Analysing the Localisation of Road Vehicles for Tracking

C. Rother[1] and H.-H. Nagel[2]
[1] Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computer Science
Royal Institute of Technology (KTH), S–10044 Stockholm, Sweden
[2] Institut für Algorithmen und Kognitive Systeme
Fakultät für Informatik der Universität Karlsruhe (TH)
Am Fasanengarten 5, D–76128 Karlsruhe, Germany
carstenr@nada.kth.se, hhn@iitb.fhg.de

Abstract

An approach for the model-based tracking of vehicles in road traffic scenes has been examined in detail in an extensive analysis. As a consequence, more effort was spend on the improvement of the initialisation phase. By applying knowledge, which is utilized in the tracking phase anyway, the initial estimation of the vehicle's position could be determined more *robust*. For this, we exploit the shape of an vehicle as well as its shadow, which is cast on the street plane. A quantitative analysis verifies the achieved improvements.

Key words

Localisation of objects, Initialisation for tracking, tracking of vehicles, model-based approach.

1 Introduction

By the use of a stationary video camera installed at a road crossing, complex traffic scenes can be recorded and analysed. Apart from individual cases, this still needs an *interactive* analysis of image sequences, which causes a considerable working effort. Therefore, this has rarely been done. By the algorithmic analysis of those digitized, monocular image sequences, traffic situations could be analysed quicker, automatically and with less costs. This could lead towards a more flexible - possibly even an individual - regulation of the traffic on a road crossing.

If the events in road traffic scenes and the illumination conditions are predictable (e.g. on a highway) or somehow verifiable (e.g. on a parking site), explicit 3D knowledge about the scene is not necessary. These approaches can be implemented relatively quickly and have rather little executing time (see for instance [2], [3] or [7]). If either the events in road traffic scenes or the illumination conditions are changing or there are occlusions in the scene, the use of explicit 3D models is advantageous. These models can be models of the fore- and background as well as models of the vehicles and their motion (see for instance [1]).

In a first step, here denoted as *initialisation phase*, the images of every individual moving vehicle have to be detected, localised and, if necessary, classified. Such a initialisation can be based on a process, which searches for characteristic configurations of edge segments in the image. Typically, these configurations are the boundaries of surfaces of the vehicle, such as the bonnet or the roof (see [8]). An alternative approach is the estimation and segmentation of fields of optical flow vectors (see for an overview [6], as well as [5, 9]). Such an alternative will be pursued here, since initial values can be estimated for the position and orientation of the vehicle and furthermore for the magnitude and direction of its velocity. These parameters describe the *state* of an individual vehicle in the 3D scene and are used as an input for the model-based *tracking phase*. We concentrate here on the task of the *initial localisation* of vehicles in the 3D scene.

2 Initial localisation of vehicle-models on the street plane

By the segmentation of an optical flow field, we obtain clusters of optical flow (of-)vectors, which are in the following denoted as 'objectmasks', whereat one 'objectmask' represents the image of one moving vehicle and its moving shadow. The transition from an 'objectmask' to the initial estimation of the state of a vehicle-model (see [5]) is based on a – very simple – assumption: the of-vectors in the image are the result of a flat, rectangular and textured plate, which is translationally moving parallel to the street plane. The rectangular plate should enclose the shape of the vertical projection of a vehicle onto the street plane. The height of the plate is assigned to the half of the average height of private vehicles. This average height of private vehicles is a *system parameter* and a-priori set to $1.4m$. The center of this rectangular plate can be considered as the reference point

of the vehicle-model on the street plane. The initial position and orientation of the vehicle-model and the magnitude of its velocity are determined by averaging the endpoints, orientations and lengths of the back-projected of-vectors on the rectangular plate[1]. For the back-projection of of-vectors, a calibrated camera is needed.

In order to examine this approach, an extensive experiment was carried out, whereat the main focus was on the quantitive analysis of the tracking phase. For the model-based tracking, the type of the vehicle was given in form of a 3D polyhedrical model. In order to separate the problems corresponding to the initialisation phase and the tracking phase respectively, the system parameter 'height' was *interactively* adjusted, whenever the initialisation values were 'obviously' poorly estimated. Similar to results of experiments described in [4], it turned out that the problem of poor initialisation values was the most frequent one among all recognized problems. A more detailed analysis of the initialisation values showed that the estimated position was the most inaccurate one among all state components of a vehicle-model. An interactive adjustment had to be carried out for 48 (circa 12%) out of 400 vehicles. Since the height of the plate directly influences the initial estimation of the *position* of an vehicle-model, a more sophisticated approach for estimating the initial position is required.

