Spatio-temporal Reasoning within a Traffic Surveillance System

A.F. Toal and H.Buxton

Queen Mary and Westfield College,
London, E1 4NS, England.
paddy@uk.ac.qmw.dcs
hilary@uk.ac.qmw.dcs

Abstract. The majority of potential vision applications such as robotic guidance and visual surveillance involve the real-time analysis and description of object behaviour from image sequences. In the VIEWS project we are developing advanced visual surveillance capabilities for situations where the scene structure, objects and much of the expected behaviour is known. This combines competences from image understanding, knowledge-based processing and real-time technology. In this paper we discuss the spatio-temporal reasoning which is of central importance to the system allowing behavioral feedback. In particular, we will elaborate the analysis of occlusion behaviour where we need knowledge of the camera geometry to invoke the occlusion region monitoring of vehicles plus knowledge of the scene geometry to maintain high-level models of possible trajectories for the occluded vehicles and to recognise the re-emerging vehicle(s).

1 Introduction and Background

A new emphasis in vision research is to produce conceptual descriptions of the behaviour of objects for dynamic scenes, for example Thibadeau 86, Nagel 87, Buxton and Walker 88, Mohnhaupt and Neumann 90, Howarth and Toal 90 in [Thibadeau '86, Nagel '88, Buxton and Walker '88], [Mohnhaupt and Neumann '90, Howarth and Toal '90]. To allow such understanding typically involves a knowledge based approach to computer vision where the knowledge specifies the models of the events and behaviours as well as the tasks that the system is to perform. The ESPRIT II VIEWS (Visual Inspection and Evaluation of Wide-area Scenes) project has the aim of demonstrating the feasibility of knowledge-based computer vision for real-time surveillance of well-structured, outdoor dynamic scenes.

The VIEWS project is supporting the development of both generic techniques and selected application systems which meet the specific requirements for airport stand surveillance, ground traffic control and road traffic incident detection. We have developed two major run-time components, the Perception Component (PC) and the Situation Assessment Component (SAC), to fulfill the two main competences required for traffic understanding: first, the ability to detect and recognise the vehicle types and trajectories in the PC; and, second, the ability to understand the situation as it develops over time in terms of vehicle behaviour and interactions in the SAC. In addition, we need a control component for real-time control of asynchronous data streams and an application support component to assist in the acquisition of scene, object and behavioral knowledge and to precompile appropriate visual strategies. The project addresses the integration of

* This research was funded as part of ESPRIT-2 P2152 'VIEWS' Project.

these components. Figure 1, illustrates a simplified functional architecture for these two main modules of our computer vision system.

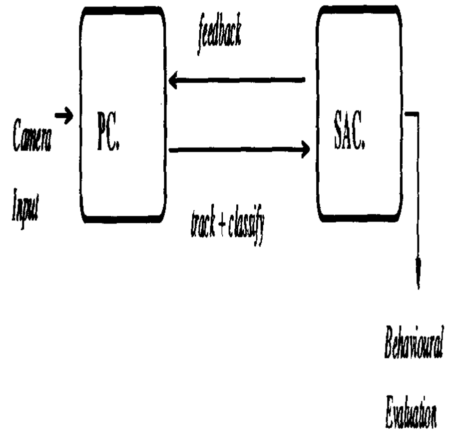

Fig. 1. *Functional Architecture*

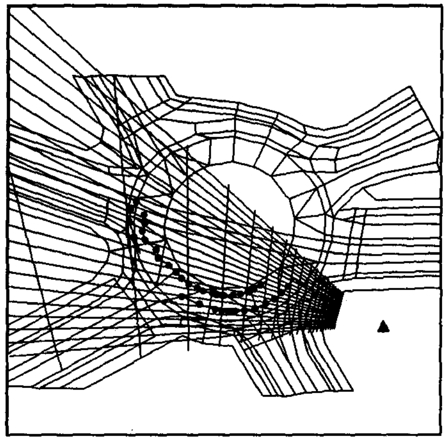

Fig. 2. *Bremer-Stern: Tracking updates*

The main exemplar running through this paper will be based around tracking of a sequence which includes the 3 frames appended to the paper. These 3 frames are from an image sequence taken at the Bremer Stern roundabout in Germany that we have used as an exemplar for the VIEWS project. Consider the vehicles where a) the lorry starts to occlude the saloon car, b) the car is fully occluded by the lorry, and c) the car starts to re-emerge from behind the lorry. The car then moves off. Processing at both the perceptual and conceptual levels needs spatio-temporal calculations. At the perceptual level, the detection, tracking and classification of the visible moving objects is performed.

