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 still constitutes a challenging task. Even if a-priori knowledge about the 3D shape of vehicles, of background structure and vehicle motion is provided, (partial) occlusion and dense vehicle queues easily can cause initialization and tracking failures. Improving the tracking approach requires numerous and time-consuming experiments. Yet, these difficulties can be eased considerably by endowing the system with a part of the qualitative knowledge, that a human observer uses in order to judge the results. In the case 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.

1 INTRODUCTION

Substantial increases in computing power and gradually stabilized subsymbolic processes have renewed interest in the potential combination of quantitative geometric and qualitative symbolic processing of information captured by images and image sequences (see, e.g., [4]). Recent systems (see, e.g., [6], [2]) have used symbolic representations in a bottom up processing fashion. Starting from conceptual primitives higher level concepts are derived from these primitives. The system discussed in this work differs from the quoted examples by coupling bottom-up *and* top-down a model-based tracking system to a symbolic component. Both parts run as separate processes which communicate with each other.

A more detailed version of this paper with additional references has been made available (see [7]).

2 SYSTEM OVERVIEW

Figure 1 shows a schematic overview of the main system components. The system `Xtrack` (see, e.g., [3]) tracks road vehicles in monocular grayvalue image sequences. Geometric knowledge of 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. Whenever an OF-segment is compatible with the appearance of a new vehicle in the field of view, a new *object candidate* is initialized. `Xtrack` estimates a state consisting of the vehicle position on the ground plane, the vehicle orientation, its speed, and the steering angle for every object candidate. State estimation is implemented through 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 orientation, whereas OF-vectors influence the orientation, speed, and steering angle estimates more strongly. For each (half-) frame processed, the tracker sends *quantitative* state information of all object candidates to the knowledge base server, which generates a *qualitative* description of the configuration. Figure 2 indicates qualitative spatial relations which subdivide the plane surrounding an object. 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. Resulting events from the evaluation are sent back through the interface to `Xtrack`.

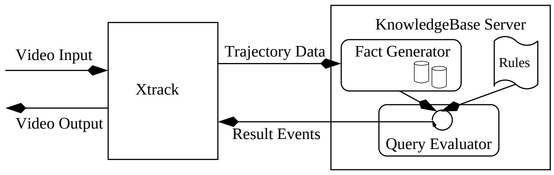


Figure 1. Collaboration of the main components

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 general we might also make all the distinctions, that are present in the RCC-8 calculus [5]. Queries in our language are 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

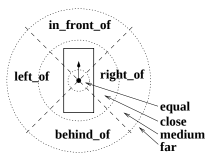
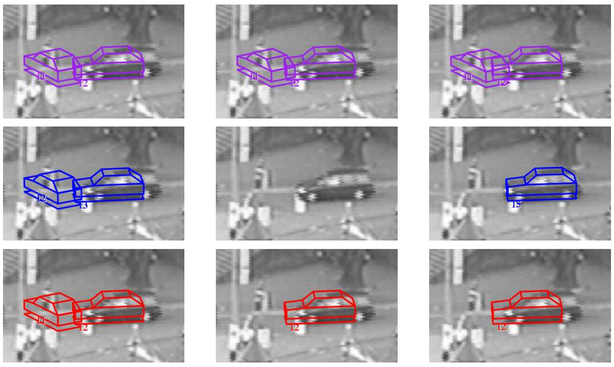


Figure 2. The model for spatial relations

---

[1] Institut für Grundlagen der Künstlichen Intelligenz, Albert–Ludwigs–Universität 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



246 # 247 # 251

Figure 3. 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`.

tracker data results in tuples of objects. 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 the spatial relations inside each frame together with a description of the object velocity.

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 candidate (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`) only a standing object candidate (i.e.: one for which $\texttt{still}(A)$ is true) will be removed in case of an overlap.

Figure 3 shows results where the tracking of a vehicle failed due to an occlusion situation, and a succeeding object candidate is driving 'through' the failed one. Based on qualitative feedback (`V1` and `V2`), this situation can be detected, and the false object candidate is removed. In `V1` both the incorrectly and the correctly tracked colliding object candidates are removed, which leads to a short interruption of tracking for the correctly tracked vehicle and a new initialization several frames later. This interruption of tracking can be avoided in experiment `V2` by removing only the *standing* object candidate. A *complete* trajectory for the succeeding vehicle can thus be computed in contrast to the previous experiment.

As shown in Figure 4, by using qualitative feedback many tracking failures can be removed, which often result from multiple initializations for large vehicles due to the fixed choice of a hatchback model for every new object candidate.

4 CONCLUSION

The qualitative feedback approach presented above for the detection and removal of *inconsistencies* between object hypotheses shows promising results. Unfortunately, there is no obvious algorithm to decide solely on this observation [4], which one is the lost vehicle; possibly both are lost during tracking. The results are improved when the tracker automatically reinitializes.

Figure 4. Tracking results at the final frame of the image sequence `stau02`. Top: experiment `V0`. Bottom: experiment `V1`. Using qualitative feedback in `V1` removes many object candidates which remain in the scene after tracking failed in experiment `V0`.

ACKNOWLEDGEMENTS

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

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] Ch. Köhler, A. Ottlik, H.-H. Nagel, and B. Nebel, 'Qualitative Reasoning Feeding Back into Quantitative Model-Based Tracking', Technical Report No. 204, Fakultät für Informatik, Albert–Ludwigs–Universität, Freiburg im Breisgau, Germany (2004) and http://cogvisys.iaks.uni-karlsruhe.de