Qualitative Reasoning Feeding Back into Quantitative Model-Based Tracking

Christian Köhler, [1] **Artur Ottlik,** [2] **Hans–Hellmut Nagel,** [2] **and Bernhard Nebel** [1]

Abstract. Tracking vehicles in image sequences of innercity road traffic scenes must be considered still to constitute a challenging task. Even if a-priori knowledge about the 3D shape of vehicles, of the background structure, and about vehicle motion is provided, (partial) occlusion and dense vehicle queues easily can cause initialization and tracking failures. A stepwise improvement of the tracking approach requires numerous and time-consuming experiments. These difficulties can be eased considerably by endowing the system with – at least part of the – qualitative knowledge which a human observer activates in order to judge the results. In the case to be reported here, a system for *qualitative reasoning* has been coupled with a *quantitative* model-based *tracking* system in order to explore the feedback from qualitative reasoning into the geometric tracking subsystem. The approach and encouraging experimental results obtained for real-world image sequences are described.

1 INTRODUCTION

Although 'making a computer see' once was considered as part of Artificial Intelligence (see, e.g., the foreword by O. Faugeras in [12]), this task developed into a discipline of its own as documented by numerous course books (e. g., [6], [12], [7]). While Computer Vision approaches evaluate raw digitized imagery predominantly numerically based on quantitative geometric methods, AI approaches tend to emphasize symbolic aspects (see, e. g., [10] or [18]). Numeric and symbolic schemes and associated representations each have their own advantages and disadvantages concerning the task to be solved.

A combination of these approaches encountered interest intermittently during past decades without, however, lasting methodological results so far. A fundamental question addresses the problem how to convert uncertainties related to measurement noise into appropriate uncertainties associated with symbolic representations. The latter have to accommodate, too, the implications of conceptual vagueness. This problem aggravates once the emphasis shifts from the evaluation of single image frames to that of entire image sequences.

Substantial increases in computing power at reasonable costs enabled researchers in the computer vision community to gradually stabilize basic signal processing and pattern recognition processes like the reliable extraction of some fairly general features even from image sequences. These developments facilitated renewed attention to the potential associated with a combination of quantitative geometric and qualitative symbolic processing of information captured by images and image sequences ([4], [15]). In this context, probabilistic methods like Hidden Markov Models, Bayesian Belief Networks, or Neural Networks (see, e.g., [2], [3]) constitute a 'natural' option for many researchers in the computer vision community, whereas declarative knowledge representation schemes and their exploitation tend to be more attractive for the AI community.

Inspired by *chronicles* (used to represent knowledge on time, events and actions in cognitive systems, see [9]), the authors of [20] used *scenarios* for the interpretation of videos. In another recent example, the generation of a textual description from a video of an innercity traffic scene relies on a *Fuzzy Metric Temporal Horn Logic* [8]. These examples use symbolic concepts in a bottom up processing fashion to build conceptual primitives in order to derive higher level concepts from these primitives. As discussed in [15], an overall system needs to cover various kinds of knowledge.

The system to be reported in the remainder of this contribution differs from others by coupling a model-based tracking system bottom-up *and* top-down to a symbolic component. Both parts run as separate processes which communicate with each other. A knowledge base server evaluates rules to check consistency of the tracking data and generates "feedback" to be transmitted back to the tracker.

2 SYSTEM OVERVIEW

Figure 1 shows a schematic overview of the main system components. The following subsections give a more detailed explanation of each block in this schematic sketch.

2.1 The Model-Based Tracker

The system `Xtrack` (see e.g. [11, 16]) tracks road vehicles in monocular grayvalue image sequences. Geometric knowledge about the observed scene such as the position of the ground plane, static objects, and models of vehicles are incorporated into the system.

At each half–frame, the system tries to detect new vehicles based on a segmentation of the Optical Flow (OF) field (see, e.g., [14]). In case an OF-segment is compatible with the appearance of a new vehicle in the field of view, a new *object candidate* is initialized. A hatchback model is assigned to each object candidate because the determination of the appropriate vehicle model does not yet work reliably enough. `Xtrack` estimates a state consisting of the vehicle position on the ground plane, the vehicle orientation, its speed, and steering angle for the new object candidate.

