Reasoning Visually about Spatial Interactions

N. Hari Narayanan and B. Chandrasekaran
Laboratory for Artificial Intelligence Research
Department of Computer and Information Science
The Ohio State University
Columbus, Ohio 43210
U. S. A.

Abstract

This paper is concerned with how diagrams can be used for reasoning about spatial interactions of objects. We describe a computational approach that emulates the human capability of predicting interactions of simple objects depicted in two dimensional diagrams. Three core aspects of this approach are a visual representation scheme that has symbolic and imaginal parts, the use of visual processes to manipulate the imaglnat part and to extract spatial information, and visual cases that encode experiential knowledge and play a central role in the generation of spatial inferences. These aspects are described and the approach is illustrated with an example. Then we show that reasoning with images is an emerging and promising area of investigation by discussing computational and cognitive research on imagery.

1 Introduction

Humans quite often make use of spatial information implicit in diagrams to make inferences. For example, anyone famlliar with the operation of gears will be able to solve the problem posed in Figure 1 by imagining the rotary motion of gear1 being transmitted to the rod through gear2, resulting in the horizontal translation of the rod until it hits the wall. In such situations humans reason about spatial interactions not only by using conceptual knowledge, but also by extracting constraints on such interactions from a perceived image. This integrated use of visual knowledge (about spatial configurations) from the diagram and conceptual knowledge (such as the rigidity or plasticity of objects involved) is a very interesting phenomenon, in this paper we illustrate a computational approach that emulates this capability for solving simple motion prediction problems.

