Matching Object Models to Segments from an Optical Flow Field

Henner Kollnig[1] and Hans-Hellmut Nagel[1,2]

[1]Institut für Algorithmen und Kognitive Systeme
Fakultät für Informatik der Universität Karlsruhe (TH)
Postfach 6980, D-76128 Karlsruhe, Germany

[2]Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB),
Fraunhoferstr. 1, D-76131 Karlsruhe, Germany
Telephone +49 (721) 6091-210 (Fax -413), E-Mail hhn@iitb.fhg.de

Abstract. The temporal changes of gray value structures recorded in an image sequence contain significantly more information about the recorded scene than the gray value structures of a single image. By incorporating optical flow estimates into the measurement function, our 3D pose estimation process exploits *interframe* information from an image sequence in addition to *intraframe* aspects used in previously investigated approaches. This increases the robustness of our vehicle tracking system and facilitates the correct tracking of vehicles even if their images are located in low contrast image areas. Moreover, partially occluded vehicles can be tracked *without* modeling the occlusion explicitly. The influence of interframe and intraframe image sequence data on pose estimation and vehicle tracking is discussed systematically based on various experiments with real outdoor scenes.

1 Introduction

Many computer vision approaches to pose refinement match model features only to stationary data features, for example edge elements or edge segments which are extracted from a *single* image frame. Applying such an approach to evaluate an image *sequence*, the temporal aspects of image sequence data appear to be insufficiently exploited.

In order to avoid matching moving objects to stationary image features that exhibit coincidentally the same gray value structure as the object image under scrutiny we do no longer match polyhedral vehicle models only to stationary image gradients which is described in more detail in [Kollnig & Nagel 95]. Rather than restricting the update step only to these *intraframe* data, we extend the update step to evaluate *optical flow* vectors as *interframe* image contributions. Optical flow estimates the apparent shift of gray value structures. In this contribution, the estimated optical flow is matched to the *motion field* (also denoted as *displacement rate*), i.e. the image plane velocity of projected scene points.

The domain of discourse of our investigations is illustrated by an image sequence recording gas station traffic (see Figure 1). The tracking of vehicle `C` is a

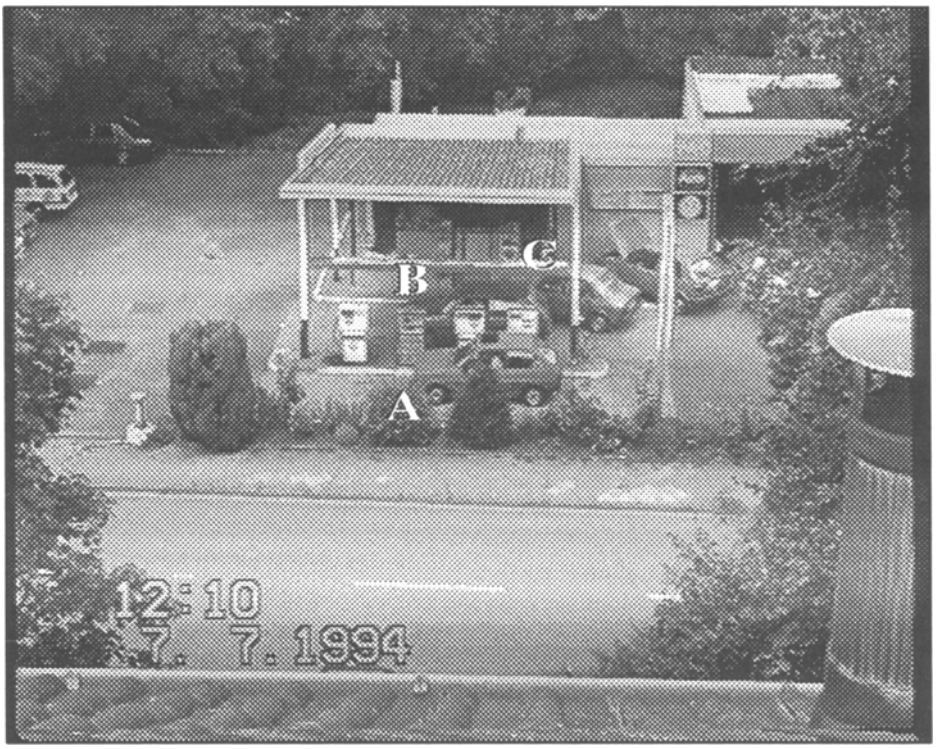

Fig. 1. 350^{th} frame of an image sequence showing gas station traffic. Vehicle A has stopped in front of the petrol pump area, a dark vehicle B has stopped on the rear lane behind the petrol pump area. An additional vehicle C is just passing the vehicle B in order to pull up to the leftmost petrol pump. This significantly occluded moving vehicle C will be discussed in more detail.

