Visual Surveillance Monitoring and Watching

Richard Howarth and Hilary Buxton

School of Cognitive and Computing Sciences,
University of Sussex,
Falmer, Brighton BN1 9QH, UK

Abstract. This paper describes the development of computational understanding for surveillance of moving objects and their interactions in real world situations. Understanding the activity of moving objects starts by tracking objects in an image sequence, but this is just the beginning. The objective of this work is to go further and form conceptual descriptions that capture the dynamic interactions of objects in a meaningful way. The computational approach uses results from the VIEWS project [1]. The issues concerned with extending computational vision to address high-level vision are described in the context of a surveillance system. In this paper we describe two systems: a passive architecture based on "event reasoning" which is the identification of behavioural primitives, their selection and composition; and an active architecture based on "task-level control" which is the guidance of the system to comply with a given surveillance task.

1 Introduction

Until recently it has been rare to find issues of control connected with computer vision (but see for example Rimey and Brown [26] and Howarth [17]). The focus tends to be on techniques that extract information from images rather than on identifying visual behaviours appropriate for visual tasks and how these operate. Ballard's landmark paper [2] identified how these two approaches could be integrated in what he called "animate vision". In this paper, we describe some of the advantages obtained by reinterpreting a pipelined, passive vision system under a more active vision approach. We use surveillance of wide-area dynamic scenes as our problem domain.

Our surveillance problem has the following simplifications that make visual understanding more tractable: we use a fixed camera that observes the activity of rigid objects in a structured domain. Examples include: a road traffic scene where the main interest is the road vehicles, and airport holding areas where we are interested in the activities of the various special vehicles that unload and service the passenger aeroplanes. We call this single viewpoint of the fixed camera the "official-observer". From this camera input we wish to obtain a description of the activity taking place in the dynamic wide-area scene, and then an understanding of the dynamic and improvised interactions of the scene objects.

[1] Thanks to project partners on ESPRIT EP2152 (VIEWS) and to EPSRC GR/K08772 for continued funding of this work.

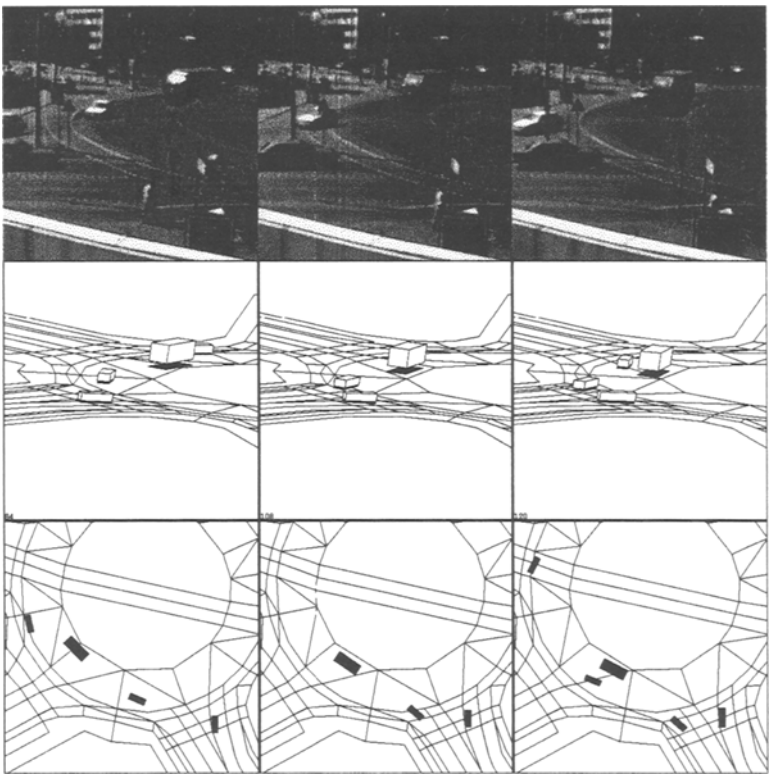

Fig. 1. Three images showing typical vehicle activity on the roundabout and how the 3D pose descriptions can be transformed to a ground-plane view.

There are constraints on the official-observer's interpretation of the objects in the scene: we only see the objects that are in the camera's field-of-view; we do not know each participant's goal (typically something like "go to place X"); and what we see is mostly reactive behaviour (rather than deeply planned).