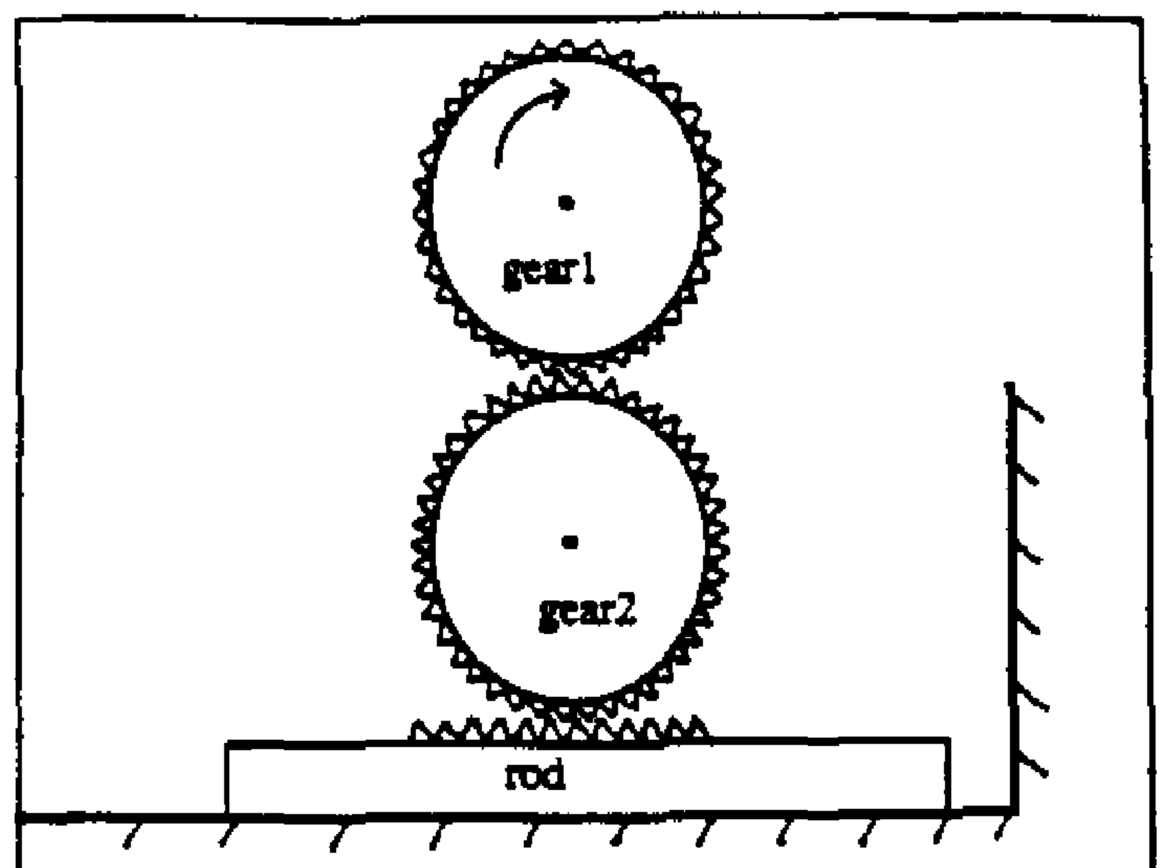

Figure 1. Predict Rod's Motion

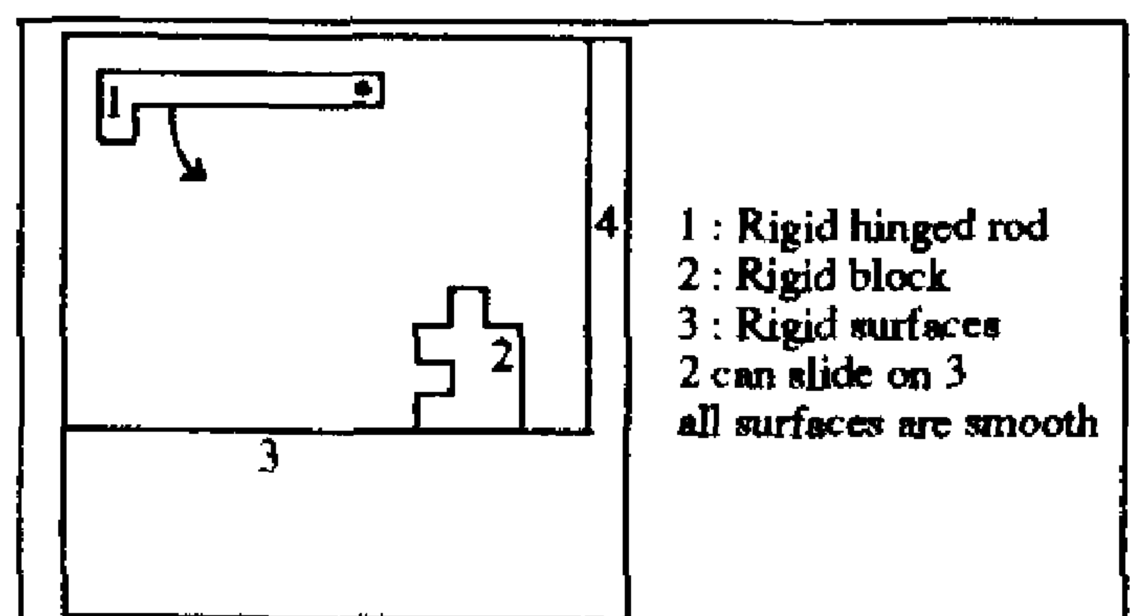

Figure 2. A Motion Prediction Problem

2 The Approach

2.1 Motlon Predlctlon Problems

The class of problems we address is the following: given a two dimensional diagram of the spatial configuration of a set of objects, one or more Initial motions of objects and relevant conceptual information about them, predict the subsequent dynamics of the configuration. Figure 2 shows a typical example.

2.2 Cognitive Inspiration

There is considerable evidence in cognitive science for the use of mental images by people when solving problems [Kosslyn, 1981]. Furthermore, Introspective reports of people when given a diagram like Figures 1 or 2 and asked to predict motions indicated that by looking at the diagram they were able to visualize the motion of one object causing that of another through physical contact. They appeared to be using (the image of) the diagram in front of them directly to simulate motions In their minds. These reports indicated the following.

(1) Given a diagram depicting the problem, humans quite rapidly focus on localities of potential interactions.

(2) People also seem to simulate or project the motion to determine the nature of interactions that will occur.

(3) For reasoning about the dynamics (e.g., how will motion be transmitted after a collision?) humans bring conceptual knowledge (e.g., gears are rigid objects) and experiential knowledge (e.g., if an object collides with another, it typically transmits motion in the same direction) to bear on the problem.

We have developed an approach that emulates these capabilities.

2.3 Representation

The specification of a motion prediction problem consists of a scene depicting the spatial configuration of the objects involved and conceptual information about their properties (see Figure 2). The spatial configuration is represented using a "visual representation" while conceptual information about object properties is represented declaratively and linked with corresponding object descriptions in the visual representation. In our computer model the visual representation of a problem specification is interactively constructed prior to problem solving, whereas in the case of humans perceptual processes deliver such representations.

