Qualitative Reasoning for Automated Traffic Surveillance

Jonathan H. Fernyhough

Division of Artificial Intelligence
School of Computer Studies
University of Leeds, Leeds, LS2 9JT
jfern@scs.leeds.ac.uk

Introduction

The simultaneous interpretation of object behaviour from real world image sequences is a highly desirable goal in machine vision (Retz-Schmidt 1988; André, Herzog, & Rist 1988). Although this is rather a sophisticated task, the complexity can be reduced in stylized domains by the use of qualitative reasoning techniques combined with a context specific spatial model of that domain.

This paper provides a review of the system we use to generate such a spatial model before presenting work in progress on the use of qualitative reasoning techniques using the model for automated traffic surveillance.

Generation of the Spatial Model

Howarth & Buxton (1992) introduced what appears to be an ideal spatial model for spatial event detection in the domain of traffic surveillance. This representation is a *region* based model where a region is defined as a (closed) 2D area of space where the spatial extent is controlled by the continuity of some property.

There are two kinds of regions: (1) *Leaf regions* form the finest granularity of region providing structure to the space. They are completely defined by the intersections of *composite regions*. (2) Concatenations of adjacent leaf regions form *composite regions* expressing areas of the same or similar behavioural significance. Different composite regions may share leaf regions (i.e. they may overlap) providing a hierarchical structure to the spatial layout. In the domain context, a composite region represents the area described by the movement of objects within the domain (i.e. a *path*.)

Unfortunately such representations of space were produced manually for each new domain: a time consuming and painstaking process. We produce a similar (2D) model automatically based on the movement of objects within the domain. The analysis of *dynamic scene* data is used to build a database of paths used by objects within the scene.

Dynamic scene data is provided as a list of objects, frame by frame, using the tracking process described in Baumberg & Hogg (1994). Visual information for the tracking process is provided through live video image sequences of the test domains from a static camera (figure 1).

The combination of the locations a tracked object occupies in its course through the domain defines the spatial extent of that object's path. Completed paths are entered into the path database. If an equivalent path already exists (i.e. another object has followed a similar course through the domain and a significant proportion of their paths match) then the new path is merged with the database path adjusting its spatial extent.

At *any* time during this training period it is possible to generate regions for the spatial model. Effectively, this halts the database generation process and uses that information to build the regions. A database verification step is used to ensure that no path equivalences have been created during update of the database. It is possible that previously unmatched paths may now be found equivalent due to the expansion of a path's spatial extent. Should any equivalent paths be identified they are merged together as before.

In order to appropriate the paths from the database that express the composite regions for the spatial model it is necessary to discard any database path with minimal use. Such paths may represent "noise" or abnormal behaviour and as this method relies on behavioural evidence it is safe to discard them.

As noted, the leaf regions are completely defined from the intersections of the spatial extents of the remaining database paths. Occasionally, adjacent paths may share small areas of common ground — perhaps from shadows or the occasional large vehicle. This can generate very small leaf regions that provide little or no benefit to the spatial model. Such small regions are removed by merging with an adjacent region and by considering the smoothness of the resulting region. To