challenge since it is significantly occluded by other vehicles as well as by stationary scene components. In addition, its image is located in a low contrast image environment. The image features which provide significant cues for the vehicle image cannot, therefore, be found only in the *spatial* gray value gradients. As Figure 2 illustrates, important information is contained in the optical flow vectors estimated from *spatio-temporal* gray value gradients. This information enables the image sequence analysis system to track partially occluded vehicles *without* explicitly modeling the occlusion.

2 Related Publications

A review of relevant literature can be found in [Sullivan 92; Koller et al. 93; Sullivan *et al.* 95; Cédras & Shah 95], for optical flow estimation see [Barron et al. 94; Otte & Nagel 94; Otte & Nagel 95]. We can, therefore, confine ourselves to recent publications.

(a) (b)

Fig. 2. An enlarged section of Figure 1. The significantly occluded vehicle C is moving from right to left. The estimated optical flow vectors shown in (b) are better cues for the vehicle image than the gray values of the current image frame (a) themselves. The optical flow vectors overlap most of the visible parts of the image of vehicle C.

[Otte & Nagel 94] recorded a three-dimensional polyhedral scene with a moving camera on a robot and compared the results of various optical flow estimation approaches with ground truth about image motion, but did not exploit the measured differences in order to update a pose estimate.

[Schirra *et al.* 87] as well as [Gong & Buxton 93] track vehicles by clustering and chaining displacement vectors. After determining optical flow and its discontinuities using non-linear diffusion, [Proesmans *et al.* 94] estimate 2D vehicle motion parameters. In distinction to our 3D pose estimation and tracking, these three groups confine themselves to techniques in the 2D image domain.

Supposing that an initial value for the object pose is available, for instance chosen interactively, [Worrall *et al.* 94] proposed a pose refinement of active models using forces in 3D without an extraction of line segments. [Tan *et al.* 94] localize vehicles without feature extraction, too, based on a 1D correlation technique. Both use a histogram voting and peak searching process in order to determine the vehicle pose whereas we compute the *Jacobian of the measurement function* to update the 3D pose with a Maximum-A-Posteriori (MAP) estimation process. Moreover, in distinction to the approach of [Worrall *et al.* 94; Tan *et al.* 94], we obtain initial pose estimates automatically by segmenting an optical flow field. The framework of our approach can be seen as analogous to the work of [Lowe 87], [Koller et al. 93], and [Kollnig *et al.* 94]. However, none of the cited approaches exploits optical flow estimates in order to update a 3D pose estimate.

3 Computing the Motion Field

The five-dimensional state vector

$$\boldsymbol{x}(t) = \Big(p_x(t), \quad p_y(t), \quad \phi(t), \quad v(t), \quad \omega(t)\Big)^T \tag{1}$$

to be estimated by our pose estimation algorithm comprises the position $\boldsymbol{p} = (p_x, p_y, 0)^T$ and orientation ϕ of the vehicle model relative to a reference (world) coordinate system in the road plane as well as the magnitudes of the translational velocity v and angular velocity ω.

In our implementation, the trajectory of the vehicle model reference point $\boldsymbol{p}(t)$ is assumed to be (locally) described by a simple circular motion model with constant magnitudes of the translational and angular velocities:

$$\boldsymbol{p}(t) = \boldsymbol{C} + \frac{v}{\omega} \begin{pmatrix} \sin\phi(t) \\ -\cos\phi(t) \\ 0 \end{pmatrix} \quad , \tag{2}$$

where $\boldsymbol{C}$ denotes the center of the circular trajectory and $\rho = v/\omega$ its radius. This model contains a straightforward movement as special case (infinite radius of the circle).

