Pragmatic Scenario Inference on Static Spatial Configurations

Dan Tappan

Department of Computer Science, Idaho State University
921 S. 8th Ave., Stop 8060
Pocatello, ID 83209-8060
tappdan@isu.edu

Abstract

The work takes static spatial configurations defined by quantitative, graphical data like the positions and orientations of nature-related objects and infers basic, high-level, pragmatic meaning of the scenario from a small set of semantic actions; e.g., the wolves are *chasing* the sheep. It uses an inheritance-based knowledge base to define contextually appropriate, case-based roles, and geometric constraint satisfaction to recognize spatial dependencies. This successful pilot study elicits semantic features of interest for follow-on investigation. It uses a quantitative survey methodology to compare the performance of the system against human subjects based on the standard information-retrieval measures of precision and recall.

Introduction

A static spatial configuration, such as in Figure 1, contains low-level, quantitative knowledge about its objects and their positions and orientations. These details are sufficient to render the image, but they do not directly provide any insight into the higher-level composition of the scene, namely what the objects might be doing individually and collectively, and why. For example, the wolves are arguably chasing the sheep.

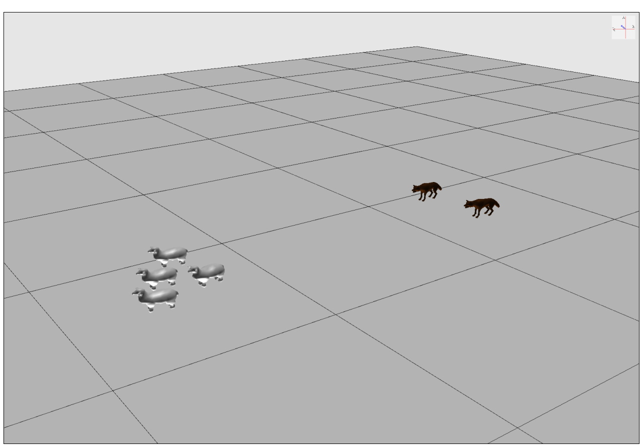

Figure 1: Sample Configuration

The goal of this work is to infer from a small set of spatially relevant semantic actions the superficial pragmatics characterizing a variety of simple predator-prey scenarios with wolves, sheep, and several supporting objects. It uses an inheritance-based knowledge base of concepts, attributes, and declarative rules to define the contextually appropriate spatial interpretation of many contexts. The underlying reasoning mechanism is non-deductive geometric constraint satisfaction.

This paper addresses a pilot study to determine the feasibility of follow-on work and to discover general semantic features it could investigate. It was formally evaluated through a quantitative survey methodology that compared the computer performance against human performance. Specifically, it used the standard measures of precision and recall from the field in information retrieval.

Related Work

This work extends the base system by Tappan (2004a, 2004b, 2008b), which generates and renders configurations from natural-language descriptions, and Tappan (2008a), which infers spatial relations from existing configurations. The approach to inferring spatial knowledge loosely draws upon other work by Neumann (1989), Walter et al. (1998), Koller et al. (1992), and Tsotsos (1985) for scene interpretation. Tversky (2000) covers in comprehensive detail many of the spatial issues that complicate the problem. Several works (Herskovits 1986; Claus et al. 1988; and Olivier and Tsujii 1994), in particular, form the basis for defining and interpreting spatial frames of reference. Most early approaches to spatial analysis adopted purely geometric solutions and did not take advantage of spatial knowledge relevant to the objects (Xu 2002; Yamada 1993). More recent work, especially in Geographic Information Systems, attempts to account for such contextual information (Peters and Shrobe 2003; Davis 1990; Egenhofer and Franzosa 1991; Frank 1992; Frank 1996; Hernández et al. 1995; Randell et al. 1992). This work follows the latter approach. Additional inspiration derives from recent work in spatial-intent recognition, case-based plan recognition, and recognition of natural scene categories (Kiefer and Schlieder 2007; Cheng and Thawonmas 2004; Lazebnik et al. 2006).

Methodology

For space reasons, the description of the system and the methodology of the study running on it are intertwined.

Spatial Configurations

A spatial configuration consists of objects in a static, two-dimensional, tabletop zoo environment. The system currently supports over 120 non-articulated objects, mostly animals and plants, selected because they exhibit great variety in their spatial characteristics and interpretations. The static aspect eliminates the effects of movement, time dependencies, and the frame problem, among others, which are indeed relevant to this work, but beyond its scope (Adorni 1984; Sowa 1991; Srihari 1994; Coyne and Sproat 2001).

