How should depiction be represented and reasoned about?

Kenneth D. Forbus, Northwestern University

Interpreting a scene requires understanding how its visual properties and context yield evidence about the spatial and conceptual properties of what it depicts. Depiction is intimately tied to spatial language, since describing a scene linguistically, or imagining a scene described in language, involves connecting linguistic and spatial knowledge. We focus here on scenes described via sketching.

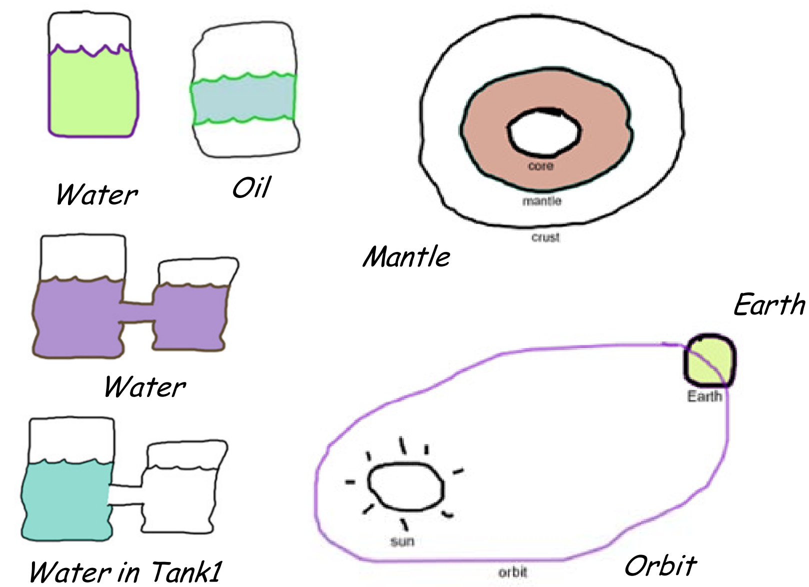

Figure 1: Examples of conceptual segmentation in depiction. The colored parts indicate what is considered to be part of the entity referred to via the given word or phrase.

A classic approach to this problem is to formulate it as constraint satisfaction (e.g. Mackworth 1977; Mulder et al 1988), typically in a specialized domain, such as maps. We believe that while constraint satisfaction is a useful approach, it represents only a piece of the puzzle. Here we describe two other approaches, both grounded in a large-scale knowledge base[1], that we believe constitute other pieces of the puzzle, and propose a corpus-gathering activity to build up via learning a broad-coverage model of depiction in sketches.

Depiction as conceptual segmentation

Understanding how to interpret sketches in meaningful ways requires knowing which parts of a diagram are meant when referred to linguistically. One important question is, is the area inside of something part of it or not? Consider the cases illustrated in Figure 1. Even though both are closed curves, the space inside the Earth is considered to be part of the Earth, while the inside of its orbit is not considered to be part of its orbit. Similarly, someone drawing a liquid in a container typically only draws the surface of the liquid, expecting that the viewer will understand what they mean because of their world knowledge. Lockwood (et al, 2008) showed that visual reasoning combined with conceptual and linguistic knowledge could be used to make such determinations. For example, since an orbit is a subconcept of `Path-Spatial`, only the line itself is considered to be part of the orbit, and not the area

[1] We use contents extracted from ResearchCyc (http://research.cyc.com) with our extensions for qualitative reasoning and analogical processing.

Dagstuhl Seminar Proceedings 10131
Spatial Representation and Reasoning in Language : Ontologies and Logics of Space
http://drops.dagstuhl.de/opus/volltexte/2011/2731

it encloses. Similarly, knowledge that the line representing a liquid is inside glyph representing a container enables the system to correctly figure out the spatial extent of the liquid. How many such conventions are needed to cover the range of diagrams that people encounter remains an open question at this point.

