Commonsense Reasoning about Moving Objects: an Elusive Goal

Leo Joskowicz
IBM T.J. Watson Research Center
P.O. Box 704, Yorktown Heights, NY 10598
E-mail: josko@watson.ibm.com

Extended Abstract

Introduction

Reasoning about how objects move and interact in space is pervasive in everyday life. It is rightly considered an essential component of intelligence. Consequently, understanding spatial reasoning and developing computational models for it has been a central concern in many fields, including cognitive psychology, mathematics, robotics, vision, and artificial intelligence. Although much progress has been made over the last decades in understanding the cognitive aspects of spatial reasoning and in solving specific tasks, such as motion planning and object recognition, computational models of commonsense spatial reasoning have been few and notoriously limited in scope. This is in sharp contrast with the relative success of formalizing other kinds of commonsense reasoning, such as temporal, qualitative, and non-monotonic reasoning.

In this paper, we identify and characterize some of the sources of difficulty and propose an explanation for the limited success in formalizing commonsense spatial reasoning. First, we evaluate four general approaches to spatial reasoning about moving objects: simulation, analytical, axiomatic, and analogical reasoning. We characterize their expected input and output, identify their underlying assumptions, and evaluate their strengths and weaknesses. We observe a wide divergence of goals and techniques, analyze their possibility of fully capturing commonsense spatial reasoning, and question the feasibility of a general purpose reasoning engine.

We conclude by suggesting that, more than any other commonsense reasoning task, spatial reasoning heavily relies on sensory-motor information and previous experience. This directly challenges the accepted wisdom in formalizing commonsense reasoning, which assumes that sensory processes preceed mental processes and can thus be sequentially decoupled. We propose revising the assumptions made about the input and output of a reasoner.

Which commonsense spatial reasoning?

For the purposes of this paper, we distinguish between *static* and *dynamic* spatial reasoning. Static spatial reasoning is concerned with spatial relations between stationary objects – is the chair to the left or the right of the table, or, can I stack two chairs on top of each other. Dynamic spatial reasoning is concerned with spatial relations of moving objects. We further distinguish between navigation tasks – how do I get from my office to the library – and other physical problems and situations – what happens if I push the glass past the border of the table. The later is the subject of this paper.

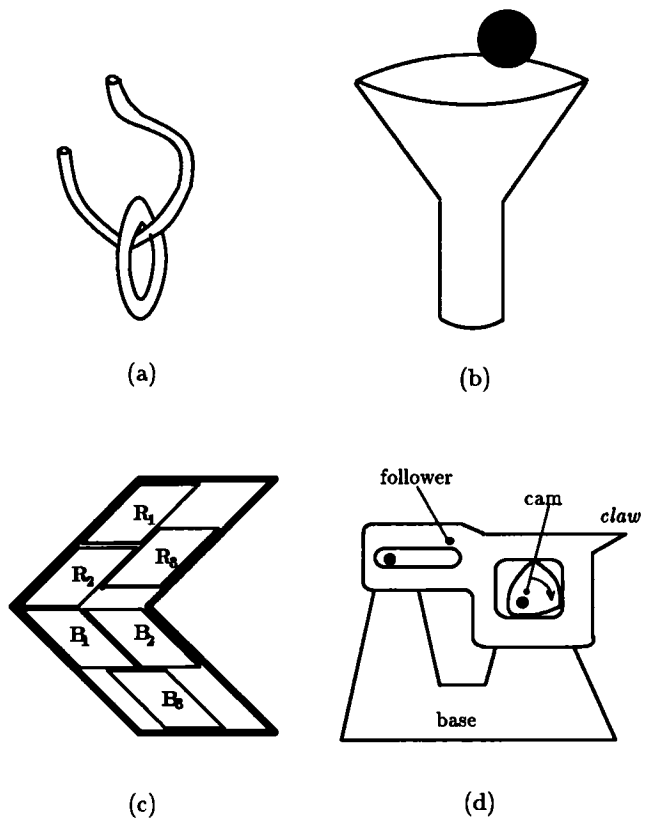

Figure 1: Commonsense spatial situations.

We will not attempt to precisely define what constitutes commonsense, or non-expert spatial reasoning.