Let $\boldsymbol{x}_m$ denote the coordinates of a point in the (vehicle) model coordinate system, $\boldsymbol{x}_w$ its position vector in the world coordinate system and $\boldsymbol{x}_c$ its coordinates in the camera coordinate system, respectively.

The trajectory of a scene point on the vehicle surface with the model coordinates $\boldsymbol{x}_m$ is given by the following equation:

$$\boldsymbol{x}_w = \boldsymbol{x}_w(\boldsymbol{x}(t), \boldsymbol{x}_m) = \boldsymbol{p} + R_{wm}(\phi)\boldsymbol{x}_m \quad ,$$

where $R_{wm}(\phi)$ denotes a orthonormal 3×3 matrix describing the rotation of the model coordinate system with respect to the world coordinate system. The 3D motion $\dot{\boldsymbol{x}}_w$ of a scene point with model coordinates $\boldsymbol{x}_m$ is given by

$$\dot{\boldsymbol{x}}_w = \dot{\boldsymbol{x}}_w(\boldsymbol{x}(t), \boldsymbol{x}_m) = v \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} + \frac{\partial R_{wm}(\phi)}{\partial \phi} \omega \boldsymbol{x}_m \quad . \tag{3}$$

The 2D image coordinates $\boldsymbol{\xi}$ of a 3D point with the world coordinates $\boldsymbol{x}_w$ are obtained by the following chain of operations:

$$\boldsymbol{x}_c = R_{kw}\boldsymbol{x}_w + \boldsymbol{t}_{kw} \quad , \tag{4}$$

$$\boldsymbol{\xi} = \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} f_x \frac{x_c}{z_c} + x_0 \\ f_y \frac{y_c}{z_c} + y_0 \end{pmatrix} . \tag{5}$$

The rotation matrix R_{kw} as well as the translation vector $\boldsymbol{t}_{kw}$ contain *external* camera parameters, whereas the focal length $(f_x, f_y)^T$ as well as the principal point $(x_0, y_0)^T$ represent *internal* camera parameters. External and internal camera parameters are to be estimated a priori by a calibration step.

The derivation of operations given in equations 4 and 5 with respect to time expresses the dependence of the 2D motion field vector $\dot{\boldsymbol{\xi}}$ at the pixel location $\boldsymbol{\xi}$ on the 3D motion $\dot{\boldsymbol{x}}_w$ of the corresponding scene point with the model coordinates $\boldsymbol{x}_m$:

$$\dot{\boldsymbol{\xi}} = \dot{\boldsymbol{\xi}}(\boldsymbol{x}(t), \boldsymbol{\xi}) = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}_c} \frac{\partial \boldsymbol{x}_c}{\partial \boldsymbol{x}_w} \dot{\boldsymbol{x}}_w(\boldsymbol{x}(t), \boldsymbol{x}_m) \quad . \tag{6}$$

Exploiting equations 4 and 5, we obtain:

$$\frac{\partial \boldsymbol{x}_c}{\partial \boldsymbol{x}_w} = R_{kw} \quad , \tag{7}$$

$$\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}_c} = \begin{pmatrix} f_x \frac{1}{z_c} & 0 & -f_x \frac{x_c}{z_c^2} \\ 0 & f_y \frac{1}{z_c} & -f_y \frac{y_c}{z_c^2} \end{pmatrix} . \tag{8}$$

In order to estimate the motion field vector $\dot{\boldsymbol{\xi}}(\boldsymbol{x}(t), \boldsymbol{\xi})$ at each pixel location $\boldsymbol{\xi}$ at time point t, we determine the model coordinates $\boldsymbol{x}_m$ of the corresponding 3D scene point by means of a ray tracing algorithm and by exploiting a priori knowledge about the scene and the current state vector $\boldsymbol{x}(t)$. Substituting equations 3, 7, and 8 into equation 6 yields the motion field vector $\dot{\boldsymbol{\xi}}$ at the pixel location $\boldsymbol{\xi}$.

Analogous considerations hold for the estimation of a motion field vector $\dot{\boldsymbol{\xi}}(\boldsymbol{x}(t), \boldsymbol{\xi})$ at a pixel location $\boldsymbol{\xi}$ which represents the image position of a shadow point cast by a vehicle, except that we have to perform *two* ray tracing steps: one to get the corresponding scene coordinates of the shadow point and a second one to get the vehicle point which is projected onto the shadow point.