The underlying representation of a configuration is a simple semantic network, which is particularly suited to this task for three reasons (Sowa 1991). First, its primary components, nodes and directional arcs, map directly to the objects, properties, and relations in a configuration. For example, Figure 2 is a semantic network that describes a wolf looking north at a sheep a little northeast of it. Second, as a straightforward computational data structure, standard graph algorithms can operate on it natively. Third, as a well-studied and commonly used formalism for artificial intelligence, it facilitates transferring knowledge representations to and from other applications (Russell and Norvig 1995; Sowa 2000).

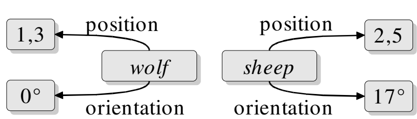

Figure 2: Semantic Network

The semantic networks derive from two sources. The first is manual specification of where the objects are and are facing. This approach is necessary to guarantee adequate coverage of particular, nuanced scenarios to test, but it is tedious for large data sets. The second source is automated scenes derived from rudimentary natural-language descriptions. It is described in detail in Tappan (2004a). The basis of its input is small, descriptive statements, such as:[1]

> There are two wolves and four sheep. The wolves are south of the sheep, near each other, facing the sheep, and midrange from the sheep. The sheep are near each other and facing away from the wolves.

Figure 1 renders one possible interpretation. Any number can be generated stochastically, which greatly reduces the amount of time needed to create tests.

1 Paraphrased somewhat for brevity and easier reading. The parser does not actually support plural nouns, irregular plurals, or comma-delimited clauses.

Annotation

A total of 20 manual and automated configurations were tagged by humans to indicate their plausible spatial interpretations; e.g., *wolves chasing sheep*. Open-ended interpretation is not the goal, so the set of tags is limited to the following perceived actions. More than one is possible per configuration.

Unary actions involve only one active object type, such as wolves. Other object types may be present, but they play a passive role. Each tag (in monospace font) is characterized in English here; the actual formalism of their definition will be covered shortly. The only unary actions currently supported are simple, collective sheep behavior:

`migrate`	sheep grouped and oriented similarly
`graze`	sheep grouped and oriented dissimilarly
`drink`	sheep positioned around (passive) water object like pond, lake, or pool

Binary actions involve two active object types. The first group reflects tactical enclosure interpretations:

`flank`	sheep grouped; 3+ wolves at base and to either side of group, likely facing it
`surround`	sheep grouped; 3+ wolves around perimeter of group, likely facing it

The next group reflects linear attack interpretations:

`conceal`	(passive) view-blocking object, like tree or rock, between wolf and sheep; wolf facing sheep, near and likely at edge of view-blocking object
`stalk`	sheep facing away from wolf; wolf facing sheep, far from sheep
`chase`	sheep facing away from wolf; wolf facing sheep, close to sheep

The final group reflects situational awareness:

`aware`	at least one sheep facing wolf
`unaware`	no sheep facing wolf
`anticipate`	all sheep facing wolf

There is also an `unknown` tag for scenarios that cannot be assessed as any of the above.

Tagging used a straightforward survey methodology: 9 computer-science undergraduates each annotated the 20 configurations with any combination of these tags. The images were available online, in color, from three consistent, fixed vantage points. For this pilot study, the surveys were not anonymous.

Manual Scenario Extraction

This initial tagging serves as manual training data to extract common spatial semantic features between similar configurations. To be effective, there must be reasonable agreement between taggers on the interpretations. A formal statistical measure like Kappa correlation is commonly used to measure intercoder reliability (Fleiss 1971). However, for simplicity, and to align with parallel work (in progress) that tries to weight the various choices, this work calculates a straightforward percentage based on the number of participants who selected a tag. This consensus-based approach stipulates that a tag with agreement below an empirically determined threshold of 65% is discarded as too ambiguous and therefore probably not computable.

Analysis of the discards suggests that disagreement is due primarily to two factors. One is the lack of articulation in the objects. For example, the states of sitting, standing, walking, running, and even sleeping all appear the same, but they can have profoundly different effects on the overall interpretation. The other is the lack of temporal cues in a single snapshot of a dynamic scene. For example, a wolf facing away from a sheep could be walking away, or it could be merely turning around.