To illustrate the difference between the passive and active approaches we will describe two systems that represent different formulations of the surveillance problem. The first is called HIVIS-MONITOR, embodying the initial design. The second system, called HIVIS-WATCHER, is a response to the problems encountered while developing our initial system. To demonstrate the different behaviour of the two systems we will use examples drawn from the road traffic domain. Here (figure 1) we illustrate this with three image frames selected from a sequence taken at a German roundabout. In this part of the sequence a number of episodic behaviours are unfolding: one vehicle leaves the roundabout; another is in an entry lane to the roundabout; also towards the rear of the image a car begins to overtake a lorry. Below the image frames we provide an illustration of the poseboxes, which are results from a model-matcher (see [9, 31] for details).

2 HIVIS-MONITOR

In our first system, HIVIS-MONITOR, we adopt a pipelined approach that reflects the general flow of data from images to conceptual descriptions. The visual processing centres around three components: extracting spatio-temporal primitives from the stream of compact encodings produced by low- and intermediate-level visual processing, detecting events from these primitives, and composing the events to form episodic sequences which are stored in an evolving database. This database is extended as new events continue, begin or end the various episodes under construction. The problem of matching behaviours to a user question is left to the query-based component that interrogates the database. At first sight, this seems an admirable system design, allowing the parallel, separate development of the perceptual processing component and the behavioural one. However, as we will see, the passive, data-driven flow of the processing causes problems for visual control.

2.1 Script-based approach

To describe the behaviour of participants in the scene we are using an ontology based upon that described by Nagel [23] to describe events, which captures the common sense notions of the terms being used. Neumann [24] provides an example set of events used in his NAOS system. Our use of events and episodes is similar in some respects to that described by Schank and Abelson [27], who compose events into scripts to describe typical behaviour of a customer at a restaurant such as entering, going to a table, ordering, eating and paying. This provides a hierarchical layering from events to episodes, and then to more complex script-like behaviours. This hierarchical decomposition and relationships between the behavioural elements can be used to define a grammar where events are terminal symbols in the language to be parsed. This approach could use syntactic methods such as attributed grammars as described by Frost [12] and Clark [6], or the island parsing described by Corrall and Hill [8].

HIVIS-MONITOR is data-driven and follows the script-based approach, constructing an interpretation of object behaviour in an evolving database that holds entries for the history of each individual object and the interactions between them. This approach reflects the flow of data from image to conceptual descriptions. Maintaining a history of the behaviour that takes place in the scene involves noting the event primitives that have been detected and then using an ongoing interpretation process to see how these events fit together. The input given to the database consists of the events and activities associated with a particular property. In addition to the functions that compute these values there are further functions that update the temporal structure by beginning, extending or ending the continuity of the value/signal for each property. To identify an episode we use a filter that matches the necessary property values.

2.2 Spatio-temporal representation

The spatial representation used by HIVIS-MONITOR is based on a model of space developed by Fleck [10, 11] for representing digitised spaces for both edge detection and stereo matching. Also, in her thesis [11], she describes how this representation can be used for qualitative reasoning and for modelling natural language semantics. The spatial and temporal representation Fleck uses and calls "cellular topology" is based on the mathematical foundation of combinatorial topology. Cellular topology uses cells to structure the underlying space and is augmented here by adding a metric (see also [16, 18]). It is to this underlying spatial representation that we can attach information about the world.

The stream of posebox data supplied by the model-matcher describes the space swept out in time by each object's path to form what we call a "conduit". The conduit is used to provide an approximation of the time at which a region is exited or entered. To do this, we extrapolate the space-time description between updates. Once we have generated the conduits, we have the problem of interpreting what they mean. If they intersect then there is a likely collision or near miss, but intersections of conduits is unusual. Other tests can be made possible by removing a pertinent dimension and testing to see if the components of the reduced model overlap, in the test for `following` behaviour we tested for an overlap with some time delay. Overtaking can be identified by ignoring the spatial dimension parallel to the objects direction of motion however, this spatial dimension should really be the 2D manifold that fits the space curve of each object's path. Mapping the conduits into one of these manifolds to perform such as test is difficult, although in principle it should be possible.

2.3 General features

We claim that HIVIS-MONITOR demonstrates typical traits of the class of traditional AI approaches we have called "script-based". In general, all script-based systems will have the following features: Maximal detail is derived from the input data. This approach obtains a description of all objects and all interactions, over the whole scene, for all the episodes it has been designed to detect; Representation is extracted first and the results are placed in an evolving database that is used to construct more abstract descriptions using hindsight. Single object reasoning is performed with ease using this approach. Simple implementation can be achieved using standard AI techniques. It is quite likely that better implementations could be developed that fulfill the script-based approach [2] but there would still be limitations.

