Using a Visual Routine to Model the Computation of Positional Relationships

Andrew Lovett (andrew-lovett@northwestern.edu)
Kenneth Forbus (forbus@northwestern.edu)
Qualitative Reasoning Group
Electrical Engineering and Compute Science, 2133 Sheridan Rd.
Evanston, IL 60208 USA

Abstract

Modeling the encoding of visual stimuli is a complex, and often ignored, problem in computational models of visual and spatial problem solving. This paper outlines a toolkit for exploring encoding for two-dimensional visual scenes, *Visual Routines for Sketches*. The utility of this approach is shown by a new model for computing positional relationships, the *Vector Symmetry* model, that explains data from seven experiments and is more parsimonious than Regier & Carlson's (2001) AVS model.

Keywords: visual perception; spatial reasoning

Introduction

A number of models have explored how people reason about visual stimuli and solve spatial problems (e.g., Carpenter, Just, & Shell, 1990; Goldstone & Medin, 1994). However, they typically do not model the processes by which stimuli are first encoded. Human perceptual processes put important constraints on what visual and spatial representations are available for reasoning. Incorporating models of the computation of visual features is an important step for creating more complete visual and spatial models.

Ullman (1984) proposed that people have access to a set of *elementary operations*, operations we can run over our visual working memory to extract information. This finite set of operations can be combined in different ways to create a near-infinite set of *visual routines* for computing different spatial features and relations.

A number of computer models have been based on the idea of visual routines. However, many of these models are designed only to solve a particular problem (e.g., Chapman, 1992; Horswill, 1995), and thus miss out on the generality promised by the original idea. Rao (1998) constructed a system for both learning and performing visual routines for solving different spatial problems. However, because his focus was on controlling a physical robot, the elementary operations in his system are often more complex and higher-level than the simple operations proposed by Ullman.

We are developing Visual Routines for Sketching (VRS) as a platform for experimenting with computational models of perception. It provides a set of low-level elementary operations, supported by the psychophysics and cognitive psychology literature. Using these operations, researchers can construct visual routines based on their theories for how a particular spatial feature is computed. These routines can be run and evaluated on two-dimensional visual scenes created in or imported into CogSketch[1] (Forbus et al., 2008), an open-domain sketch understanding system.

This paper uses VRS to implement a model for the computation of positional relations. Positional, or projective, relations describe the location of one object, the *target*, relative to another, the *referent*, in a visual scene. A number of researchers (Logan & Carlson, 1996; Hayward & Tarr, 1995; Gapp, 1995; Regier & Carlson, 2001) have studied how people compute these relations. Regier and Carlson demonstrated several different factors that independently contribute to participants' assessments of whether a target is "above" a referent. They built a mathematical model which predicted all these factors and correlated closely with human data.

While the Regier and Carlson model helped reveal what factors people consider in computing positional relations, it does not describe the actual processes used by humans in performing the computation. Here we show that a parsimonious VRS model can achieve similar results on Reiger and Carlson's data.

We begin with a brief introduction to VRS. We then summarize prior research on positional relations. We show how VRS can be used to construct a new, simple model of positional relations, the *Vector Symmetry* model. We then test our model's ability to match human results on the full set of seven experiments run by Regier and Carlson (2001). Finally, we conclude and discuss future work.

Visual Routines for Sketching

Visual Routines for Sketching (VRS) is built into the CogSketch sketch understanding system. Users can create stimuli in CogSketch either by drawing with a pen or by importing shapes built in PowerPoint. VRS works directly with the *ink* of the sketch, the lines representing the edges of each object. Thus, it avoids edge segmentation issues.

Basic Representation

Ullman (1984) suggested that the human perceptual system uses a bottom-up, parallel approach to build an initial *basic representation* of the visual world. VRS computes a basic representation via two steps: First, the ink is projected onto a retinotopic map, a simplification of V1 in the primary visual cortex which represents the orientation of any edges

[1] Available for download at:
http://silccenter.org/projects/cogsketch_index.html

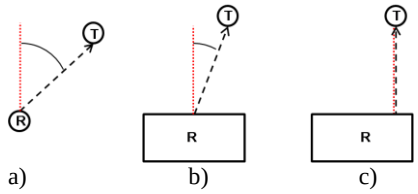


Figure 1: Ratings for "above" depend on the target's location relative to the reference (a), relative to the reference's center-of-mass (b), and relative to the reference's proximal point (c).

at each location in the image. This produces a set of *edge activations* at various locations. Second, edge activations are grouped together to form *contours*. This step is based on the contour integration literature (Yen & Finkel, 1998; Li, 1998), which suggests that there is a parallel process in which individuals group edges together based on the Gestalt grouping principles of good continuation and closedness. To these principles we add the constraint of uniform connectedness (Palmer & Rock, 1994).

