Modelling English Spatial Preposition Detectors

Connor Smith, Allen Cybulskie, Nic Di Noia, Janine Fitzpatrick, Jobina Li, Korey MacDougall, Xander Miller, Jeanne-Marie Musca, Jennifer Nutall, Kathy Van Bentham, and Jim Davies

Institute of Cognitive Science, Carleton University,
1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6
jim@jimdavies.org

Abstract. In this paper we present five algorithms for the detection of spatial relationships within an image: above/below, adjacent to, occlusion, between, and close to.

Keywords: natural language, spatial relationships, cognitive modeling

The use of spatial relations has importance in understanding what people take away from diagrams, and the generation of scene descriptions.

We present five computational spatial-relation detectors intended to functionally model human perceptions corresponding to English words: above/below, adjacent_to, occlusion, between, and close_to. We model them after high-level representations found in language.

According to [1], spatial terms parse the space around some reference object(s) into regions, with some regions being more prototypical characterizations than others. Fuzzy logic captures this vagueness, describing the truth value of a spatial proposition with a real number between 0 and 1.

We use image data from an online game called Peekaboom [2], which consists of labels associated with point clouds in images. That is, we know "where" the label is in each image. For every pair of labels our system uses the detectors to generate spatial propositions with corresponding fuzzy values. For example, if a cup is above a table, our system might output a fact such as (`cup above-below table .9`).

Above/Below. The computations we use are based on Bloch's [3] fuzzy approach.

The target object is assigned a status of 'above' or 'below' the reference object based on whether its centroid's y-axis has a lesser or greater value than that of the reference object's y-axis value. A slope is then found for a line connecting the target and reference object's centroids using the rise and run. If the run equals 0 then the line is perfectly vertical and the relationship receives a belief score of 1.0. Otherwise, an angle of deviation is calculated by taking the arctangent of the line slope; this represents the angle between the line and the horizon. We calculate the truth value by applying: the vertical trigometric, sin^2, function of the arctangent (sin^2(angle)).

Adjacenct_to. People intuitively see two objects as adjacent if some edge of object A is near, or touches, some edge of object B [4]. Additionally, we hypothesize another

A.K. Goel, M. Jamnik, and N.H. Narayanan (Eds.): Diagrams 2010, LNAI 6170, pp. 328–330, 2010.
© Springer-Verlag Berlin Heidelberg 2010

relevant factor: that object area in relation to the shortest straight line that can be drawn between the facing edges of the objects. Intuitively we may judge two very small objects as not adjacent if the distance between them is much larger than the area of one or both of the objects.

If the objects overlap, or if there is an object between the target and reference, the truth value returned is 0.0. If the shortest distance between the objects is <= 1 (pixels), a truth value of 1.0. A percentage is then derived from a ratio of minimum distance between the objects over the size of the target object. A truth value is then returned as such (ratio to truth value): <2.0 to 1.0, <6.0 to 0.8, <12.0 to 0.5, <18.0 to 0.2, and >20.0 to a value of 0.0.

Occlusion. A contour junction in occlusion is where two contours meet and one of them appears to abruptly end. Most contour junctions are called T-junctions, where the stem is the occluded contour and the crossbar is the occluding contour [5]. Closure is when the continuation of a stem from one junction closes off a region by connecting to the stem of another junction; non-closure is when the stems do not close off a region [6].

There are three cases of occlusion to account for: A) where a smaller object occludes a larger object, and the smaller object is completely contained within the larger object's convex hull, resulting in no contour junctions. If one of the objects has all of its coordinates located within the convex hull of the other, it is returned as the occluding object. B) is when occlusion occurs but neither object is completely contained within the other. Lines of the convex hulls are checked to determine whether they partially share line functions – if so, the algorithm looks to find a line that forms a junction with the shared line. The object with the line that forms a junction with the shared line is returned as the occluded and the other as the occluding. C) is somewhat similar to Case B, however it deals with instances of closure, where the occluded lines of an object are perceived to (or do) meet at a point.

Truth values are then output based on the percentage of the occluded object's area that the occluding object occupies. Example: 90% = 0.9, 80% = 0.8, etc.