2.4 Limitations

HIVIS-MONITOR has the following limitations:

[2] Achievements of the project are illustrated by the video [3] of the ESPRIT project VIEWS, and by Corrall and Hill [7, 8], King et al. [21] and Toal and Buxton [28]

- It is passive in its processing, operating a simple control policy, that is, not affected by changes in the perceived data.
- It is not real-time because the construction of the results database is an off-line process, and does not send feedback to any form of intermediate-level visual processing. This means that there is a problem getting timely recognition of perceived object activity.
- Unbounded storage is required because any pieces of data contained in the results database might be needed later either to compose some more abstract description or to be accessed by the user to answer a query. Since we do not retract what we have seen or the episodes that we have identified, the database structure is monotonically increasing in size.
- Multiple object reasoning is difficult within the global coordinate system used to express pose positions. A solution to this is needed because contextual knowledge is not enough to analyse the interactions, although it does provide a context for interpretations.
- The computation performed by HIVIS-MONITOR is mainly dependent upon the number of objects in the input data, i.e., it is *data-dependent.*
- This model is inflexible because it only deals with known episodes. Within the constraints of the predicates provided (language primitives that describe events and activities), new behavioural models can be added. However, defining new predicates may be difficult.
- The addition of new operators increases the number of tests performed on all the objects in the scene. For a single object operator there is a $O(n)$ increase, for most binary object operators there is a $O(n^2)$ increase, and for most multiple object operators the increase is polynomial with a maximum of $O(n^n)$, where n is the number of objects in the scene.
- The behavioural decomposition does not take into consideration the temporal context in which the events have occurred, which contributes to the process of interpretation. It is possible that the selection of the "correct" episode description is not possible due to only seeing part of an episode.

2.5 Discussion

From these features and limitations we can identify the following key problems: computation is performed to obtain results that may never be required; and as the database of results increases in size, the performance of the system will degrade. It might be possible to address these by extending the script-based approach however, we will not take this evolutionary route. Instead we will investigate a more situated approach. This new approach differs greatly from the passive, data-driven script-based approach and requires a complete reformulation of the problem to obtain an active, task-driven situated solution.

3 Reassessment

To begin this reformulation we first consider the use of more local forms of reasoning in terms of the frame-of-reference of the perceived objects, the spatial

arrangements of these objects and the use of contextual indexing from knowledge about the environment. In HIVIS-MONITOR a global extrinsic coordinate system was assumed. By taking a global view we comply with a commonly held Western view of how to represent space in a map-like way as opposed to the egocentric approach described by Hutchins [20] as being used by the Micronesians to perform navigation. The absolute coordinate system also fits well with the concept of the optic-array (see Gibson [13] and Kosslyn et al. [22] for details), if we can consider placing a grid over the ground-plane to be analogical to the optic-array of the perceiver. This representation would allow reasoning to be performed that does not need full object recognition with spatial relationships represented in terms of the optic-array's absolute coordinates (in some respects this is like the video-game world used by Agre and Chapman [1] where the "winner-takes-all" recognition mechanism (see Chapman [5] and Tsotsos [29]) allows objects and their positions to be identified by key properties, such as, colour and roundedness).

In contrast to this global viewpoint, when reasoning about the behaviour of each scene object it would be useful if the representation of the properties related to each object could be described in its own relative coordinate system. However, this involves recognising each object to the extent that an intrinsic-front can be identified together with its spatial extent. This requirement places the need for a more sophisticated understanding of how the image data present in the optic-array relates to how objects exist in the environment. In our surveillance problem we can obtain the pose-positions of the scene objects via model-matching making local reasoning attractive, although its extra cost in terms of the complexity of intermediate-level vision should be noted. The **local-form** is representation and reasoning that uses the intrinsic frame-of-reference of a perceived object (exocentric with respect to the observer). The **global-form** is representation and reasoning that uses the perceiver's frame-of-reference, which operates over the whole field-of-view (egocentric with respect to the observer). The global-form is not a public-world since it, like the local-form, only exists for the perceiver. We are not dealing with representing a shared world in terms of each participant. The suitability of each HIVIS-system is detailed in table 1, indicating the extent of the reformulation for the surveillance problem.