The boundary is very fuzzy and highly dependent on previous experience. Instead, we will focus on systems' capabilities: the more types examples a reasoner can handle, the better! Fig. 1 shows four examples of commonsense spatial reasoning scenarios about moving objects: (a) a ring and a hook: can the ring get off the hook? (b) a ball and a funnel: what will be the final position of the ball? (c) a puzzle: the upper blocks cannot mix with the lower blocks (d) the film advance mechanism of a motion-picture camera: how does it work?

Four spatial reasoning paradigms

In this section, we briefly review the four main general computational approaches to spatial reasoning about moving objects: simulation, analytical, axiomatic, and analogical reasoning. We deliberately exclude cognitive psychology because our focus is computational (but see [Ph81] and [Ko81]), and special-purpose approaches such as path planning and vision because commonsense reasoning requires generality (but see the conclusion for the role of vision).

Simulation

A direct and visually useful approach to spatial reasoning is simulation. A simulator takes as input a geometric model of the objects, their initial positions, and the forces acting on them. It constructs the motion equations for each object and numerically integrates them at subsequent time snapshots, while testing for object collisions. The output is the object positions at each point in time, which can be graphically displayed as a "movie" of the objects moving as time progresses.

Existing simulators, both commercial [Ha84] and academic [Ho89] place relatively few constraints on the shape and number of objects in the scene. State of the art simulators can handle relatively complex situations, such as a chair falling in a staircase, an oscillating chain, and a car suspension. The output, when displayed as a sequence of object positions, is visually meaninful to humans. It is cognitively appealing as it matches the type of imagery produced by mental simulations. However, simulation has many drawbacks for automating commonsense reasoning. First, it does not support generalization because it only provides a solution for a specific initial condition. For example, a simulation cannot determine that the upper and lower blocks in Fig. 1(c) will never mix. Second, the results of a simulation are almost always too detailed to support reasoning at different levels of abstraction. A simulator will determine the trajectory of the ball going through the funnel in Fig. 1(b), but this description will be very detailed. Third, the output either requires extensive interpretation or is simply inappropriate to support tasks other than specific prediction. Fourth, simulators cannot handle uncertainty, approximate shapes, or detect behavioral ambiguities.

With the advent of faster computers and more sophisticated techniques, simulators will be capable of handling complicated situations, such as pouring sand in a bucket. However, it is unlikely that simulation as we know it now will eventually fully automate commonsense reasoning (although it might play an important role).

Analytical reasoning

Analytical reasoning proposes to symbolically reason about the constraints that objects impose on each other's motions. Given a geometric description of the objects, an analytic reasoner formulates the algebraic constraints that prevent objects from interpenetrating. Any objects' motions must satisfy these constraints. The nature of the constraints and the dimensionality of the motion space – also called the *configuration space* of the objects – depends on the geometry of the objects and their degrees of freedom. Spatial reasoning is thus reduced to deriving the topology and properties of the objects' underlying configuration space. Configuration spaces are the first principles theory of motion because they explicitly relate the objects' shapes and positions to their possible motions.

The power of the analytical method lies in their ability to find closed-form algebraic solutions (or their approximations) that concisely characterize the global motion properties of a set of objects. This severely constraints the shapes, degrees of freedom, and number of objects in a scene, thus reducing the scope of the method. When applicable, analytical reasoning supports a wide variety of tasks, including abstraction, generalization, expert and commonsense reasoning, and even simulation. The most succesful applications are in motion planning [Lo83], where very few objects move, e.g., Fig. 1(c), and reasoning about mechanical devices such as transmissions and door locks, where object motions are highly constrained, e.g., Fig. 1(d) [Fo87, Fa90, Jo89a, Jo89b, Jo90b].

Analytical reasoning can support an interesting variety of commonsense situations within a somewhat narrow scope. It cannot deal with objects with highly unconstrained motions, such as a chair falling in a staircase, nor make generalizations about many moving objects. Currently, it does not support inferences about classes of objects and can only deal with very simple dynamics or very simple situations, such as a swinging pendulum. Future research will most likely attempt to broaden the scope by finding appropriate approximations to object constraints and dynamics. For commonsense reasoning, analytical techniques are probably best used as "motion and geometry experts" for specialized situations.

