Geolocating Orientational Descriptions of Landmark Configurations

James Pustejovsky[1], Marc Verhagen[1], Anthony Stefanidis[2], Caixia Wang[2]

[1] Department of Computer Science
Brandeis University, Boston, MA
{jamesp, marc}@cs.brandeis.edu
[2] Center for Geospatial Intelligence
George Mason University, Fairfax, VA
{astefani, cwangg}@gmu.edu

Abstract. In this paper we outline how to translate verbal subjective descriptions of spatial relations into metrically meaningful positional information, and extend this capability to spatiotemporal monitoring. Document collections, transcriptions, cables, and narratives routinely make reference to objects moving through space over time. Integrating such information derived from textual sources into a geosensor data system can enhance the overall spatiotemporal representation in changing and evolving situations, such as when tracking objects through space with limited image data. We focus on landmark identification, since it proves to be a more tractable problem than open-domain image recognition.

Keywords: Spatial language, geolocating, spatial configurations, landmarks.

1 Introduction

The relation between language and space has long been an area of active research. Human languages impose particular linguistic constructions of space, of spatially-anchored events, and of spatial configurations that relate in complex ways to the spatial situations in which they are used. Establishing tighter formal specifications of this relationship has proved a considerable challenge and has so far eluded general solutions. One reason for this is that the complexity of spatial language has often been ignored. In much earlier and ongoing work, language is assumed to offer a relatively simple inventory of terms for which spatial interpretations can be directly stated. Examples of this can be found not only in accounts that focus on formalizations of particular tasks, such as path and scene descriptions, navigation and way-finding, but also in foundational work on the formal ontology of space, on qualitative spatial calculi, and on cognitive approaches.

Visual information in human experience is frequently accompanied by a linguistic description of the image or scene. Consider, for example, the image in Figure 1. If the goal is to identify the region of the image where one should look for the lost keys, one first must identify the correct tree. If this image is automatically segmented using a stock library of images for trees and entrances (Millet et al., 2005; Hollink et al., 2004), several candidate regions for “tree” and “entrance” will be identified. Each candidate region may then be ranked with respect to how likely it is to correspond to a tree or an entrance, producing two ranked lists of candidate regions, T = (T1; T2; : : :)

and E = (E1;E2; : : :), where Ti are the candidate regions for “tree”, Ti ranks higher than Ti+1, and Ej are the candidate regions for “entrance”. The associated verbal description invokes the “left of” relation, thereby restricting the search for the appropriate pair of candidate regions by imposing the corresponding spatial constraint: LEFT_OF(Ti;Ej). The (Ti;Ej) pairs that do not satisfy the specified spatial relation are given lower ranking, thus increasing the likelihood of identifying correctly the relevant region in the image.

Fig. 1: Speaker: *See the tree to the left of the entrance? I dropped my keys under that tree.*

Over the past decade, image annotation has been the focus of attention within several research areas, in particular, in the context of content-based image retrieval (CBIR). Some research, including the work done within the TRIPOD project at Sheffield, examines the different ways that geo-referenced images can be described (Edwardes et al., 2007), though different approaches, such as the ESP Game can also be used to address this problem. Much of the work on text-based image retrieval has relied on extracting information about the image from image captions, as well as the surrounding text and related metadata, such as filenames and anchor text extracted from the referring web pages, as for example, in Yahoo!’s Image Search. Another kind of image annotation data has become available with the rise of “citizen geography”. User-annotated geo-referenced digital photo collections allowing for image content labeling and annotation are being generated in distributed environments, such as Flickr and GoogleEarth. Images are indexed with user-supplied labels that typically form a particular language subset (Grefenstette, 2008). Under such schemes, however, detailed image content annotation is not provided. A notable exception is the “Flickr notes” feature that allows users to annotate regions within images. This and other adaptations of the Fotonotes image annotation standard and the associated software provide an opportunity for detailed annotation of images with both captions and extended free text associated with each annotated image region.