HIVIS-MONITOR would be useful for off-line query of behaviour, whereas in HIVIS-WATCHER, by asking the question first, we remove the importance of the results database because we are no longer providing a query-based system. This removes the need to remember everything and solves the problem of the monotonically increasing database because in HIVIS-MONITOR it is difficult to know when something can be forgotten. The development of a more situated approach in HIVIS-WATCHER is part of the adoption of a more local viewpoint that uses a deictic representation of space and time. In some applications, using HIVIS-MONITOR and processing all scene objects might be necessary however, in cases where it is not, the HIVIS-MONITOR approach is ungainly. In the surveillance problem where we are inherently concerned with the "here-and-now" (the evolving contexts of both observer and scene objects), it is important

HIVIS-MONITOR	HIVIS-WATCHER	illuminates
off-line/pipelined	on-line	immediacy
structured	purposive	approaches
global	local	viewpoint
maximal detail	sufficient detail	investigation
passive	active	control
unlimited resources	limited resources	complexity
extract representation first	ask question first	timeliness
answer question from representation data	answer question from scene data	memory cost
data dependent	task dependent	propagation

Table 1. This table summarises the comparison between the two HIVIS-based systems, with the illuminates column describing what each row is about.

to form a consistent, task relevant interpretation of this observed behaviour. By taking a deictic approach in HIVIS-WATCHER we don't name and describe every object, and we register only information about objects relevant to task. By doing this the information registered is then proportional to properties of interest and not the number of objects in the world.

4 HIVIS-WATCHER

In HIVIS-WATCHER we remove the reliance on the pipelined flow of data and instead use feedback to control the behaviour of the system. By making the perceptual processing and behavioural interpretation in the HIVIS-systems more tightly coupled we provide a more active control that can direct the processing performed by the system to those elements that are relevant to the current surveillance task. Deictic representation plays an important role in this framework because it supports attentional processing with emphasis placed on the behaviour of the perceiver as it interprets the activity of the scene objects rather than just representing the behaviour of the scene objects on their own.

4.1 Situated approach

Background details to the situated approach is given in Howarth [17] and its role in perceptual processing is further described in [4]. In HIVIS-WATCHER we have three separate elements: the "virtual-world" which holds data about the world, the "peripheral-system" which operators that access the world, the "central-system" which controls system behaviour. The peripheral-system is based on Ullman's [30] visual routine processor following the approach described by Agre and Chapman [5]. Horswill [15] describes a real-time implementation of such a visual routine processor. Both HIVIS-systems employ event detection operators however, the key difference is that in the HIVIS-WATCHER peripheral-system

operators are not run all the time, they are only run when selected by the task-level control system.

We have separated the operators in the peripheral-system into preattentive ones that are global, simple, and of low-cost and attentive ones which are applied to a single object and are more complex. The preattentive operators are used to guide application of attentive ones. Example preattentive operators include gross-change-in-motion which is described below, and mutual-proximity which is described in Howarth and Buxton [19]. The motivation behind the preattentive and attentional cues chosen here, was their potential usefulness on low-level data such as the identification of possible objects from clustering flow-vectors (see Gong and Buxton [14] where knowledge about a known ground plane is used to develop expectations of likely object motion). Once we have these coarse descriptions, and if they comply with the preattentive cue, then they would become candidates for further attentional processing such as model-matching (or some other form of object-recognition) to obtain aspects about the object. Basically, once we have found where something interesting is, we then try and work out what it is.

There are two types of marker in HIVIS-WATCHER. The agent type are all used by the same cluster of rules that identify events concerning changes in velocity, type-of-spatial-region-occupied, relative-position-of-other-local-objects, etc.. These rules represent the observer's understanding of typical-object-behaviour. The kernel type are each run by different sets of rules to fulfill some specific purpose, for example, the `*stationary-marker*` is allocated to any object that has recently stopped moving, by a perceiver-routine that is interested in objects that have just stopped.

4.2 An implementation of perceiver routines

To illustrate how perceiver routines work we will describe the routines associated with the official-observer looking for the presence of giveway behaviour. As mentioned above, the preattentive cue identifies any gross-change-in-motion (i.e., instances where an object changes state between `stationary` and `moving`. We can use this to initiate detection when, for example, a vehicle stops at a junction. Because this task of looking for giveway behaviour requires the identification of a number of distinct stages that involve different scene participants. The perceptual task of the official-observer involves three important entities: the first two correspond to the two roles in the giveway episode and are denoted by Stationary for *the-stationary-vehicle*, and Blocker for *the-vehicle-that-Stationary-is-giving-way-to*; and the third is denoted CA for *the-conflict-area* (a special region). When the two roles of Stationary and Blocker have been found, an area of mutual conflict, CA, can be identified (the space infront of Stationary and through which Blocker will pass). This area links Stationary to its cause. All that remains is to determine that Stationary is giving way to approaching traffic, and exhibits no other plausible behaviour (e.g., broken down, parked).

