

Stereo Vision Based Ego-Motion Estimation with Sensor Supported Subset Validation

Jan Horn, Alexander Bachmann, and Thao Dang
Institut für Mess- und Regelungstechnik
University of Karlsruhe (TH), Germany
{horn,bachmann,dang}@mrt.uka.de

Abstract— We propose a method to reliably estimate the motion of a dynamic stereo camera system in the three dimensional world where observations are disturbed by high portions of independently moving objects. Robustness of the estimation process is achieved by applying an additional visual sensor. The system consists of a stereo vision sensor, acquiring optical flow and depth information of the scene and a camera with its optical axis oriented perpendicular to the road surface, measuring the speed over ground of the camera-equipped vehicle. The fusion approach presented in this paper combines the motion estimates of the two sensors and applies an efficient random sampling scheme that evaluates the distribution of motion patterns in the scene. The goal of the sampling scheme is to separate the observations into alien and ego-motion portions used in the subsequent step to extract the ego-motion of the camera system. The presented setup of the two visual sensors in combination with the observation sampling scheme increases robustness of the overall system.

I. INTRODUCTION

Reliably estimating the motion of a dynamic camera system is becoming increasingly important in any vision system deployed to complex and manifold environments. Knowledge about the relative motion of a camera with respect to the static environment represents a cornerstone for subsequent dynamic scene analysis and future developments in advanced driver-assistance systems. Existing approaches for vision-based motion discrimination can be roughly divided into two categories:

The first group relies on the analysis of *monocular image sequences* (see [1] for example). Here, 2D motion of object points in the image plane is connected to the underlying 3D motion of the object in the scene by a number of constraints and restrictions to overcome inherent ambiguities associated with optical flow. One popular constraint is, that the observed motion vectors originate from the motion of a plane relative to the camera. In traffic scenes, this plane usually corresponds to the road plane. Other approaches reduce the number of degrees of freedom in the motion model to overcome ambiguities.

The second group applies *stereo motion analysis*, where the ambiguities mentioned above can be dissolved by

The authors would like to thank the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for the financial support within the special research project SFB/Tr 28 "cognitive automobiles".

incorporating depth information of the scene [2], [3].

Both, mono and stereo approaches assume rigid body motion with observations originating exclusively from one region or object performing the same motion in the 3D world. This basic prerequisite is very difficult to accomplish if it comes to real traffic scenarios where the rigidly moving background can be overlaid by a unknown number of objects varying in velocity, size, motion, appearance, etc. (see fig. 1).



Fig. 1. Top: Image from a real traffic scene. Bottom: Corresponding 2D motion (flow vectors). The approaching vehicles move relatively to the rigid world and consequently introduce model violating flow vectors.

A meaningful estimate of the ego-motion of the camera system relative to the static environment implies the elimination of outliers and observations originating from independently moving objects in the scene. Several approaches exist to separate image regions that are static relative to the camera system from image regions with different motion profiles, but mostly they suffer from the limited availability of salient image features and the difficulty to differentiate between statistical outliers and model errors like other moving cars.

Therefore in this paper we present a system based on the stereo motion paradigm that is extended by a second, complementary optical sensor measuring the speed over ground of the camera-equipped vehicle (see fig.2). The speed over ground sensor (in the sequel called SOG-sensor) is unaffected by independently moving objects in the scene but prone to measurement inaccuracies due to pitch and roll motion of the vehicle. In [4], the complementary sensor system has been shown to provide robust and highly available ego-motion estimates. But still, the implicit requirement here is that the majority of observations originates from the static background of the scene and not from other moving objects like cars.

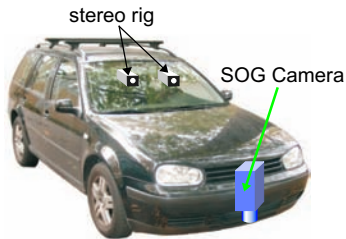


Fig. 2. The vehicle is equipped with a stereo rig behind the windscreen and a camera to measure speed over ground by directly observing the road.

As an addition, the main contribution of this paper is the extension of the existing system by an efficient random sampling scheme that evaluates the distribution of motion patterns in the scenery and classifies the observations into alien and ego-motion portions. The connection between 2D-motion in the image plane and 3D-motion in the world is derived based upon the well known Longuet-Higgins equations [5]. Thus, the enhanced system is capable of reliably measuring the 6-degree of freedom motion of the camera system even if the scene is highly dominated by distinct moving objects. The main benefits of the presented system are a high robustness of the motion estimates and improved self assessment possibilities.

The remainder of the paper will be organized as follows. Section II briefly introduces the SOG-sensor that directly observes the road as a moving surface to obtain precise speed vector information. In sec. III the stereo vision sensor that provides measurements of the optical flow and disparity is described. These observations are used to determine the 3D motion of the camera system relative to the static environment. In sec. IV we show the validation technique that eliminates observations caused by moving objects in the scene. In sec. V real image sequences are analyzed to demonstrate the performance of the proposed technique. Section VI sums up before sec. VII gives an outlook towards future work.