Tracking takes place in a prediction–update–cycle realized by a Kalman-Filter. The update step estimates a new state based on Edge Elements (EEs) and OF-vectors in the image region surrounding an object candidate. EEs mainly influence the estimation of position and

[1] Institut für Grundlagen der Künstlichen Intelligenz, Albert–Ludwigs–Universtät Freiburg, Georges–Köhler–Allee, Geb. 52, D-79110 Freiburg, Germany. email: {ckoehler|nebel}@informatik.uni-freiburg.de

[2] Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik, Universität Karlsruhe (TH), Postfach 6980, D-76128 Karlsruhe, Germany. email: {ottlik|nagel}@iaks.uni-karlsruhe.de

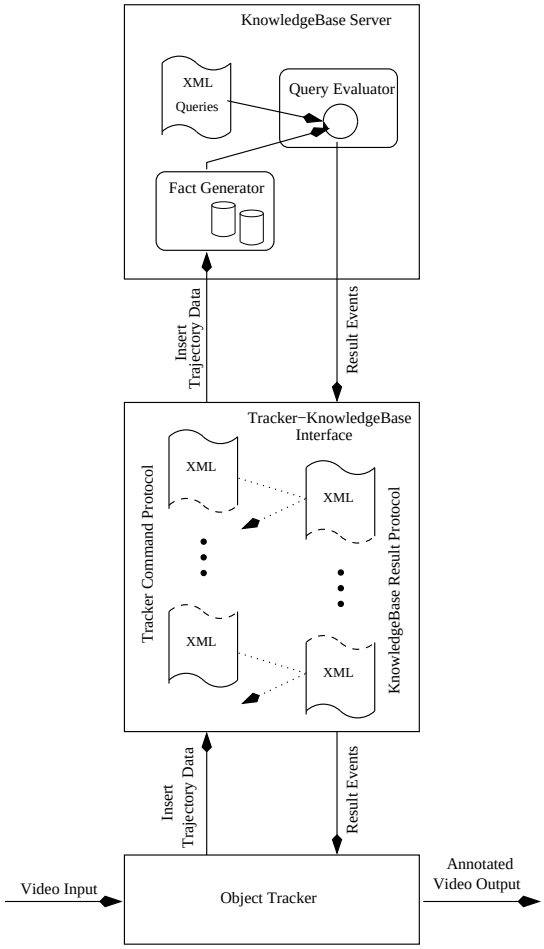


Figure 1. Collaboration of the main components

orientation, whereas OF-vectors tend to affect the orientation, speed, and steering angle estimates more strongly.

2.2 The Interface between Tracker and Knowledge Base Server

The interface between the system components is based on an abstraction from the trajectory data provided by `Xtrack`. All communication between the two components shown in Figure 1 is marked up in a synchronous XML Protocol. For each (half-) frame processed, a list of tracked objects is inserted into the knowledge base. Events resulting from the evaluation of qualitative queries are sent back to the tracker through the interface.

2.3 The Qualitative Knowledge Base Server

The tracker provides a tuple (x, y, v, θ, i, t) for each object i in each frame t. (x, y) are the coordinates of its centroid on the ground plane with respect to an arbitrary but fixed world coordinate system. v is its measured velocity and θ an angle with respect to an arbitrary but fixed reference direction. These tuples are then processed in order to generate a qualitative description of the configuration in each frame.

Figure 2 indicates the qualitative spatial relations which subdivide the plane surrounding an object.

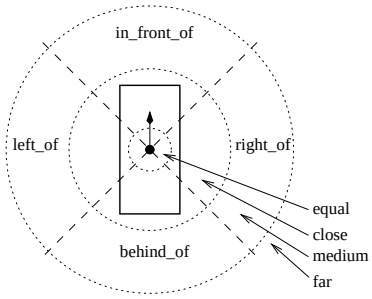


Figure 2. The model for spatial relations

The parameters for the generation of qualitative facts are shown in Figure 3.

Rectangular Shape	
length	$3.8m$
width	$1.5m$

Distance	
equal	$[0m, 1m[$
close	$[1m, 5m[$
medium	$[5m, 15m[$
far	$[15m, \infty m[$

Velocity	
still	$[0m/s, 1m/s[$
slow	$[1m/s, 3m/s[$
fast	$[3m/s, \infty m/s[$

Orientation	
in_front_of	$]-45°, +45°[$
behind_of	$]+135°, +180°] \cup]-180°, -135°[$
right_of	$[-135°, -45°]$
left_of	$[+45°, +135°]$

Figure 3. Parameters for generating the qualitative facts

The fact generator processes the tracker data to build qualitative relations for each pair of objects in the current frame. Subsequently, a set of rules is evaluated based on the qualitative facts derived from the tracker data of the current frame.