Some of the updates for the model based tracking of the occlusion frames are displayed in figure 2. This figure displays vehicles as points projected onto the groundplane of the roundabout which is a complex environment segmented into road, bicycle and tram lanes (cutting across the middle of the roundabout). Figure 2 also shows the relationship of the camera field of view to the groundplane, this is important for occlusion reasoning, as will be discussed later.

The information extracted by the PC is given as updates at different levels of analysis and speed of processing (but always as estimates with respect to the ground plane) to the SAC. The basic form is as $< label, position, time >$ updates. The SAC then processes the information first, at the level of events such as stopping/starting and entering/exiting regions with an additional local check on the spatio-temporal continuity. Second, the global consistency checking constructs and maintains the space-time histories using behavioural constraints in space and time from the context of other vehicles in the locality. Finally, the SAC can also perform selected behavioral prediction and feedback which, for this full occlusion, must maintain the occlusion relationships for recognition of the re-emerging vehicle and give feedback both within the SAC and to the PC so that the relabelling of the vehicle and constructing a consistent history can be performed.

Any vision based tracking system must incorporate some means of overcoming occlusion problems. Even a simple task like counting the number of vehicles passing through the scene requires occlusion reasoning so that re-emerging vehicles are not counted twice.

In this paper, we will present results from a case study where the role of the SAC is primarily to act as a high level, long term memory keeping explicit representations of total occlusions occurring in dynamic scenes and utilizing behavioural knowledge to aid relabelling.

2 Analogical Representation

The analogical spatial representation underlying the behavioural models described later is a flexible, multi-purpose representation which maintains the structure of the world in a usable manner. The key requirements for the static knowledge in this representation are: (1) the conversion of the ground plane geometry into meaningful regions; and (2) the description of the connectivity of these regions. The objective of the surveillance is to reason about the vehicle in the scene, so we also require a means of representing dynamically: (3) spatial extent and edges of vehicle ground plane bounding boxes to determine occupied regions; and (4) velocity and orientation to derive behaviour. When we also include vehicle interactions, we require (5) inter-vehicle orientation and inter-vehicle distance.

The analogical representation provides a uniform framework in the SAC, incorporating the static scene knowledge and dynamic vehicle histories. The *static* knowledge is attached to regions. Regions, on the groundplane, are constructed from cell sets. For example, these can express: (1) the types of vehicle that use a region (eg roads, cycle-lane, tram-line); (2) regions of behavioural significance (eg giveway, turning); (3) direction information (eg lane leading onto roundabout); and (4) the basic connectivity of the leaf regions.

Cells are also used to represent *dynamic* vehicle histories. The analogical spatio-temporal representation is shown in figure 3. Calculation based on relating vehicle histories can fully capture the semantics of behaviours like 'following', 'overtaking', 'crossing', 'queueing' etc. An example of dynamic relation 'following' is given in figure 5 (overlap in spatial history within some time delay). Given the dynamic and static knowledge bases, we can usually determine what a vehicle is doing.