II. SPEED OVER GROUND SENSOR

Speed over ground (SOG) sensing within this work is performed and described like in [4], [6], [11].

In the past, several optical techniques have been proposed to measure speed over ground of vehicles. One widely spread is the one already described in [7] and [8]. Since image processing equipment has become much more powerful in recent years, image based approaches offer themselves to measure speed over ground. This means, that a camera acquires an image sequence of the road. The road is modeled as a textured surface that moves relatively to the camera, i.e. the vehicle. In [9] an approach based upon impulse train illuminated images is proposed to acquire the velocity vector of a moving textured surface. Many other algorithms known from image registration can also be used to precisely measure the velocity vector. These are described in e. g. [10].

The velocity vector \mathbf{v}_{sog} is measured by calculating the translational shift $\mathbf{t} = (t_x, t_y)^T$ between two images of the road using image registration techniques (fig. 3, [11], [6]). The images are called b_1 and b_2 in the sequel. We assume that these images are acquired with a frame rate f_L and that the road only moves in the plane orthogonal to the optical axis of the camera.

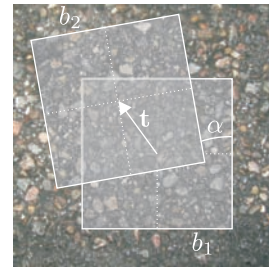


Fig. 3. The speed over ground sensor estimates the shift \mathbf{t} between two images of the moving road to measure the speed vector \mathbf{v}_{sog} .

Based upon a roto-translational motion model with translation $\hat{\mathbf{t}} = (\hat{t}_x, \hat{t}_y)^T$ and rotation $\hat{\alpha}$, the transformation of a single point $\mathbf{x}_i = (x_i, y_i)^T$ in image b_1 to the corresponding point $\hat{\mathbf{x}}'_i = (\hat{x}'_i, \hat{y}'_i)^T$ in image b_2 can be written as¹:

$$\begin{pmatrix} \hat{x}'_i \\ \hat{y}'_i \end{pmatrix} = \begin{bmatrix} \cos \hat{\alpha} & \sin \hat{\alpha} \\ -\sin \hat{\alpha} & \cos \hat{\alpha} \end{bmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} \hat{t}_x \\ \hat{t}_y \end{pmatrix}. \quad (1)$$

Since the images b_1 and b_2 need to overlap, very high frame rates f_L have to be used. Consequently the rotation $\hat{\alpha}$ between the images is very small and we can linearize with respect to the angle $\hat{\alpha}$ by setting $\cos \hat{\alpha} = 1$ and $\sin \hat{\alpha} = \hat{\alpha}$:

$$\begin{pmatrix} \hat{x}'_i \\ \hat{y}'_i \end{pmatrix} = \begin{bmatrix} 1 & \hat{\alpha} \\ -\hat{\alpha} & 1 \end{bmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} \hat{t}_x \\ \hat{t}_y \end{pmatrix}. \quad (2)$$

Taking N points into consideration, a linear observation

¹Please consider fig. 3 as a sketch, since the strict definition of the sign and orientation of \mathbf{t} might be different than in fig. 3.

model can be assumed:

$$\underbrace{\begin{pmatrix} \hat{x}'_0 - x_0 \\ \hat{y}'_0 - y_0 \\ \vdots \\ \hat{x}'_{N-1} - x_{N-1} \\ \hat{y}'_{N-1} - y_{N-1} \end{pmatrix}}_{=: \Delta \hat{\mathbf{x}}} + \hat{\mathbf{e}} = \underbrace{\begin{bmatrix} 1 & 0 & y_0 \\ 0 & 1 & -x_0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & y_{N-1} \\ 0 & 1 & -x_{N-1} \end{bmatrix}}_{=: \mathbf{H}} \underbrace{\begin{pmatrix} \hat{t}_x \\ \hat{t}_y \\ \hat{\alpha} \end{pmatrix}}_{=: \hat{\mathbf{p}}}. \quad (3)$$

Or written in matrix form:

$$\Delta \hat{\mathbf{x}} + \hat{\mathbf{e}} = \mathbf{H} \hat{\mathbf{p}}. \quad (4)$$

For an isotropic road texture, it can be assumed that the observations are uncorrelated and have the same variance. Consequently, a standard least-squares estimation is sufficient to calculate the translation $(\hat{t}_x, \hat{t}_y)^T$ and the rotation $\hat{\alpha}$:

$$\hat{\mathbf{p}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \Delta \hat{\mathbf{x}}. \quad (5)$$

The velocity vector \mathbf{v}_{sog} can then be calculated by multiplying the translation vector $\hat{\mathbf{t}}$ with the frame rate f_L of the camera:

$$\mathbf{v}_{sog} = \begin{pmatrix} v_{xsog} \\ v_{ysog} \end{pmatrix} = f_L \cdot \hat{\mathbf{t}}. \quad (6)$$

This speed over ground estimate is passed to the observation validation stage described in sec. IV. The index "sog" indicates a "speed over ground" sensor based estimate.