2.3.1 The Query Language

In the literature on spatial knowledge representation, one finds a wide variety of possible qualitative knowledge representation schemes. Although they all address different aspects of space, such as topology, direction, size etc., they use all a common formal framework. All schemes support binary (and possibly unary) relations that are usually JEPD, i.e., jointly exhaustive and pairwise disjoint. From a formal point of view, these relation systems are usually the atomic relations of sub-structures of Tarskian [19] relation algebras [13, 5].

In our query language, we use predicates for qualitative distance, qualitative intrinsic orientation (see Figure 2) as well as topological descriptions restricted to `overlap(X,Y)` and `disjoint(X,Y)`. We do not need a richer vocabulary for topological relations because

these are all the possible relationships that can meaningfully hold between two objects in our domain. In the general case, we might also consider making all the distinction that are present in the RCC-8 calculus [17].

Queries in our languages are what has been termed conjunctive queries in database theory. In other words, it is a conjunction of logical atoms. Some of the variables that appear in the query can be existentially quantified effectively projecting this variable away. Evaluating such a query over the knowledge base of qualitative descriptions generated from the tracker data results in tuples of objects.

For example, the query

$$\exists X\colon \texttt{close}(A, X) \wedge \texttt{right\_of}(A, X) \wedge \texttt{slow}(A),$$

returns all objects A that are *right of* and *close* to some other object X and at the same time move *slowly*.

In addition to purely spatial queries, our system can also evaluate spatio-temporal queries, where the temporal dimension is described using Allen's [1] interval calculus. However, in the application described here, it is enough to consider spatial relations inside each frame together with a description of the object velocity.

One interesting observation is that satisfiability reasoning that is usually thought to be at the heart of qualitative spatial reasoning does not play a prominent role here. It is still important for meta reasoning, e.g., for deciding whether a query is satisfiable at all or when deciding query containment, but does not play a role for query evaluation on the object level.

3 EXPERIMENTS AND RESULTS

Experiments have been carried out on the `stau02` image sequence, which can be downloaded from http://i21www.ira.uka.de/image_sequences. It consists of 2050 half-frames in which 28 vehicles are visible.

The first experiment (identifier: `V0`) has been performed without any qualitative feedback. In the second experiment (`V1`) simple results of qualitative reasoning are exploited in the tracking loop. All object candidates that are overlapped on the ground plane by another object candidates, i.e., all object candidates returned by

$$\exists X\colon \texttt{overlap}(A, X),$$

are removed from the tracking loop. In a further experiment (`V2`) this feedback is modified. In case of an overlap only a standing object candidate, i.e., one for which $\texttt{still}(A)$ is true, will be removed from the tracking process.

In Figure 4 results are shown in case the tracking of a vehicle fails and a succeeding object candidate is driving 'through' the failed one. Tracking of the preceding object candidate failed because the corresponding vehicle drove slowly into an occlusion. In such a situation, OF estimation is difficult which led to an imprecise estimation of the vehicle speed and as a consequence to a tracking failure. The succeeding object candidate does not fit to the vehicle precisely because the model had been initialized before the vehicle had completely entered the field of view. Based on qualitative feedback (`V1` and `V2`), this situation can be detected and the false object candidate is removed (center and bottom row). Since in `V1` both the incorrectly and the correctly tracked colliding object candidates are removed, a short interval occurs within which the tracking of the succeeding vehicle is interrupted (half-frames #247–#250). In half-frame #251, a new object candidate is initialized for this vehicle. This interruption of tracking can be avoided in experiment `V2` by just removing the standing object candidate from the set of two object candidates which virtually collide. In this experiment, therefore, a complete trajectory for the succeeding vehicle can be computed whereas two *independent* trajectory parts have been created in the previous experiment. The latter experiment thus preserves the identity of the second vehicle even while it traverses the location where the first vehicle was lost.