We separate the giveway episode into five routines that use region-based-prediction and perceiver level coordination. These routines are:

- Notice-stopping-object, which on completion generates **event-gw1**. The gross change in motion from moving to stationary allocates an agent and prompts the question "why is vehicle stationary?".
- Look-for-path-blocker, which on completion generates **event-gw2**. To be blocking it does not need to be physically in the way, it can also block by having "right-of-way" such that its path *will* block.
- Work-out-conflict-area, which on completion generates **event-gw3**. Having predicted the paths of Stationary and Blocker above, intersect them to find the mutually shared conflict area, CA.
- Watch-for-enter-conflict-area, which on completion generates **event-gw4**. In order to determine whether Stationary gives way, wait until Blocker has passed through CA.
- Notice-starts-to-move, which on completion generates **enter-gw5**. We then observe if Stationary moves. The gross change in motion from stationary to moving reallocates an agent to Stationary.

The five routines given above order and, as a continuous sequence, describe a temporal sequence of perceiver activity that identifies a giveway episode.

4.3 Results

Here we compare the effect of using two different tasks: "look for likely overtaking behaviour" and "look for likely giveway behaviour", to illustrate how HIVIS-WATCHER is able to solve the problem of identifying overtaking encountered in HIVIS-MONITOR. Also, we see how changing the observation task given to HIVIS-WATCHER noticeably alters the performance of the system (results are displayed in figure 2– figure 4).

Overtaking The purpose of this example is to show that HIVIS-WATCHER can pick out a pair of vehicles that are performing an overtaking episode. To do this we will use the policy "attend to likely overtaking and ignore likely following". The missing entries are because of the occlusion where no mutually proximate objects are visible to the observer. The vehicle shapes given in outline denote uninteresting peripheral objects, the number near each vehicle is its index reference (or buffer slot number), and the vehicle outlines that have marker shapes "attached" to them are selected objects that have been allocated an agent.

During frames 96 and 108 one of the vehicles occludes the other from the camera. The camera's field-of-view which affects the contents of the frame updates because we are dependent upon what is visible from the camera position not what is visible from the overhead view. By frame 132 overtaking is positively identified. A comparison between this policy and the similar one for "attend to likely following and ignore likely overtaking", together with more implementation details, is given in Howarth and Buxton [19].

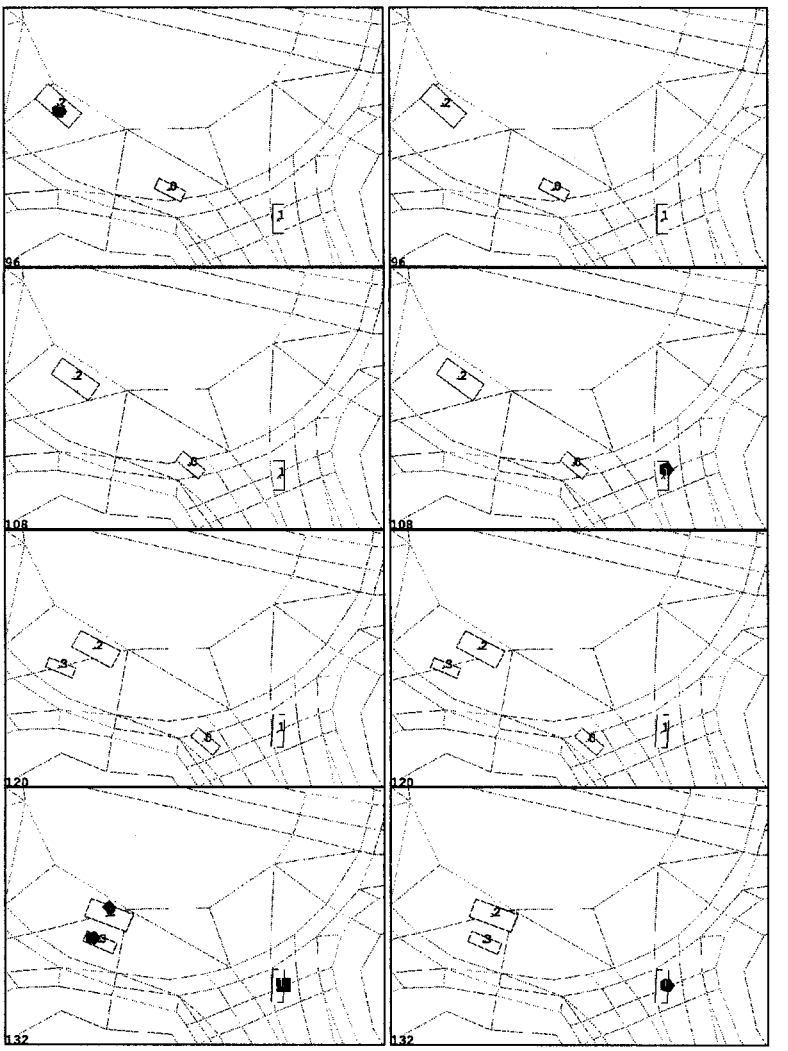

Fig. 2. Part 1 (overtaking policy left, giveway policy right).

