure 5: Linked contour points and recovered surfaces of revolution from Contour Explorer EP (left to right wine bottle, coke bottle, beer bottle, Orangina bottle).

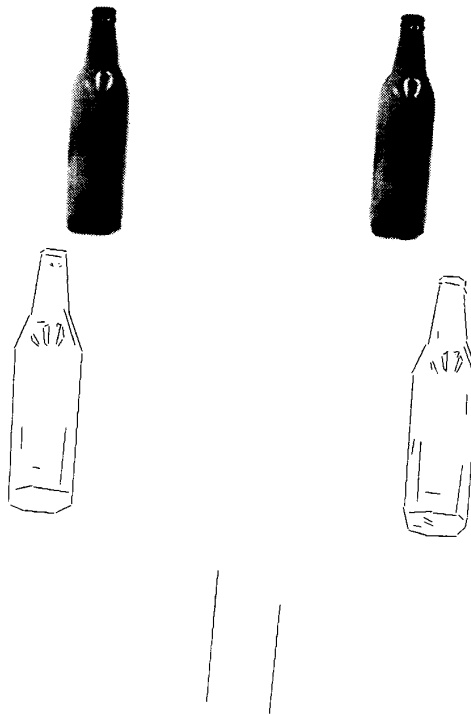


Figure 6: a) Image of beer bottle b) linear features c) matched features