

Reasoning About Grasping

S. A. Stansfield

Sandia National Laboratories
Albuquerque, N. M. 87185

Abstract

The promise of robots for the future is that of intelligent, autonomous machines functioning in a variety of tasks and situations. If this promise is to be met, then it is vital that robots be capable of grasping and manipulating a wide range of objects in the execution of highly variable tasks. A current model of human grasping divides the grasp into two stages, a precontact stage and a postcontact stage. In this paper, we present a rule-based reasoning system and an object representation paradigm for a robotic system which utilizes this model to reason about grasping during the precontact stage. Sensed object features and their spatial relations are used to invoke a set of hand preshapes and reach parameters for the robot arm/hand. The system has been implemented in PROLOG and results are presented to illustrate how the system functions.

I. Introduction

Most robots today are little more than advanced automation, limited to simple, repetitive tasks in highly structured environments. But the promise of robotics for the future is that of autonomous machines capable of carrying out complex tasks in unknown or partially specified worlds – robotic explorers, for example, functioning in space and beneath the sea, or intelligent machines capable of entering hazardous environments to perform rescue and clean-up operations. Such robots will need not only to reason about their environment, but also to act upon it in intelligent and flexible ways.

For this reason, the design of general purpose hands and grasping algorithms has begun to receive attention. Not only is grasping of vital importance to our goal of intelligent, autonomous robots, but it will also find utility in such near-term industrial applications as flexible manufacturing and dextrous assembly. Ultimately, it will provide the robot with a mechanism for interfacing with and learning about its world, independent of human programmers.

Previous research into robotic grasping has been predominantly analytical. Unfortunately, while such research provides us with useful insights into the mechanisms of grasping, it is often difficult to apply to “real world” situations. This is due to the fact that the analysis itself often becomes unwieldy when too many factors are taken into

consideration. Therefore, simplifying assumptions about the world must be made. For example, grasps are often assumed to be planar. They are often modeled as individual contacts on an object, sometimes with the further constraint that the contacts be point only, or that there be no friction involved in the grasp. Unknowns and uncertainties in the environment are usually not taken into account. There are no mechanisms within these models for automatically integrating information about the task or the perceived state of the world. And finally, the difficulty in controlling a multiple-degree-of-freedom robot hand capable of providing such contacts – and the inherent inaccuracies of such a device – are usually not considered.

Recently, psychologists and cognitive scientists have also become interested in the grasping problem. Their observations on human grasping provide a number of potentially useful insights for the researcher in robotic grasping. For example, it has been noted that humans tend to use a predetermined set of grasp configurations in the initial stage of a grasp [Jeannerod78, Lyons85]. Often, several fingers are coupled, reducing the degrees of freedom within the system [Iberall87]. High-level knowledge about the task, the object to be grasped, and the perceived state of the world affect grasp choice and execution [Arbib83, Cutkosky87, Klatzky86]. And finally, perceptual information is utilized in all stages of the grasp. What all of this offers to robotic grasping is a mechanism for simplifying the synthesis and control of a grasp, while making fewer limiting assumptions about the world in general.

This research addresses the design and implementation of a robotic system that incorporates these ideas. The goal is robotic grasping of generic, or basic level, objects. If our robot is familiar with the object “screwdriver,” for example, then we would like it to be able to grasp *any* screwdriver which it may encounter during the execution of some task – even a screwdriver which it may not have previously encountered. To accomplish this, we need not only to “partition” the grasp into a set of achievable subtasks, but also to consider fundamental questions concerning the synergism of motor-interaction and high-level knowledge as mediated by perception. Theories of human grasping and manipulation offer guidance in both the former and the latter, and we propose to use these as a foundation for our design.

This paper presents a rule-based system to be used by

a robot to reason about grasping in the precontact stage. The object representation paradigm upon which this rule-base operates is also described. Both the reasoning and the representation have been implemented in PROLOG, and we present some examples as a way of illustrating the operation of the system. Finally, we discuss briefly how the system will be extended in the future.

II. A Two-stage Model of Grasping

One model of human grasping proposed by Jeannerod [Jeannerod78], Arbib, et al. [Arbib83], and Tomovic, et al. [Tomovic87], divides the grasp into two stages. In the initial, precontact stage, visual information and high-level knowledge are used in a feedforward manner to drive hand preshaping, hand/wrist orientation, and a ballistic reach toward the object. In the second stage of the grasp, haptic (tactile plus kinesthetic) information is used in a feedback manner to adjust and fine-tune the grasp. Both Arbib and Tomovic hypothesize complex, high-level control mechanisms involving both perceptual and cognitive functions, as well as low-level "reflex" actions.

