Reasoning About Sketches Using Context, Domain Knowledge, and Interaction with the User

Aaron Adler
Raytheon BBN Technologies
10 Moulton Street
Cambridge, MA 02138
aadler@bbn.com

Abstract

Visual information can be communicated using informal sketches of the sort used by people in design conversations. These sketches can be captured using TabletPCs, however, they can be hard or impossible to understand without additional context, domain knowledge, and interaction with the user. We illustrate the utility of these components with examples and then describe a system, MIDOS, that uses these components to reason about a simple mechanical design.

Introduction

Informal sketching is a visual representation that enables people to communicate complex ideas or situations to each other. People have created drawings for thousands of years, but only relatively recently with computers, starting with (Sutherland 1963). Digital sketches (e.g., on TabletPCs) allow not only the capture of the final drawing but the sequence of timed points used to create the drawing. This allows for recognition of the sketch into high-level components that can be used to understand the sketch. For example, a digital sketch of an electronic circuit can easily be turned into a simulation and then further refined. Figure 1 is a sketch of an AC/DC transformer with red and blue annotations showing current flow. (The annotations and large gaps between components in this sketch would be difficult for current recognition systems to handle.) Considerable progress has been made in creating effective sketch-based user interfaces for a variety of applications, including web design (Newman et al. 2003), circuit analysis (Gennari et al. 2005), education (Forbus et al. 2008), and organic chemistry (Ouyang and Davis 2007).

This paper discusses two areas relevant to Visual Representations and Reasoning: sketch understanding and multimodal representations and reasoning. Sketching represents visual information well, but understanding and reasoning about sketches and the visual information they represent also requires: context, domain knowledge, and interaction with the user. We discuss each of these ideas, briefly describe a working system, MIDOS, that uses these ingredients, and describe how MIDOS reasons about the multimodal information.

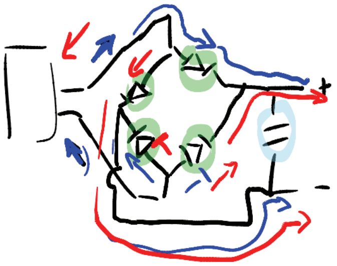


Figure 1: A sketch of a circuit. Red and blue ink are used to illustrate two different current paths in the AC/DC transformer.

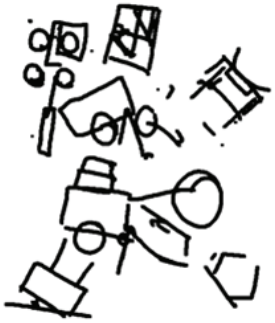


Figure 2: A sketch that is difficult to reason about without the accompanying speech.

Context

Several user studies that we've conducted (Adler and Davis 2009; Adler 2009) have examined how people use sketches to convey visual ideas in conversations. In many domains, however, sketches alone are inadequate for effective communication. This is particularly true in engineering design, where informal sketches are invariably incomplete and are typically accompanied by a verbal description that fills in the "blanks." Additional context information may come from speech, gesture, or text; in this paper we focus on speech. The speech and sketching that occur in conversations has motivated work on combining these modalities (e.g., (Cohen et al. 1997)).

Sketching is typically used for communicating geometric and spatial information, such as the shape or location, while speech is used for describing device behavior or properties. That speech, although informal, can convey a considerable amount of information and additional context. Research has shown that the combination of speech and sketching provides more information than either modality alone (Bischel et al. 2009). This extra information is critical to being able to reason about the sketch.

Figure 2 is an illustrative example of a sketch that is difficult to make sense of without the accompanying speech. The sketch makes considerably more sense when it is compared to the robot it describes (Figure 3). The lower portion of the sketch is a side view of the robot with the large circle corresponding to the three large wheels in the photograph. Other parts of the sketch are different views of the robot or its components.

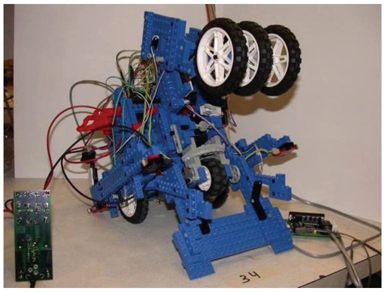


Figure 3: The robot that was sketched in Figure 2.