Giveway To illustrate the need for local and global viewpoints we use the policy "look for likely giveway behaviour". HIVIS-WATCHER uses three attentional markers to perform the giveway detection routine, and the events correspond to the five routines described in section 4.2. The frames 108, 132–156 describe the allocation of `*agent2*` cued by gross-change-in-motion. At frame 120 the vehicle moved again, before the motion-prior was altered from `moving` to `stationary` by the agency operator `change-motion-prior!`. The value of motion-prior has changed by frame 168 because the object ceases to have an interesting motion property. Frame 192 shows the results from the region path predictions that

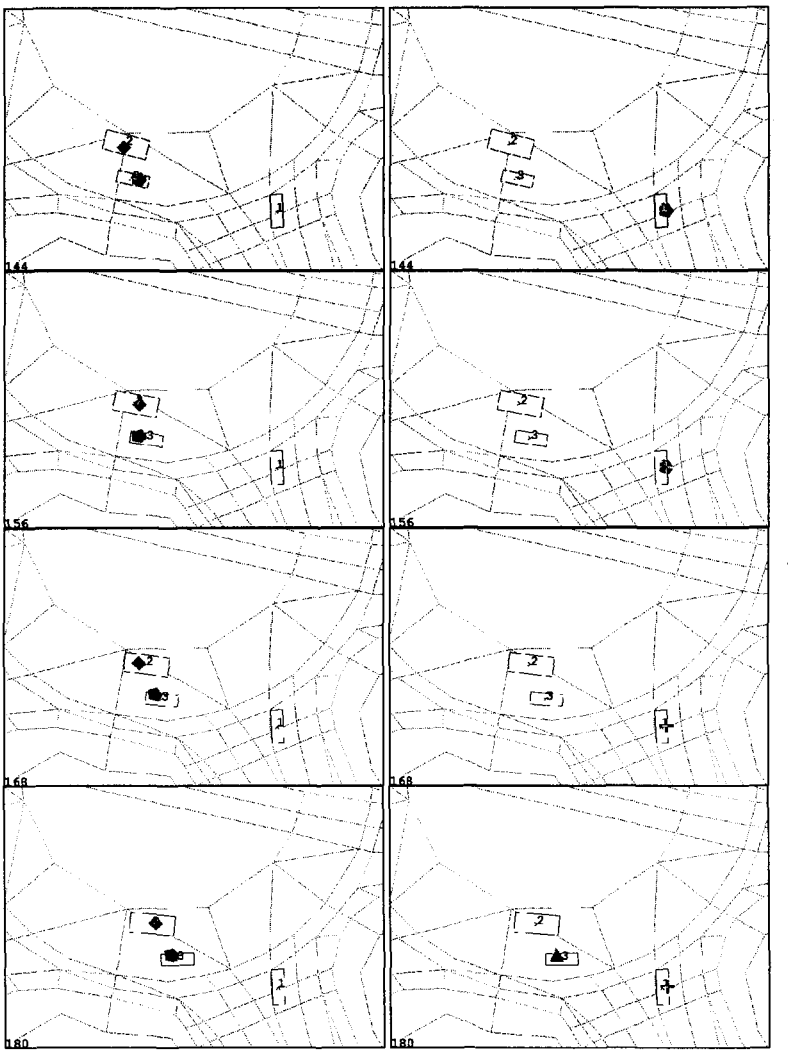

Fig. 3. Part 2 (overtaking policy left, giveway policy right).

generate the contents of the kernel activation planes. Frames 204–258 display the activation plane. Frame 228 shows the removal of `*head-marker*` following a successful intersection.