The goal of scenario extraction is to characterize the kinds of details that contribute to different interpretations. They are informally organized into three categories.

Constraint Satisfaction The underlying reasoning formalism, to be discussed shortly, uses geometric constraint satisfaction. Some features of interest map directly to it:

- *visibility*: can the wolf see the sheep, based on field of view, range, and visual acuity (possible degradation over range)?
- *accessibility*: can the wolf get to the sheep it sees?
- *boundary conditions*: when do states apply and not apply, and what kind of transition is there between the two? For example, is going from not visible to visible abrupt or smooth?
- *scale and range*: behavior may be based on size or scope; e.g., wolves far from sheep may be more cautious than those close to them, so as not to alert the sheep.

Behavioral Roles High-level interpretation of objects in concert requires an understanding of their case-based roles (Turner 1998; Cheng and Thawonmas 2004):

- *disabler*: an object that hinders an interpretation; e.g., an uncrossable stream between the wolf and the sheep.
- *enabler*: an object that helps an interpretation; e.g., a bridge over the stream, or a tree for concealment.
- *neutral*: an object that either party can use, but it favors neither; e.g., a wall for hiding.
- *inert*: an object that plays no discernible role; e.g., clouds.
- *outward action*: what an object can and cannot do to other objects; e.g., a wolf can attack sheep but cannot attack more than one simultaneously.
- *inward action*: what an object can and cannot have done to it by other objects; e.g., sheep can be attacked by more than one wolf simultaneously, but not by another sheep.

Superficial Planning There is not enough information to do substantial planning currently, but some elements seem promising:

- *necessary and sufficient conditions*: how objects initiate the trajectory for a chain of events; e.g., a wolf shows up, then the sheep flee. They do not flee without the wolf. Thus, to tag a configuration without a wolf as `chase` makes no sense.
- *utility*: what the objects value; e.g., a wolf "wins" by killing sheep, and sheep "win" by not being killed by wolves.
- *Individual vs. collective outcomes*: how interpretations differ according to scope; e.g., killing one sheep is bad for the individual, but it may allow others to escape, which is good for the collective.

Knowledge-Base Augmentation

This informal characterization of semantic features is not adequate for an automated computational approach. The algorithm needs to be able to infer substantial unstated details about objects from commonsense background knowledge. A knowledge base provides this support.

Existing Knowledge Base The system that this work extends generates a set of plausible images from a restricted class of English sentences describing zoo-related scenes. It employs an inheritance-based, declarative knowledge base of over 120 physical concepts, each of which either inherits its attributes and rules for spatial interpretation from its ancestors, or it defines/overrides them itself. Figure 3 is a highly simplified abstraction, which Tappan (2004a) formally defines in detail.

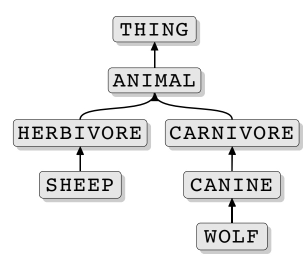

Figure 3: Simplified Knowledge Base

An attribute defines whether a concept exhibits a particular spatial behavior; e.g., whether a concept has a canonical front, which generally corresponds to its having a face or eyes. As objects and concepts are not articulated, any head is always fixed in line with the orientation of the body. This simplification eliminates the need to determine the configuration of body parts; e.g., the body of the dog is oriented north, but it is looking east.

A rule specifies when a particular spatial relation, like `near`, applies from one object to another. It uses a formalism of geometric fields that describe a collection of cells in a two-dimensional, top-view, polar projection centered around the object (Yamada et al. 1992; Yamada 1993; Gapp 1994; Olivier and Tsujii 1994; Freska 1992). Experimentation suggests that 32 sectors and 100 rings similar to Figure 4 are sufficient for the current domain of concepts and relations. Each cell defines a small subregion of the projection that can be conditionally inspected for the presence of other objects.

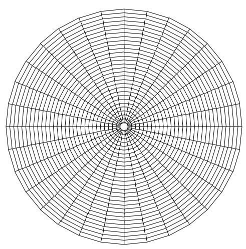

Figure 4: Available Fields

Any combination of selected cells among the 3,200 available is valid, but in practice, only variations of two types define all spatial relations: *wedges* for position and orientation relations, and *rings* for distance relations. Figures 5a and 5b show respective examples of the relations `front-of` and `far-from` for object `c`, which is facing the direction of the arrows.