An approach to robotic grasping would implement this model in the following way: In the precontact stage of the grasp, a vision system (or a combination of vision and exploratory touch) would be used to extract object features relevant to the task of grasping. These features would include location of the object in space, segmentation of the object into components, and the determination of the relations among these components. High-level reasoning would then utilize this information, as well as knowledge about the task and the perceived state of the world, to select an appropriate hand preshape and to drive the reach toward the target. In the postcontact stage, the hand would make contact with the object and relevant haptic features would be extracted and used to fine-tune the grasp.

This paper concerns itself with the precontact stage. We are interested primarily in the following question: What is the proper knowledge-base and object representation paradigm to allow the system to generate the appropriate set of grasp preshapes and grasp/reach parameters for the given object? We do not currently address the issue of what the optimal set of hand preshapes is. Lyons [Lyons85] has done such an analysis. He proposes a set of hand configurations useful for grasping, and the three which we have chosen to implement are similar to those proposed by him.

The research environment in which this work is being carried out consists of a Salisbury robot hand under position and force control; a PUMA 560 robot arm upon which the hand is mounted; a vision system; and a Symbolics 3645 running PROLOG to provide the high-level programming environment.

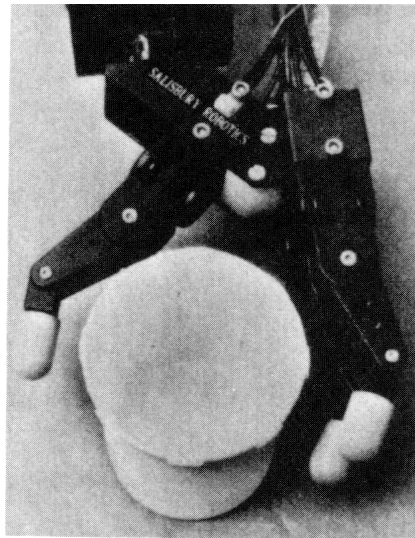


Figure 1: Wrap hand preshape.

III. Hand Preshaping and Reach

We have chosen to implement the following three grasp preshapes: wrap, pinch, and grip. Figures 1 - 3 show these preshapes as implemented using the Salisbury hand. Each grasp is described by the number of virtual fingers which it uses (one, two, or three) and by the type of contact which the fingers make with the object (point or extended). The concept of virtual fingers is due to Arbib and Iberall [Arbib83]. Virtual fingers involve the use of one or more real fingers to implement a grasp. The number of real fingers used depends upon the parameters of the object or component to be grasped. In this work, we will fix the number of real fingers to be used per virtual finger and will not yet attempt to do a dynamic mapping. Thus, a wrap involves two virtual fingers, the object is enclosed, and the contacts are extended. A grip involves three virtual fingers, the object is held at the fingertips, and the contacts are point.

In the precontact stage of the grasp, the system uses information extracted from the object to generate a set of grasp preshapes and to determine parameters for both the hand shape and the reach. Currently, the only parameter of the hand shape with which we are concerned is the distance between the fingers. The high-level reasoning module checks that the selected object component will fit into the span of the hand for the chosen grasp. Parameters of the reach are: the approach plane, that is, the plane perpendicular to the axis along which the arm/hand will approach the target; the target point, that is, the point on the component which the approach axis will intersect; and the rotation of the wrist/hand to provide proper placement of the fingers on the component.

The hand shapes and reach parameters are generated by a set of rules which operate on an object representation scheme which we call the *spatial polyhedron*. Because this

representation is the driving force behind the high-level reasoning within the system, we will present it briefly below. We will then return to our discussion of reasoning for grasping and present some examples.

IV. Object Representation

The spatial polyhedral representation was developed as a mechanism for representing generic, or basic level, objects for active tactile exploration and categorization. It is a feature-based representation which includes both definitional information about an object in the form of a hierarchy of frames, and spatial information in the form of a set of *aspects* [Koenderink79]. All information is symbolic, and the goal is to represent objects in such a way that generic objects such as "cup" may be modelled and reasoned about. An additional requirement is that the representation must be "practical" enough to allow its use within a robotic system equipped with real sensors. The representation is presented fully in [Stansfield87], and its successful use for categorizing basic objects explored by an active robotic system is shown. In this paper, we show how it may be used by a reasoning system to generate grasps. To that end, we will describe the representation informally and present an example.