III. STEREO VISION SENSOR

This section summarizes stereo vision based ego-motion estimation originally developed in our earlier work [4].

Motion determination of a moving vehicle based on stereo vision as presented in this work can be largely interpreted as the task of resolving the visual 2D motion in the image into its components in 3D space. To achieve this, the 3D information retrieved from a stereo image pair is gathered with the motion of corresponding image points between frames captured at different time steps, in the sequel called displacement or optical flow $(u, v)^T$. Note that 3D information, namely the disparity Δ , is determined with respect to the right camera. The 2D motion analysis (i. e. optical flow calculation) also operates on the images of the right camera. In this contribution optical flow $(u, v)^T$ is determined with a hierarchical Lukas-Kanade-algorithm as proposed in [12], [13].

For every 3D scene point $\mathbf{X} = (X, Y, Z)^T$, the respective position \mathbf{x} in the image plane is denoted by

$$\mathbf{x} = \begin{pmatrix} y \\ z \end{pmatrix} = \Pi(\mathbf{X}) = -\frac{f}{X} \begin{pmatrix} Y \\ Z \end{pmatrix}. \quad (7)$$

It is assumed that the stereo camera setup is fully calibrated [15] and all entities within the images are given in normalized coordinates, i. e. we suppose that the images are acquired by normalized cameras with focal lengths $f = 1$

and the image center located at coordinates $\mathbf{c} = (0, 0)^T$. The depth information of a scene can be recovered by measuring the horizontal displacement of two corresponding points in a rectified stereo image pair ([16]). Given the image coordinates y_L in the left and y_R in the right image along the epipolar line, the depth X of the scene point and its disparity are related by

$$X = \frac{b}{\Delta}, \quad (8)$$

where $\Delta = (y_R - y_L)$ is the disparity and b indicates the base length of the stereo rig. Figure 4 depicts the model of a stereo rig. Please note that the disparity Δ , the image displacement $(u, v)^T$, and image coordinates are specified in normalized form. The choice of the coordinate systems is derived from vehicular technology and might deviate from the image processing context of the stereo vision sensor. The x -axis is oriented in longitudinal direction of the car. Consequently image coordinates of the stereo vision sensor are y and z .

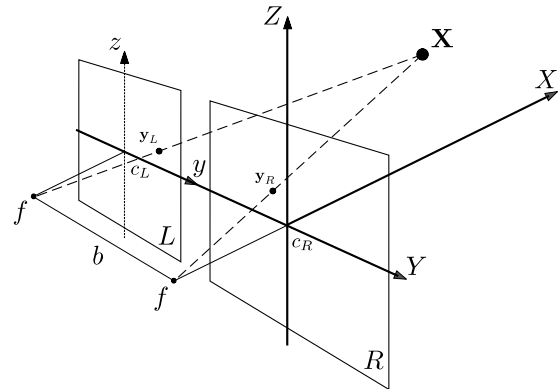


Fig. 4. The model of the stereo rig. The coordinate system of the right camera coincides with the global coordinate system. For any 3D scene point \mathbf{X} , its projection $\Pi(\mathbf{X})$ into the image plane can be described by eq. (7).

Since we consider a rigid motion model, the system is fully specified by the translational velocity $\mathbf{T} = (v_x, v_y, v_z)^T$ and rotational velocity $\mathbf{\Omega} = (\omega_x, \omega_y, \omega_z)^T$. With the Longuet-Higgins equations [5], the relation between 3D velocity and 2D displacement can be adequately described. Recalling eq. (7) and defining $(\dot{y}, \dot{z})^T = (u, v)^T$ as the motion of an image point across the image plane, the Longuet-Higgins equations for the case of translational and rotational rigid motion can be written

$$\dot{y} = u = \frac{\dot{Y}}{X} - \frac{Y}{X^2} \dot{X} = \left(-\frac{v_y}{X} - \omega_z + \omega_x \cdot z \right) - y \left(\frac{v_x}{X} - \omega_y \cdot z + \omega_z \cdot y \right), \quad (9)$$

$$\dot{z} = v = \frac{\dot{Z}}{X} - \frac{Z}{X^2} \dot{X} = \left(-\frac{v_z}{X} - \omega_x \cdot y + \omega_z \right) - z \left(-\frac{v_x}{X} - \omega_y \cdot z + \omega_z \cdot y \right). \quad (10)$$

By substituting eq. (8) into eqns. (9) - (10) and separating the translational from the rotational motion components, the

2D image displacement (u_i, v_i) of image point i caused by the 3D motion of the camera rack can therefore be written