Figure 5 illustrates a problem for the tracker with unmodeled occlusion by a tree at the image boundary. Since no OF-vectors can be estimated after a vehicle begins to become occluded by the tree, the corresponding object candidate's speed is reduced and thus the object candidate comes to a halt at the position of the tree. These circumstances lead to an accumulation of lost object candidates at the left image boundary and in the top left image corner in experiment `V0` where vehicles are occluded by trees. Using qualitative feedback, the accumulation is avoided as it is shown in the center and bottom row of Figure 5. In experiment `V1`, *both* object candidates are removed at half-frame #1571 after the second object candidate collided with the first (incorrectly tracked) object candidate. The tracking failure of the last object candidate cannot be detected, therefore, in half-frame #1635. Using the feedback as set up in experiment `V2` has the effect that an already failed standing object candidate is removed, but not the approaching one. This latter one fails several half-frames later due to the occlusion situation.

Due to the fixed choice of a hatchback model which is used for any new object candidate, large moving vehicles in the image sequence lead to initialization of several object candidates. Since these object candidates do not properly fit the vehicle, tracking often fails. With qualitative feedback these incorrectly tracked object candidates are removed. As a consequence, less incorrectly tracked object candidates remain which can be seen in Figure 6. An unsuspected consequence of this result is the fact that `Xtrack` runs significantly faster in combination with the qualitative reasoning module because it no longer has to take the 'corpses' into account.

4 CONCLUSION

The approach presented above shows promising results – despite the fact that the qualitative terms exploited for reasoning are still remarkably simple. So how can this actually happen? The knowledge represented in qualitative terms (basically a 'true' or 'false' for the intersection test of two oriented rectangles) reveals an inconsistency between object hypotheses established on the basis of quantitative geometric results provided by the tracker.

The assumption is essential that no two vehicles can be at the same time at the same place. It is the only 'clue' from which we can abduce that the tracker lost at least one of the 'virtually colliding' vehicles. Unfortunately, there is no obvious algorithm [15] to decide on the basis of just this observation which one is the lost vehicle; possibly both are lost during tracking. Automatic reinitialization, therefore, improves tracking results. Furthermore, the assumption that vehicles lost during tracking no longer move leads to a better explanation of the overall developments in the scene recorded by the video to be evaluated. The second experiment provides evidence that the proposed assumption holds in most cases. This result can be understood to imply that the qualitative reasoning rules described above contain a 'grain of truth'.

ACKNOWLEDGEMENTS

We gratefully acknowledge partial support of this work by the European Union (Project CogViSys, IST–2000-29404).