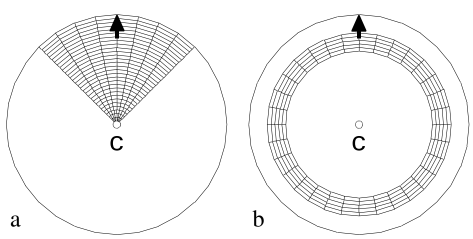

Figure 5: Sample Wedge and Ring Fields

Each concept in the knowledge base has access to the attributes and rules for its spatial interpretation. These rules define the 42 distance, orientation, and position relations in Tables 1 through 3, respectively. For space reasons, Tables 2 and 3 omit 35 other relations prefixed with `direct`, which specifies a narrower interpretation with the same general meaning; e.g., `direct-front-of` would fan out less to the sides. The interpretation of appropriateness depends on certain *ad hoc* generalities of the concept. For example, the relation `near` is closer (in absolute terms) for a mouse than it is for an elephant due to their differences in magnitude (Hernández 1994; Olivier and Tsujii 1994; Stevens and Coupe 1978).

Many relations in Table 3 have both local and global forms, which respectively apply in the intrinsic (or object-centered) and deictic (or viewer-centered) frames of reference (Herskovits 1986). For example, the intrinsic relationship *in front of the dog* specifies a region outward from the dog's face, but the deictic relationship *in front of the tree* specifies a region outward from the tree to the position of the viewer, which is not stated.

inside	midrange-from
outside	far-from
adjacent-to	at-fringe-of
near	

Table 1: Distance Relations

facing	facing-west
facing-away-from	facing-northeast
facing-north	facing-northwest
facing-south	facing-southeast
facing-east	facing-southwest

Table 2: Primary Orientation Relations

local-front-of	global-front-right-of
local-back-of	global-back-left-of
local-left-of	global-back-right-of
local-right-of	between
local-front-left-of	north-of
local-front-right-of	south-of
local-back-left-of	east-of
local-back-right-of	west-of
global-front-of	northeast-of
global-back-of	northwest-of
global-left-of	southeast-of
global-right-of	southwest-of
global-front-left-of	

Table 3: Primary Position Relations

Each concept has access to its contextually applicable rules that map fields to relations. For example, this (slightly abridged) rule returns the set of all objects that have an object (instance) of this concept in their `near` field:

```
(RELATION near
 (FIELD-MUST-CONTAIN ?b.field-near ?self))
```

This rule returns the set of all objects that are in the `front` field of this object, if it has a canonical front:

```
(RELATION facing
 (TRUE ?self.has-canonical-front
  (FIELD-MUST-CONTAIN ?self.field-front ?b)))
```

And this rule,

```
(RELATION in-back-of
 (OR
  (TRUE ?b.has-canonical-front
   (FIELD-MUST-CONTAIN ?b.field-back ?self))
  (FALSE ?b.has-canonical-front
   (FIELD-MUST-CONTAIN ?b.field-north ?self))))
```

returns the set of all objects subject to the following criteria:

- the other object has a canonical front (e.g., `WOLF`) and this object is in its `back` field; or,
- the other object does not have a canonical front (e.g., `TREE`) and this object is in its `north` field.

These conditional cases account for the deictic and extrinsic frames of reference, respectively (Tappan 2004b; Herskovits 1986). The latter extends the intrinsic frame by fixing the position of the viewer; e.g., *in front of the tree (as seen from the north).*

The final element of this stage combines the explicitly stated knowledge from the semantic network with the implicitly inferred background knowledge from the knowledge base. Figure 6 depicts a simplified example of this process: objects *wolf* and *tree* link to concepts `SHEEP` and `WOLF`, respectively. Thus, *wolf* has access to the rules about itself and, through inheritance, also to its ancestor concepts `CANINE`, `CARNIVORE`, `ANIMAL`, and `THING`. The same process holds for *sheep*. It is important to note the distinction between an *object*, which is a unique instance in the configuration, and a *concept*, which is a shared set of attributes and rules that all instances of it must have in common. For clarity, this distinction is rendered typographically through italics and capitalized monotype font, respectively.