Domain Knowledge

Domain knowledge is also an important component to understand a sketch. A critical fact in one domain might be irrelevant in another. Figure 1 is illustrative of the informal sketches that users draw. To understand the sketch, one must understand the set of symbols it contains (capacitors, diodes, etc.) as well as the connections between them (wires). The capacitor, circled in light blue, should have wires connected to it forming a "T" shape. Similarly there are other gaps in the sketch between components that should be connected. In contrast, the crossing wires in the circuit diagram, shown in detail in Figure 4, mean that the crossing wires are *not* connected. In this sketch, some crossing lines indicate disconnected wires and some disconnected lines indicate a connection. As this example illustrates, inferring even a simple property such as connectedness, requires domain specific knowledge.

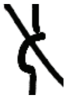


Figure 4: Two wires that intersect but are not connected illustrating the domain knowledge necessary to understand the sketch in Figure 1.

Interaction with the User

Even with the first two ingredients, domain knowledge and context, it is possible that the user draws something that is truly ambiguous. In these cases, no matter how much reasoning the computer does about the sketch, it will not be able to determine the correct answer.

For example, if the user is drawing part of a circuit diagram and draws the symbol shown in Figure 5(a), even a person would have trouble determining if it is a battery (Figure 5(b)) or a capacitor (Figure 5(c)). In situations like this, a person would just ask a question about the symbol. For example, a person might circle the ambiguous symbol and ask "Is this a capacitor or a battery?"

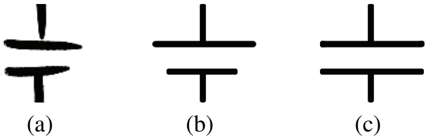


Figure 5: An ambiguous circuit component (a) that could either be a battery (b) or a capacitor (c).

MIDOS

A computer should take the same approach: first it should reason about all the components that it can and then it should ask the user questions about the rest of the sketch. MIDOS, Multimodal Interactive DialOgue System, is a multimodal dialogue system which uses sketching and speech for input and output to engage the user in a symmetric multimodal dialogue about a simple mechanical device. Here we give a brief description of MIDOS to highlight how it uses context, domain knowledge, and interaction with the user. More details can be found in (Adler and Davis 2009; Adler 2009).

The domain for MIDOS is simple, Rube-Goldberg style, mechanical devices. The whimsical "egg cracker" in Figure 6 (from (Narayanan, Suwa, and Motoda 1995)) is an example of the sort of device MIDOS can simulate. When the (green) stopper on the left is pulled up, the spring-and-block pushes the second block off the edge of the platform. That block falls, causing the platform below to rotate counterclockwise. This causes the triangular knife to move downward, pushing the egg into the frying pan. Although this description may seem complete, there are in fact several details missing that are required if we want the device to behave as desired. For example, initially the knife must be in balance with the rotating platform.

MIDOS attempts to determine the interaction between the components. It contains a qualitative physics simulator, providing domain knowledge, that allows it to interpret the device and ask a question whenever it finds that it cannot predict the next qualitative state of the device. The user answers these questions using a combination of sketching and speech. MIDOS knows that it has understood the device correctly when it has correctly simulated the device.

There is no fixed script or set of fields to fill in; somewhat like a human observer the system asks just what it needs

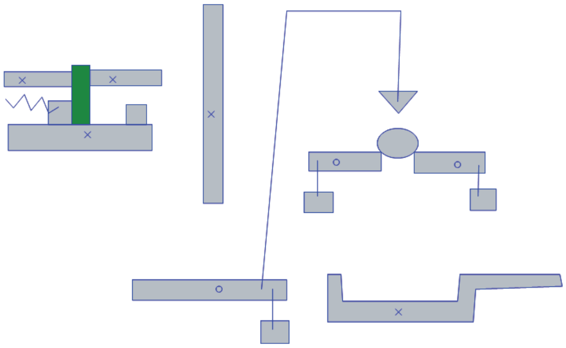


Figure 6: A typical device, an egg cracker, that MIDOS can discuss (the stopper is highlighted in green for purposes of identification). Devices are built from a small vocabulary of objects including: bodies, springs, pulleys, weights, pivots (drawn as small circles), and anchors (drawn as small 'x's).

to know to understand the device being designed (where by "understand" we mean being able to predict what will happen next). These questions are multimodal and use a coordinated combination of sketching and speech. MIDOS uses the additional context the speech provides to understand the user's visual input. By asking even a simple, but appropriate, question, the system can gather additional information. The next section describes how the system reasons about this input.