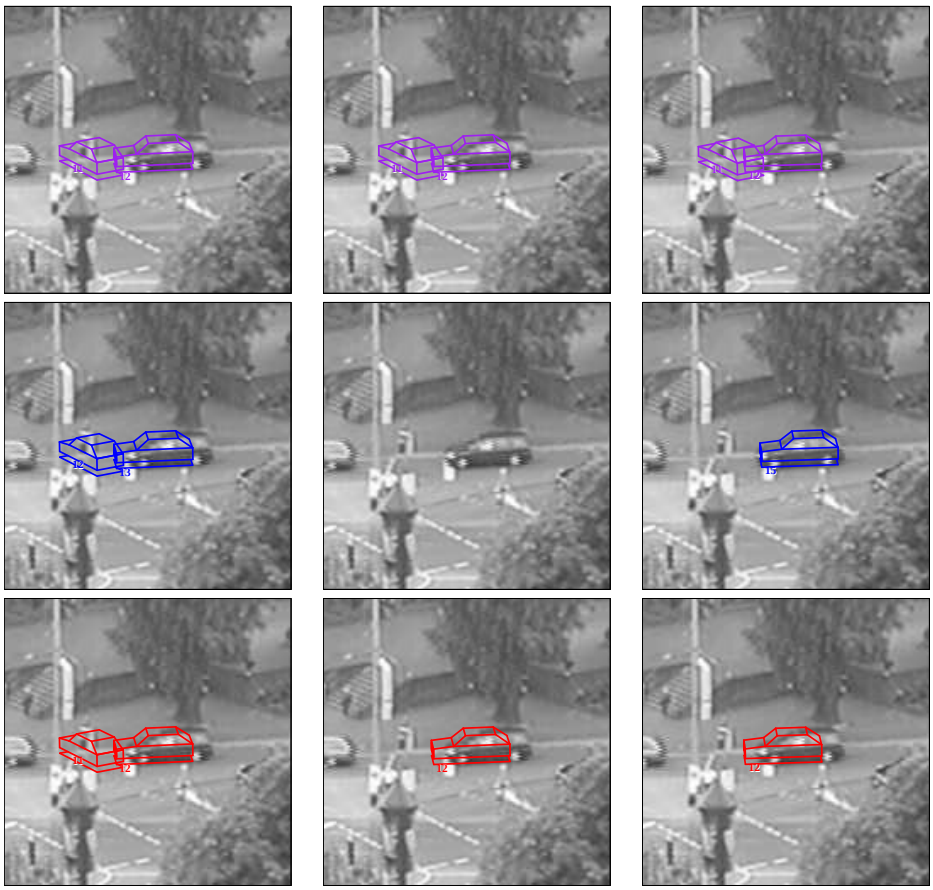


Figure 4. System behavior in case of a collision between a failed and a correctly tracked object candidate. Top row: experiment `V0` (without any feedback). Center row: experiment `V1`. Bottom row: experiment `V2`. Using qualitative feedback, the failed object candidate is removed. In `V1` the correctly tracked object candidate is also removed and reinitialized four half-frames later, whereas in experiment `V2` the second vehicle is tracked without any interruption.

REFERENCES

[1] J. F. Allen, 'Maintaining Knowledge about Temporal Intervals', *Communications of the ACM* **26**:11 (1983) 832–843.

[2] H. Buxton and S. Gong, 'Advanced Visual Surveillance using Bayesian Networks', in Proc. *Workshop on Context–Based Vision (In Conjunction with Fifth International Conference on Computer Vision (ICCV95))*, 19 June 1995, Cambridge/MA; J. L. Mundy and T. Strat, (Eds.), IEEE Computer Society Press, Los Alamitos/CA (1995).

[3] A. J. Howell and H. Buxton, 'Active vision techniques for visually mediated interaction', in Proc. *International Conference on Pattern Recognition (ICPR'02)*, 11-15 August 2002, Quebec City, Canada; C. S. R. Kasturi and D. Laurendeau (Eds.), IEEE Computer Society Press, Los Alamitos/CA (2002), pp. 296 - 299.

[4] A. G. Cohn, D. R. Magee, A. Galata, D. C. Hogg, and S. M. Hazarika, 'Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction', in *Spatial Cognition III*; C. Freksa, W. Brauer, C. Habel, and K.F. Wender (Eds.), LNCS 2685, Springer-Verlag, Berlin, Heidelberg, New York (2003), pp. 232-248.

[5] Ivo Düntsch, 'A tutorial on relation algebras and their application in spatial reasoning'. Tutorial given at *Conference on Spatial Information Theory (Cosit'99)*, 25–29 August 1999, Stade, Germany; http://www.cosc.brocku.ca/~duentsch/archive/relspat.pdf.

[6] O. Faugeras: *Three-Dimensional Computer Vision - A Geometric Approach*. The MIT Press, Cambridge/MA and London/UK 1993.

[7] O. Faugeras and Q.-T. Luong: *The Geometry of Multiple Images*. The MIT Press, Cambridge/MA and London/UK 2001.

[8] H.-H. Nagel, R. Gerber, and H. Schreiber, 'Deriving textual descriptions of road traffic queues from video sequences', in Proc. *15th European Conference on Artificial Intelligence (ECAI-2002)*, 21–26 July 2002, Lyon, France; F. van Harmelen (Ed.), IOS Press, Amsterdam (2002), pp. 736–740.

[9] M. Ghallab, 'On chronicles: Representation, on-line recognition and learning', in Proc. *Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*, 4–8 November 1996, Cambridge/MA; L. C. . Aiello, J. Doyle, and S. C. Shapiro (Eds.), Morgan-Kauffman, San Mateo/CA (1996), pp. 597–606.