Mental representation of visual information and Its relation to the phenomenon of mental imagery have been the focl of considerable research in cognitive science [Biederman, 1990; Finke, 1989; Kosslyn, 1981; Pyiyshyn, 1981]. A central issue here is the question of how mental imagery that appears to be analogic in nature can arise from underlying representations that are considered to be pro positional. One hypothesis regarding this issue [Chandrasekaran and Narayanan, 1990] is that representations for different sensory modalities are operated upon by interpreters that provide privileged operations specific to that modality, including binding the symbols in the representation to perceptual primitives in the corresponding sensory domain. Thus our answer to the above question is that symbolic representations of visual information are interpreted by mechanisms that are specialized to the visual modality and which provide operations tailored to this modality. These operations construct mental Images using perceptual primitives In the visual domain. Visual representations in our computer model are therefore structured as multi-level hierarchies that contain *imaginal* descriptions and *symbolic* descriptions of the object configuration. Each level of the hierarchy contains a symbolic description and an imaginal description of the configuration at a certain resolution. The symbolic description is built from parametrized shape primitives like circles, rectangles, etc., whereas the imaginal description is a two dimensional pixel array of fixed width and height in which a configuration is depicted by object boundaries and is implemented as a bitmap. In the rest of the paper we will use the term "diagram" to refer to this boundary-based rendering of the object configuration. Thus the visual representation is dual (symbolic and imaginal) in nature. The two types of mental representations (surface images and deep encodings) that Kosslyn [1981] proposes reflect a similar duality.