3 Improvement of the initial position estimation

This 'plate-model' neglects both the true *shape* of a vehicle and its shadow on the street plane. Since knowledge about the *type* of the vehicle and an illumination model of the scene are utilized in the tracking phase anyway, this knowledge can already be applied in the initialisation phase. How can we achieve this ?

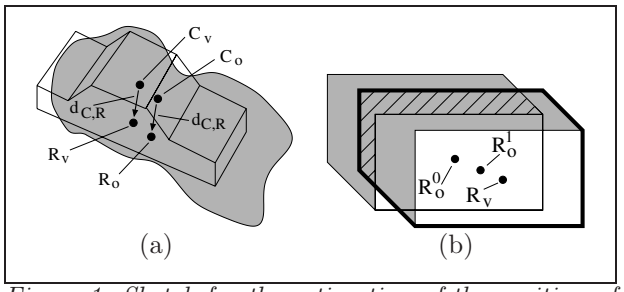

Figure 1: Sketch for the estimation of the position of a vehicle with (a) and without (b) the consideration of its shadow.

Firstly, we consider the estimation of the position of a vehicle, which casts *no shadow* on the street plane. In figure 1(a), C_v denotes the center of gravity of all visible surfaces of the model, which are projected in the image. The projection of the vehicle's reference point in the 3D scene into the image is denoted as R_v. On the basis of a coarse estimation of the vehicle's position with the old approach, which was described in chapter 2, a displacement vector $d_{C,R}$ can be determined. When we add this displacement vector $d_{C,R}$ to the center of gravity C_o of the 'objectmask' (drawn in grey colour), we obtain R_o, which should be a better approximation of the reference point of the vehicle in the image. It can be seen in figure 1(a) that the quality of the new reference point R_o depends on how good the 'objectmask' covers the image of the vehicle-model. Finally, a better reference point of the vehicle in the 3D scene can be determined by back-projecting R_v onto the street plane.

Since the shadow of a driving vehicle moves corresponding to the vehicle, we obtain of-vectors in its shadow area as well, which are in magnitude and direction comparable to the ones of the vehicle. Therefore, these of-vectors are included in the 'objectmask' as well, which can cause an inaccurate estimation of the position with the old approach. Taking the shadow into account, we suggest an iterative process, in which the of-vectors, which are both in the 'objectmask' and the shadow of the vehicle, are removed from the 'objectmask'. Figure 1(b) shows the image of a cuboid seen from the 'bird's eye view' with the reference point R_v. The bold polygon displays the 'objectmask' of the cuboid, which ideally comprises of the image of the cuboid and the shadow (drawn in grey colour). With this 'objectmask' and the method described above we obtain a first estimation of the reference point R_o^0. The image of the cuboid with the reference point R_o^0 and its shadow is depicted as well. The hatched area displays this part of the 'objectmask', which includes the shadow of the cuboid. Therefore, the of-vectors of this part of the 'objectmask' can be removed. With the resulting 'objectmask', we can predict a new reference point R_o^1, which is closer to the true reference point R_v. An iteration of this process ideally removes the whole shadow area of the 'objectmask' and consequently improves the initial estimation of the Position R_o of the vehicle.

An alternative method of removing the shadow would be the incorporation of the shadow as a surface, which is part of the vehicle-model. Thus, with the above method, which considers no shadow, and a new displacement vector $d_{C,R}$, we would directly obtain a new reference point R_o. The reason, why we prefer an iterative process is that this process gives better results when the shadow, because of various reasons, is not visible or detectable in the image.

4 Experimental results

Firstly, we consider the 48 vehicles, for which the system parameter 'height' has to be adjusted interactively in order to obtain a satisfying estimation of their position. A more detailed analysis showed that for 20 vehicles the difference between the 'true' height

[1]It is assumed that the initial orientation of the vehicle-model is the same as the direction of its velocity.

of the vehicle and the a-priori height of $1.4m$ was more than $0.2m$. Figure 2 (left) demonstrates an example, for which the initial position was poorly estimated ("−") with the a-priori height of $1.4m$. The interac-

Figure 2: The vehicle no. 26 and its vehicle-model, which was initialised in half-frame 2140 of the image sequence "stau03". The result with the old method and without interactive adjustment is shown in the left image. The middle image depicts the result with the old method and with interactive adjustment. The right image shows the result with the new automatic method.