2 Geolocating Descriptions of Landmark Configurations

While such efforts as those discussed above are useful metadata encodings over images, there remain significant problems with unconstrained object recognition. Hence, in this paper, we will focus on linguistic descriptions of *landmark configurations*. Landmarks are visually identifiable objects with fixed spatial locations, which carry semantic meaning for large groups of individuals. They are typically large man-made or physical structures (e.g. buildings, communication antennas, hills) and play an important role in navigation and wayfinding decisions (see e.g. Werner et al, 1998; Steck and Mallot, 2000). For example, routes can be expressed as sequences of landmarks (Duckham et al., 2010) and paths connecting them. The saliency of different landmarks can be expressed in terms of their perceptive, cognitive, and contextual value (Caduff and Timpf, 2008). In this section we address their role for geolocating an observer describing their relative orientational properties in his/her view of a scene.

Let us consider the scene depicted in Fig. 2, taken from Google StreetView, of the intersection of Huntington and Mass. Avenues in Boston. In it we can identify reference landmarks, namely three buildings: Horticultural Hall (HC), Prudential Center (PC) and the Christian Science Monitor building (CSM). It also comprises various other objects, for example a white van, a black truck, and a red car. Our interest is in geolocating the observer of this scene by using orientational descriptions of the relative appearance of the landmarks contained in it.

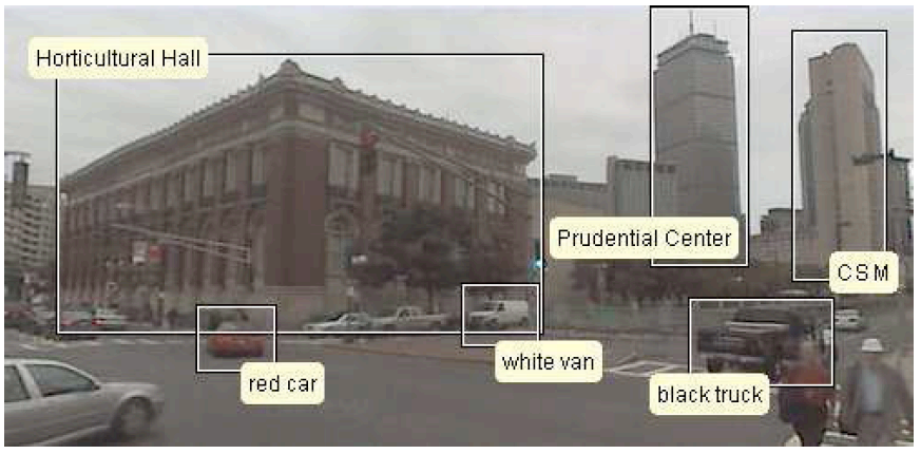

Fig. 2 Landmarks and objects identified in a ground-view.

Assuming that a narrator is familiar with these three landmarks, he/she could describe the scene as follows:

I see the SW and SE sides of Horticultural Hall, and
to the right of it I see the SW and SE sides of the Prudential Center, and
to the right of it I see the Christian Science Monitor Building.

In this situation the narrator has described the scene through three types of statements:

- *explicit reference to specific landmarks*, positioning the scene in their vicinity,
- *explicit description of orientational properties* expressing the relative positions of these landmarks in an observer-centric system[1], and
- *implicit visibility declarations*, whereby she indicates that she can observe specific façades of landmark buildings.

The orientational properties are modeled using ISO-Space (Pustejovsky et al., 2011). ISO-Space distinguishes two major types of elements: entities and relations. Entities include *location*, *spatial entity*, *motion*, *event* (or *spatial state*), and *path*. The two main relations between these entities are the *distance relation* and the *qualitative spatial relation*, which can be either a topological or a relative spatial relation.

Relations such as "to the right of" are annotated as a relative spatial relation between two elements, the figure and the ground, and the viewer perspective is accounted for by two further attributes on the link tag: *rframe*, with values absolute, relative and intrinsic, and *viewer*, which contains a variable indexed to the viewer (Levinson, 2003, Freksa 1992, Ligozat, 1998). Using the three kinds of information above (landmarks, relative positions and visibility declarations), we can identify the three landmarks in a GIS (Fig. 3), and proceed to estimate the location of the observer through a series of view analysis and visibility polygon overlays as we describe below.