Between. The algorithm finds the centroid and area for all three objects. If the target hull overlaps either reference object by 50% the algorithm returns a truth value of 0.0. Otherwise, the convex hull of betweenness (β_{CH}) is generated - defined as per [7]. This joins the reference objects' hulls together and then returns a hull representing the space between them. If the target object has no overlap with either reference object and its centroid falls within β_{CH}, a truth value of 1.0 is returned.

If the overlap of the target convex hull with a reference object is between 0-50%, the detector returns a value calculated by the formula: 1 - (greatest overlap ratio/2). The greatest overlap refers to the greatest ratio of a reference object's convex hull which is overlapped by that of the target object.

If the target centroid does not fall within β_{CH}, the algorithm: (i) calculates the proportion of the target hull that overlaps with β_{CH} to obtain $overlap_{ConvexBetween}$, and (ii) draws a circular field of betweenness, β_{Circ}, between the reference objects, and determines the proportion of the target hull that overlaps with β_{Circ} to obtain $overlap_{CircBetween}$. The diameter of β_{Circ} is the shortest distance, x, between the reference objects, and its centroid is the midpoint of x. The truth value is the greater of $overlap_{ConvexBetween}$ and $overlap_{CircBetween}$.

Close_to. We propose two hypotheses for 'close to' detection: The ad hoc unit scaling hypothesis and the framed space comparison hypothesis.

The ad hoc unit scaling hypothesis states that the distance between object A and object B is judged by estimating how many copies of A fit in the space between the objects (Robert Thomson, personal communication, February 13, 2009). This predicts that the order of presentation of the objects will affect the distance judgment. This is important when the two objects vary largely in size.

The framed space comparison hypothesis states that while objects remain the same distance apart in absolute terms, the larger the image frame, the closer the objects appear to each other.

First, an ad hoc unit scale judgement is made by finding the closest points between the target and reference. The line between these points is then extended through the target object to find the ad hoc unit. This unit is compared to the distance between the objects by finding a distance to ad hoc unit ratio. A belief value is then assigned from the ratio. Example (ratio to truth value): <=1.0 to 1.0, <2.0 to 0.9, <2.9 to 0.8, <3.1 to 0.5, <4.1 to 0.2, and otherwise a 0.0 truth value.

The second function makes a frame-space comparison judgement. It finds the area between the objects and calculates a ratio of the area to the overall area of the image. This ratio is then assigned to a truth value. Example (ratio to truth value): <0.020 to 1.0, <0.025 to 0.8, <0.033 to 0.5, <0.05 to 0.2, and otherwise a value of 0.0.

The two judgements are united using a function which takes the belief value of each judgement, assigns a weight to them, and then outputs their combined belief value.

By modeling spatial relationships after spatial terms in natural language we are able to capture linguistic parameters associated with different spatial relations as well as some of the perceptual mechanisms behind them.

References

1. Hayward, W.G., Tarr, M.J.: Spatial Language and Spatial Representation. Cognition 55, 39–84 (1995)
2. von Ahn, L., Liu, R., Bloom, M.: Peekaboom: a Game for Locating Objects in Images. In: Proceedings of the Special Interest Group on Computer-Human Interaction Conference on Human Factors in Computing Systems, pp. 55–64. Association for Computing Machinery, Montreal QC (2006)
3. Bloch, I.: Fuzzy Spatial Relationships for Image Processing and Interpretation: a Review. Image Vision and Computing 23, 89–110 (2005)
4. Rosenfeld, A., Klette, R.: Degree of Adjacency or Surroundedness. Pattern Recognition 18(2), 169–177 (1985)
5. Sedgwick, H.A.: Visual Space Perception. In: Goldstein, E.B. (ed.) Blackwell Handbook of Perception. Blackwell Publishers Ltd., Massachusetts (2001)
6. Boselie, F., Wouterlood, D.: A Good-Continuation Model of Some Occlusion Phenomena. Psychological Research 54, 267–277 (1992)
7. Bloch, I., Colliot, O., Cesar Jr., R.M.: On the Ternary Spatial Relation "Between". IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics 36, 312–327 (2006)