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \mathbf{H}_i \cdot (v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)^T, \quad (11)$$

with

$$\mathbf{H}_i = \begin{bmatrix} \frac{-y_i \cdot \Delta_i}{b} & -\frac{\Delta_i}{b} & 0 & z_i & y_i \cdot z_i & -1 - y_i^2 \\ \frac{z_i \cdot \Delta_i}{b} & 0 & -\frac{\Delta_i}{b} & -y_i & 1 + z_i^2 & -y_i \cdot z_i \end{bmatrix}. \quad (12)$$

Stacking eq. (12) for N combinations of flow (u_i, v_i) and disparity Δ_i observed at the image coordinates (x_i, y_i) provides a linear observation model with $2N$ equations:

$$\underbrace{\begin{pmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ u_n \\ v_n \end{pmatrix}}_{=: \hat{\mathbf{u}}} + \hat{\mathbf{e}} = \underbrace{\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_3 \end{bmatrix}}_{=: \mathbf{H}} \begin{pmatrix} v_{xs} \\ v_{ys} \\ v_{zs} \\ \omega_{xs} \\ \omega_{ys} \\ \omega_{zs} \end{pmatrix} = \mathbf{H} \begin{pmatrix} \hat{\mathbf{T}}_s \\ \hat{\mathbf{\Omega}}_s \end{pmatrix}. \quad (13)$$

In eq. (13), the error vector $\hat{\mathbf{e}}$ accounts for errors and the additional index "s" indicates a "stereo" vision based velocity estimate. Neglecting the fact that the disparities Δ_i are subject to errors, the ego-motion parameter vector $(\hat{\mathbf{T}}_s, \hat{\mathbf{\Omega}}_s)^T = (v_{xs}, v_{ys}, v_{zs}, \omega_{xs}, \omega_{ys}, \omega_{zs})^T$ can be estimated by a straightforward least squares estimation² that minimizes $\hat{\mathbf{e}}^T \hat{\mathbf{e}}$.

$$\begin{pmatrix} \hat{\mathbf{T}}_s \\ \hat{\mathbf{\Omega}}_s \end{pmatrix} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \hat{\mathbf{u}} \quad (14)$$

It is evident from the matrix \mathbf{H}_i in eq. (12) that the translational components v_{xs} , v_{ys} and v_{zs} are not observable for a disparity $\Delta \rightarrow 0$. Therefore, we discard observations with a disparity that is below a certain threshold.

The calculation of the ego-motion vector $(\hat{\mathbf{T}}_s, \hat{\mathbf{\Omega}}_s)^T$ by the pseudo inverse technique in eq. (14) would yield acceptable results in case there are no outliers. But due to

- 1) errors in optical flow calculation,
- 2) errors in disparity calculation and
- 3) **alien motion in the image caused by other moving objects**

a robust technique is required that replaces least squares estimation. An M-Estimator ([17], [18]) or a RANSAC-Estimator ([19]) have proven to yield good estimates for the ego-motion $(\hat{\mathbf{T}}_s, \hat{\mathbf{\Omega}}_s)^T$ even if there are other moving objects in the scene.

But even the robust estimation approaches require – roughly speaking – that the majority of flow-disparity-combinations is caused by the ego-motion and not by other moving objects like other moving cars, pedestrians or shaking trees in the wind for instance. Especially in dense

²The framerate and calibration parameters must also be taken into account to obtain velocity units for $(\hat{\mathbf{T}}_s, \hat{\mathbf{\Omega}}_s)^T$.

traffic it is very unlikely that the majority of observations is only caused by the ego-motion. Therefore, the next section describes a SOG-sensor supported random sampling scheme that is capable of eliminating observations (u_i, v_i, Δ_i) that arise from other moving objects.

IV. SENSOR SUPPORTED SUBSET VALIDATION

In sec. III it was described how the ego-motion of a stereo camera rig can be estimated in case the motion in the scene is dominated by the ego-motion. Now, SOG sensing devices like described in sec. II already provide part of the ego-motion information: Speed over ground sensors provide longitudinal and transversal components of the velocity, namely v_{xsog} and v_{ysog} . In the following, this additional information is applied to separate ego-motion caused observations from alien motion caused observations. This technique allows a stereo vision based ego-motion estimation even in dense traffic.

Figure 5 shows a block diagram of the proposed approach. As described in sec. II, the SOG-sensor provides 2D speed over ground (v_{xsog}, v_{ysog}) by analyzing a high speed video sequence of the road. The stereo vision sensor provides optical flow (u_i, v_i) and disparity Δ_i observed at image coordinates (x_i, y_i) .

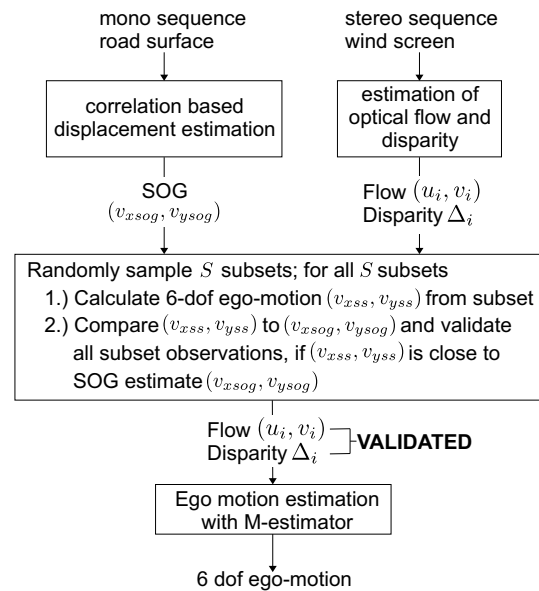


Fig. 5. Block diagram of the overall system.