Incremental Representation

Ullman proposed that there are a set of elementary operations that can be applied serially to the basic representation. By combining these operations into visual routines, an individual can both gather information and update the representation, thus producing an *incremental representation*. In VRS there are three key elementary operations, inspired by Ullman's proposal, which gather data and add elements to the incremental representation:

1) *Curve tracing* traces along consecutive edge activations. It produces a *curve*, a new grouping of activations which may lie along one or multiple contours.

2) *Scanning* begins at one location and moves forward in a fixed direction. It produces a straight curve representing the line scanned over.

3) *Region coloring* fills in the area between curves and contours, creating a new *region*.

All three operations can be constrained in several ways, e.g., curve tracing along a region, region coloring along a curve, or scanning between two points. These operations can be used to gather new information, detecting what other elements lie along a curve or within a region. The elements they produce can also be queried to access their attributes, such as the size of an element, the center of an element, the curvedness of a curve, or the orientation of a straight curve.

Current State of VRS

At present, VRS contains the elementary operations described above, as well as others for marking locations, inhibiting elements, and grouping elements to form *objects* (Kahneman et al., 1992), mid-level representations that serve as a bridge between the visual and the conceptual. However, we are still in the process of determining the full set of operations and the ways they can interact. Eventually, we hope to develop a simple coding language which will allow users to build their own visual routines by combining elementary operations in novel ways.

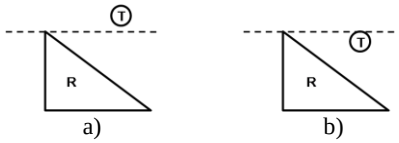


Figure 2: Targets lying above the grazing line (a) receive much higher "above" scores than targets lying on or below the grazing line (b).

Positional Relations

The positional relations most commonly studied are above/below and left/right. For simplicity, we will use the "above" relation for all examples throughout this section. However, in most cases researchers have studied either "above" and "below," or all four of the relations together.

Positional relations are typically studied in an assessment task. Participants are shown a proposition, such as "X is ABOVE Y," followed by a visual scene containing X and Y. They then state whether the proposition is true or rate the proposition on a numerical scale. Much of the research based on this paradigm (Logan & Sadler, 1996; Hayward & Tarr, 1995; Gapp, 1995) suggests that participants' ratings are based on the orientation of a line drawn from the referent to the target (see Figure 1a). If this line is perfectly vertical, the example is an ideal instance of "above." As the angle between this line and a vertical reference line increases, the ratings decrease at a linearly rate. As the angle approaches 90 degrees, the ratings drop more sharply, approaching 0 for a target that lies directly beside or even below the referent.

Studies by Regier and Carlson (2001) teased apart four different factors and showed that each one contributed independently to assessments of positional relations. The first is *center-of-mass* orientation, i.e., the target's deviation from directly above the referent's center-of-mass (Figure 1b). Importantly, this is distinct from the second factor, *proximal* orientation. The proximal orientation describes the target's location relative to the closest point on the referent (Figure 1c). The third factor is the *grazing line*, the horizontal line at the level of the topmost point of the referent (see Figure 2). As the target approaches and then falls below the grazing line, ratings for "above" fall sharply—this explains the nonlinearity as the angle between the referent and the target approaches 90 degrees.

The final factor is an interaction between center-of-mass orientation and the distance between the referent and target (see Figure 3). When the target is far above the referent, deviations in the center-of-mass orientation will result in a

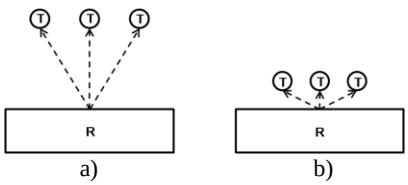


Figure 3: Targets far above the referent (a) differ more in their "above" ratings than targets immediately above the referent.

noticeable drop in "above" ratings. In contrast, when the target lies immediately above the referent, especially for a wide referent, changes in the center-of-mass orientation will have little effect on "above" ratings.