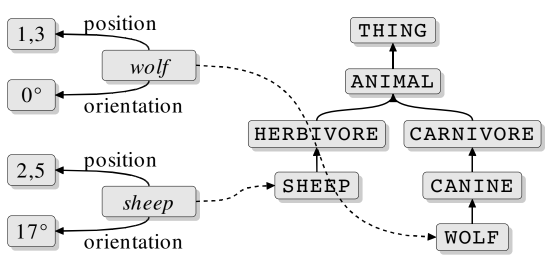

Figure 6: Semantic Network Linked to Knowledge Base

Extended Knowledge Base The knowledge base in the base version of this system targets how one object relates to another on an individual, one-to-one basis. The scenarios to be classified in this work are on a collective basis, which requires one-to-many, many-to-one, and many-to-many relationships. For some semantic features, this extension requires merely adding additional spatial relations. For example, this rule,

```
(RELATION migrating-with
 (AND
  (IS-CONCEPT ?self collective-animal)
  (IS-CONCEPT ?b    collective-animal)
  (FIELD-MUST-CONTAIN
   (RANGE ?b field-adjacent field-midrange) ?self)
  (FIELD-MUST-CONTAIN
   (RANGE ?self field-adjacent field-midrange) ?b)
  (SIMILARITY ?b.azimuth ?self.azimuth 0.7)))
```

returns the set of all objects subject to the following criteria:

- this object and the other object are both descendants of `COLLECTIVE-ANIMAL` in the knowledge base; and,
- this object and the other object are in any range field from `adjacent` to `midrange` of each other; and,
- this object and the other object are facing generally in the same direction.

Defining collective concepts is also straightforward: through the existing restricted multiple inheritance (no conflicts allowed), `SHEEP` and `WOLF` in Figures 3 and 6 now additionally inherit from the new `COLLECTIVE-ANIMAL`, which maintains this new, shared `migrating-with` rule.

Not all the manually identified semantic features would be so straightforward to incorporate, of course. For the 12 tags in this study, however, extending the knowledge base is relatively easy.

Automatic Classification

The knowledge base provides the contextually appropriate, computable background knowledge to identify which of its relations apply between which objects in a configuration. This geometric inference process is documented throughly in Tappan (2004b, 2008a). It generates a substantial number of inferences, which correspond to unstated spatial dependencies. For example, Figure 7 shows experimental results from Tappan (2008a) for related work, where this number ranged from 27 inferences for 3 objects to 602 for 10 objects.

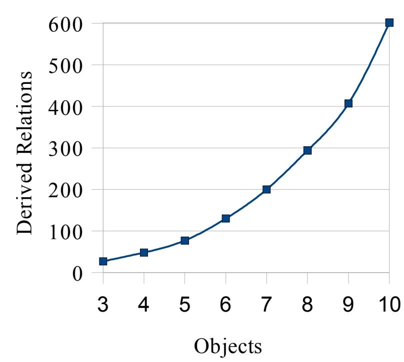

Figure 7: Inferred Relations

Defining relations for pragmatic spatial features is an iterative, experimental process. For each change, the original 20 tagged configurations were run against the updated knowledge base to determine the effectiveness at inferring any of the tags. If these results unsatisfactorily deviated from expectation, as defined in the next section, the knowledge base was manually tweaked, and the process was repeated. This approach constitutes supervised learning in machine-learning terms (Harter and

Hert 1997). The goal is to tailor the knowledge base by hand to perform well on data it has already seen, referred to as the *training set*.

Experiments

Distilling the essence of semantic features through manual training is admittedly subjective, arbitrary, and *ad hoc*. The true test of effectiveness is actually in how well the approach performs on data it has *never* seen, referred to as the *test set*. To this end, an additional 20 configurations were generated as described earlier.

These new configurations were combined with the original ones and randomly shuffled. The original participants then tagged this set as described earlier. The time between the original and subsequent tagging was three weeks to control for any residual familiarity with the originals.

Results and Discussion

For both the training and test sets, performance was evaluated according to the agreement between the results from the human taggers and the computational approach. For this approach to be effective, variation in the human performance must be considered because not every human tagged the same configurations the same way. Thus, if humans cannot determine a consistent answer, it might be unfair to expect a computer to do so.

Controlling for human inconsistency was a two-dimensional process. Lateral agreement, which was already discussed for the training set, is defined as how closely the tags for each configuration within either set agree among all the participants.

