

A Framework for Multi-Domain Sketch Recognition

Christine Alvarado and Randall Davis

MIT Artificial Intelligence Laboratory
 {calvarad,davis}@ai.mit.edu

Introduction

We present a multi-domain architecture for sketch recognition systems that we believe will make these systems both easier to construct and more robust in operation. At the heart of our approach is a method that combines high-level, domain-specific information with low-level, domain-independent recognizers.

We aim to build a recognition framework that can be applied naturally and efficiently to a variety of domains, yet takes advantage of the power that comes from context. We are motivated by the success of using domain-knowledge for speech understanding, and inspired by the design of the Hearsay-II System (Erman *et al.* 1980). Hearsay-II combined knowledge at various levels of the speech interpretation process, including the syllable, word, and phrase levels, to generate and choose from multiple interpretations of a spoken utterance. We believe a similar architecture can be effective in sketch understanding.

There are a variety of issues to be addressed in such an undertaking; this paper focuses on just one of them: using domain-specific knowledge to guide recognition.

Knowledge Representation

The system combines three types of domain-specific knowledge to aid recognition:

Domain-Specific Patterns Patterns particular to a given domain, defined using a shape description language to specify the component shapes that make up the pattern and the geometric properties of and between these components (e.g. Figure 1).

Temporal Context Information about the order in which the domain-specific patterns, as well as strokes that make up those patterns, are likely to be drawn.

Spatial Context Information about configurations of domain-specific patterns that are likely to occur.

The three knowledge sources are combined in a Bayesian network framework. Each domain-specific pattern specification is translated into a fragment of a Bayesian network, with a node for that pattern connected to nodes for each of the low-level shapes and

their properties (Figure 1). Conditional probability tables are constructed taking into consideration how likely the low-level shapes are to occur both within and outside of that pattern. Priors on the root nodes are influenced by the temporal and spatial context in which the shape is hypothesized to occur. Note that the shapes and properties at the bottom are “observed” with some degree of confidence depending on how well the data fit the shape or property.

Recognition Algorithm

Recognition involves mapping a set of patterns to the user’s strokes. Our algorithm generates a number of possible *interpretations*—a mapping from a set of strokes to a single high-level pattern—by combining bottom-up pattern activation with top-down interpretation. These interpretations are then pruned using the notion of “islands of certainty” developed in Hearsay-II.

There are four steps in our recognition algorithm:

1. **Bottom-up Step:** As the user draws, the system parses the strokes into ovals, lines, and arcs using a domain-independent recognition toolkit developed in previous work (Sezgin 2001). New interpretations are hypothesized by instantiating the Bayesian network fragments that specify high-level patterns that include these low-level shapes, even if not all the sub-components of the pattern have been recognized.
2. **Top-down Step:** The system then identifies the missing subcomponents and attempts to reinterpret strokes that are temporally and spatially proximal to the proposed shape to fulfill the role the missing component. If, for example, the system had detected a body arc and two wires of the and-gate in Figure 1, it would try to reinterpret spatially and temporally adjacent strokes as lines to complete the body.
3. **Ranking Step:** Based on previously interpreted parts of the sketch, the system identifies temporal and spatial context for the newly recognized patterns and propagates conditional probabilities through the network using prior probabilities for the root nodes influenced by context. The system then explores sets of interpretations for the user’s strokes starting with the highest ranked individual interpretation (an island of certainty) using a best-first-search method until it generates n possible sets.

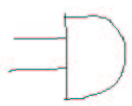
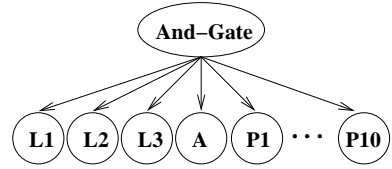
Sketch	Description	Network Fragment
	<p>DEFINE AND-GATE</p> <p><i>L1, L2, L3</i>: line L1 L2 L3;</p> <p><i>A</i>: arc A</p> <p><i>P1</i>: parallel L1 L2</p> <p><i>P2</i>: same-horiz-position L1 L2</p> <p><i>P3</i>: same-length L1 L2</p> <p><i>P4</i>: connected A.p1 L3.p1</p> <p><i>P6</i>: connected A.p2 L3.p2</p> <p><i>P5</i>: meets L1.p2 L3</p> <p><i>P7</i>: meets L2.p2 L3</p> <p><i>P8</i>: semi-circle A1</p> <p><i>P9</i>: orientation(A1, 180)</p> <p><i>P10</i>: vertical L3</p>	

Figure 1: The description of an and-gate symbol includes the properties and low-level shapes that compose it. Each of these shapes and properties becomes a node in a Bayesian network fragment.