Axiomatic reasoning

In axiomatic, or rule-based reasoning, object shapes and their motions are described by (logical) first-order

formulae. Object situations and interactions are described by axioms or rules consisting of antecedents and consequents. Antecedents specify situations and object properties which must hold for the consequences to be true. Conclusions are derived using a theorem prover or inference engine. The effectiveness of axiomatic reasoning depends on expressivity of the objects' and situations' languages and on the deductive power of the domain axioms. A classical example of axiomatic spatial reasoning is the program STRIPS, reasoning about stacking and unstacking blocks in the blocks world domain.

Axiomatic reasoning can in principle handle situations with potentially many objects and classes of shapes. However, it is extremely hard to develop ontologies and axioms, even for moderately complex spatial domains [Da90]. Predicates, which can adequately represent classes of object shapes, are inappropriate to describe elaborate shapes, such as a chair and their behaviors, such as bouncing or rolling. Rules tend to be highly interdependent and context sensitive. In addition, it is extremely difficult to determine if all the logically valid conclusions are also physically valid without exhautively enumerating them. The resulting systems are typically narrow in scope and potentially brittle (however, see [Ka91] for a recent promising ontology).

Recent research in axiomatic reasoning has captured interesting commonsense situations, such as an object falling inside a funnel [Da90] (Fig. 1(b)). However, it is very unlikely that axiomatic reasoning can be made general-purpose enough to support even a moderate variety of object shapes and situations. It is best suited for situations where the objects' geometry has been abstracted and the interactions between objects are simple and have been predefined. Thus, axiomatic reasoning is appropriate, for example, for dealing with a variety of puzzles and stylized situations, such as in Fig. 1(c), but not for reasoning about mechanical devices, such as Fig. 1(d)

Analogical reasoning

An interesting departure from the previous three approaches is analogical or diagrammatic reasoning. In this paradigm, the objects in a scene are described by a set of spatially adjacent pixels corresponding to object chunks, rather than symbolically by their names and geometry. The laws of motion are enforced by locally passing messages between adjacent object pixels. The messages indicate the pixels' next position in space. The pixels' position changes, corresponding to motions are then interpreted and symbolic information, such as contact changes and surface alignments, can be derived. Symbolic reasoning (typically axiomatic) can then take place, and further experiments or situations can be tried.

We find three prominent examples of spatial analogical reasoning in the litterature. Gardin and Meltzer's program [Ga89] simulates interesting commonsense situations, such as pouring water on a glass or pulling with a flexible string by passing messages between adjacent "molecules", or coarse object pixels. Funt's WHISPER program [Fu77] predicts the event sequence of collapsing block towers. It uses a retina-like structure to interpret and report the results of pixel movements caused by physical laws acting on the objects' pixels. Narayanan and Chandrasekaran [Na90] go a step further in integrating visual and symbolic processing by proposing a tighter interaction and the use of previously seen cases to determine what will happen next.

Analogical reasoning uses a different input from the previous three approaches. In doing so, it attempts to emulate human capabilities by bridging the gap between visual perception and symbolic reasoning. In principle, it can handle objects of arbitrary shapes with highly unconstrained motions. Because objects are typically described at a coarse resolution by few pixels, simulations using diagrammatic representations tend to be more efficient than traditional simulations – there are no equations to numerically integrate, and no collisions to detect globally. The resulting behaviors generally capture the behavior of a class of objects (due to the coarseness of the representation) and are mostly qualitatively correct. Moreover, by choosing different resolutions for describing the objects, some general conclusions can be derived.

Because analogical reasoning requires the ability to simulate motions, it shares some of the drawbacks of simulation. It is currently difficult to assess the exact scope and extent of analogical methods, especially with respect to situations as in Fig. 1. In a sense, analogical methods are the most open-ended because their architecture can accommodate many variants. A crucial question is understanding and properly balancing the relation between visual simulation and symbolic processing.

Evaluation

The previous survey shows the wide span and variation of existing spatial reasoning techniques. Currently, no single one satisfactorily addresses, even partially, the scope of commonsense spatial reasoning. Although hybrid reasoning techniques, such as combining simulation and analytical approaches, or axiomatic and analogical approaches, can broaden the scope of applicability, we belive that the problem is more fundamental and raises basic methodological questions.