Longitudinal agreement is defined as how consistent each participant was between the original and subsequent tagging of the same configurations. This measure was intended to verify that the participants—students who knew there are no true right or wrong answers—took the task seriously and gave consistent answers. It also moderately controlled for possible survey fatigue, where participants grow tired of answering questions and put less effort into later ones (Porter et al. 2004). The system does not use this measure, but it appears to be helpful in informally interpreting the salience of the results.

The tags produced by the computational approach and the human participants can agree or disagree in four ways, as indicated in Table 4.

Type		Computer	Human
true positive	(TP)	present	present
true negative	(TN)	absent	absent
false positive	(FP)	present	absent
false negative	(FN)	absent	present

Table 4: Possible Tag Agreements

The ultimate performance measure is based on the standard approach for information retrieval (Harter and Hert 1997):

- *precision* is the accuracy or relevance of the classifications; i.e., the probability that a configuration will be classified with a correct tag. It is defined as the number of true positives divided by the sum of the true and false positives.
- *recall* is the completeness or coverage of the classifications; i.e., the probability that all the relevant configurations will be found given a tag. It is defined as the number of true positives divided by the sum of the true positives and false negatives.

Training Set

As Table 5 shows, overall agreement on the 20 configurations in the training set is perfect: 100% precision and 100% recall for all tags. These results are not surprising, however, because the knowledge base was painstakingly tailored to match these configurations precisely. This decision may have actually overfitted the data and degraded the results of the subsequent test set (Tetko et al. 1995).

Tag	TP	TN	FP	FN	Precision	Recall
`migrate`	3	17	0	0	1.0	1.0
`graze`	4	16	0	0	1.0	1.0
`drink`	2	18	0	0	1.0	1.0
`flank`	3	17	0	0	1.0	1.0
`surround`	5	15	0	0	1.0	1.0
`conceal`	6	14	0	0	1.0	1.0
`stalk`	7	13	0	0	1.0	1.0
`chase`	13	7	0	0	1.0	1.0
`aware`	11	9	0	0	1.0	1.0
`unaware`	5	15	0	0	1.0	1.0
`anticipate`	2	18	0	0	1.0	1.0
`unknown`	0	20	0	0	n/a	n/a
sum	61	179	0	0		
average					1.0	1.0

Table 5: Training Results

Test Set

The test set evaluates the computational performance on the unseen configurations. The previously seen configurations were removed because there is no point in retesting them, and they would skew the results in the positive direction. As Table 6 shows, overall agreement on the 20 remaining new configurations in the test set is respectable: 70% precision and 70% recall. Although these results are not statistically significant given the small sample size, they do suggest that this proof-of-concept work has promise.

Tag	TP	TN	FP	FN	Precision	Recall
migrate	4	13	2	1	0.7	0.8
graze	4	14	1	1	0.8	0.8
drink	3	16	0	1	1.0	0.8
flank	4	11	2	3	0.7	0.6
surround	4	9	2	5	0.7	0.4
conceal	3	12	3	2	0.5	0.6
stalk	3	11	2	4	0.6	0.4
chase	9	9	1	1	0.9	0.9
aware	10	8	1	1	0.9	0.9
unaware	11	5	2	2	0.8	0.8
anticipate	3	4	8	5	0.3	0.4
unknown	0	20	0	0	n/a	n/a
sum	58	132	24	26		
average					0.7	0.7

Table 6: Test Results

Discussion

Even for this small tag set, many independent and dependent relationships are apparent. Some tags are disjoint. For example, `graze` and `migrate` are not compatible given the small number of sheep. For a larger number, it is conceivable that one subgroup could be grazing while the other is migrating, but the configurations were not set up this way. Other tags may have overlaps. For example, `graze` and `drink` are compatible, as are `flank` and `surround`. Others may exhibit logical entailment. For example, `stalk` implies that the wolves are `aware` of the sheep, but the sheep are `unaware` of the wolves. Some have real-world semantic inconsistencies. For example, `chase` implies that each group should be aware of the other, but `aware` requires at least one sheep to be facing the wolves. In practice, fleeing sheep will not behave this way. Feature rules based on orientation appear to be more effective than those based on range, possibly because it is difficult to specify how distance affects their applicability. Finally, follow-up analysis with the participants suggests that false positives appear to be more plausible than false negatives, perhaps because it may be easier to justify the possible presence of a semantic action than to claim its complete absence.