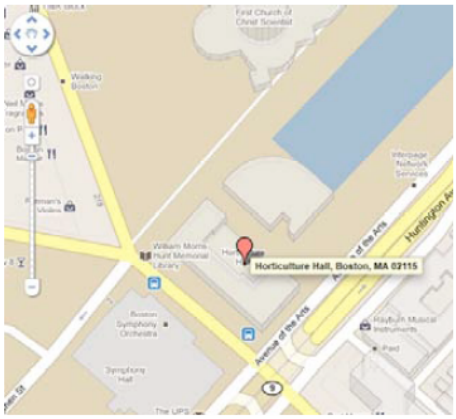

Fig. 3 Map location of Horticultural Hall.

For every visibility statement we can identify a *visibility zone* through *viewshed analysis*, using the local GIS information (Kim et al., 2004). The 2D visibility zone of a specific façade (or any other object in space) is the locus of all points from which at least a part of this façade is visible. For example, in Fig. 4, the visible zone of façade

[1] An alternative would be to use the intrinsic orientation of the landmark, in which case "to the right" would be interpreted relative to the landmark and not relative to the observer. Clearly, both options would need to be explored down-stream.

F_1 is shown as the gray-shaded area. From any point outside this area it would be impossible to see façade F_1.

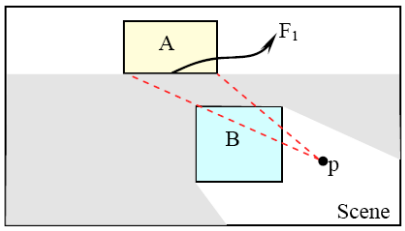

Fig. 4. The visibility zone (gray shaded area) for façade F_1 of Building A.

Each additional visibility statement introduces additional visibility zone information, and the location of the narrator can be eventually determined through the intersection of the corresponding visibility zones through polygon clipping techniques, such as Weiler-Atherton (1977). Fig. 5 shows the implementation of this process for the scene of Fig. 1, through a progressive assessment of visibility conditions for HC, CSM, and PC.

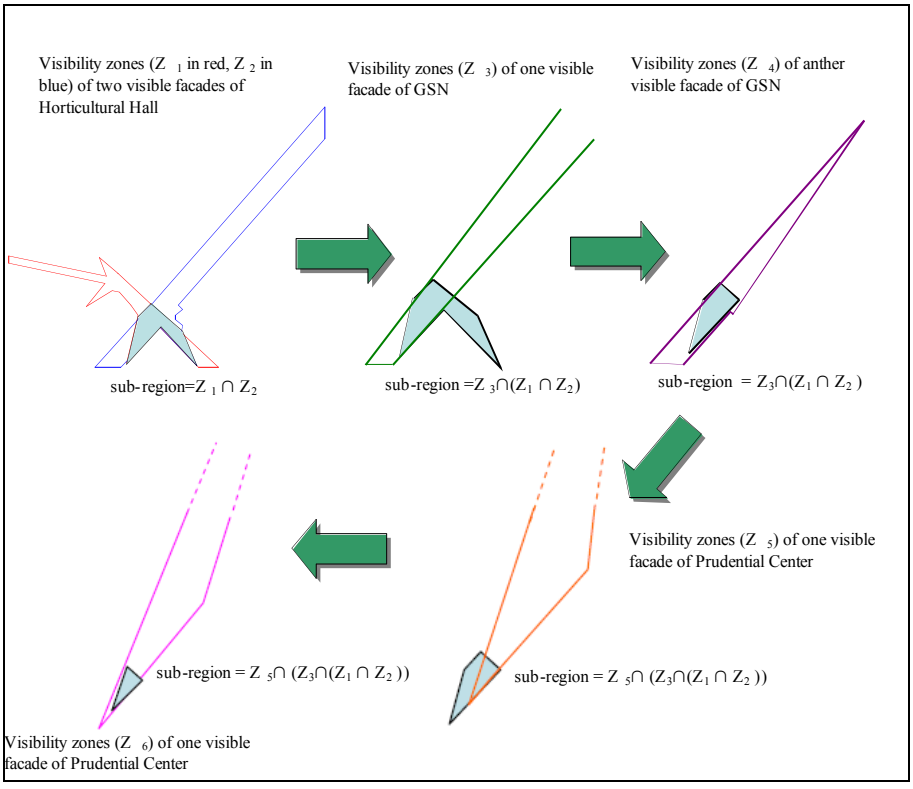

Fig. 5. The progressive visibility intersection process

The narrator position estimated through the process visualized in Fig. 5 is shown on the local map in Fig. 6, marked as a red triangle. The triangle corresponds to all positions from which the narrator would have a view of our scene that would be comparable to the one depicted in Fig. 2 in terms of the orientational relationships of the three depicted landmarks.