The most interesting property of this representation is that it simultaneously provides abstract symbolic descriptions of an object configuration and directly captures, in the imaginal descriptions, specific spatial information about the object configuration (the extent of contact between two surfaces, for example). The justification for our decision to structure the symbolic descriptions in terms of parametrized shape primitives stems from shape representation theories that utilize primitives like geons [Biederman, 1990] and generalized cones [Marr and Nishihara, 1976], Multiple levels of description are provided in the representation to allow visual processes to operate at different levels of resolution.

2.4 Reasoning

The basic model of reasoning Is as follows. The system goes through a sequence of deliberative states. This sequence corresponds to the changes that the initial object configuration undergoes due to motion and interaction of objects. Each deliberative state represents a particular configuration that the objects assume at some point during this evolution of behavior. What distinguishes a deliberative state from other states is that it represents a configuration in which an interaction (such as collision) has occurred that will change the subsequent behavior of objects. The term deliberative refers to the necessity of "deliberation" that arises at these states in order to predict future behavior of objects. A significant characteristic of this deliberation Is the combined use of perceptual information from the diagram and conceptual knowledge relevant to the situation.

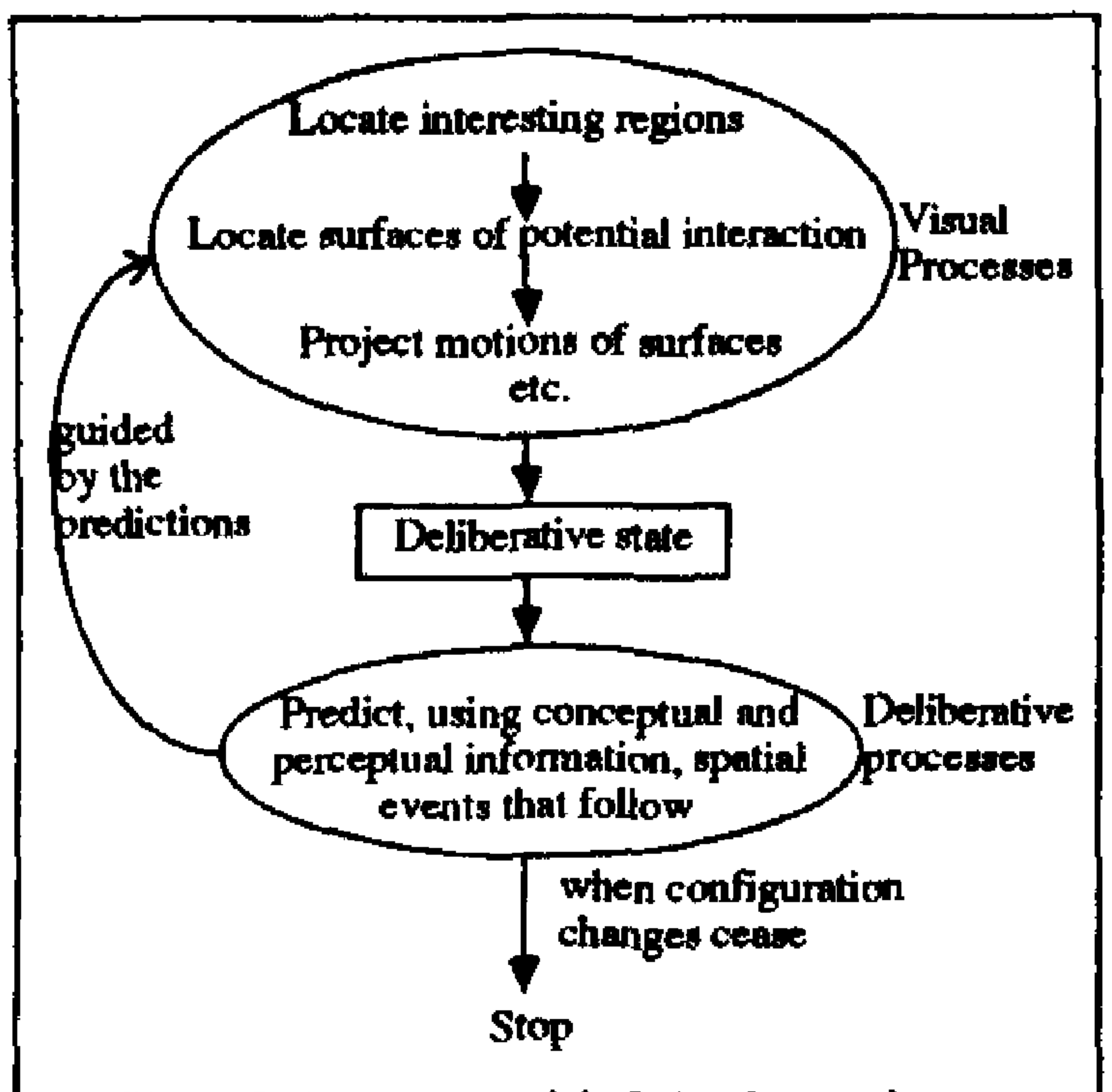

Figure 3. A process model of visual reasoning

The transition between deliberative states is accomplished by two groups of processes, in one, purely visual operations such as attention-focussing, scanning, boundary-tracking, and contact-detection are used to identify significant aspects of the current object configuration from the diagram (e.g., locating interesting regions, identifying surfaces of potential Interaction, etc.), to reason about how the configuration will evolve (e.g., project surface motions), and to detect the next deliberative state. A deliberative state is detected by watching out for certain events as the configuration depicted in the diagram changes. The establishment of a new contact between objects, the elimination of a previously existing contact, the establishment of a new support relationship between objects, and the removal of a support relationship are some examples of events which indicate a deliberative state. Motivations behind and justifications for these processes derive from the extensive literature on mental imagery, some of which are discussed in section 3, and the work of Chapman 11990] and Uliman (19851 on visual routines. This group of processes corresponds to the human cognitive process of "imagining".

The second group of processes accomplish the aforementioned task of deliberation. Here knowledge about how interacting physical objects tend to behave under various conditions is used to predict the behavior of objects following the current deliberative state. We take a specific position on the form in which this knowledge is available and the way In which it is utilized. This is described next. A process model of visual reasoning Is shown in Figure 3.