Instead of estimating ego-motion with all N observations by a robust estimator, we validate observations by comparing them with the measurements of the SOG-Sensor³. To achieve this, we sample S randomly selected subsets of all N observations forcing three different, non-collinear

³Mathematically, there are $2N$ observations (see eq. (4)). In this paper one combination of flow (u_i, v_i) and disparity Δ_i is called one observation.

observations in a subset. Additionally, each subset consists of a minimum number of observations so that – again roughly speaking – outliers cannot “hide” in a mass of observations. A minimum number of three observations is needed to estimate⁴ 6-dof ego-motion. This yields six equations corresponding to the six parameters to be calculated. Let $(\hat{\mathbf{T}}_{ss}, \hat{\mathbf{\Omega}}_{ss})^T := (v_{xss}, v_{yss}, v_{zss}, \omega_{xss}, \omega_{yss}, \omega_{zss})^T$ be the ego-motion calculated with a subset; the index “ss” indicates a “subset” based estimation. Since the number of unknown parameters is equal to the number of equations the calculation of $(\hat{\mathbf{T}}_{ss}, \hat{\mathbf{\Omega}}_{ss})^T$ is basically a matrix inversion:

$$\begin{pmatrix} \hat{\mathbf{T}}_{ss} \\ \hat{\mathbf{\Omega}}_{ss} \end{pmatrix} = \mathbf{H}^{-1}(u_1, v_1, u_2, v_2, u_3, v_3)^T. \quad (15)$$

All three observations in this subset are classified as valid if the speed over ground estimated from this subset is close to the speed over ground provided by the stereo vision sensor. Stated in other words: The distance between $(v_{xss}, v_{yss})^T$ and $(v_{xsog}, v_{ysog})^T$ must be less than a threshold Δv to validate all observations within a subset:

$$\sqrt{(v_{xss} - v_{xsog})^2 + (v_{yss} - v_{ysog})^2} < \Delta v \Rightarrow \text{Valid Subset} \quad (16)$$

As a last step, ego-motion is estimated with all *validated* observations by an M-Estimator ([17], [18]) using a Huber-Function based weighting scheme. In the next section the performance of the proposed measurement principle is shown on real image sequences.

V. EXPERIMENTAL RESULTS

A. Experimental Setup

For the experimental tests we used a stereo rig with 0.3 meters basewidth (fig. 6(a)). Stereo images are acquired with a framerate of 10 Hz whereas the SOG camera works at 500 Hz framerate, since sequent images need to overlap to perform correlation based displacement estimation. Additionally, since the transversal component of the velocity v_y depends on the distance from the center of gravity of the car according to the bicycle car model, we mounted the stereo vision sensor right above the SOG camera (see fig. 6(b)). Consequently a transformation of the stereo vision based v_{ys} or v_{ys} into the coordinate system of the SOG via a yaw rate sensor is not necessary.

B. Non-moving ego-vehicle with transversal traffic in the field of view

As a first maneuver, the ego-vehicle was parked and the SOG-sensor provided $v_{xsog} = v_{ysog} = 0$. During recording the stereo sequence a car drives from right to left through the image. Figure 7 shows an image of the sequence at a time instant t_0 and the corresponding disparities.

The subset validation scheme described in sec. IV has been applied to the corresponding images at the same time

⁴Using as many equations as parameters to be estimated is actually a “calculation” not an estimation.

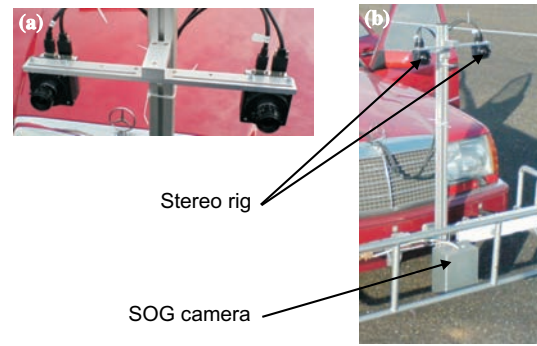


Fig. 6. Measurement setup mounted on the test car. (a) Stereo rig. (b) Overall camera system.