4. **Pruning Step:** The system accepts any interpretations that have probability above a threshold and eliminates any interpretations not appearing in the sets generated in step three. All other interpretations are deemed possible and are considered in relation to the user’s next strokes when step one repeats.

Discussion

This approach provides seamless integration of bottom-up and top-down recognition. Bayesian networks are a natural tool for allowing low-level information to influence expectation of high-level components and in turn other low-level patterns.

Our system combines three types of knowledge by allowing the spatial and temporal context to alter the prior probabilities of the root nodes in the Bayesian network. Exactly how context should influence these probabilities is a non-trivial question at the heart of our approach that we will explore through experimentation.

A further challenge is how to enter the knowledge into the recognition system in the first place. Grammars are tedious to write and in previous work we found that explicitly specifying context, while possible, is difficult. We are currently investigating ways to learn the grammars (as (Do & Gross 1996) has attempted), as well as the temporal and spatial information, from examples.

Related Work

Shape description grammars were introduced formally by Stiny and Gips (Stiny & Gips 1972) and have been used mainly for generation of patterns. While they fell out of favor for pattern generation, we believe they are a feasible approach for recognition because under-constrained productions are acceptable for recognition, but not for generation.

Other sketch recognition systems include those developed by Landay and Meyers (Landay & Myers 2001), Do and Gross (Do & Gross 1996), Forbus *et. al.* (Forbus, Ferguson, & Usher 2001) and Stahovich (Stahovich 1999). Each system copes with recognition ambiguity in a different way. Our previous work (Alvarado & Davis 2001) uses context to disambiguate between multiple interpretations of a sketch, but is still driven by low-level recognition accuracy. The work described here differs from previous systems in its ability to allow high-level

interpretations to guide low-level recognition accuracy.

A limited amount of work in using top down information to guide real-world visual interpretation exists, including (Bienenstock, Geman, & Potter 1997). An advantage of sketch data over real-world image data is that sketches are highly stylized, so the problem of locating (but not recognizing) low-level shapes is lessened.

Current Status

We are still implementing this algorithm and have not yet tested it. We have a complete low-level, domain independent recognition tool-kit, as well as a code library for recognizing geometric properties of and relations among shapes. To test these ideas we intend to implement this system on an initial domain (e.g., a restricted set of symbols from digital electronics) and report the results at the Workshop.

References

- Alvarado, C., and Davis, R. 2001. Resolving ambiguities to create a natural sketch based interface. In *Proceedings of IJCAI-2001*.
- Bienenstock, E.; Geman, S.; and Potter, D. 1997. Compositionality, mdl priors, and object recognition. In M. C. Mozer, M. I. Jordan, T. P., ed., *Advances in Neural Information Processing Systems 9*. MIT Press. 838–844.
- Do, E. Y.-L., and Gross, M. D. 1996. Drawing as a means to design reasoning. *AI and Design*.
- Erman, L.; Hayes-Roth, F.; Lesser, V.; and Reddy, R. 1980. The hearsay-ii speech-understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys* 12(2):213–253.
- Forbus, K.; Ferguson, R.; and Usher, J. 2001. Towards a computational model of sketching. In *IUI '01*.
- Landay, J. A., and Myers, B. A. 2001. Sketching interfaces: Toward more human interface design. *IEEE Computer* 34(3):56–64.
- Sezgin, T. M. 2001. Early processing in sketch recognition. Master’s thesis, Massachusetts Institute of Technology.
- Stahovich, T. F. 1999. Learnit: A system that can learn and reuse design strategies. In *1999 ASME Design Engineering Technical Conference Proceedings*.
- Stiny, G., and Gips, J. 1972. Shape grammars and the generative specification of painting and sculpture. In Freiman, C. V., ed., *Information Processing 71*. North-Holland. 1460–1465.