We believe that the knowledge humans bring to

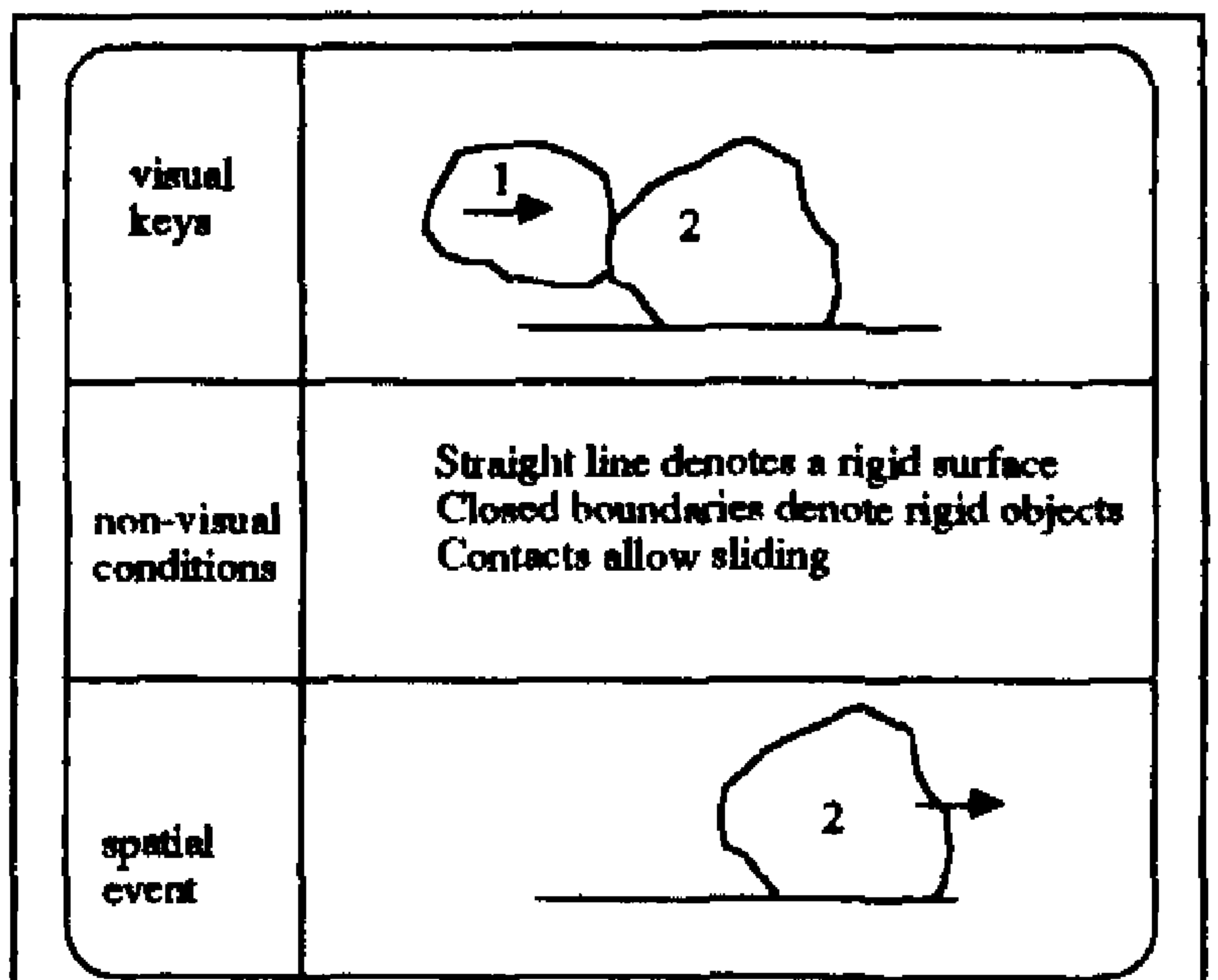

Figure 4. A Visual Case

bear on making spatial inferences in similar situations is mostly acquired through experience, and so In the computer model experiential knowledge has been given a central role In deciding how to proceed from a deliberative state. Experiential memory is considered to be an organized and indexed collection of cases [Schank, 1982] and case-based reasoning is a computational paradigm for modelling the role of experience in problem solving (Kolodner and Simpson, 1989]. Therefore, representational structures called "visual cases" have been developed to encode knowledge applicable at deliberative states and to facilitate Inferencing. Each case represents a typical spatial event. Since cases represent experiential knowledge, they may not be logically parsimonious or mutually exclusive. A visual case has three parts. One Is information about spatial configurations to which the case is applicable. Cases are called "visual" because this information is visual in nature and is the "key" by which relevant cases get selected during reasoning. It may also be viewed as an "abstract" image that depicts the essential aspects of configurations to which the case is applicable. Because of this abstractness a case can be matched with a variety of specific configurations. This property obviates the need for a large number of cases. The second part is non-visual Information that qualifies the visual part further and it is used for deciding the applicability of a case to a particular situation. The third part is a predicted event affecting objects In the spatial configuration represented by the case. This event may specify a state change (e.g., a directional force being applied on an object), a continuous change (e.g., an object moving in a particular direction), etc

Humans are skilled at blending perceptual and conceptual Information in generating spatial Inferences. To illustrate this, first consider your