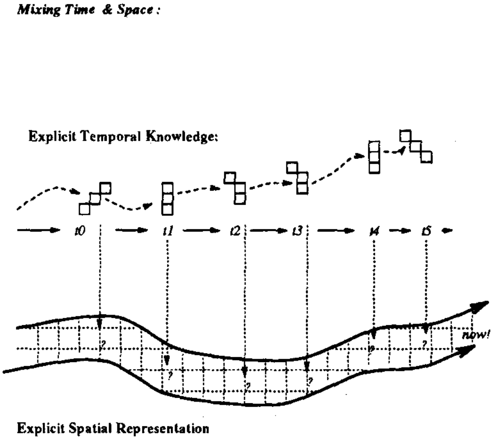

Fig. 3. *Spatio-Temporal Representation*

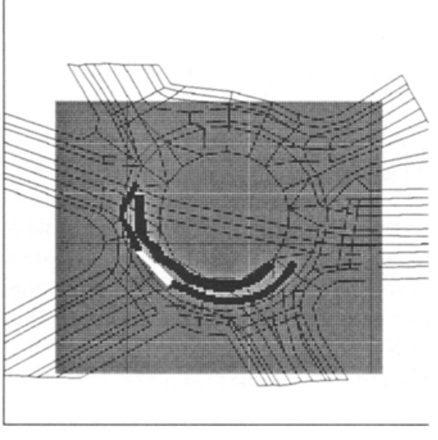

Fig. 4. *Completed S-T Histories*

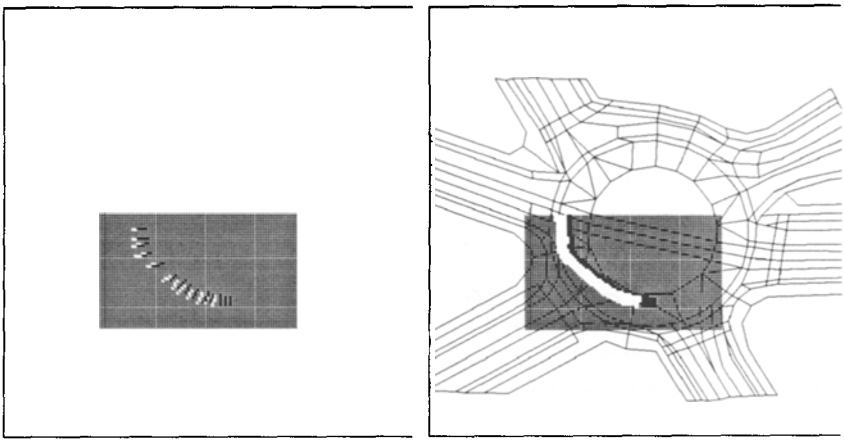

Fig. 5. *Spatial & Temporal Histories for 'follow'*

3 Occlusion

The Perception Component incorporates a range of vision tracking competencies. Its primary task is motion focused tracking and classification of vehicles. The specific tracking competencies are detailed elsewhere ([Sullivan et al. 90]). The VIEWS configuration we are considering is as follows:

- 'Gated' initialization of tracking: Static areas of the image where vehicles are known to appear are highlighted for track initialization.
- PC < 1second vehicle memory: The PC is dealing with video rate input, and once a vehicle track is lost attempts to relocate the vehicle are swiftly curtailed.
- Motion cued tracking: pixel grouped motion in the image is the primary cue for processing. Vehicles which halt and remain stationary for extended periods may be 'forgotten'. This problem may be handled as a ***virtual occlusion*** by the techniques presented here.

In the case of tracking being lost by total occlusion, some recovery mechanism is needed so any *new* tracks initialized by the PC are correctly labelled as continuations. In the short term, tracking algorithms attempt recovery by focusing processing, directed by calculations extrapolating the "lost" vehicles velocity. Occluded vehicles manoeuvre unpredictably, this makes simple velocity based calculation a poor approximation in the long term.

The SAC maintains *occluded-by* relationships. The boundary of a lead vehicle in the image defines the area of potential emergence. The SAC occlusion reasoning aids correct relabelling after emergence and indicates to the PC which vehicles to monitor for boundary deformations caused by emerging vehicles. Occlusions are so common these techniques are developed to be of low cost.

3.1 The example

Examine the example occlusion sequence, frames are shown at the end of the paper and tracks plotted in figure 2. A saloon travels behind a lorry, is occluded for some time and finally emerges. When the saloon emerges, it is labelled as a new track by the PC. The minimum competences for the SAC are to spot the occlusion occurring, maintain the occlusion relationships, relabel the saloon correctly upon emergence and complete its Spatio-temporal history.

While the reconstruction of the specific $< label, position, time >$ track is impossible, the form of the spatio-temporal history can be bounded in a useful manner (see figure 4).

3.2 Flagging Occlusion

A cheap robust method of candidate generation is needed, several are possible (eg. from the merging of coherent motion 'envelopes' generated by the primary tracker). Early experiments operated well and cheaply by casting potential *shadows* as shown in figure 6. Currently more accurate partial occlusion relations developed in the 3D model matcher are used. The model matcher is described in Marslin, Sullivan & Baker '91 [Marslin et al. '91] (see also: ECCV-92).