4 Update Step

Let $\boldsymbol{x}_k = \boldsymbol{x}(t_k)$ denote the state vector (see equation 1) to be estimated at halfframe time point t_k. We adopt the usual dynamic system notation (see, e.g., [Gelb 74]) denoting by $(\hat{\boldsymbol{x}}_k^-, P_k^-)$ and $(\hat{\boldsymbol{x}}_k^+, P_k^+)$, respectively, the estimated state vectors and their covariances before and after an update which incorporates the measurement at halfframe time point t_k.

Let an initial guess $\hat{\boldsymbol{x}}_0^-$ about state vector $\boldsymbol{x}_0$ be provided by a data driven motion segmentation step as described by, e.g., [Bouthemy & François 93; Gong & Buxton 93; Proesmans *et al.* 94; Kollnig *et al.* 94] or by a predicted estimate $\hat{\boldsymbol{x}}_k^-$ exploiting the state vector $\hat{\boldsymbol{x}}_{k-1}^+$ at the previous halfframe time point t_{k-1} and a state transition function with respect to a vehicle motion model. A view sketch is then generated, i.e. a set of model edge segments, by projecting edges of a 3D polyhedral vehicle model from the scene into the image plane and by removing

invisible edge segments by a hidden-line algorithm. At each pixel location $\boldsymbol{\xi}$ of the frame at time point t_k, let $h_{||\nabla g||}(\boldsymbol{x}_k, \boldsymbol{\xi})$ denote the synthetic gradient norm as described by [Kollnig & Nagel 95] and let $(h_u(\boldsymbol{x}_k, \boldsymbol{\xi}), h_v(\boldsymbol{x}_k, \boldsymbol{\xi}))^T = \dot{\boldsymbol{\xi}}(\boldsymbol{x}_k, \boldsymbol{\xi})$ denote the motion field according to equation 6. The measurement function

$$\boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\xi}) = \begin{pmatrix} h_{||\nabla g||}(\boldsymbol{x}_k, \boldsymbol{\xi}) \\ h_u(\boldsymbol{x}_k, \boldsymbol{\xi}) \\ h_v(\boldsymbol{x}_k, \boldsymbol{\xi}) \end{pmatrix} \tag{9}$$

is nonlinear in the state vector $\boldsymbol{x}_k$ since the perspective projection is nonlinear.

At each pixel location $\boldsymbol{\xi}$ of the halfframe at time point t_k, we estimate

$$\boldsymbol{z}_k(\boldsymbol{\xi}) = \begin{pmatrix} ||\nabla g_k(\boldsymbol{\xi})||_2 \\ u(\boldsymbol{\xi}) \\ v(\boldsymbol{\xi}) \end{pmatrix} \tag{10}$$

the Euclidean norm $||\nabla g_k||_2$ of the gray value gradient as well as the optical flow $(u, v)^T$.

We assume that the measurement $\boldsymbol{z}_k(\boldsymbol{\xi})$ at the current halfframe time point t_k at the pixel location $\boldsymbol{\xi}$ is equal to the measurement function $\boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\xi})$ plus white Gaussian measurement noise $\boldsymbol{v}_k$ with covariance R_k:

$$\boldsymbol{z}_k(\boldsymbol{\xi}) = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\xi}) + \boldsymbol{v}_k. \tag{11}$$

Assuming the state vector $\boldsymbol{x}_k$ is normally distributed around the estimate $\hat{\boldsymbol{x}}_k^-$ with covariance P_k^-, a MAP estimation can be stated as the minimization of the following objective function:

$$\begin{aligned} \frac{1}{2n} \sum_{\boldsymbol{\xi}} \Big(\boldsymbol{z}_k(\boldsymbol{\xi}) - \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\xi})\Big)^T R_k^{-1} \Big(\boldsymbol{z}_k(\boldsymbol{\xi}) - \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{\xi})\Big) \\ + \frac{1}{2} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-\right)^T P_k^{-1} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-\right) \rightarrow \min_{\boldsymbol{x}_k} \end{aligned} \tag{12}$$

resulting in an update step of an *iterated extended Kalman Filter* (IEKF) [Bar-Shalom & Fortmann 88; Gelb 74]. n denotes the size of the image region to be summed over. In our actual implementation, each term of the sum is weighted by the confidence (the estimated singular value) of the corresponding optical flow vector.