The AVS Model of Positional Relations

Regier and Carlson's Attentional Vector-Sum Model (AVS) for positional relations assessment consists of two components: the vector sum and the grazing line. The vector sum component computes vectors from every point along the referent to the target (see Figure 4). It takes a weighted sum of the orientations of these vectors, with the distribution of the weights depending on the proximity of the target to the referent. The summed orientation is compared to the vertical reference line (for "above") to determine angular deviation. The second component is the grazing line, which looks at the height of the target compared to both the topmost point and bottommost point of the referent. A sigmoid function is applied to these heights and averaged.

Regier and Carlson evaluated their model, along with three other models, on a set of seven studies designed to test the influence of the four factors described previously on "above" ratings. While all four models showed a strong correlation with human ratings, only the AVS model correctly predicted that all four factors would affect the ratings. Regier and Carlson argued this demonstrated that the AVS model best described how humans compute positional relations.

AVS is a strong mathematical model of the factors that contribute to assessing positional relations. However, we believe it does not describe the cognitive processes used by humans in computing positional relations. Firstly, the vector sum component requires computing a large number of vectors. Evidence from curve tracing (Jolicoeur et al., 1986) suggests that individuals move their attention along a line in a serial manner, and that the trace is slowed if there are other distractor lines nearby. Thus, drawing a large number of lines between points along the referent and the target would be a serial process requiring a significant amount of time. It is unclear how else these vector orientations could be computed.

Secondly, the grazing line component is underspecified. While people might use the heights of the topmost and bottommost points of the referent, it is unclear what processes are used to compute these points.

An Alternative Model: Vector Symmetry

We believe Regier and Carlson's results can be explained using a different, simpler model: the Vector Symmetry Model (VS). Like the AVS model, the VS model requires computing vectors from the referent to the target. However, the VS model computes vectors from only two points: the leftmost and rightmost points along the upper surface of the referent (see Figure 4). The model then examines the symmetry of these vectors' orientations about the y-axis, as measured by the difference between vector A's orientation and vector B's orientation reflected across the y-axis. If they are perfectly symmetric, the stimulus is an ideal example of "above." As the symmetry deviates, the model gives lower ratings for "above."

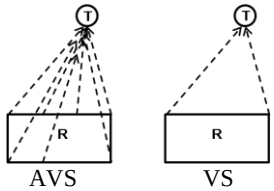


Figure 4: Vectors computed for the AVS and VS models.

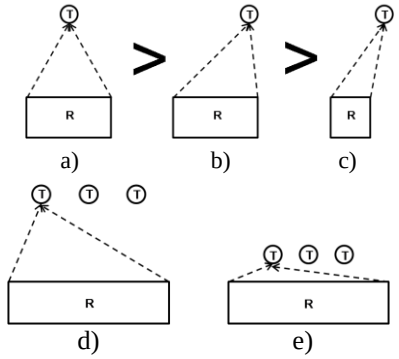


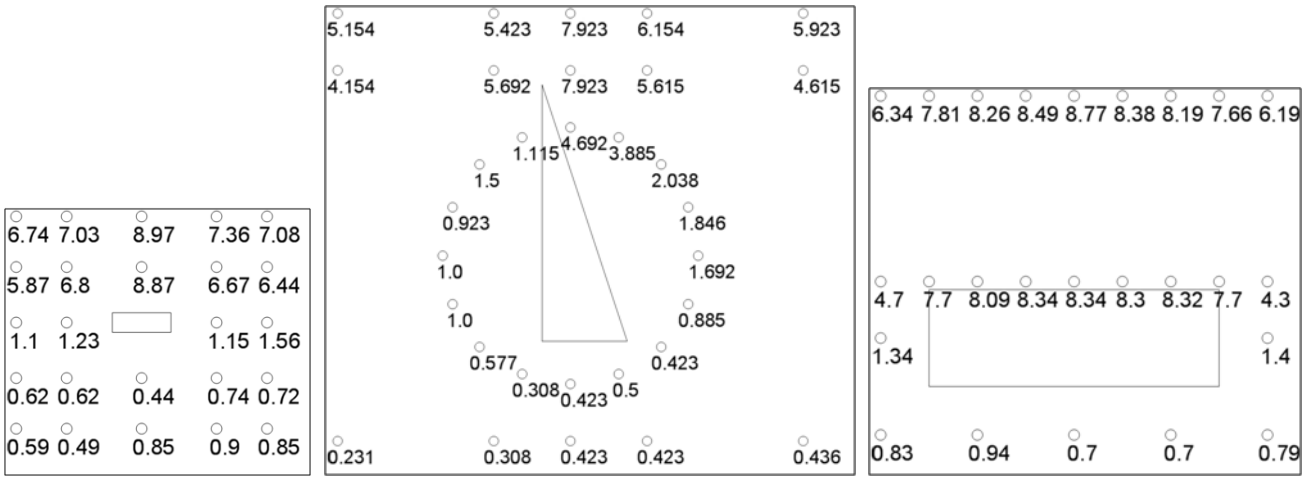
Figure 5: Using the VS model. As center-of-mass orientation changes, (a)->(b), symmetry decreases. As proximal orientation changes, (b)->(c), symmetry decreases. For far away targets (d), changes in center-of-mass orientation result in noticeable drops in symmetry. For near targets (e), changes in center-of-mass orientation have little effect on symmetry.