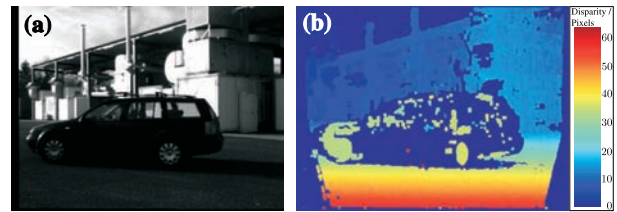


Fig. 7. (a) Image of the analyzed sequence. (b) Disparity calculated from stereo images.

instant t_0 . Please keep in mind that images from other time instances are used for optical flow calculation. Figure 8, which shows the results of the subset validation stage, confirms the performance of the presented technique. The flow vectors originating from a moving object – the car in the field of view – are eliminated and not considered in the subsequent motion estimation process.

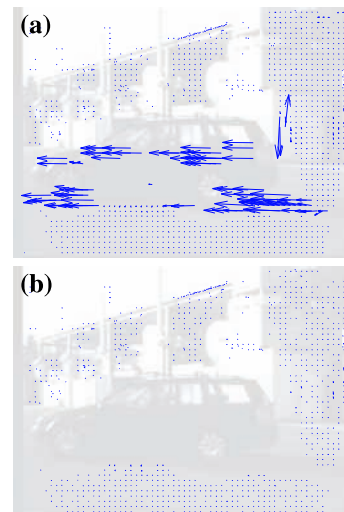


Fig. 8. (a) Flow vectors that belong to observations that have been part of a subset. (b) Validated flow vectors representing ego-motion of the ego-car.

The sensor supported random sampling stage is also well described by fig. 9. It shows the distribution of the calculated subset ego-motions (v_{xss}, v_{yss}) . Ego-motions (v_{xss}, v_{yss}) from valid subsets are marked with black triangles. Figure 8

also depicts the used threshold $\Delta v = 0.2$ meters per second.

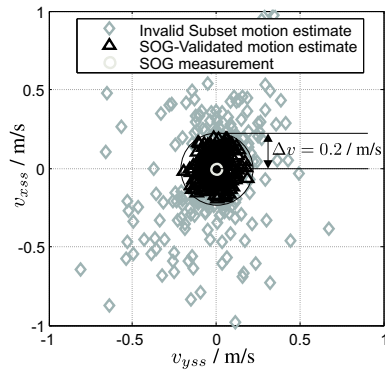


Fig. 9. Distribution of the calculated subset ego-motions ($v_{x,ss}, v_{y,ss}$). All observations in a subset that yield an estimate "close" to the SOG-estimate are classified as valid. What "close" means is determined by the threshold Δv .

C. Straight on driving ego-vehicle with a vehicle in front performing slalom

As a second maneuver we drove straight on with the ego-vehicle. While the ego-vehicle accelerates to approximately 2.8 m/s and then decelerates, the vehicle in front performs a slalom. Figure 10 shows an image of the sequence at a time instant $t_0 \approx 32$ s and the corresponding disparities.

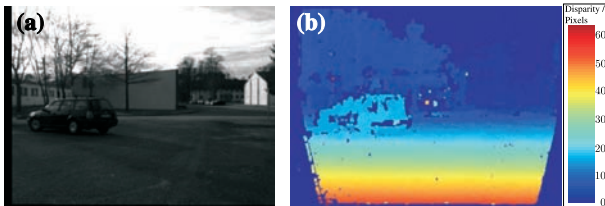


Fig. 10. (a) Image of the analyzed sequence at $t_0 \approx 32$ s. (b) Disparity calculated from stereo images.

Figure 11(a) shows the flow vectors of the observations before and (b) after sensor supported subset validation.

By validating the observations in a subset with the measurements from the SOG-sensor only valid observations remain (fig. 11(b)). Observations that belong to the moving car are discarded while the ego-vehicle is also moving.

Figure 12 shows the measurements of the SOG-sensor during the maneuver. The time instant $t_0 \approx 32$ s for the images shown in figs. 10 and 11 is marked with a dashed line. There are some systematic errors in the SOG measurements, since we did not use a telecentric lens [11].

At $t_0 \approx 32$ s the measurements from the SOG-sensor are taken into consideration for subset validation, which is shown in fig. 13. Again the threshold Δv was 0.2 m/s as indicated

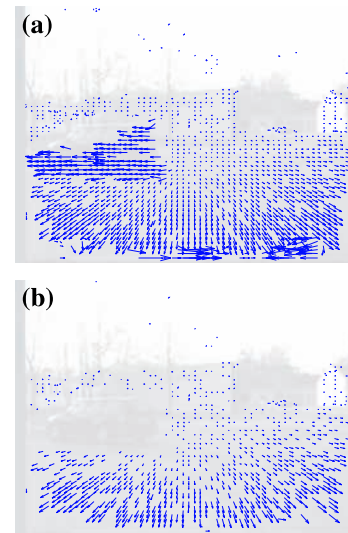


Fig. 11. (a) Flow vectors that belong to observations that have been part of a subset. (b) Validated flow vectors representing ego-motion of the ego-car.