Objects within the system are defined hierarchically. An object is made up of a set of invariant components (a cup consists of a body and a handle, for example.) Each component is described by a set of features (the body consists of a curved surface, a planar surface, and a rim contour.) The features are described by view-dependent parameters (the handle viewed from above has a different appearance than it does when viewed from the side.) In addition, as one moves around an object, different components will come into or leave one's view. This information is embodied in the spatial polyhedron, which may be described informally as follows.

Imagine an object at the center of an n -sided polyhedron. If the object were to be viewed, or sensed, along a line normal to each face of this polyhedron, then certain components and features of the object would be "viewable", while all others would not. Slight changes in attitude as the viewer moves around the object will not result in any new features coming into view. When the viewer has moved sufficiently, however, then he will be sensing the object from a different "perspective" (or face of the spatial polyhedron) and different components and features will be viewable. Thus we model an object by mapping to each face of the spatial polyhedron all of the features which we expect to be "viewable" along that face. This mapping consists of a list of these features and their appearance from the specified view. Thus the faces of the spatial polyhedron represent different aspects of the object. Figure 4 shows a psuedo-PROLOG implementation of this representation for a glass. The face predicates implement the spatial polyhedron.

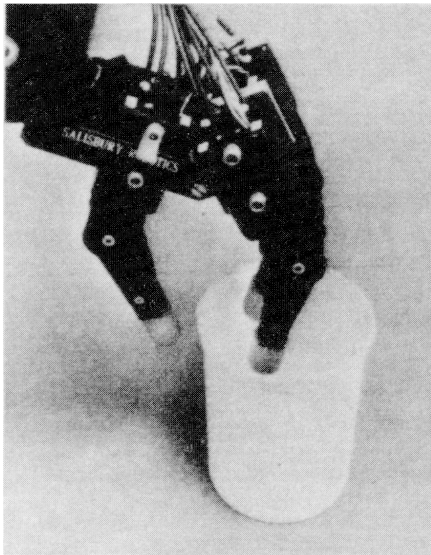


Figure 2: Pinch hand preshape.



Figure 3: Grip hand preshape.

```

object(glass, bounding_volume, dimension)
component(glass,body,bounding_volume)
face(glass, body, contour:(rim,curved,
    bounds_on_radius),side1)
face(glass, body, surface:(nonelastic,
    noncompliant,smooth,planar,
    bottom_surface),countour:(border,
    curved,bounds_on_radius),side2)
face(glass, body, surface:(nonelastic,
    noncompliant,smooth,curved,
    side_surface),side3)
face(glass, body, surface:(nonelastic,
    noncompliant,smooth,curved,
    side_surface),side4)
face(glass, body, surface:(nonelastic,
    noncompliant,smooth,curved,
    side_surface),side5)
face(glass, body, surface:(nonelastic,
    noncompliant,smooth,curved,
    side_surface),side6)

```

Figure 4: Representation of glass.

V. Reasoning About Grasping

The object database to be used by the system consists of a set of generic objects represented as described above. It is assumed that vision, or a combination of vision and exploratory touch, has been used to extract the pertinent features and relations from the object to be grasped and that this information has been integrated into a symbolic representation also accessible to the system.

Stansfield [Stansfield87] describes in detail a robotic system which utilizes vision and touch to explore and categorize objects. The output of such a system serves as the input to the one described here. Reasoning for grasping consists of a set of rules concerning which features and relations must be present in the explored object in order to invoke a particular grasp preshape and to generate the grasp/reach parameters.

Reasoning about grasps works as follows: An attempt is first made to categorize the object as one of those known to the system. It is not necessary that the robot recognize an object in order to grasp it. Indeed, the robot should be capable of grasping objects which it does not recognize. However, if an object is recognized, then assumptions may be made about those portions which have not been sensed (that the back-side of a glass is a curved surface, for instance.) These assumptions may then be used by the system when generating grasps. If the object is not recognized, then the system will only generate grasps which place the fingers on sensed portions of the object, since it assumes that nothing is known about the unsensed portions. When there are multiple hypotheses concerning object identity, the system chooses the simplest. (The simplest object is the one with the fewest number of unsensed

```

wrap_right_body(object)
IF
    right_face has curved surface AND
    right_face has no other components AND
    front_face has curved surface AND
    back_face has curved surface AND
    distance between front_face and
        back_face fits within span of hand THEN
        valid grasp for object is
        wrap for component body.
        approach target is center of
        curved surface from right.
        oppositions are curved surface
        in front and curved surface in back.