Future Work

This pilot study is intended to direct a more detailed study. A number of extensions are under consideration:

- Incorporating a stronger statistical definition of agreement.
- Weighting the computational results to reflect the human results so multiple interpretations can be ranked by preference.
- Including some degree of articulation in the objects.
- Supporting more than one configuration snapshot to provide some degree of temporal progression in a scenario.
- Adding more breadth and depth to the objects and relations under study.
- Considering deeper plan extraction, namely strategic, tactical, and operational elements, for a top-down decomposition from *what* the objects are doing to *how* they are doing it (Azarewicz et al. 1989).
- Running simulations to determine empirical values for some of the *ad hoc* choices in the knowledge base (Tappan 2008c).
- Performing sensitivity analysis to determine how certain properties transition from one value to another.
- Allowing the participants to control the vantage points dynamically. The rendering engine allows complete control over the perspective, but this functionality is not available online. It may facilitate additional interpretations.

Conclusion

This pilot study considered the feasibility of adding higher-level, collective pragmatic analysis of objects to the existing lower-level, individual analysis in the base system. It showed respectable results within a tightly confined environment. These results are not statistically significant due to the small sample size, but they are promising. A full study would undoubtedly uncover many more implicit semantic and pragmatic dependencies.

References

Adorni, G., Di Manzo, M., and Giunchiglia, F. 1984. Natural Language Driven Image Generation. In Proceedings of COLING-84, 495-500. Stanford, CA.

Azarewicz, J., Fala, G., and Heithecker, C. 1989. Template-based multi-agent plan recognition for tactical situation assessment. In Proceedings of Fifth Conference on Artificial Intelligence Applications, 247-254.

Cheng, D. and Thawonmas, R. 2004. Case-based plan recognition for real-time strategy games. In Proceedings of 5th Game-On International Conference, 36-40.

Claus, B.; Eyferth, L.; Gips, C.; Hörnig, R.; Schmid, U.; Wiebrock. S.; and Wysotzki, F. 1988. Reference Frames for Spatial Inference in Text Understanding. In Freksa, C., Habel, C., and Wender, K., eds., Spatial Cognition—An interdisciplinary approach to representing and processing spatial knowledge, 214-226.

Coyne, B. and Sproat, R. 2001. WordsEye: An Automatic Text-to-Scene Conversion System. In Proceedings of SIGGRAPH-01, 487-496, Los Angeles, CA.

Davis, E. 1990. *Representations of Commonsense Knowledge*. San Mateo, CA: Morgan Kaufmann.

Egenhofer, M. and Franzosa, R. 1991. Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems* 5(2): 161-174.

Fleiss, J. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5): 378–382.

Frank, A. 1992. Qualitative Reasoning about Distances and Directions in Geographic Space. *Journal of Visual Languages and Computing* 3(4): 343-371.

Frank, A. 1996. Qualitative Spatial Reasoning: Cardinal Directions as an Example. *International Journal of Geographical Information Systems* 10(3): 269-290.

Freska, C. 1992. Using Orientation Information for Qualitative Spatial Reasoning. In Frank, A., Campari, I., and Formentini, U., eds., Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, LNCS 639, Springer-Verlag, Berlin.

Gapp, K. 1994. Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space. In Proceedings of AAAI-94, 1393-1398, Seattle, WA.

Harter, S. and Hert, C. 1997. Evaluation of information retrieval systems: approaches, issues, and methods. *Annual Review of Information Science and Technology* 32: 3-94.

Herskovits, A. 1986. *Language and Spatial Cognition: An interdisciplinary Study of the Prepositions in English*. Cambridge, UK: Cambridge University Press.

Hernández, D. 1994. *Qualitative Representation of Spatial Knowledge*. Berlin: Springer-Verlag.

Hernández, D.; Clementini, E.; and Di Felice, P. 1995. Qualitative Distances. In Proceedings of Third European Conference on Spatial Information Theory, 45-58, Semmering, Austria.

Kiefer, P. and Schlieder, C. 2007. Exploring Context-Sensitivity in Spatial Intention Recognition. In Proceedings of 1st Workshop on Behavior Monitoring and Interpretation, 102-116.