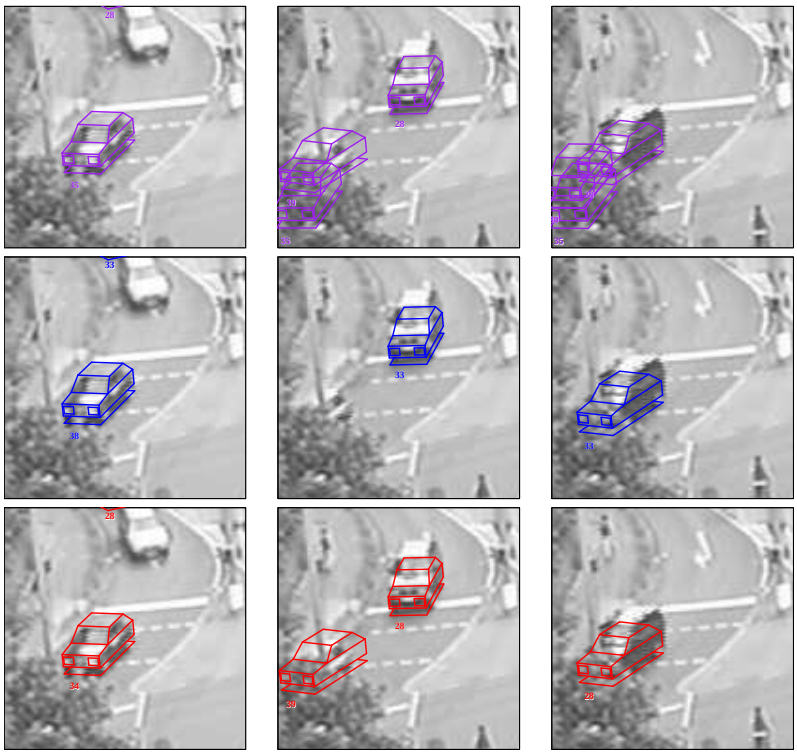
[10] G. Görz (Hrsg.), 'Einführung in die Künstliche Intelligenz', Addison-Wesley, Bonn, Reading/MA 1993.

[11] M. Haag and H.-H. Nagel: *Combination of Edge Element and Optical Flow Estimates for 3D–Model–Based–Vehicle Tracking in Traffic Image Sequences*. International Journal of Computer Vision **35**:3 (1999) 295–319.

[12] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge/UK 2000.

[13] Peter B. Ladkin and Roger D. Maddux, 'On binary constraint problems', *Journal of the ACM*, **41**:3 (1994) 435–469.

[14] M. Middendorf and H.-H. Nagel: *Estimation and Interpretation of Discontinuities in Optical Flow Fields*. In: Proc. 8th International Conference on Computer Vision (ICCV 2001), 9-12 July 2001, Vancouver/BC, Canada, Vol. I, IEEE Computer Society: Los Alamitos/CA, USA 2001, pp. 178-183.



1400 # 1571 # 1635

Figure 5. Accumulation of object candidates at the left image boundary. Top row: experiment `V0`. Center row: experiment `V1`. Bottom row: experiment `V2`.

Figure 6. Tracking results at the final frame of the image sequence `stau02`. Left: experiment `V0`. Right: experiment `V1`. Using qualitative feedback in `V1` removes many object candidates which remain in the scene after tracking failed in experiment `V0`.

[15] H.-H. Nagel, 'Reflections on cognitive vision systems', in Proc. *Third International Conference on Computer Vision Systems (ICVS 2003)*, 1–3 April 2003, Graz, Austria; J. L. Crowley, J. H. Piater, M. Vincze, and L. Paletta (Eds.), LNCS 2626, Springer-Verlag, Heidelberg, Berlin, New York (2003), pp. 34–43.

[16] A. Ottlik and H.-H. Nagel: *On Consistent Discrimination Between Directed and Diffuse Outdoor Illumination.* In: Proceedings of the 25th DAGM-Symposium (DAGM'03), 10-12 September 2003, Magdeburg, Germany; B. Michaelis and G. Krell (Ed.), LNCS 2781, Springer-Verlag, Berlin·Heidelberg·New York, 2003, pp. 418-425.

[17] D. A. Randell, Z. Cui, and A. Cohn, 'A spatial logic based on regions and connection', in Proc. *Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, 25–29 October 1992, Cambridge/MA; B. Nebel, C. Rich, and W. Swartout (Eds.), Morgan Kaufmann, San Mateo/CA (1992), pp. 165–176.

[18] S. Russell and P. Norvig: *Artificial Intelligence – A Modern Approach.*

Prentice-Hall, Inc.: Upper Saddle River, NJ 1995.
[19] A. Tarski, 'On the calculus of relations', *Journal of Symbolic Logic* **6**:3 (1941) 73–89.
[20] V.-T. Vu, F. Brémond, and M. Thonnat: 'Automatic video interpretation: A novel algorithm for temporal scenario recognition', in Proc. *18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, 9–15 August 2003, Acapulco, Mexico; G. Gottlob and T. Walsh (Eds.), Morgan Kaufmann, San Mateo/CA (2003), pp. 1295-1302.