```

Figure 5: Rule for invoking wrap preshape for a body from the right.

components.)

The rule-base for grasping consists of a set of rules concerning which features of an object must be present, and what the relations among these features must be, in order for a particular hand preshape to be valid for grasping the sensed object. If these conditions are met, then an additional set of rules for generating the parameters of the grasp and reach are invoked. Figure 5 shows the rule for invoking a wrap grasp of the body of an object from the right. The necessary conditions are that there be a set of three adjacent curved surfaces; that the approach surface be unoccluded by other components of the object; and that the body fit within the span of the hand. All information concerning which features of the object are present, and what the parameters of these features are for a given aspect, comes directly from the spatial polyhedron, as do the target approach plane and the oppositions. The target point is currently predetermined for each feature type. Oppositions determine the placement of the virtual fingers and the rotation of the wrist. The term *opposition* is borrowed from Iberall [Iberall87]. She uses the term to describe the way in which the hand can apply forces around a given object for a given task. We use the term in essentially the same way. Oppositions are to some extent implicit in the relations between pairs of faces of the spatial polyhedron. The type of feature and its parameters, as well as the type of hand preshape under consideration, determine the choice of oppositions.

VI. Examples and Discussion

Example 1: Glass.

In our first example, the input object is a glass. The system generates the following output:

```

Object hypothesis is: glass
If object is glass then these
components are missing: none

```

Object hypothesis is: mug
If object is mug then these
components are missing: handle

I'm going to assume that the object is glass

The object may be grasped as follows:

Use hand preshape Wrap for component Body
Approach target is the center of the curved
surface from the right

Opposition 1: curved surface from the front
Opposition 2: curved surface from the back

Use hand preshape Wrap for component Body
Approach target is the center of the curved
surface from the left

Opposition 1: curved surface from the front
Opposition 2: curved surface from the back

Use hand preshape Wrap for component Body
Approach target is the center of the curved
surface from the front

Opposition 1: curved surface from the right
Opposition 2: curved surface from the left

Use hand preshape Pinch for component Body
Approach target is point on rim from top
Opposition 1: inside of chosen rim point
Opposition 2: outside of chosen rim point

Use hand preshape Grip for component Body
Approach target is center of contour from top
Opposition 1: contour from left
Opposition 2: contour from right
Opposition 3: contour from front

Since the object is assumed to be a glass, the system generates the wrap preshape for the body in three different configurations (approach from the front, the left, and the right). This is because it assumes a curved surface behind the object upon which it may place fingers. In addition, grip and pinch preshapes are generated for the rim from above.

Example 2: Unknown Object.

When the glass and mug objects are removed from the database, and the same sensed object is input, the system can no longer categorize the object as one with which it is familiar. Therefore, it generates the following:

I don't recognize the object
I'm not making any assumptions
about object identity

The object may be grasped as follows:

Use hand preshape Wrap for component Body
Approach target is the center of the curved
surface from the front

Opposition 1: curved surface from the right
Opposition 2: curved surface from the left

Use hand preshape Pinch for component Body
Approach target is point on rim from top

Opposition 1: inside of chosen rim point
Opposition 2: outside of chosen rim point

Use hand preshape Grip for component Body
Approach target is center of contour from top
Opposition 1: contour from left
Opposition 2: contour from right
Opposition 3: contour from front

The wrap preshapes which approach from the left and the right are not generated this time because they would require that a finger be placed on the unexplored back-side of the object.

Example 3: Mug.