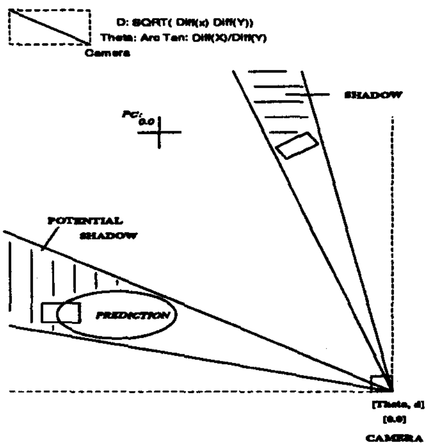

Fig. 6. *< theta, depth >*

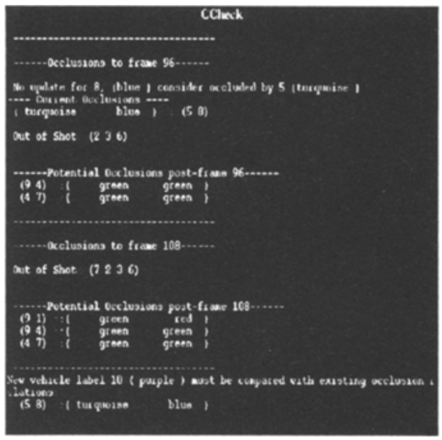

Fig. 7. *SAC Occlusion Output*

3.3 During Occlusion

In the example, partial occlusion of the saloon by the lorry (see first frame), develops into full occlusion (see frame-2). After the update for frame-96 the SAC generates the message shown in figure 7 and asserts the relationship *occluded-by(096, 5,* $< 8 >$*)*. The lorry is 'object5' the saloon 'object8'. The first field is either the time from which the occlusion is active, or a range indicating the period of time the occlusion is considered to have lasted. Once an *occluded-by* relationship is created the SAC must be able to reason about its maintenance and development.

During a typical occlusion development:

- Vehicles may emerge from occlusion.
- Vehicles may join the occlusion.
- The lead vehicle may itself become occluded.

- A vehicle may leave shot never having emerged from occlusion. Once occluding vehicle leaves shot, all occluded vehicles must do so.
- When not themselves becoming totally occluded, other vehicles may enter into partial occlusion relationships with the lead vehicle and then separate. In this case, it is possible that occluded vehicles move from one vehicle 'shadow' to another.

In the next stage, 'relabelling', it will be argued that behavioral knowledge is required to help disambiguate labelling problems as vehicles emerge. This is because a total occlusion may be of arbitrary length and over a long time almost any reordering of vehicles could occur. No processing is expended on reasoning about the development of the relative positions of vehicles *until* they emerge. Behavioral relationships established prior to the occlusion eg. *following*(car_a,car_b), are stored for consideration during relabelling. But the consideration of the development of relationships is not useful *during* the actual time of occlusion.

A grammar is used to manage the development of occlusions. Where $t_a : t_b$ is the time range from t_a to t_b and $v_j ::< list >$ represents the union of v_j and $< list >$. $< list > \ominus v_j$ represents $< list >$ with v_j removed. Where $\otimes$ is an exclusive-or such that *occluded-by*(t_j,v_m,car_a) $\wedge$ *occluded-by*(t_j,v_l,car_a) cannot be true. Unless otherwise stated t_i is before t_j and where used $t_i < t_k < t_j$.

Existence................. *Does occlusion exist?*
occluded-by(t_i,v_l,$\emptyset$) $\equiv$ no occlusion

Emergence... *What may emerge next?*
occluded-by(t_i,v_l,$< list >$) $\rightarrow$ next_emerge(t_j,v_l) $\in < list >$
If occluded-by(t_i,v_l,$< list >$) $\wedge$
emerge-from(t_j,v_l,v_{old})
$\rightarrow$ occluded-by(t_j,v_l,$< list > \ominus v_{old}$) $\wedge$
occluded-by($t_i : t_j$,v_l,v_{old})

Joining... *Has a vehicle joined existing occlusion?*
occluded-by(t_i,v_l,$< list >$) $\wedge$
occluded-by(t_j,v_l,v_{new}) $\rightarrow$
occluded-by(t_j,v_l,$v_{new} ::< list >$)

Lead Occluded... *Has an occluding vehicle been occluded?*
occluded-by(t_i,v_l,$< list >_l$) $\wedge$
occluded-by(t_k,v_{new},$< list >_{new}$) $\wedge$
occluded-by(t_j,v_{new},v_l)
$\rightarrow$ occluded-by(t_j,v_{new},$v_l ::< list >_{new} ::< list >_l$)
t_i &t_k are not ordered.t_j follows both

