D2.4 A qualitative spatial inference service for a visual agent

A number of partners, particularly Leeds and Hamburg are well known internationally for their work on qualitative spatial reasoning. This work has certainly greatly influenced the direction and goals of a number of parts of CogVis. However, we did not choose to focus on the production of qualitative spatial techniques and tools in the abstract, but rather to integrate them into the rest of the cognitive vision systems that we built. This can be seen in deliverables attached where we have chosen to show how qualitative spatial reasoning techniques can be incorporated, in a complementary way, to more traditional machine vision. In the rest of the introduction to this deliverable, we summarise the three papers which form the reports comprising this deliverable. The first two have corresponding implementations, the last is of a more foundational nature and is currently a purely theoretical analysis, though it is based on a formalism (description logic) which has many practical instantiations and applications.

1. In the first piece of work (Report 1), which has been accepted to appear in ECAI-04 [2] and has been submitted to a special issue of Image and Vision Computing [1] (it is this longer version which is attached as a deliverable, along with the Prolog code (see file d2-4-enhanced-tracking-code.tgz)), we show how qualitative spatio-temporal reasoning can be used to improve the accuracy and reliability of tracking. We assume that our domains of interest involve objects which move in space over time; thus the ability to track objects over time is critical. We do not wish to assume the existence of a ready made library of object models. At the lowest level therefore, we have built a tracker[4] which uses foreground, background and motion models. Each pixel is *e*xplained either as background, belonging to a foreground object or as noise.

   In brief, the process is as follows: blobs are tracked and features are extracted from a bounding box. Features (e.g. a colour histogram) are clustered in a learning phase to form a profile for each object from training data. These profiles can then be used to compute the *r*aw likelihood of an object being one of these learned profiles.

   However, in general, there is no certainty that the most likely hypothesis so computed will indeed be the correct object. Moreover, in the case of occlusion, the blob may in fact correspond to multiple objects. This

is of course a classic computer vision problem. Our approach to this is to deliberately under-segment, then form a qualitative spatio-temporal description which abstracts the essential aspects of the video sequence, and then reason over this high level representation, applying commonsense notions of continuity to disambiguate the initial frame-by-frame low level classification hypotheses. In other words, we deliberately increase vagueness, by drawing bounding boxes sufficiently large to be sure that they enclose all spatially local objects (rather than risking cutting some in half), and then reason about box occupancy explicitly, relying on later (or possibly earlier) data in which the relevant objects are not occluding each other to provide more secure knowledge.

The lower level tracking and recognition systems explicitly model the possibility of ambiguity and error by assigning probabilities for the presence of objects within bounding boxes in each video frame. This output is passed to a reasoning engine which constructs a ranked set of possible models that are consistent with the requirements of object continuity. The final output is then a globally consistent spatio-temporal description of the scene which is maximally supported by probabalistic information given by classifier.

We show experimentally that tracking is improved, particularly for multi-object boxes (i.e. those where occlusion has occurred). Further experimentation with other sequences/domains is ongoing.

2. In very new work (Report 2) [5] in collaboration with DIST, we are developing a framework for the integration of a spatial inference service with an active vision system using a steerable webcam, and three levels of attention.

   A symbolic representation is formed from lower level continuous models. A particular qualitative spatial calculus is used in which scene objects are used to create a coordinate frame, within which the relative position of other objects can be described using predicates such as "leftof" and "rightof". The choice of particular calculus is not crucial to the overall framework; the important idea is that this module provides a service in which qualitative spatial descriptions are computed and stored, and then used to help drive a symbolically learned spatio-temporal attention mechanism: our hypothesis is that modelling spatial relationships qualitatively provides a much more powerful model for expectation than modelling absolute positions.

3. The third item in this deliverable is work of a theoretical nature ad-

dressing the problem of how to develop formal languages for representing, and reasoning about moving spatial scenes [3]. The challenge is to find an appropriate tool for putting together, on the one hand, existing languages for the representation of time, and existing languages for the representation of space. This achieved via the well-known Description Logic (DL) ALC augmented with a concrete domain D, and known as ALC(D). The nodes of ALC(D) interpretations (the set of possible worlds, in modal logic terminology) are time points or time intervals, depending on whether we use discrete or continuous time; the roles are temporal relations on points or on intervals, again depending on whether we use discrete or continuous time; and the concrete domain is spatial, and is generated by any RCC8-like spatial relation algebra. The paper (Report 3) [3] presents these ideas in more detail and includes a discussion on the formalisation of continuous spatio-temporal change (a key component of the first component of the deliverable D2.4 cited above).

References

[1] B. Bennett, D. R. Magee, A. G. Cohn, and D. C. Hogg. Enhanced tracking and recognition of moving objects by reasoning about spatio-temporal continuity. Submitted to special issue of IVC on Cognitive Vision, 2004.

[2] B. Bennett, D. R. Magee, A. G. Cohn, and D. C. Hogg. Using spatio-temporal continuity constraints to enhance visual tracking of moving objects. In L. Saitta, editor, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-04)*. ECCAI, IOS Press, 2004. to appear.

[3] A. Isli. A family of qualitative theories for continuous spatio-temporal change as a spatio-temporalisation of alc(d) - first results. In *Proc. of ECAI-02 Workshop on Spatial and Temporal Reasoning*, pages 81–86, Lyon, France, 2002.

[4] D. R. Magee. Tracking multiple vehicles using foreground, background and motion models. *Image and Vision Computing*, 20(8):581–594, 2004.

[5] C. Needham, D. Magee, and S. Rao. A framework for attention and spatio-temporal inference for an embodied visual agent. Draft Paper, 2004.