Multimodal Reasoning

MIDOS reasons about the user's multimodal response to questions in part by using the fact that the answer is (most likely) in response to the question it asked. The user is not restricted in what they can say or draw, so the system must be flexible in interpreting the answers. MIDOS accomplishes this by using a set of expected answers (templates) to the question. Each template is associated with a function that will resolve to an answer to the question. Simple questions may resolve to a yes or no answer while more complex questions may resolve to a function on the strokes the user drew.

The user's sketched and spoken inputs are then separately compared to the expected answers in the templates. Finally, the matches for speech and sketching are compared to determine if they are consistent. If the system determines that the modalities are consistent, it can update its physics model. If it determines that the answer are not consistent, it can ask a follow up question.

Table 1 is a partial illustration of the reasoning for sketching and speech input. In this case, the user is responding to a question: "Will this spring expand or contract?" These examples illustrate a successful match as well as two cases where the user will be prompted for further information because of missing or conflicting information. Using this technique, MIDOS allows the user freedom to describe the simple mechanical device while still obtaining enough information to understand their response. Using the physics simulator and asking questions, MIDOS can simulate the operation of the device.

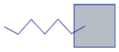
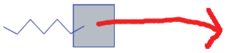
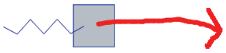
User Sketching	User Speech	Consistency Check Result
	"It moves in this direction"	Insufficient
	"It contracts"	Conflict
	"It expands"	Success

Table 1: A visual summary of possible consistency check results.

Conclusion

User sketches contain rich visual information. Three key ingredients to understanding the sketches are context, domain information, and interaction with the user. MIDOS illustrates how these ideas can be used in an actual system.

MIDOS does not attempt to understand the full details that the user provides in response to the questions. These detailed responses are a good target for further research on visual representations and reasoning. Future work should study how to interpret the user's open ended responses and investigate ways to use this additional information to benefit the users.

Acknowledgements

This work was supported in part by NSF Award 0729422. The author thanks Fusun Yaman and the anonymous reviews for their helpful feedback. Also thank you to Professor Randall Davis and the members of the Multimodal Understanding Group at CSAIL for all of their help and advice.

References

Adler, A., and Davis, R. 2009. Symmetric multimodal interaction in a dynamic dialogue. In *2009 Intelligent User Interfaces Workshop on Sketch Recognition*. ACM Press.

Adler, A. D. 2009. *MIDOS: Multimodal Interactive DialOgue System*. Ph.D. Dissertation, M.I.T.

Bischel, D.; Stahovich, T.; Peterson, E.; Davis, R.; and Adler, A. 2009. Combining speech and sketch to interpret unconstrained descriptions of mechanical devices. In *Proceedings of the 2009 International Joint Conference on Artificial Intelligence (IJCAI)*.

Cohen, P. R.; Johnston, M.; McGee, D. R.; Oviatt, S. L.; Pittman, J.; Smith, I.; Chen, L.; and Clowi, J. 1997. QuickSet: Multimodal interaction for distributed applications. In *Proceedings of Mutimedia '97*, 31–40. ACM Press.

Forbus, K.; Usher, J.; Lovett, A.; Lockwood, K.; and 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*.

Gennari, L.; Kara, L. B.; Stahovich, T. F.; and Shimada, K. 2005. Combining geometry and domain knowledge to interpret hand-drawn diagrams. *Computers and Graphics* 29(4):547 – 562.

Narayanan, N. H.; Suwa, M.; and Motoda, H. 1995. *Hypothesizing Behavior from Device Diagrams*. MIT Press. chapter 15, 501–534.

Newman, M. W.; Lin, J.; Hong, J. I.; and Landay, J. A. 2003. DENIM: An informal web site design tool inspired by observations of practice. *Human-Computer Interaction* 18(3):259–324.

Ouyang, T. Y., and Davis, R. 2007. Recognition of hand drawn chemical diagrams. In *Proceedings of AAAI*, 846–851.

Sutherland, I. B. 1963. Sketchpad, a man-machine graphical communication system. *Proceedings of the Spring Joint Computer Conference* 329–346.