Leave Scene... *Anything occluded leaves show with lead!*
occluded-by(t_i,v_l,$< list >$) $\wedge$
left_shot(t_i,v_l) $\rightarrow$
left_shot(t_i,$< list >$)

Visible Interaction... *If shadows meet occluder may change!*
occluded-by(t_i,v_l,$< list >$) $\wedge$
(partially-occludes(t_k,v_l, v_m) $\vee$
partially-occludes(t_k,v_m, v_l))
$\rightarrow$ occluded-by(t_j,v_l,$< list >_l$)
$\otimes$ occluded-by(t_j,v_m,$< list >_m$)
where $< list >_l ::< list >_m \equiv < list >$

The SAC maintains the consistency of the options with constraint reasoning. Updates from the PC are compared to these possibilities. In our example, the initial relationship generates expectations:

$$\textit{occluded-by}(096, 5, <8>) \rightarrow \text{next\_emerge}(t_j,5) \in <8>$$

Later the creation of a new track for 'vehicle-10', matches the suggested emergence of the occluded saloon (see figure 7). This matches the following rule:

$$\text{occluded-by}(096,5,<8>) \wedge \text{emerge-from}(108,5,8)$$
$$\rightarrow \text{occluded-by}(108,5,\emptyset) \wedge \text{occluded-by}(096:108,5,<8>)$$

In the example the relabelling is unique ($10 \equiv 8$) and the no additional occlusion relationships exist ($\rightarrow \emptyset$). The relabelling and the fact the lorry no longer occludes anything are passed to the PC. This relabelling case is the simplest possible. More complex cases and behavioural evaluation, will be considered in the following section.

3.4 Emergence

The remaining occlusion reasoning in the SAC can be summarised as two capabilities:

(1) Relabel While the actual moment of emergence is not possible to predict, this does not mean no useful inferences can be made. When a *new* track is initialized, indicated by a new vehicle label in the VIEWS system, it is compared with currently maintained occlusion relationships. If it matches, the new label is replaced by the matching previous label for both the PC and SAC.

(2) History Completion As complete histories are preferable for behavioral evaluation, the SAC needs access to the last and the newest position so the complete history is 'extruded' as shown in figure 4 to fill in the trajectory of the occluded vehicle.

Two more advanced considerations:

1. If a vehicle is known to be occluding another, currently only the identities are passed to the PC. With both enough geometric knowledge (eg. the ground to the right of the lorry is grass not road) and behavioral knowledge (eg. that a vehicle is 'turning off'), some consideration of *where* on the lead vehicles boundary a vehicles may emerge is possible. Currently only preliminary results of such a study are available and they will not be presented here.
2. Complex cases of history completion occur. Eg. where a vehicle turns a corner and a simple direct link of updates crosses pavement or grass. Analogical representations are well suited to this form of path completion problem. Cells store localised knowledge about what *type* of space they represent (eg. roadway or grass). Deforming a suggested completion to conform to expectation is possible with techniques reported elsewhere (eg. [Steels '88, Shu and Buxton '90]).

4 More complex relabelling

Consider the case where two vehicles (car_a and car_b) are occluded by the same lorry. When the PC indicates a new vehicle track the following cases, generated by the grammar given earlier, are all initially plausible, since the 'emergence' may be caused by the vehicles motion relative to the camera or each other:

occluded-by(lorry, $<car_a, car_b>$*)* $\rightarrow$ *visible(lorry)* $\wedge$

1. *next_emerge(lorry,*car_a*)* $\wedge$ *occluded-by(lorry,* $<car_b>$*)*
2. *next_emerge(lorry,*car_a*)* $\wedge$ *occluded-by(*car_a*,* $<car_b>$*)*

3. $next\_emerge(lorry, car_b) \wedge occluded\text{-}by(lorry, < car_a >)$
4. $next\_emerge(lorry, car_b) \wedge occluded\text{-}by(car_b, < car_a >)$

The temporal fields are omitted here for simplicity.

Although all cases must be considered, the SAC should provide some ordering on the options. This is where behavioral knowledge may be utilized. Eg. where a $following(car_a, car_b)$ relationship was established prior to the occlusions, the default for the SAC is to maintain behavioral relationships unless given specific contradictory evidence, and interpret the evolving scene accordingly.