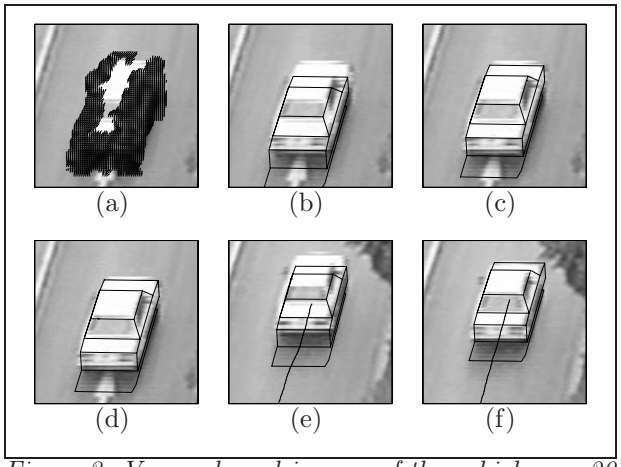

Figure 3: Very enlarged images of the vehicle no. 20 of the image sequence "stau01". Image (a) shows the image of the vehicle with the superimposed 'objectmask' (of-vectors are drawn in black). It is visible that the 'objectmask' covers not only the image of the vehicle, but also part of its shadow. The initial position estimation with the old method in the half-frame 1100 (image (b)) is average ("0"). A better ("+") position of the vehicle (image (c)) can only be obtained by the decrease of the system parameter 'height' from 1.4m to 1.0m, which is in this case incorrect. The new method automatically calculates a good result (image (d)). Furthermore it can be seen that the improved position estimation with the new approach effects the results of the tracking phase. The result of the tracking phase after 80 half-frames started with the initialisation corresponding to image (b) and image (d) respectively, which are shown in image (e) and image (f) respectively.

tive increase of the parameter 'height' from $1.4m$ to the 'true' height of $1.8m$ results in a good ("+") initial position estimation. The new method achieves automatically a comparably good result. The shadow, which was cast by a vehicle on the street plane, could also be identified as a reason for an inaccurate localisation of the vehicle-model (see figure 3).

Table 1 summarizes the initialisation results, which have been determined by the old and new approach

Table 1: Quantitative assessment of the initialisation and tracking results for the position estimation on the basis of the old approach and the new approach respectively. The symbol '+' means good, '0' average and '−' poor or failed initialisation results and tracking results respectively.

	48 vehicles with interactive adjustment			152 vehicles without interactive adjustment		
Assessment	+	0	−	+	0	−
Init. (old met.)	25	17	6	77	53	22
Init. (new met.)	32	9	7	86	42	24
Track. (old met.)	27	5	16	118	15	19
Track. (new met.)	28	7	13	117	20	15

respectively. Note, the results with the old approach were assessed after the interactive adjustment. The table additionally includes the initialisation results of 152 vehicles, for which no interaction with the old approach was necessary. By the comparison of the first two lines of the table, we can state that the new, automatic approach is slightly superior to the old approach. The new method terminated in the average after 5 iterations.

On the basis of these initialisation values, the *tracking* was started for all 200 vehicles. The third and fourth line of table 1 show that the tracking results for the 200 vehicles are almost the same. In figure 4, an example is presented, in which the initialisation results of the two different approaches were unequal, but the quality of the tracking results after 48 half-frames were nearly identical.

5 Discussion and future work

In order to solve a complex task, such as the tracking of vehicles in road traffic scenes, simplifying assumptions are inevitable for a first design of the system. After accomplishing this sub-goal, these assumptions have to be questioned.

A quantitive analysis verified that the disregard of the 'true' height of a vehicle and its shadow could lead towards an inaccurate estimation of the vehicle's position. Since knowledge about illumination conditions and the type of a vehicle are used in the tracking phase anyway, this knowledge was integrated in a new method, which is part of the initialisation phase. By the comparison of the old and new approach it turned out that the new method provides equivalent tracking results without the necessity of interactive adjustments.

Apart from the transition from an interactive to a more automatic initialisation phase, the question has to be asked, if an expensive initialisation phase is necessary, since the following, first actualisation step of

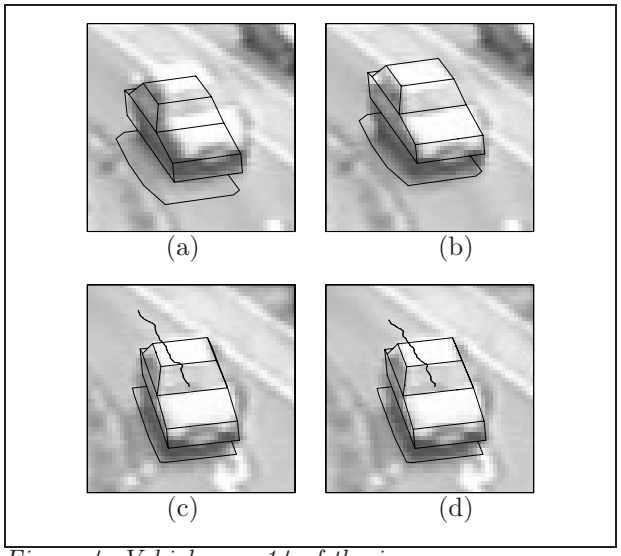

Figure 4: Vehicle no. 14 of the image sequence "stau01", whereat the initial position was calculated with the old approach (a) and new approach (b) respectively. The result of the racking phase after 48 half-frames is depicted in the image (c) (started with the initialisation of image (a)) and in the image (d) (started with the initialisation of image (b)).