The first question is the desirability and feasibility of a general-purpose spatial reasoner. Different situations require different levels of generality. Instead of building a reasoner that is capable of handling a vast variety of situations, we could be better off building a collection of reasoners that handle well-defined classes and tasks, and then integrate them together as necessary. A major challenge of this approach is identifying those classes and tasks that achieve the right

balance between generality and specificity. We believe that this is a most promising direction, based on partial evidence of the success of special purpose reasoners in other related fields such as robotics (in particular Brooks' work), vision, and geometric modelling.

A direct consequence of revising the "general-purpose reasoner" paradigm is reconsidering what are the appropriate forms of inputs and outputs and the assumptions made about them. Justifying these assumptions is crucial, for they can have major consequences on the methods' development. Also, comparisons between methods must take into account these assumptions.

More than any other commonsense reasoning task, spatial reasoning heavily relies on sensory-motor information. This directly challenges the accepted wisdom in formalizing commonsense reasoning, which assumes that sensory processes preceed mental processes and can thus be sequentially decoupled. Other areas in commonsense reasoning can very reasonably assume that symbolic reasoning about, say agents' beliefs, starts with a set of axioms from which conclusions are extracted. But is this appropriate for commonsense spatial reasoning? Simulation and analytical methods assume the input is a symbolic geometrical description of the objects. Axiomatic methods assume objects are described by predicates. Diagrammatic methods start with pixels. Will all of them be eventually necessary?

An interesting question is the role that vision and previous experience plays in commonsense spatial reasoning. Their computational counterparts, high-level machine vision and case-based reasoning might offer interesting computational insights to automating commonsense spatial reasoning. It is encouraging to note that recent research in diagramatic representations is beginning to address these issues.

References

Chandrasekaran, B. and Narayanan, H., "Towards a Theory of Commonsense Visual Reasoning"

Davis, E. *Representations of Commonsense Knowledge*, Morgan-Kauffman, 1990.

Davis, E. "A Logical Framework for Commonsense Deductions of Solid Object Behavior", *Artificial Intelligence in Engineering* **3**(3), 1988.

Faltings, B., "Qualitative Kinematics in Mechanisms", *Artificial Intelligence* **44**, 1990.

Forbus, K., Nielsen, P. and Faltings, B. "Qualitative Kinematics: a Framework" *Proceedings of IJCAI-87*, Milano, Italy, 1987.

Funt, B., "Problem-Solving with Diagrammatic Repesentations" *Artificial Intelligence* **13**(3), 1980.

Gardin, F. and Meltzer, B., "Analogical representations of Naive Physics" *Artificial Intelligence* **38**(2), 1989.

Hopcroft J., and Hoffman C., "The NEWTON Project", Cornell University, 1990.

Haugh, E., (Ed), *Computer Aided Analysis and Optimization of Mechanical Systems Dynamics*, Springer-Verlag, New York, 1984.

Joskowicz L. and Sacks, E., "Computational Kinematics", *Artificial Intelligence*, **51**(1), 1991.

Joskowicz L., "Simplification and Abstraction of Kinematic Behavior" *Proceedings of IJCAI-89*, Detroit, 1989. Reprinted in D. Weld and J. DeKleer eds, *Readings in Qualitative Reasoning about Physical Systems*, Morgan Kauffman.

Joskowicz L., "Mechanism Comparison and Classification for Design", *Research in Engineering Design* **1**(3), 1990.

Kaufman, S., "A Formal Theory of Spatial Reasoning", *Proceedings of IJCAI-91*, Morgan Kaufman, 1991.

Kosslyn, S., "The medium and the mesage in mental imagery: a theory" *Psychological review* **88**, 1981.

Narayanan, H. and Chandrasekaran, B., "Reasoning Visually about Spatial Interactions" *Proc. of IJCAI-91*, Sydney, Australia (1991).

Lozano-Pérez, T., "Spatial Planning: A Configuration Space Approach", *IEEE Transactions on Computers* **32**(2), 1983.

Pylyshyn, Z., "The imagery debate: analogue media versus tacit knowledge", *Psychological review* **88**, 1981.