Our final example illustrates a set of grasps generated for an object with multiple components. The object is a mug with the handle visible and to the left of the body. The system generates the following:

Object hypothesis is: mug
I'm going to assume that the object is mug

The object may be grasped as follows:

Use hand preshape Wrap for component Body
Approach target is the center of the curved
surface from the right

Opposition 1: curved surface from the front
Opposition 2: curved surface from the back

Use hand preshape Pinch for component Body
Approach target is point on rim from top
Opposition 1: inside of chosen rim point
Opposition 2: outside of chosen rim point

Use hand preshape Grip for component Body
Approach target is center of contour from top
Opposition 1: contour from left
Opposition 2: contour from right
Opposition 3: contour from front

Use hand preshape Wrap for component handle
handle is on the left
Approach target is the center of the part
from the left
Opposition 1: part from the front
Opposition 2: part from the back

Since the handle is on the left side of the body, the only wrap preshape generated for the body is an approach from the right. This is because the handle would be in the way if a wrap preshape were attempted from any other approach direction. The system has also generated a wrap preshape for the handle of the mug.

VII. Summary and Future Work

In this paper, we have presented a rule-based reasoning system and an object representation paradigm for reasoning about the precontact stage of a grasp. During this stage, object features and high-level knowledge are used in a feedforward manner to generate a hand preshaping, wrist/hand orientation, and ballistic reach toward the object. The spatial polyhedron was introduced as a mechanism for representing and reasoning about generic objects. Rules which utilize the information embodied in this representation were used to generate a set of grasp preshapes and reach parameters for the given object.

Future work in this research will continue in two directions. First, the set of grasps generated by the system must be pruned down to one, which will be sent to the robotic devices. The grasp choice is currently left to the operator. An obvious extension to the high-level reasoner presented here would prune the grasps based on the task and the perceived state of the world. Thus, if the task was to "pour from the glass", then any grasp which placed the hand over the cavity would be pruned. Such knowledge could be embodied in a set of rules utilized by a task planner. The perceived state of the world will affect grasp choice, as well. A grasp which requires the hand to approach the object from the side would be pruned if the object were surrounded on all sides by other objects.

The reasoning might also be extended to allow for use of multiple hands: the system currently generates a set of grasps for the object independent of the devices. Rules which map these grasps to different hands in a coordinated manner might be used to allow the system to grasp objects which are too large for a single hand.

The second direction in which this research will proceed is to extend the grasp procedure to include the postcontact stage. This will involve bringing the hand (equipped with tactile sensors) into contact with the object and then utilizing a robotic tactile perception system, such as that described in Stansfield [Stansfield87], to fine-tune the grasp.

References

- [Arbib83] Arbib, M., T. Iberall, and D. Lyons. "Coordinated Control Programs for Movements of the Hand." COINS TR 83-25. University of Massachusetts, Amherst. August, 1983.
- [Cutkosky87] Cutkosky, M. and P. Wright. "Modeling Manufacturing Grips and Correlations with the Design of Robotic Hands." *Proceedings of the IEEE Conference on Robotics and Automation*. Raleigh, N.C. April, 1987. pp. 1533-1539.
- [Iberall87] Iberall, T. "The Nature of Human Prehension: Three Dextrous Hands in One." *Proceedings of the IEEE Conference on Robotics and Automation*. Raleigh, N.C. April, 1987. pp. 396-401.
- [Jeannerod78] Jeannerod, M. "Intersegmental Coordination During Reaching at Natural Visual Objects." In *Attention and Performance, IX*. Long, J. and A. Baddley, Eds. Lawrence Erlbaum. Hillsdale, N.J. 1978.
- [Klatzky86] Klatzky, R., B. McCloskey, S. Doherty, J. Pellegrino, and T. Smith. "Hand Shaping and Object Processing." Cognitive Science TR 8605. University of California, Santa Barbara. 1986.
- [Koenderink79] Koenderink, J. and A. van Doorn. "The Internal Representation of Solid Shape with Respect to Vision." *Biological Cybernetics*. Vol. 32 (1979). pp. 211-216.
- [Lyons85] Lyons, D. "A Simple Set of Grasps for a Dextrous Hand." *Proceedings of the IEEE Conference on Robotics and Automation*. St. Louis, Mo. April, 1985. pp. 588-593.
- [Stansfield87] Stansfield, S. "Visually-guided Haptic Object Recognition." *Ph. D. Thesis*. University of Pennsylvania. December, 1987.
- [Tomovic87] Tomovic, R., G. Bekey, and W. Karplus. "A Strategy for Grasp Synthesis with Multifingered Robot Hands." *Proceedings of the IEEE Conference on Robotics and Automation*. Raleigh, N.C. April, 1987. pp. 83-89.