4.4 General features

The traditional separation made in cognitive science between input and central systems provides a description of the two tightly coupled components in HIVIS-WATCHER. The input system obtains object aspects, while the central system

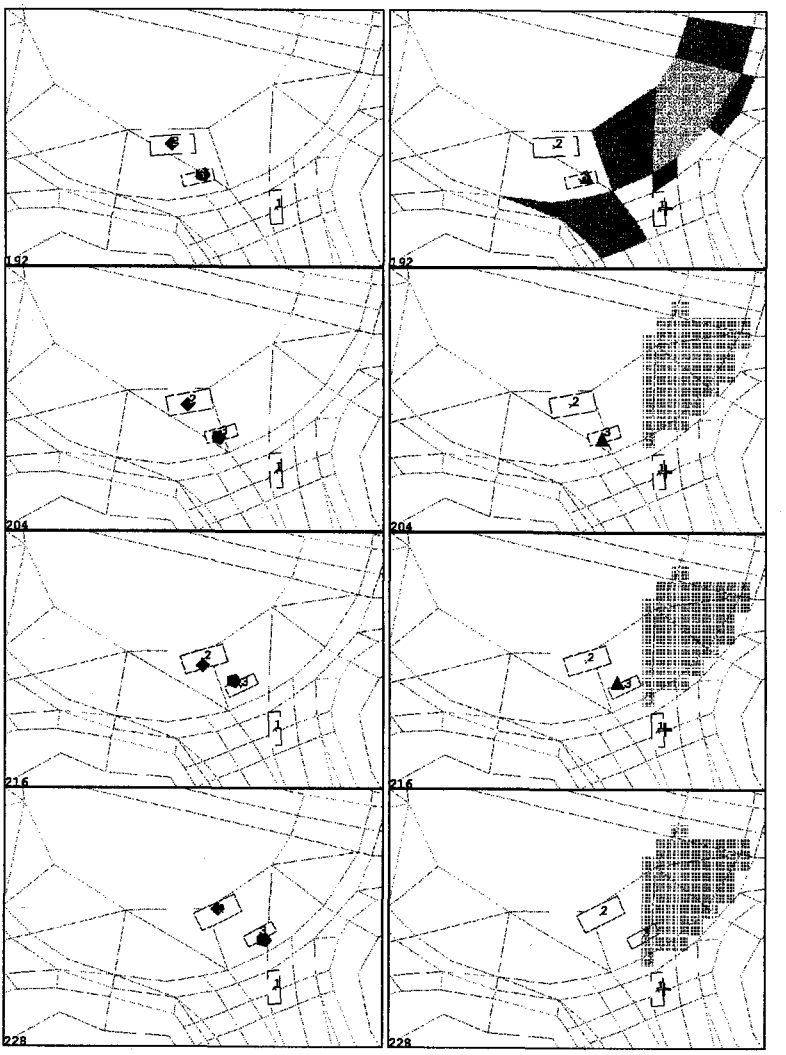

Fig. 4. Part 3 (overtaking policy left, giveway policy right).

controls which objects should be attended so their aspects can fulfill a given surveillance task. The separation of preattentive and attentive processing, and the use of a task directed central mechanism here provides what is needed for the official-observer to watch out for selected behaviours. HIVIS-WATCHER thus provides timely surveillance information about what is happening in the scene.

5 Conclusion

The main benefit of HIVIS-WATCHER over HIVIS-MONITOR is that its task orientedness reduces runtime representation, reasoning and complexity. In HIVIS-WATCHER: (1) the deictic representation has simplified the computational model of behaviour; (2) the situated approach has taken into account both the evolving context of the dynamic scene objects and also the task-oriented observer's context; (3) the use of selective attention provides a more viable form of real-time processing.

Other key points of this paper concern: (1) the distinction between script-based and more situated approaches; (2) the separation and integration of global and local reasoning in the context of a single official-observer, together with the illustration of how both play complementary roles in developing different levels of understanding; (3) the propagation of reasoning in the "here-and-now" through to the control mechanism in order to reflect the reactive quality of dynamic object behaviour.

Current work is addressing two important issues. The first concerns ways to control perceptual processing so that the task-level knowledge will influence when model-matching is performed. The second concerns learning the behavioural information, removing the hand coded element in choosing preattentive and attentive cues in HIVIS-WATCHER. Although this research has been illustrated by using data from a road-traffic surveillance, the intention is that the general framework should be applicable to other domains.