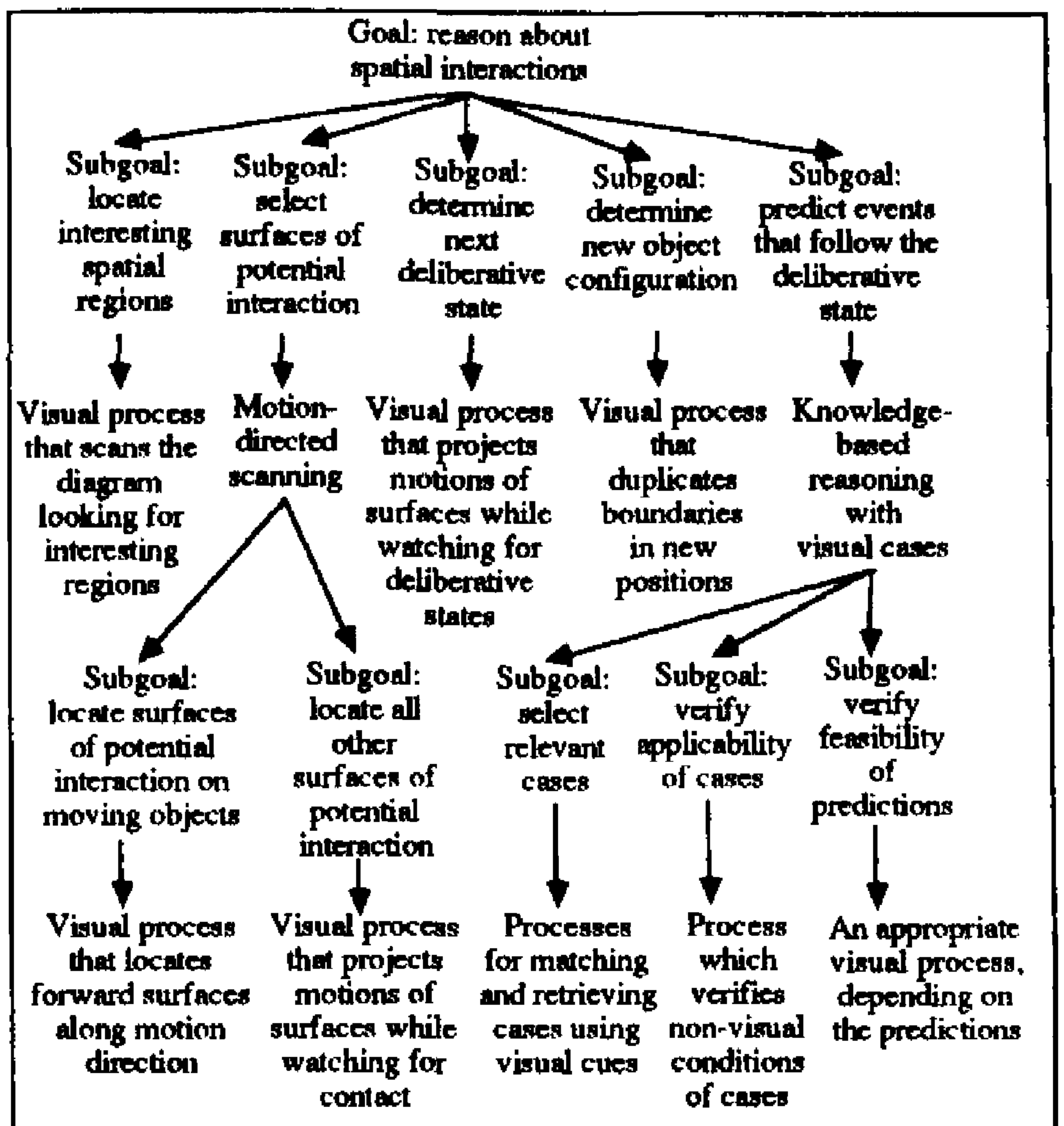

Figure 5. Execution trace for the example

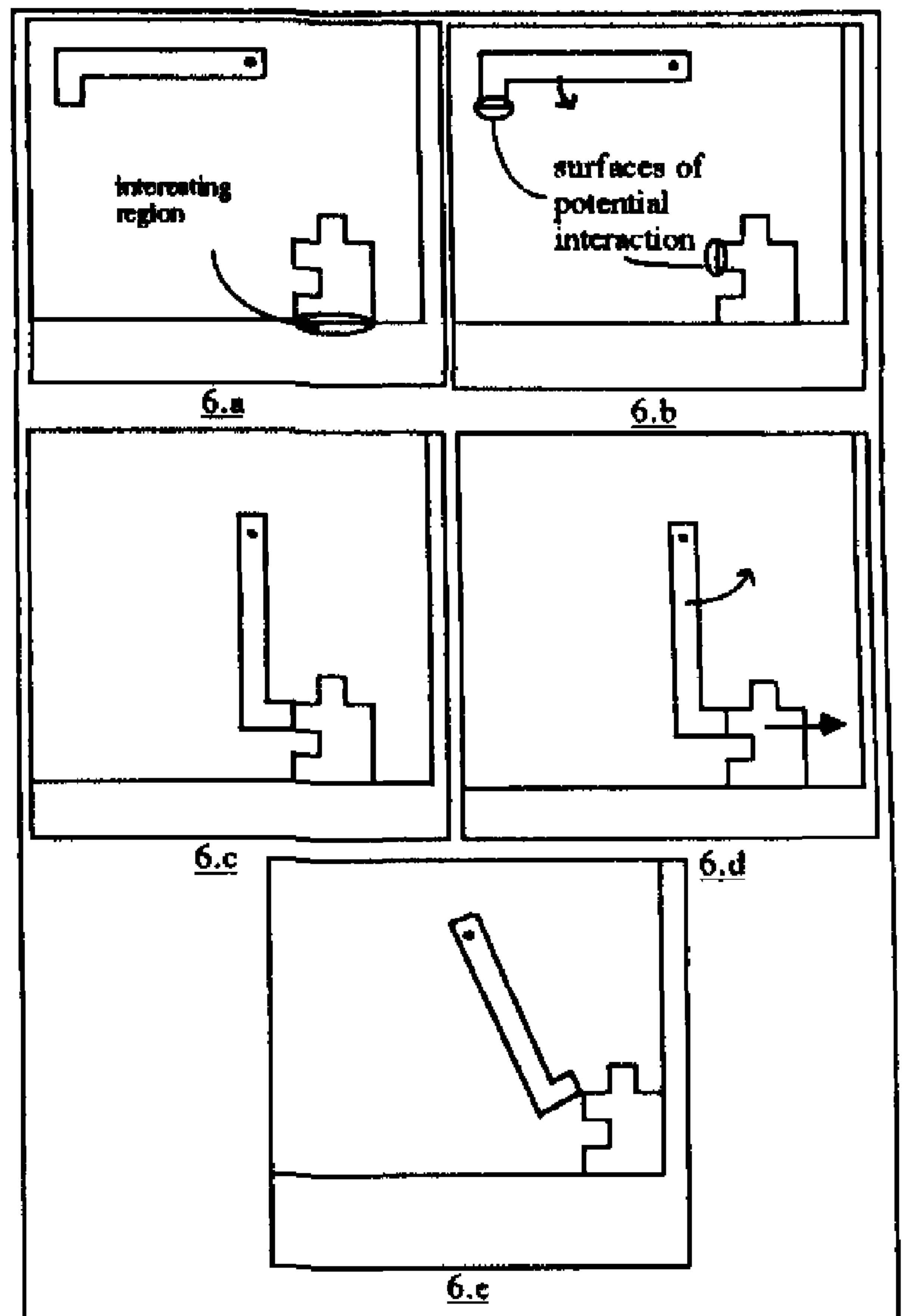

Figure 6. Diagram Snapshots

prediction about the motion of ob|ect2 after object 1 collides with it, given the problem specification of Figure 2, and then notice how this prediction will change if the specifications were changed to indicate that object1 is non-rigid (say, made of rubber) and that object2 is fixed on surface3. The visual and non-visual parts of a case explicitly capture this aspect. Thus the intent of visual cases is to represent simple chunks of experiential knowledge about spatial events that humans typically have, and to model the blending of conceptual and perceptual information in making spatial inferences. An example of typical knowledge about spatial events is "a rigid object resting on a rigid flat surface, when collided by a moving rigid object, will tend to slide in the same direction". The schematic of a corresponding visual case is shown in Figure 4.