Koller, H.; Heinze, N.; and Nagel, H.-H. 1992. Algorithmic Characterization of Vehicle Trajectories from Image Sequences by Motion Verbs. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 90-95, Maui, Hawaii.

Lazebnik, S.; Schmid, C.; and Ponce, J. 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Neumann, B. 1989. Natural Language Description of Time-Varying Scenes. In Waltz, D., ed., Semantic Structures: Advances in Natural Language Processing, 167-207. Hillsdale, NJ: Lawrence Erlbaum.

Olivier, P., and Tsujii, J. 1994. A computational view of the cognitive semantics of spatial prepositions. In Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, NM.

Peters, S. and Shrobe, H. 2003. Using Semantic Networks for Knowledge Representation in an Intelligent Environment. In Proceedings of 1st Annual IEEE International Conference on Pervasive Computing and Communications, Ft. Worth, TX.

Porter, S., Whitcomb, M., and Weitzer, W. 2004. Multiple surveys of students and survey fatigue. *New Directions for Institutional Research* 121: 63-73.

Randell, D.; Cui, Z.; and Cohn, A. 1992. A Spatial Logic based on Regions and Connection. In Proceedings of 3rd International Conference on Knowledge Representation and Reasoning, 165-176, San Mateo, CA.

Russell, S. and Norvig, P. 1995. *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Sowa, J. 2000. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole.

Sowa, J., ed. 1991. *Principles of Semantic Networks: Explorations in the Representation of Knowledge by Computers*. New York. NY: Academic Press.

Srihari, R. 1994. Computational Models for Integrating Linguistic and Visual Information: A Survey. *Artificial Intelligence Review* 8: 349-369.

Stevens, A. and Coupe, P. 1978. Distortions in Judged Spatial Relations. *Cognitive Psychology* 13: 422-437.

Tappan, D. 2004a. Knowledge-Based Spatial Constraint Satisfaction. In Proceedings of Florida Artificial Intelligence Research Society International Conference, Special Track on Spatio-Temporal Reasoning, Miami Beach, FL.

Tappan, D. 2004b. Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions. Ph.D. diss., New Mexico State Univ., Las Cruces.

Tappan, D. 2008a. Augmentation of Explicit Spatial Configurations by Knowledge-Based Inference on Geometric Fields. In Proceedings of 2nd International Conference on Knowledge Generation, Communication and Management, Orlando, FL.

Tappan, D. 2008b. Knowledge-Based Spatial Reasoning for Scene Generation from Text Descriptions. In Proceedings of Association for the Advancement of Artificial Intelligence, Chicago, IL.

Tappan, D. 2008c. Monte Carlo Simulation for Plausible Interpretation of Natural-Language Spatial Descriptions. In Proceedings of WorldComp International Conference on Artificial Intelligence, Las Vegas, NV.

Tetko, I., Livingstone, D., and Luik, A. 1995. Neural network studies: Comparison of Overfitting and Overtraining. *Journal of Chemical Information and Computer Sciences* 35: 826-833.

Tsotsos, J. 1985. Knowledge Organization and its Role in Representation and Interpretation for Time-Varying Data: the ALVEN System. *Computational Intelligence* 1: 16-32.

Turner, R. 1998. Context-Mediated Behavior for Intelligent Agents. *International Journal of Human-Computer Studies* 48(3): 307-330.

Tversky, B. 2000. Levels and structure of spatial knowledge. In Cognitive Mapping: Past, present and future, Kitchin, R. and Freundshuh, S., eds. London and New York: Routledge.

Walter, I.; Lockemann, P.; and Nagel, H-H. 1988. Database Support for Knowledge-Based Image Evaluation. In Proceedings of 13th Conference on Very Large Databases, 3-11, Brighton, UK.

Xu, K.; Stewart, J.; and Fiume, E. 2002. Constraint-Based Automatic Placement for Scene Composition. In Proceedings of Human-Computer Interaction and Computer Graphics, 25-34, Calgary, Canada.

Yamada, A. 1993. Studies on Spatial Description Understanding Based on Geometric Constraints Satisfaction. Ph.D. diss., Univ. of Kyoto, Japan.

Yamada, A., Yamamoto, T., Ikeda, H., Nishida, T., and Doshita, S. 1992. Reconstructing Spatial Image from Natural Language Texts. In Proceedings of COLING-92, 1279-1283, Grenoble, France.