References

1. Philip E. Agre and David Chapman. Pengi: An implementation of a theory of activity. In *Sixth AAAI Conference*, pages 268–272. AAAI Press, 1987.
2. Dana H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
3. Hilary Buxton and others. VIEWS: Visual Inspection and Evaluation of Wide-area Scenes. *IJCAI-91 Videotape Program*, Morgan Kaufmann, 1991.
4. Hilary Buxton and Shaogang Gong. Visual Surveillance in a Dynamic and Uncertain World. *Artificial Intelligence*, 78:371–405, 1995.
5. David Chapman. *Vision, Instruction and Action.* The MIT Press, 1991.
6. Anthony N. Clark. Pattern recognition of noisy sequences of behavioural events using functional combinators. *The Computer Journal*, 37(5):385–398, 1994.
7. David R. Corrall, Anthony N. Clark, and A. Graham Hill. Airside ground movements surveillance. In *NATO AGARD Symposium on Machine Intelligence in Air Traffic Management*, pages 29:1–29:13, 1993.
8. David R. Corrall and A. Graham Hill. Visual surveillance. *GEC Review*, 8(1):15–27, 1992.
9. Li Du, Geoffery D. Sullivan, and Keith B. Baker. Quantitative analysis of the viewpoint consistency constraint in model-based vision. In *Fourth ICCV*, pages 632–639. IEEE Press, 1993.
10. Margaret M. Fleck. Representing space for practical reasoning. *Image and Vision Computing*, 6(2):75–86, 1986.

11. Margaret M. Fleck. *Boundaries and Topological Algorithms.* PhD thesis, MIT AI Lab., 1988. AI-TR 1065.
12. R.A. Frost. Constructing programs as executable attribute grammars. *The Computer Journal*, 35(4):376–387, 1992.
13. James J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin Company, 1979.
14. Shaogang G. Gong and Hilary Buxton. On the expectations of moving objects. In *Tenth ECAI Conference*, pages 781–784, 1992.
15. Ian Horswill. Visual routines and visual search: a real-time implementation and an automata-theoretic analysis. In *Fourteenth IJCAI Conference*, pages 56–62. Morgan Kaufmann, 1995.
16. Richard J. Howarth. *Spatial Representation, Reasoning and Control for a Surveillance System.* PhD thesis, QMW, University of London, 1994.
17. Richard J. Howarth. Interpreting a dynamic and uncertain world: high-level vision. *Artificial Intelligence Review*, 9(1):37–63, 1995.
18. Richard J. Howarth and Hilary Buxton. An analogical representation of space and time. *Image and Vision Computing*, 10(7):467–478, 1992.
19. Richard J. Howarth and Hilary Buxton. Selective attention in dynamic vision. In *Thirteenth IJCAI Conference*, pages 1579–1584. Morgan Kaufmann, 1993.
20. Edwin Hutchins. Understanding micronesian navigation. In G. Dedre and A.L. Stevens *Mental Models*, pages 191–225. Lawrence Erlbaum Associates, 1983.
21. Simon King, Sophie Motet, Jérôme Thoméré, and François Arlabosse. A visual surveillance system for incident detection. In *AAAI workshop on AI in Intelligent Vehicle Highway Systems*, pages 30–36. AAAI Press, 1994.
22. Stephen M. Kosslyn, Rex A. Flynn, Jonathan B. Amsterdam, and Gretchen Wang. Components of high-level vision: a cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition*, 34:203–277, 1990.
23. Hans-Hellmut Nagel. From image sequences towards conceptual descriptions. *Image and Vision Computing*, 6(2):59–74, May 1988.
24. Bernd Neumann. Natural language descriptions of time-varying scenes. In David L. Waltz, editor, *Semantic Structures: Advances in Natural Language Processing*, pages 167–206. Lawrence Erlbaum Associates, 1989.
25. Nils J. Nilsson. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research*, 1:139–158, 1994.
26. Raymond D. Rimey and Christopher M. Brown. Control of selective perception using Bayes nets and dicision theory. *International Journal of Computer Vision*, 12(2/3):173–207, April 1994.
27. Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, 1977.
28. Andrew F. Toal and Hilary Buxton. Spatio-temporal reasoning within a traffic surveillance system. In G. Sandini, editor, *Computer Vision -ECCV'92*, pages 884–892. Springer-Verlag, 1992.
29. John K. Tsotsos. Toward a computational model of attention. In T. Papathomas, C. Chubb, A. Gorea, and E. Kowler, editors, *Early Vision and Beyond*, pages 207–218. The MIT Press, 1995.
30. Shimon Ullman. Visual routines. In Steven Pinker, editor, *Visual Cognition*, pages 97–159. The MIT Press, 1985.
31. Anthony D. Worrall, Geoffery D. Sullivan, and Keith B. Baker. Advances in model-based traffic vision. In *British Machine Vision Conference 1993*, pages 559–568. BMVA Press, 1993.