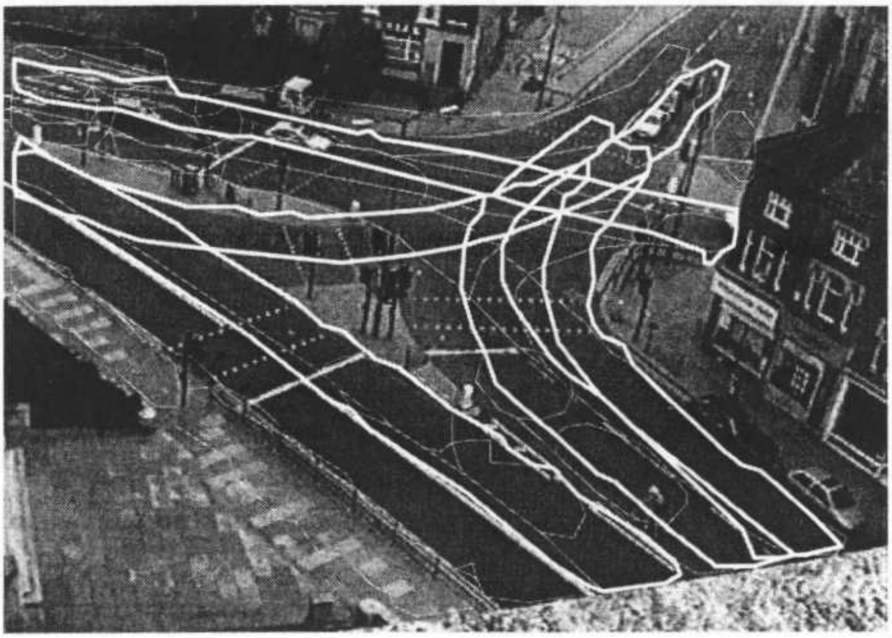

Figure 1: Road junction displaying identified leaf regions and a selection of composite regions.

complete the spatial model, it is necessary to discover the union of leaf regions which make up each composite region. Figure 1 shows results from one of our test domains. More complete details of this region generation process can be found in Fernyhough, Cohn, & Hogg (1996).

Event Determination

Event recognition provides a significant challenge to high-level vision systems. Nagel (1988) outlines several applications connecting vision and natural language systems to provide retrospective descriptions of analysed image sequences. Typically the vision component provides a 'geometric scene description' (GSD) describing the spatial structure and object locations within the scene at each interval. A generic event model (Neumann & Novak 1983), characterizing a spatio-temporal representation of an event can be matched against the GSD to recognize instances of that event which can then be expressed in natural language.

Generic event models are usually provided manually as part of the *a priori* knowledge for the system. We are attempting to build such models from the analysis of object movements and interactions over extended periods.

Considering just an object's location, it is already possible to predict the course an object will take through the scene just from its initial location. The spatial model is based on typical object movement and lane following tends to be the most typical behaviour. Similarly, lane changing can be discerned from a change in the predicted course an object will take.

Further analysis involves building object histories containing qualitative information concerning object location (in terms of the occupied leaf regions) and that object's association with related objects (i.e. those objects in the same or an intersecting composite region.) These relationships are modeled using a combination of the qualitative spatial calculi to be found in Cui, Cohn, & Randell (1992) and Schlieder's (1995) ordering information.

Object histories can then be analysed to find similarities which will provide the basis for scene specific event models allowing the recognition of more complex events (e.g. overtaking and giving-way.)

Conclusion

We have demonstrated how a context specific (2D) model of space can be created automatically based on the movement of objects within a scene. Although the generation of the spatial model relies on quantitative knowledge it is not necessary for the current work. Using the spatial model we can represent object locations and interactions qualitatively with the intention of simplifying the analysis of object histories used to generate event descriptions.

References

André, E.; Herzog, G.; and Rist, T. 1988. On the simultaneous interpretation of real world image sequences and their natural language description: The system soccer. In *Proc. ECAI-88*, 449–454.

Baumberg, A. M., and Hogg, D. C. 1994. Learning flexible models from image sequences. In *European Conference on Computer Vision*, volume 1, 299–308.

Cui, Z.; Cohn, A.; and Randell, D. 1992. Qualitative simulation based on a logical formalism of space and time. In *Proceedings of AAAI-92*, 679–684. Menlo Park, California: AAAI Press.

Fernyhough, J. H.; Cohn, A. G.; and Hogg, D. C. 1996. Generation of semantic regions from image sequences. In *Fourth European Conference on Computer Vision*. ECCV. To appear.

Howarth, R. J., and Buxton, H. 1992. An analogical representation of space and time. *Image and Vision Computing* 10(7):467–478.

Nagel, H. H. 1988. From image sequences towards conceptual descriptions. *Image and Vision Computing* 6(2):59–74.

Neumann, B., and Novak, H.-J. 1983. Event models for recognitions and natural language description of events in real-world image sequences. In *Proceedings of the Eighth IJCAI Conference*, 724–726.

Retz-Schmidt, G. 1988. A replai of soccer: Recognizing intentions in the domain of soccer games. In *Proc. ECAI-88*, 455–457. Munich: Pitman.

Schlieder, C. 1995. Reasoning about ordering. In Frank, A. U., and Kuhn, W., eds., *Spatial Information Theory*, Lecture Notes in Computer Science No. 988, 341–349. Semmering, Austria: COSIT'95.