5 Results

5.1 Downtown Intersection Sequence

We discuss the results of our experiments with an image sequence illustrated by its first frame in Figure 3.

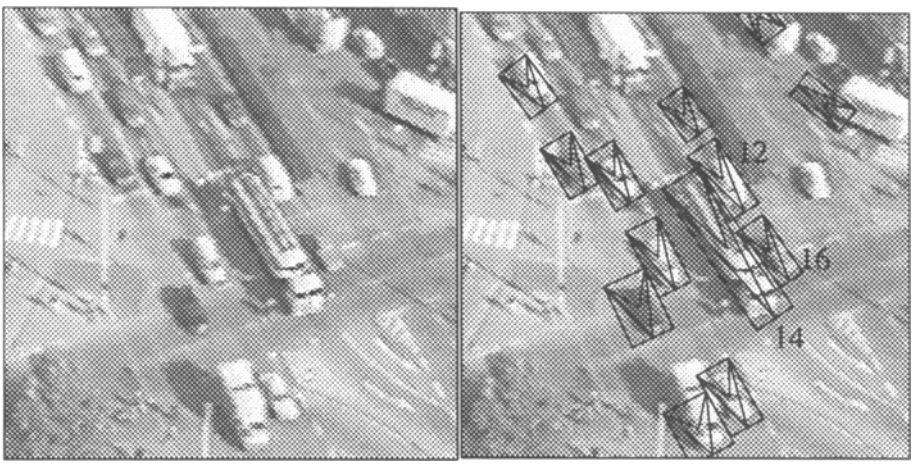

Fig. 3. An enlarged section of the 3^{rd} frame of an image sequence recording downtown intersection traffic. The location and shape of detected moving image regions are described by an enclosing rectangle with one side parallel to the direction of motion. The lines inside the rectangle form an arrow in the direction of motion. By means of an off-line calibration process, these vehicle hypotheses are backprojected from the 2D image plane into the 3-D world. The resulting pose estimates are used to initialize a Kalman-Filter tracking process. The vehicles are referred to by numbers indicated in this frame.

In order to systematically analyze the influence of the optical flow measurements on the pose update and tracking, we first exploit only the optical flow, neglecting a match of image gradients, i.e. ignoring the first component $h_{||\nabla g||}(\boldsymbol{x}_k, \boldsymbol{\xi})$ of the measurement function given in eq. 9. A potential deficiency of the approach based purely on the new measurements is thus not covered by the robustness of the previously used approach [Kollnig & Nagel 95].

Although we could track a remarkable number of vehicles exploiting only optical flow estimates as measurements (see, too, Section 5.3), the object #12 could not be correctly tracked in this mode.

In order to assess the state–estimations, we compare the root of the estimated state covariance diagonal elements (standard deviations) for object #12 for three different tracking methods (see Figure 4): using only image gradients in the measurement function (first row in Figure 4), using only optical flow (second row), and finally combining these two approaches (last row). The three diagrams in the *left* column show the temporal development of the estimated standard deviations for the (estimated) *position* of the object in the road plane. The middle diagram shows, that the estimated position is not very accurate if we exploit only optical flow measurements: the estimated standard deviations remain in the vicinity of their initial values. This is taken as an indication that we cannot rely only on the optical flow measurements. The estimated motion

field depends on the vehicle position, but the vehicle images are too small in the image sequence under scrutiny for the motion field to vary significantly along a vehicle image. Combining the two kinds of measurements, the position estimation becomes accurate even without modeling occlusion, see Figure 4 (bottom, left). The *orientation* can be better estimated by exploiting optical flow than image gradients, see Figure 4 (*right* column). In the latter case (first row), the estimated orientation oscillates in the initial phase with a significant amplitude. In both cases (position and orientation), the combination of image gradients and optical flow yields the lowest values for the estimated standard deviations and thus yields the most reliable state estimations. Similar considerations hold for the estimation of the remaining state components (speed and angular velocity).

Figure 5 demonstrates the evaluation of the entire sequence with superimposed results.