After a deliberative state is detected, visual cases *are* retrieved and *applied to predict events* that follow. The retrieval of cases relevant to the spatial configuration in the diagram is based on visual cues. From among the retrieved cases, applicable ones are selected by using information about object properties (which is available as part of problem specification) to verify the non-visual parts of the cases. Events predicted by the applicable cases are further pruned by verifying, through visual processes, their feasibility in the current object configuration. The remaining events serve to guide subsequent steps of reasoning. Since a case brings conceptual knowledge to bear on visual reasoning, this mechanism of inference may be viewed as a computational realization of cognitive penetrability or the influence of tacit knowledge on mental imagery [Pylyshyn, 1981].

2.5 An Example

In this section we present a problem solving episode in some detail. The specification of the problem, which Includes a depiction of an initial configuration of objects, an initial motion and relevant non-visual properties of the objects, is shown in Figure 2. The goal Is to predict all resulting motions by reasoning about spatial interactions that will occur among the objects. In our computer model control of reasoning is done by a procedure that generates goals and subgoals, and activates relevant processes to achieve them. Thus an execution trace will appear as a tree consisting of goals, subgoals, and processes. The goal generation follows the process model in Figure 3.

Figure 5 shows a partial execution trace for this example. "Reason about spatial interactions" is the top level goal and it has four subgoals as shown. Consider the first subgoal "locate interesting spatial regions". There is a set of heuristic criteria to locate interesting regions, one of which is that regions representing touching surfaces of multiple objects are interesting. The visual process corresponding to this subgoal focuses on each object in turn, tracks its boundary, and looks for regions that satisfy the criteria. In this example it finds the bottom surface of ob)ect2 as shown in Figure 6.a. Next, surfaces that have the potential for interaction are located (Figure 6.b shows the surfaces identified for the current problem) and another process projects the motion of moving surfaces that are identified to have interaction potential while watching for the occurrence of deliberative states. The first deliberative state detected is the configuration in which contact occurs between objects 1 and 2 and the diagram is modified to depict this configuration, as shown in Figure 6.c. The next subgoal is to predict subsequent dynamics of this configuration and this is accomplished by the application of visual cases. Three visual cues (the presence of a rotating object and a stationary object, and the occurrence of a collision between the two) are used to retrieve cases, and visual keys of cases are matched with the current configuration by inspecting its visual representation. The availability of symbolic descriptions as well as diagrams In the visual representation allows the matching of visual keys to proceed at an abstract level without recourse to techniques like template matching. A visual case similar to the one shown in Figure 4 (except that the moving object is undergoing rotation) is found to be relevant and applicable, and the event that it predicts is found to be feasible in the current configuration. Non-visual conditions associated with this case are similar to those in Figure 4 and are easily verified from the problem specification. The resulting prediction is shown in Figure 6d. As the process model shows, after this step the entire cycle is repeated and in the next detected deliberative state object2 has collided with surface4. This time a case that predicts cessation of motion gets applied and Figure 6.e shows the final configuration.

3 Related Work

In this section we present computational and cognitive research which touches upon imagery, in support of our contention that imaginal reasoning is an emerging research area that is highly promising. Cognitive scientists have demonstrated not only the powerful role of imagery in human problem solving but also the advantages of incorporating similar reasoning capabilities in computer programs. For example, Larkin and Simon [1987] persuasively argue for the computational advantages afforded by diagrammatic representations and perceptual inferences that such representations support, for solving physics and geometry problems. Koedinger and Anderson [1990] describe a model of geometric proof problem solving In which parsing of a diagram of the problem to detect specific diagram configurations is a key step. These configurations then cue relevant schematic knowledge for proceeding with the proof. Visual cases represent a generalization of this idea. Logicians have also noted the power of visual representations. Barwise and Etchemendy [1990] illustrate the role of visual representations in mathematical reasoning through a program called Hyperproof which allows the user to reason using sentential and pictorial forms of Information.

Despite intuitively compelling evidence for the use of imagery by humans, there has not been much work In artificial intelligence toward endowing machines with a similar capability. An early program that utilized diagrams was WHISPER [Funt, 1977] which addressed rotation, sliding, and stability of blocks-world structures. More recently, work on using pictorial or "analogical" representations for simulating the behavior of strings and liquids in space has been reported [Gardin and Meltzer, 1989]. Shrager [1990] describes a computational model of understanding laser operation in which reInterpretation processes utilize event depictions in an iconic memory as well as In a proposItional memory. The research on computational modelling of the cognitive process of spatial reasoning with diagrams [Narayanan, 1991] is yet another step towards realizing the full potential of Imaginal reasoning by computers.

4 Concluding Remarks

We have described a novel approach to reasoning about spatial Interactions, Since our aim in this paper has been to provide the reader with an overview of all significant aspects of visual reasoning within the limited space available, the descriptions have been necessarily schematic in character. Further details on components of this approach - structure of visual representations, how visual processes are composed from basic visual operations, indexing and adaptation of visual cases, the computer program that implements this model, etc. - can be obtained from [Narayanan, 1991].