an iterative extended Kalman filter (IEKF) takes into account these effects anyway. With this, the actualisation step of an IEKF is based on the assumption that the position component of the state of a vehicle-model has a *normal distributed* error around its estimated value. However, the new approach of the initialisation phase attacks *systematic* errors, which depend on the geometry of the examined objects, i.e. the vehicles, as well as illumination effects, i.e. the cast of a shadow. If the difference between the estimated and the 'true' position is within the 'scope' of the IEKF, this error can be compensated (see figure 4). On the other hand, if the estimated position is out of the 'scope' of the IEKF (see figure 3), the tracking will fail immediately or in the progress of the tracking phase. In the latter case, the identification of the true reason for the failure, can be very difficult.

Furthermore, the new method for estimating the vehicle's position can be easily extended regarding a new estimation of the vehicle's orientation and velocity respectively. There are two reasons, why we didn't integrate this so far. Firstly, an extensive analysis showed that the estimation of the vehicle's position is the most inaccurate one among all state components of a vehicle-model. Secondly, in order to verify the improvement of a complete system, only very small changes should be carried out.

Acknowledgment

We thank M. Haag for his careful reading of a draft as well as for his help with the experiments. Part of these investigations have been financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

References

[1] E. D. Dickmanns: *Expectation-based, Multi-focal, Saccadic Vision for Perceiving Dynamic Scenes (EMS-Vision)*. In 'Dynamische Perzeption', S. Posch und H. Ritter (Hrsg.), Proceedings in Artificial Intelligence Vol. 8, infix-Verlag Sankt Augustin 1998, pp. 47–54.

[2] S. Gil, R. Milanese, and T. Pun: *Comparing Features for Target Tracking in Traffic Scenes*. Pattern Recognition **29**:8 (1996) 1285–1296.

[3] W.E.L. Grimson, C. Stauffer, R. Romano, and L. Lee: *Using Adaptive Tracking to Classify and Monitor Activities in a Site*. Proc. CVPR'98, June 1998, Santa Barbara, CA, pp. 22–29.

[4] M. Haag and H.-H. Nagel: *Beginning a Transition from a Local to a More Global Point of View in Model-Based Vehicle Tracking*. In Proc. 5th European Conference on Computer Vision (ECCV'98), 2–6 June 1998, Freiburg/Germany; H. Burkhardt and B. Neumann (Eds.), Lecture Notes in Computer Science LNCS 1406 (Vol. I), Springer-Verlag Berlin, Heidelberg, New York 1998, pp. 812–827.

[5] D. Koller, K. Daniilidis, and H.-H. Nagel: *Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes*. International Journal of Computer Vision **10**:3 (1993) 257–281.

[6] A. Mitiche and P. Bouthemy: *Computation and Analysis of Image Motion: A Synopsis of Current Problems and Methods*. International Journal of Computer Vision **19**:1 (1996) 29–55.

[7] V. Rehrmann: *Object Oriented Motion Estimation in Color Image Sequences*. In Proc. 5th European Conference on Computer Vision (ECCV'98), 2–6 June 1998, Freiburg/Germany; H. Burkhardt and B. Neumann (Eds.), Lecture Notes in Computer Science LNCS 1406 (Vol. I), Springer-Verlag Berlin, Heidelberg, New York 1998, pp. 704–719

[8] T.N. Tan, G.D. Sullivan, and K.D. Baker: *Model-Based Localisation and Recognition of Road Vehicles*. International Journal of Computer Vision **27**:1 (1998) 5–25.

[9] P.H.S. Torr and A. Zisserman: *Concerning Bayesian Motion Segmentation, Model Averaging, Matching, and the Trifocal Tensor*. In Proc. 5th European Conference on Computer Vision (ECCV'98), 2–6 June 1998, Freiburg/Germany; H. Burkhardt and B. Neumann (Eds.), Lecture Notes in Computer Science LNCS 1406 (Vol. I), Springer-Verlag Berlin, Heidelberg, New York 1998, pp. 511–527.