Like the vector sum component of the AVS model, the VS model predicts three of the four factors presented by Regier and Carlson: center-of-mass orientation, proximal orientation, and the interaction between distance and center-of-mass (see Figure 5). Like AVS, it requires a separate component to explain the grazing line. However, this component is computable from these two vectors. In cases where either the leftmost or rightmost point also lies along the referent's grazing line (in Figure 4, both do), that point's vector will approach the horizontal orientation as the target approaches the grazing line. Thus, the VS model uses the individual orientations of its two vectors to detect the height of the target relative to the referent's grazing line.[2]

Computing Positional Relations in VRS

The Vector Symmetry model requires only two input values: the orientations of the vectors from the top rightmost and top leftmost points of the referent to the target. We compute these values via two visual routines, described here in simplified form:

[2] In cases where neither of the points lies along the referent's grazing line, a third vector from the referent might need to be computed. Because this is not the case in any of the data currently being evaluated, we defer this question to a future time.



a) Experiment 2-wide b) Experiment 6 c) Experiment 7

Figure 6: Stimuli for three of the Regier & Carlson (2001) experiments, as entered into CogSketch. The numbers are human ratings for each position as an instance of "above."

Find objects in the visual scene
1) **Region Coloring**: Color the ground, locate any contours in it. 2) **Curve Tracing**: Trace each contour to determine whether it is a closed shape. 3) **Region Coloring**: If a contour is a closed shape, color the area inside it to identify its interior. 4) **Object Creation**: Make an *object* for each curve and the accompanying interior region.

Computing vectors for positional relations
1) **Scanning**: Scan from the referent's center upward to locate a point *pointTop* along the top of the referent's surface. 2) **Curve Tracing**: Trace the referent's curve clockwise and counter-clockwise from *pointTop* to find the rightmost and leftmost points along its top (*pointR* and *pointL*). 3) **Scanning**: Scan from *pointR* and *pointL* to the target's center to produce two new curves, *curveR* and *curveL*. 4) **Attribute Access**: Sample the orientation of the two curves to produce two orientation values, *oR* and *oL*.

Which object is the referent and which object is the target is currently indicated by labeling the sketch. We plan to explore doing this automatically via a visual routine that, for example, compared the sizes of the two objects.

Formula for "Above" Ratings

As stated above, the VS model contains two components: vector symmetry and a grazing line estimate. Vector symmetry is computing by reflecting the orientation oR about the y-axis and comparing it to oL:

SymmetryDist = X-Reflection (oR) – oL

where a **SymmetryDist** of 0 indicates perfect symmetry.

The second component, the grazing line, is also computed from oR and oL. Studying the results for Experiments 6 and 7 from Regier & Carlson (see Figure 6), we noted that when a target does not lie directly above the referent, i.e., it lies to the left or to the right, its "above" ratings fall sharply as it approaches the grazing line, and they approach 0 as it falls below the grazing line. However, when the target is directly above the referent, it receives high ratings even when it is barely above the grazing line (Experiment 7), and the ratings drop at a slower rate as it falls below the grazing line (Experiment 6). Based on this observation, we decided to apply a grazing line penalty only for targets which approach the grazing line but are not directly above the referent, i.e., when oR points right, away from the referent, or when oL points left, away from the referent. However, it is still necessary to apply a penalty for targets lying directly above a referent that fall below the grazing line, i.e., targets that fall below either pointR or pointL. Therefore, we use the following formulae:

Height(o) = Degrees of o above the horizontal
Penalty = *One-Down-Penalty* if one vector points down
Two-Down-Penalty if both vectors point down
0 otherwise

Rating = ((SymmetryDiff * *Slope*) – Penalty) * Sigmoid(Minimum(Height(oR), Height(oL)), *Height-Gain*)

Here we only consider **Height(**oR**)** if oR points right, away from the referent. Thus, **Height** only plays a role if the target is not directly above the referent. These formulae have four free parameters:

1. *Slope*: the cost as the vectors deviate from symmetry
2. *Height-Gain*: a gain value for the sigmoid function applied to the height
3. *One-Down-Penalty*: a fixed cost for having one vector point downwards
4. *Two-Down-Penalty:* a fixed cost for having both vectors point downwards

The assumption of a fixed cost applied when one or both vectors point downward is simplistic but seems to be a reasonable first approximation.

Experiment

We evaluated the VS model by simulating the results from the seven Regier and Carlson experiments. We also ran the model on an "above" rating experiment by Logan and Sadler (1996) which used small objects that might be treated as point masses for both the targets and the referent. We programmatically generated stimuli in CogSketch which were at locations identical to those used in the experiments.

We followed Regier and Carlson in fitting our model to the Logan and Sadler study to determine the values for the VS model's free parameters, and then using those values to evaluate it on the other seven experiments. We fit the model by performing an exhaustive, breadth-first search over all combinations of reasonable values for the free parameters, returning the set of values that resulted in the highest correlation between the model and Logan and Sadler's results. Correlations were R^2 computed via linear regression.

One parameter of VS, *One-Down-Penalty*, only applies when pointR and pointL are at different heights and the target lies between them. Thus, this parameter could not be determined based on the Logan and Sadler study, in which all referents were small and symmetric. Therefore, once the other three parameter values had been determined, we determined the value of this parameter by fitting the model to the results of Regier and Carlson Experiment 5, one of their experiments which used an asymmetric referent in which pointR and pointL were at different levels. Overall, three of their experiments used such a referent: 4, 5, and 6. Thus, the parameter fit to Experiment 5 could be evaluated on the other two experiments.

Results

The results of the eight simulations are given in Table 1. R^2 is a measure of the proportion of variance in one variable that is explained by another. As the table shows, the VS model correlates well with human performance on every experiment, achieving an R^2 above .90 in all cases. However, the correlation values are typically slightly below the correlations for the AVS model.

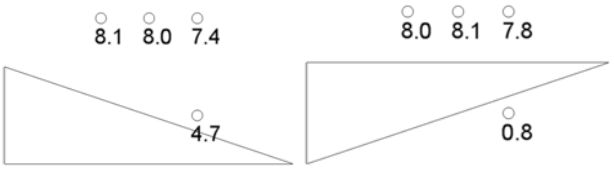
Each of the seven Regier & Carlson experiments was designed to test one of the four factors in positional relations outlined earlier. As Table 1 shows, VS's performance qualitatively matched the effects of those factors in almost all cases, failing only on the second part of Experiment 4. None of the models which Regier and Carlson compared to the AVS model fared as well on these qualitative tests.

Discussion

Overall, the VS model performed quite well on the eight experiments, matching or nearly matching the AVS model in most cases, despite using considerably less information, i.e., the two vector values. However, we believe two weaknesses of the model should be addressed. Firstly, the model's correlations, while high, were generally under the AVS model. This was particularly noticeable in Experiment 4-Upright Triangle and Experiment 5. There are two possible reasons for the lower correlations on these problems: (1) These involved asymmetric shapes—a triangle and an "L" shape—so participants might have been less likely to consider vector symmetry when computing "above." (2) These are two of the problems on which some targets lay between pointL and pointR, meaning that vectorL pointed down while vectorR pointed up. Thus, it may be that our grazing line component, which merely deducts a fixed cost in such cases, is insufficient. We suspect that our simplistic grazing line component may have weakened the model overall in its performance vs. the AVS model.

Table 1: Simulation results.