5.2 Gas Station Image Sequence

In a second experiment, we have tested our new approach by means of an image sequence recorded at a gas station (see Figure 1). Due to space limitations, we focus on the tracking of vehicle `C` which turns out to be the most challenging case: it is partially occluded by stationary scene components and by the vehicles `A` and `B`. Moreover, vehicle `C` is moving in a low contrast image area under the gas station roof. The tracking is shown in detail in Figure 6. The left column shows the result of the updated pose estimation, while the right column depicts the estimated optical flow field. The resulting trajectories are shown in Figure 7.

5.3 Testing Benchmark Image Sequences

In order to assess our approach, we tried to track vehicles only on the basis of optical flow matching in image sequences in which we are able to track all moving vehicles with our former approach (i.e. only using the image gradient).

Based only on optical flow estimates, we could correctly track 11 out of 12 moving objects in the Durlacher Tor sequence and 10 out of 12 moving objects in the Ettlinger Tor sequence. In the latter sequence, we could even track the partially occluded vehicle moving in front of a bus, while we got problems with our former approach.

Moreover we noticed, that by exploiting only optical flow vectors, a rough initial estimation for the *orientation* of the bus can be corrected in 3 iteration steps, significantly less steps than those which are necessary in the approach exploiting only image gradients [Kollnig & Nagel 95]. However, by exploiting only optical flow, the *position* component in the driving direction is not correctly updated.

6 Conclusion

As far as we know, we present the first approach in which exploiting *interframe* image data leads to an improvement of tracking objects in the 3D scene domain