The advantages of using diagrams in this approach arise from the property that spatial information such as obstacles to motion or pathways that guide motion are directly evident in Images. Our approach is not only Intuitive, but flexible as well.

Objects which have irregular shapes that will make their algebraic representations complex can be represented and reasoned about In the same way as regular objects If diagrams are used.

As Forbus and colleagues rightly point out [Forbus *et al.,* 1987] there can be no purely qualitative method for spatial reasoning. What is required is to integrate qualitative and quantitative methods so that qualitative ones provide approximate solutions that serve to focus the application of quantitative methods to only those aspects of the initial solutions that require more precision or further refinement. With this goal in mind, we are currently investigating the integration of visual reasoning with other qualitative and quantitative methods [Narayanan and Chandrasekaran, 1991].

Acknowledgments

A discussion with Stephen Kosslyn provided inspiration during the early stages and Jeff Shrager and Janet Kolodner helped clarify many issues during the course of this research. We also thank the anonymous referees for their suggestions. This research is supported by DARPA & AFOSR contract # F-49620-89-C-0110 and by AFOSR grant # 890250.

References

[Barwlse and Etchemendy, 1990] J. Barwlse and J. Etchemendy. Visual information and valid reasoning. In W. Zimmerman, (Ed.), *Visualization in Mathematics,* Mathematical Association of America, Washington D. C, 1990.

[Biederman, 1990] I. Biederman. Higher-level vision. In D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, (Eds.)r *An Invitation to Cognitive Science, Volume 2: Visual Cognition and Action,* MIT Press, Cambridge, MA, 1990.

[Chandrasekaran and Narayanan, 1990] B. Chandrasekaran and N. H. Narayanan. Integrating imagery and visual representations. In *Proc. Annual Conference of the Cognitive Science Society,* pp. 670-678, Boston, MA, 1990.

[Chapman, 1990] D. Chapman. Intermediate vision: architecture, implementation, and use. Technical Report TR-90-06, Teleos Research, Palo Alto, CA, 1990.

[FInke, 1989] R. A. Finke. *Principles of Mental Imagery,* MIT Press, Cambridge, MA, 1989.

[Forbus *et al,,* 1987] K. D. Forbus, P. Nielsen, and B. Faltfngs. Qualitative kinematics: a framework. In *Proc. 10™ International Joint Conference on Artificial Intelligence,* pp. 430-436, Milano, Italy, 1987.

[Funt, 1977] B. V. Funt. WHISPER: a problem solving system utilizing diagrams and a parallel processing retina. In *Proc. 5th International Joint Conference on Artificial Intelligence,* pp. 459-464, Cambridge, MA, 1977.

[Gardln and Meltzer, 1989] F. Gardin and B. Melfcer. Analogical representations of naive physics. *Artificial Intelligence Journal,* 38:139-159, 1989.

[Koedinger and Anderson, 1990] K. R. Koedinger and J. R. Anderson. Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognitive Science,* 14:511-550, 1990.

[Kolodner and Simpson, 1989] J. L. Kolodner and R. L. Simpson. The MEDIATOR: analysis of an early case-based problem solver. *Cognitive Science,* 13:507-549, 1989.

[Kosslyn, 1981] S. M. Kosslyn. The medium and the message In mentaf Imagery: a theory. *Psychological Review,* 88:46-66,1981.

[Larkin and Simon, 1987] J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science,* 11:65-99, 1987.

[Marr and Nishihara, 1976] D. Man and H. K. Nishihara. Representation and recognition of the spatial organization of three dimensional shapes, AI Memo 377, Artificial Intelligence Laboratory, MIT, Cambridge, MA, 1976.

[Narayanan, 1991] N. H. Narayanan. A study of representations and processes *tor* reasoning visually about spatial interactions. Doctoral Dissertation, Department of Computer & Information *Sconce,* The Ohio State University, Columbus, OH, forthcoming.

[Narayanan and Chandrasekaran, 1991] N. H. Narayanan and B. Chandrasekaran. Integration of qualitative and quantitative methods in visual reasoning. In *Proc. 2nd Conference on Al, Simulation and Planning in High Autonomy Systems,* pp. 272-278, Cocoa Beach, FL, 1991.

[Pylyshyn, 1981] Z. W. Pylyshyn. The Imagery debate: analogue media versus tacit knowledge. *Psychological Review,* 88:16-45,1981.

[Schank, 1982] R. Schank. *Dynamic Memory: A Theory of Learning in Computers and People.* Cambridge University Press, New York, 1982.

[Shrager, 1990] J. Shrager. Commonsense *perception* and the psychology of theory formation. In J. Shrager and P. Langely, (Eds.), *Computational Models of Scientific Discovery and Theory Formation.* Morgan Kaufmann, San Mateo, CA, 1990.

[Ullman, 1985] S. Ullman. Visual routines. In S. Pinker, (Ed,), *Visual Cognition.* MIT Press, Cambridge, MA, 1985.