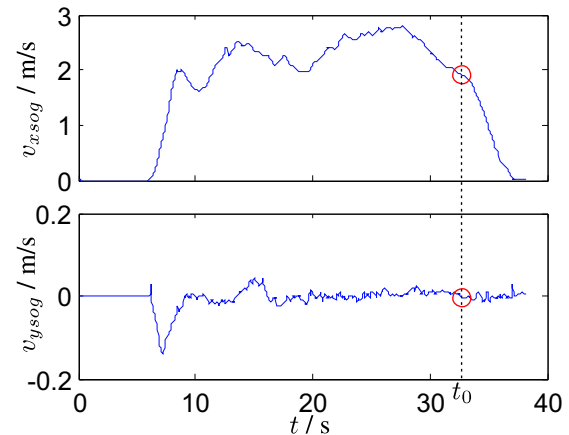


Fig. 12. Measurements $v_{x,sog}$ and $v_{y,sog}$ of the SOG-sensor during the maneuver. The circles indicate the SOG measurements for $t_0 = 32$ s.

in fig. 13.

Figure 14 shows the results of stereo vision based estimation of the ego-motion components v_{xs} , v_{ys} and ω_{zs} by the scheme described in sec. IV and the number of valid observations over time. The outliers in the velocity components coincide with frames where the number of valid subsets is very small. This is the case here because the SOG-sensor provides measurements with very high systematic errors at these points of time.

From fig. 14 we see, that ego-motion is correctly measured like preset by the maneuver although there was another moving car in the scene. There are some outliers in the measurements but they can be detected by analyzing the number of valid observations (fig. 14(d)). During the maneuver and especially during the weak transversal motion at $t \approx 8$ s the estimated yaw rate ω_{zs} is nearly proportional to v_{ys} here, since the maneuver we performed was not

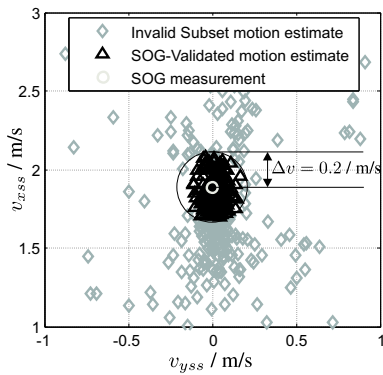


Fig. 13. Distribution of the calculated subset ego-motions $(v_{x_{ss}}, v_{y_{ss}})$.

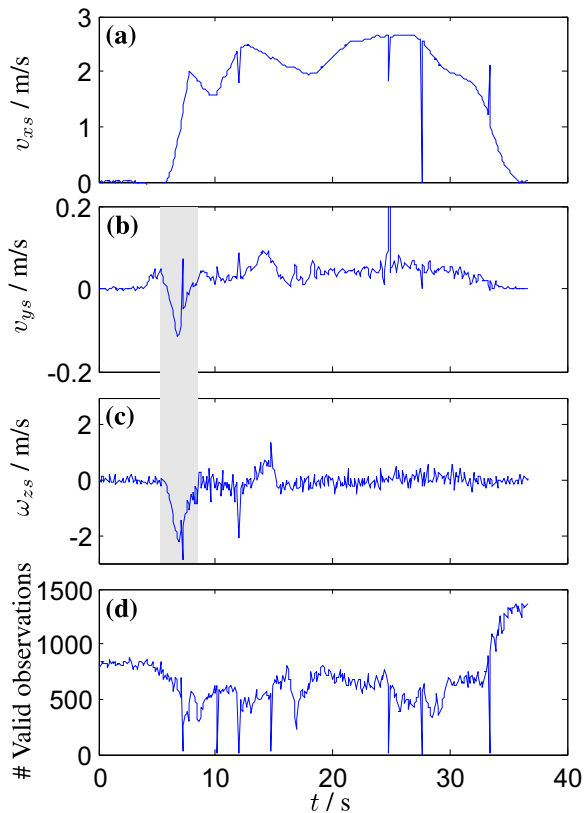


Fig. 14. Results of stereo vision based estimation of the ego-motion. (a) Longitudinal velocity $v_{x_{ss}}$. (b) Transversal velocity $v_{y_{ss}}$. (c) Yaw rate $\omega_{z_{ss}}$. (d) Number of valid observations. The gray box highlights a weak transversal motion.

highly dynamic and therefore the sideslip angle in the center of gravity was approximately zero.

In summary, we can conclude that the proposed validation scheme allows a robust ego-motion estimation and additionally a classification of observations u_i , v_i and Δ_i into an ego-motion caused and an alien motion caused cluster.

VI. SUMMARY AND CONCLUSION

We described an ego-motion estimation system that performs a fusion of the measurements from a stereo vision camera system with the measurements from a special speed over ground sensor. A random sampling technique was used to validate flow and disparity observations so that non-ego-motion caused observations are detected and discarded for further estimation of ego-motion.