The job of the SAC is *twofold*:

1. To produce consistent histories and alert the PC to any inconsistency in data provided.
2. To produce ongoing behavioural evaluations; both for the end user and to supply a source of knowledge for defaults in the production of consistent histories.

In the introduction an example of 'following' behaviour was demonstrated, which could be deduced by comparing the analogical spatial and temporal histories (ie. the car 'following' must overlap the trail of the lead car in the same temporal sequence). Such definitions have proved a very natural expression of behavioural concepts. Others which have been investigated include: turning (off one road-instance onto another), crossing (of two vehicles), giveway (at a junction), potential intersection (for risk evaluation) and overtaking.

We have demonstrated techniques which allow the production of consistent spatio-temporal histories under conditions of total long term occlusion. The final completed histories in figure 4 are only a limited example, but even they can allow for the final deduction that the saloon car, having been occluded by the lorry, finally *overtakes* it. The techniques used to produce these final behavioural comparisons will be reported in [Toal and Buxton '92].

5 Conclusion

In summary we have discussed some of the competences developed for specific application of road traffic surveillance. The main competence elaborated here is to is to deal with total occlusions, in order to develop consistent spatio-temporal histories for use in behavioral evaluation eg. following, queue formation, crossing and overtaking. Specifically we have proposed a grammar for the handling of occlusion relationships that allows us to infer correct, consistent labels and histories for the vehicles. This was illustrated in with a simple example, although the information made available is sufficient to deduce that the saloon car, having been occluded by the lorry *overtakes it*. In addition we also described, the analogical representation that supports the contextual indexing and behavioural evaluation in the scene.

Work is continuing on more complex behavioural descriptions and their interaction with more complex occlusions as well as generating more constrained expectations of vehicles on emergence. In particular, it is important to look at the trade-offs between behavioural and occlusion reasoning, decision speed and the accuracy of predicted re-emergence.

6 Acknowledgements

We would especially like to thank Simon King & Jerome Thomere at Framentec for much discussion of ideas about the SAC and behavioural reasoning and Geoff Sullivan and Anthony Worrall at Reading University for discussion of the interaction with the PC and we would like to thank Rabin Ezra for access to his boundless knowledge about postscript. In addition, we would thank all the VIEWS team for their work in putting together this project and the ESPRIT II programme for funding.

References

[Buxton and Walker '88] Hilary Buxton and Nick Walker, *Query based visual analysis: spatio-temporal reasoning in computer vision, pages 247-254 Image and Vision Computing 6(4), November 1988.*

[Fleck '88a] Margaret M. Fleck, *Representing space for practical reasoning, pages 75-86, Image and Vision Computing, volume 6, number 2, May 1988.*

[Howarth and Toal '90] Andrew F. Toal, Richard Howarth, *Qualitative Space and Time for Vision, Qualitative Vision Workshop, AAAI-90, Boston*

[Mohnhaupt and Neumann '90] Michael Mohnhaupt and Bernd Neumann, *Understanding object motion: recognition, learning and spatiotemporal reasoning, FBI-HH-B-145/90, University of Hamburg, March 1990*

[Nagel '88] H-H. Nagel, *From image sequences towards conceptual descriptions, pages 59-74, Image and vision computing, volume 6, number 2, May 1988.*

[Steels '88] Luc Steels, *Step towards common sense, VUB AI lab. memo 88-3, Brussels, 1988.*

[Thibadeau '86] Robert Thibadeau, *Artificial perception of actions, pages 117-149, Cognative science, volume 10, 1986.*

[Sullivan et al. 90] Technical Report *D102* "Knowledge Based Image Processing", G. Sullivan, Z. Hussain, R. Godden, R. Marslin and A. Worrall. Esprit-II P2152 'VIEWS', 1990.

[Marslin et al. '91] R. Marslin, G.D. Sullivan, K. Baker, *Kalman Filters in Constrained Model Based Tracking, pp371-374, BMVC-91*

[Shu and Buxton '90] C. Shu, H. Buxton *"A parallel path planning algorithm for mobile robots", proceedings: International Conference and Automation, Robotics and Computer Vision, Singapore 1990*

[Toal and Buxton '92] A.F. Toal, H. Buxton *"Behavioural Evaluation for Traffic Surveillance using Analogical Reasoning and Prediction"* in preparation.

This article was processed using the LaTeX macro package with ECCV92 style