Model	Qualitative Test	R^2	Adj. R^2
Logan & Sadler			
AVS	------	.963	.959
VS	------	.965	.965
Experiment 1	Proximal		
Tall Rectangle	Orientation		
AVS	**pass**	.996	.995
VS	**pass**	.985	.984
Wide Rectangle			
AVS	**pass**	.994	.993
VS	**pass**	.970	.969
Experiment 2	Center-of-Mass		
Tall Rectangle	Orientation		
AVS	**pass**	.993	.992
VS	**pass**	.980	.980
Wide Rectangle			
AVS	**pass**	.995	.993
VS	**pass**	.977	.975
Experiment 3	Center-of-Mass		
Tall Rectangle	Orientation		
AVS	------	.984	.983
VS	------	.969	.968
Wide Rectangle			
AVS	**pass**	.995	.993
VS	**pass**	.980	.980
Experiment 4	Center-of-Mass		
Upright Triangle	Orientation		
AVS	**pass**	.991	
VS	**pass**	.959	
Inverted Triangle			
AVS	**pass**	.990	
VS	**fail**	.999	
Experiment 5	Grazing Line		
L shape			
AVS	**pass**	.976	.975
VS	**pass**	.907	.906
Experiment 6	Grazing Line		
Tall Triangle			
AVS	**pass**	.930	.919
VS	**pass**	.930	.928
Experiment 7	Distance/Center-of-		
Wide Triangle	Mass Interaction		
AVS	**pass**	.965	.958
VS	**pass**	.959	.956



a) Upright b) Inverted

Figure 7: "Above" ratings for Regier & Carlson Experiment 4.

The second weakness of the model is that it failed one qualitative test: the center-of-mass orientation effect in Experiment 4-Inverted Triangle (see Figure 7b). It failed to predict that the upper left target would receive a slightly higher score than the upper right target. However, we observe that: (1) The effect, while statistically significant, is quite small. There is a larger effect for center-of-mass in the Upright case (Figure 7a), wherein the VS model does show the predicted effect. (2) This experiment contained only eight stimuli, the four target locations for the two triangle types. Given so few stimuli, and given that the top three targets for the Inverted Triangle are so similar, a few participants may have used a more sensitive strategy to provide better contrast. They may have looked directly at the orientation between the referent's center-of-mass and the target, or considered the relative width of the referent directly below each target.

Conclusion

As the results show, the VS model strongly correlates with human "above" ratings on eight experiments. It correctly predicts all four factors contributing to "above" ratings, as given by Regier and Carlson. The VS model does not correlate quite as well as Regier and Carlson's AVS model. However, the VS model takes only two vector orientations as its input, while the AVS model uses many vector orientations, as well as the height of the target relative to the topmost and bottommost points of the referent. The strong performance of the VS model with only two vector orientations supports the hypothesis that these two vectors are used by humans in assessing positional relations.

Because we have implemented the VS model using visual routines, we can use it to make novel predictions about the computation of positional relations. The scanning process can be disrupted by the presence of other curves between the referent and the target. Therefore the VS model predicts that distractors lying between the referent and the target, particularly if they lie along the scan lines used to compute the VS model's two vectors, will disrupt the process of computing positional relations. While Carlson and Logan (2001) have argued against an effect of distractors between the target and the referent for letter stimuli, we are currently evaluating this prediction with simpler stimuli, basic color patches.

This paper illustrates how Visual Routines for Sketches can be used to implement and evaluate a perceptual model. In the future, we hope to make VRS generally available, so that other researchers can use it to explore the computations underlying perception.

Acknowledgments

This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC). We thank Terry Regier and Laura Carlson for providing the data so we could replicate their original studies.

References

Carlson, L. A., & Logan, G. D. (2001). Using spatial terms to select an object. *Memory & Cognition, 29*(6), 883-892.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review, 97*, 404-431.

Chapman, D. (1992). Intermediate vision: Architecture, implementation, and use. *Cognitive Science, 16*, 491-537.

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 5th Eurographics Workshop on SBIM.

Gapp, K. (1995). Angle, distance, shape, and their relationship to projective relations. In Proceedings of the 17th Annual Meeting of the Cognitive Science Society.

Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 29-50.

Hayward, G. H., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39-84.

Horswill, I. (1995). Visual routines and visual search: A real-time implementation and an automata-theoretical analysis. In Proceedings of *IJCAI '95*.

Jolicoeur, P., Ullman, S., & MacKay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory and Cognition, 14*(2), 129-140.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175-219.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation, 10*(4), 903-940.

Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and Space*.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review, 1*(1), 29-55.

Rao, S. *Visual routines and attention.* (Doctoral dissertation, MIT, 1998).

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General, 130*(2), 273-298.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

Yen, S. C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research, 38*, 719-741.