The described subset validation scheme improves stereo vision based ego-motion estimation especially when there are many moving objects in the scene. In addition the technique is also capable of assessing if there are enough valid observations to estimate ego-motion reliably. The information gathered in the validation stage can also be used as a pre-segmentation of the observed scene. It is possible to distinguish regions from the rigid world and moving objects like cars or pedestrians. This can for example be used in path planning stages: The planned path should not intersect with the trajectory of a moving object.

Since – like for most optical sensors – bad weather is an issue, the SOG Sensor has been tested in various weather conditions. For example, even when the road surface is slightly wet, the SOG sensor still provides reliable estimates for the velocity vector \mathbf{v}_{sog} . In contrast, if there is a lot of water on the road with wind-caused waves, a malfunction is rather likely. But as the SOG sensor is self assessing [11] the malfunction can be detected and the corresponding measurements can be discarded. A malfunction could also be detected by analyzing the number of valid observations of the overall system (see. fig. 14(d)). Few valid observations are an indicator for a non valid measurement.

VII. FUTURE WORK

Since the focus of this paper was the sensor supported random sampling technique described in sec. IV, we estimated ego-motion at a certain time instant without using information from the past. Most of the motion parameters are not subject to sudden changes. Therefore, a tracking of the ego-motion parameters with a Kalman-Filter ([20]) can for example be applied to improve the overall system.

Flow and disparity estimation are self assessing. This means, that for each observation (u_i, v_i) and Δ_i we can calculate a quality measure and improve the estimation process by weighting an observation (u_i, v_i) and Δ_i by its quality measure. The SOG-Sensor is also self assessing and its quality measure can for example be applied to adapt the threshold Δv to the current reliability of the SOG-Sensor.

Moreover, a more theoretic analysis of the approach is necessary. Currently, parameters like the number S of subsets and the threshold Δv have only been determined heuristically or experimentally. In addition a more profound analysis of observability must be performed to study under

what circumstances a rotational motion can be reliably distinguished from a translational motion.

Further tests on real traffic scenes are also necessary. The results in sec. V only showed examples where the majority of observations arises from ego-motion. An analysis of sequences in very dense traffic is required to assess how the proposed algorithm performs when the majority of observations is caused by other moving objects and only a minority of the observations is valid.

REFERENCES

- [1] G. P. Stein, O. Mano, and A. Shashua, "A robust method for computing vehicle ego-motion," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2000.
- [2] U. Franke, and S. Heinrich, "Fast Obstacle Detection for Urban Traffic Situations," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 3, pp. 173–181, 2002.
- [3] T. Dang, C. Hoffmann, and C. Stiller, "Chapter: Visuelle mobile Wahrnehmung durch Fusion von Disparitaet und Verschiebung," in *Fahrerassistenzsysteme*, M. Maurer and C. Stiller, Eds. Springer, 2004, pp. 21–42.
- [4] J. Horn, A. Bachmann, and T. Dang, "A fusion approach for image-based measurement of speed over ground," in *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006.
- [5] H. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. Royal Society London*, vol. 208, pp. 385–397, 1980.
- [6] J. Horn, "Bildbasierte vektorielle Geschwindigkeitsmessung texturierter Oberflaechen," *tm - Technisches Messen*, vol. 72, no. 10, pp. 556–565, 2005.
- [7] A. Zomotor, "Ein korrelationsoptisches Verfahren zur direkten Messung von instationaeren Schwimm- und Schraeglaufwinkeln an Kraftfahrzeugen," *ATZ - Automobiltechnische Zeitschrift*, vol. 77, no. 7/8, pp. 213–218, 1975.
- [8] *Correxit®-SL Non-contact optical sensor for slip free measurement of longitudinal and transversal dynamics*, Corrsys-Datron Sensorsysteme GmbH, 2001. [Online]. Available: www.corrsys-datron.com
- [9] J. Horn, "Analysis of impulse train illuminated images for 2d velocity measurement," in *Proceedings of IEEE International Conference on Image Processing*, 2004.
- [10] H. Haussecker and H. Spies, "Chapter: Motion," in *Computer Vision and Applications*, B. Jaehne and H. Haussecker, Eds. Academic Press, 2000, pp. 309–395.
- [11] J. Horn, "Zweidimensionale Geschwindigkeitsmessung texturierter Oberflaechen mit flaechenhaften bildgebenden Sensoren," Ph.D. thesis, University of Karlsruhe, 2006.
- [12] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *DARPA Image Understanding Workshop*, 1981, pp. 121–130.
- [13] J. Bouget, "Pyramid implementation of the Lucas Kanade feature tracker," Tech. Rep. included in the OpenCV library, Tech. Rep., 2002, <http://www.intel.com/research/mrl/research/ovencv/>.
- [14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [15] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [16] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.
- [17] D. Forsyth and J. Ponce, *Computer vision: A modern approach*. Prentice Hall, 2003.
- [18] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.
- [19] A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," in *Communications of the ACM*, vol. 24, no. 6, 1981.
- [20] G. Welch and G. Bishop, "An introduction to the Kalman filter," in *SIGGRAPH*, ser. Course 8, 2001.