From visual to conceptual relationships

Given the kinds of entities that appear in a sketch, the relationships between the ink that depicts them suggests possible relationships between the entities themselves. In Figure 2, for example, the fact that the glyph representing the wheel touches the glyph representing the ground suggests that the wheel itself is above the ground and touching it. In general, there are quite a large number of relationships in the knowledge base that are *a priori* plausible given just the visual relationship between pieces of ink: When one glyph is inside another, there are over 150 possible relationships, and when one glyph is touching another, there are over 200 possible relationships. This number drops somewhat when further constrained by taking into account what the glyphs are intended to represent (e.g., wheel, ground), to 122 relationships on average, but finding the best relationship is still a daunting problem. A useful way to tackle this problem is via analogical reasoning (Forbus et al 2005). That system used a corpus of sketches, drawn by several people, who had used the sketching system's interface to supply the correct conceptual interpretation for the visual relationships that it automatically found. Given a new sketch, the system used analogical retrieval to find a similar prior sketch and analogical mapping to make specific suggestions for conceptual interpretations of visual relationships. This allowed it to provide suggestions 54% of the time, with an accuracy of 66%. We view this as a promising method for accumulating interpretation knowledge via examples, and believe that even more robust performance can be achieved by using analogical generalization.

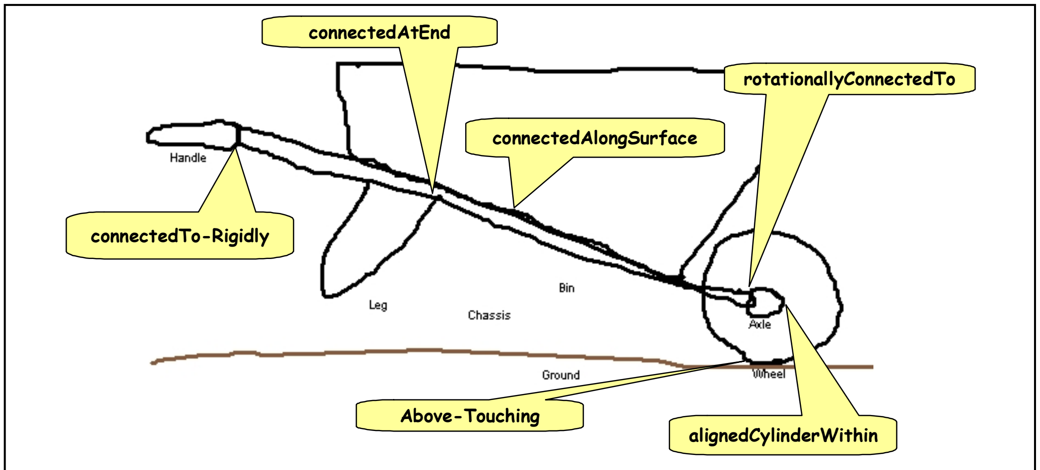

Figure 2: Conceptual relationships between the entities depicted based on the visual relationships between the glyphs that represent them

Accumulating depiction knowledge via corpus gathering and analysis

The sheer numbers of types of objects in the world and relationships between them makes modeling depiction a daunting challenge. Given its scale, crowd-sourcing via a game appears to be the only practical approach (von Ahn, 2006). CogSketch provides a useful platform for doing the reasoning underlying such a game, because it contains a model of visual processing and conceptual knowledge that provides a useful starting point for accumulating more knowledge. For an on-line game, one possibility is to gather ink and natural language using a lightweight application, for offline processing and learning via CogSketch.

Some aspects of depiction seem more amenable to this technique than others. For example, consider learning how the parts of something are depicted. One can imagine asking someone to draw something, for example a cat, while naming each part. (The cover story we are planning to use involves teaching an alien about our planet.) Similarly, asking someone to color in what is being referred to in a sketch when using a linguistic label for parts of it (e.g., examples like those in Figure 1, which were automatically generated by CogSketch) is a reasonable thing to expect people to do. However, selecting an appropriate relationship between the parts (e.g., examples like in Figure 2) will require substantially more natural interaction with players, to avoid asking them to understand the underlying ontology. Finding tradeoffs that make games attractive for players, while yielding high quality data, is a difficult challenge.

References

- Forbus, K., Usher, J. and Tomai, E. (2005). Analogical learning of visual/conceptual relationships in sketches. *Proceedings of AAAI-05*.
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. In *Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling*. Annecy, France.
- Lockwood, K., Lovett, A., Forbus, K., Dehghani, M., and Usher, J. (2008). A Theory of Depiction for Sketches of Physical Systems. *Proceedings of QR 2008*.
- Mackworth, A. 1977. On reading sketch maps. *Proceedings of IJCAI 1977*.
- Mulder, J., Mackworth, A., and Havens, W. 1988. Knowledge structuring and constraint satisfaction: The Mapsee approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**(6), pp 866-879.
- von Ahn, L. 2006. Games with a purpose. *IEEE Computer*, pp 96-98.