Fig. 6. The estimated location of the narrator, indicated as a red triangle, and the views used to estimate it.

3 Conclusion

In this paper, we discuss the integration of multi-source data analysis for spatial knowledge extraction from images. In particular, we focused on the specific contribution of verbal subjective descriptions of spatial relations involving orientation, and how these can be translated into metrically interpretable positional statements within a GIS environment. We concentrated on the more tractable subproblem of landmark identification. Orientational information in language was modeled with ISO-Space annotation, providing both qualitative spatial relations and anchored GPS values, once geolocating is performed.

This work is ongoing research aimed to allow for the integration of information available from different sources, better addressing the evolving needs of the geoinformatics community. Our preliminary results suggest that scene content information provided by verbal description can be mapped faithfully to metrically grounded information. As this is preliminary work, there are clearly many details to be worked out. For example, we have not yet precisely defined how orientational relations are used to identify a landmark in a GIS, especially with anonymous landmarks, a problem exacerbated when an ambiguity between intrinsic and observer-based relative orientation cannot be easily resolved.

One of the ultimate goals of this research is the development of algorithms that take an image and accompanying verbal utterances and maps these to a partition of a 2D grid. This application would be tuned to deal with more natural utterances than the somewhat stilted verbal descriptions given with Fig. 2 above.

4 Acknowledgements

This research was funded under the NURI grant HM1582-08-1-0018 by the National Geospatial Agency.

References

Caduff D., Timpf S. On the Assessment of Landmark Salience for Human Navigation. Cognitive Processing, 9(4), 249-267 (2008).

Duckham M., Winter S., Robinson M. Including Landmarks in Routing Instructions. J. Location Based Services, 4(1), 28-52 (2010).

Edwardes, R. Purves, S. Birche, and C. Matyas. Deliverable 1.4: Concept ontology experimental report. Technical report, TRIPOD Project (2007).

Freksa, Christian. Using orientation information for qualitative spatial reasoning. In A. Frank, I. Campari, and U. Formentini, eds, Theories and methods of spatiotemporal reasoning in geographic space, pages 162–178. Springer, Berlin, (1992).

Grefenstette, G. Comparing the Language Used in Flickr, general Web Pages, Yahoo Images and Wikipedia. In OntoImage 2008, LREC, pages 6–11, (2008).

Hollink, L., G. Nguyen, G. Schreiber, J. Wielemaker, B. Wielinga, and M. Worring. 2004. Adding spatial semantics to image annotations. In Proceedings of 4th International Workshop on Knowledge Markup and Semantic Annotation, 3rd International Semantic Web Conference.

Kim Y.-H., Rana S., Wise S. Exploring Multiple Viewshed Analysis using Terrain Features and Optimisation Techniques. Computers & Geosciences, 30(9-10), pp. 1019-1032 (2004).

Levinson, S. C. Space in Language and Cognition, Cambridge University Press, (2003).

Ligozat, G. Reasoning about cardinal directions. *Journal of Visual Languages and Computing*, 9:23-44. (1998).

Millet, C., I. Bloch, P. Hede, and PA Moellic. Using relative spatial relationships to improve individual region recognition. In Proc. 2nd Eur. Workshop Integration Knowledge, Semantics and Digital Media Technology, pages 119–126. (2005).

Pustejovsky, J., J. Moszkowicz, and M. Verhagen. ISO-Space: The Annotation of Spatial Information in Language, in Proceedings of ISA-6: ACL-ISO, Oxford, England, (2011).

Steck S., Mallot, H.: The Role of Global and Local Landamarks in Virtual Environment Navigation. Presence, 9(1), 69-83 (2000).
Weiler K., P. Atherton. Hidden Surface Removal using Polygon Area Sorting. ACM SIGGRAPH Computer Graphics, 11(2), 214-222 (1977).
Werner S., Krieg-Brueckner B., Mallot H., Schweizer K., Freska C.: Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation. In: Jarke M., Pasedach K., Phl K. (eds.) Informatik '97, pp. 41-50, Springer Verlag (1997).