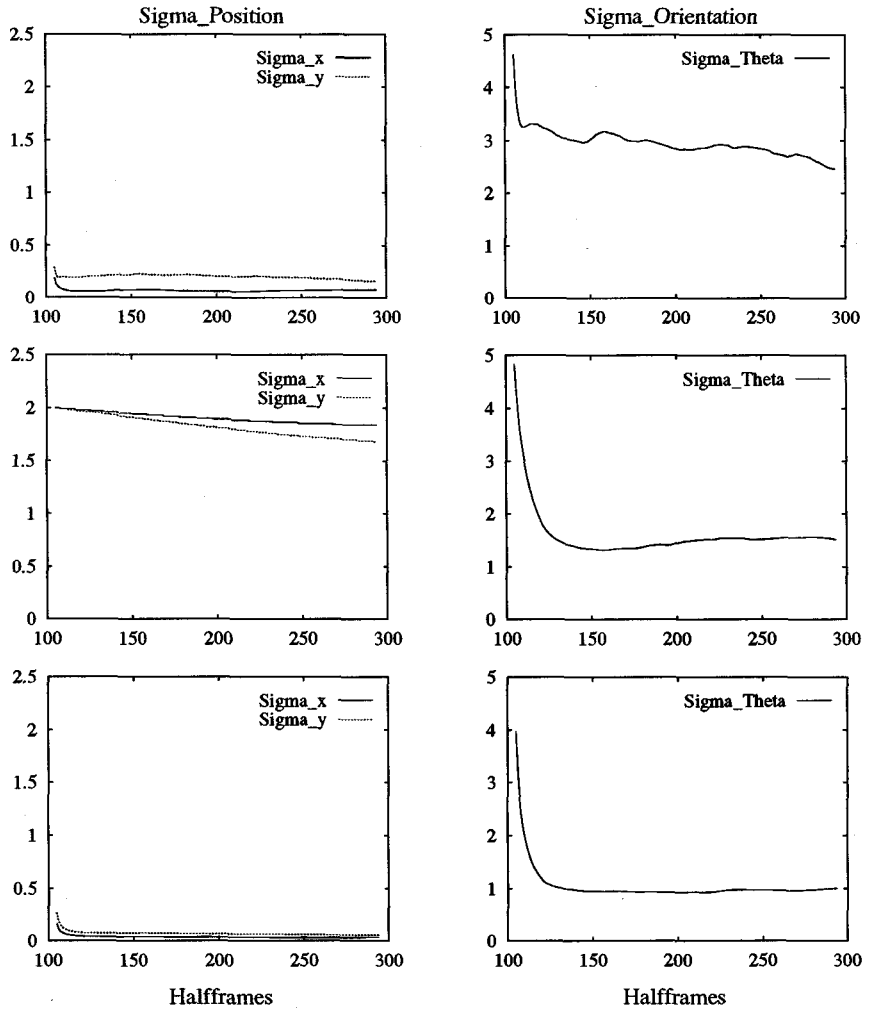

Fig. 4. Estimated standard deviation (root of the diagonal state covaraince matrix) for position (left column) and orientation (right column) of vehicle #12. In the first row, only image gradients are used in the measurement function. Here we had to explicitly model the occlusion in order to track the object properly. In the second row only optical flow is used *without* explicitly modelling the occlusion. In the last row, image gradients and optical flow are combined in the measurement function. In the case of the single use of optical flow, the estimated *position* is not very accurate: the estimated standard deviations remain in the vicinity of their initial values (second row, left), while the single use of image gradients results in a more reliable estimate (top left diagram). On the contrary, the *orientation* can be estimated better by means of optical flow than with image gradients. In both cases (position and orientation) the combined use of image gradients and optical flow reduces the estimated standard deviations in distinction to the isolated use of either measurement (last row).

Fig. 5. The results of uninterrupted tracking of the objects recorded in the downtown intersection sequence until halfframe time point 250^{th}.

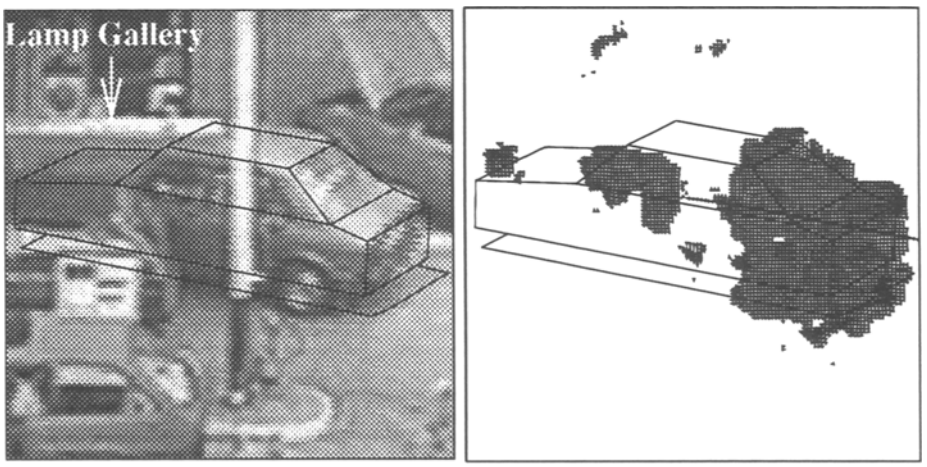

Fig. 6. Evaluation of the image of vehicle C of the gas station sequence. It is occluded by other cars and by stationary scene components, for instance a lamp gallery. The left image shows the updated pose estimate, while the right image depicts the estimated flow vectors for halfframe #700. The occluding scene components can be recognized in the optical flow field although they are not modeled explicitly.

Fig. 7. Computed vehicle trajectories for vehicles A, B, and C.

compared to using *intraframe* image data only, for instance image gradients. The influence of both kinds of measurements on the pose estimation is discussed systematically in different real-world image sequences. The robustness of the new approach has been demonstrated on a considerable range of vehicles which are partially occluded and can be correctly tracked *without* modeling the occlusion explicitly.

Acknowledgment

We thank M. Haag for his help in the preparation of the final version of this contribution and H. Damm for recording the gas station image sequence.

References

[Bar-Shalom & Fortmann 88] Y. Bar-Shalom, T. E. Fortmann, *Tracking and Data Association*, Academic Press, Inc., Boston/MA, Orlando/FL, and others, 1988

[Barron et al. 94] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, Performance of Optical Flow Techniques, *International Journal of Computer Vision* **12** (1994) 43-77

[Bouthemy & François 93] P. Bouthemy, E. François, Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence, *International Journal of Computer Vision* **10** (1993) 157–182.

[Cédras & Shah 95] C. Cédras, M. Shah, Motion-Based Recognition: A Survey, *Image and Vision Computing* **13**:2 (1995) 129–155.

[Gelb 74] A. Gelb (ed.), *Applied Optimal Estimation*, MIT Press, Cambridge/MA, 1974.

[Gong & Buxton 93] S. Gong, H. Buxton, From Contextual Knowledge to Computational Constraints, in *Proc. British Machine Vision Conference*, Guildford/UK, Sept. 21-23, 1993, pp. 229–238.

[Koller et al. 93] D. Koller, K. Daniilidis, H.-H. Nagel, Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes, *International Journal of Computer Vision* **10** (1993) 257–281.

[Kollnig *et al.* 94] H. Kollnig, H.-H. Nagel, and M. Otte, Association of Motion Verbs with Vehicle Movements Extracted from Dense Optical Flow Fields, in J.-O. Eklundh (ed.), *Proc. Third European Conference on Computer Vision ECCV '94,* Vol. II, Stockholm, Sweden, May 2-6, 1994, Lecture Notes in Computer Science **801**, Springer-Verlag, Berlin, Heidelberg, New York/NY, and others, 1994, pp. 338–347.

[Kollnig & Nagel 95] H. Kollnig and H.–H. Nagel, 3D Pose Estimation by Fitting Image Gradients Directly to Polyhedral Models, *Proc. Fifth International Conference on Computer Vision ICCV '95*, Cambridge/MA, 20–23 June 1995, pp. 569–574

[Lowe 87] D.G. Lowe, Three-Dimensional Object Recognition from Single Two-Dimensional Images, *Artificial Intelligence* **31** (1987) 355–395.

[Otte & Nagel 94] M. Otte, H.–H. Nagel, Optical Flow Estimation: Advances and Comparisons, *Proc. Third European Conference on Computer Vision ECCV '94*, Vol. I, Stockholm / Sweden, 2–6 May 1994, J.-O. Eklundh (ed.), Lecture Notes in Computer Science **800**, Springer-Verlag Berlin Heidelberg New York/NY 1994, pp. 51–60.

[Otte & Nagel 95] M. Otte, H.–H. Nagel, Estimation of Optical Flow Based on Higher–Order Spatiotemporal Derivatives in Interlaced and Non–Interlaced Image Sequences, *Artificial Intelligence* **78** (1995) 5–43

[Proesmans *et al.* 94] M. Proesmans, L. Van Gool, E. Pauwels, A. Oosterlinck, Determination of Optical Flow and Its Discontinuities Using Non-Linear Diffusion, in J.-O. Eklundh (Ed.), *Proc. Third European Conference on Computer Vision ECCV '94,* Vol. II, Stockholm, Sweden, May 2-6, 1994, Lecture Notes in Computer Science **801**, Springer-Verlag, Berlin, Heidelberg, New York/NY and others, 1994, pp. 295–304.

[Schirra *et al.* 87] J.R.J. Schirra, G. Bosch, C.K. Sung, G. Zimmermann, From Image Sequences to Natural Language: A First Step towards Automatic Perception and Description of Motion, *Applied Artificial Intelligence* 1 (1987) 287–307.

[Sullivan 92] G.D. Sullivan, Visual Interpretation of Known Objects in Constrained Scenes, *Philosophical Transactions Royal Society London* (B) **337** (1992) 361–370.

[Sullivan *et al.* 95] G.D. Sullivan, A.D. Worrall, and J.M. Ferryman, Visual Object Recognition Using Deformable Models of Vehicles, in *Proc. Workshop on Context-Based Vision*, 19 June 1995, Cambridge/MA, pp. 75–86.

[Tan *et al.* 94] T.N. Tan, G.D. Sullivan, K.D. Baker, Fast Vehicle Localization and Recognition without Line Extraction and Matching, in *Proc. British Machine Vision Conference*, York/UK, Sept. 13-16, 1994, pp. 85–94.

[Worrall *et al.* 94] A.D. Worrall, G.D. Sullivan, and K.D. Baker, Pose Refinement of Active Models Using Forces in 3D, in J.-O. Eklundh (ed.), *Proc. Third European Conference on Computer Vision (ECCV '94),* Vol. I, Stockholm, Sweden, May 2-6, 1994, Lecture Notes in Computer Science **800**, Springer-Verlag, Berlin, Heidelberg, New York/NY, and others, 1994, pp. 341–350.