Vision-Based Online-Calibration of Inertial Gaze Stabilization

Alois Unterholzner, Michael Rohland, Michael Schweitzer and Hans-Joachim Wuensche

Abstract— **Active gaze stabilization is of vital importance for the use of high resolution tele-cameras in autonomous vehicles. Small aperture angles together with large focal lengths cause high sensitivity to rotational vehicle motion induced e.g. by bumps or braking. Due to large latencies in image processing only gaze stabilization based on inertial sensors is fast enough to ensure stable images. As this is a feed-forward control, imperfections of the sensor or the stabilizing actuator may result in undesireable image motion. To further enhance image stabilization we propose a novel vision-based online-calibration of the inertial angular rate sensor of our camera platform. Thus we are able to incorporate visual feedback into the gaze stabilization, while keeping the high bandwith of the inertial sensor.**

I. Introduction

The use of high resolution tele-cameras enhances the perceptional capabilities of autonomous vehicles. It enables both the early detection of distant objects and provides detailed images of points of interest in the vehicles surrounding. But tele-cameras are more sensitive to disturbances – especially rotational vehicle motion – than wide angle cameras.

To obtain stable images and keep objects of interest within camera scope active gaze stabilization is necessary. In [1] and [2] inertial measurements of the rotational vehicle motion are used for active image stabilization in autonomous vehicles.

This is a pure feed-forward control, since image-based information is not used for the stabilization. Residual image motion may be induced by errors of the inertial angular rate sensor, such as sensor offset, wrong or changing sensor sensitivity, errors in the mounting position of the sensor or imperfections in the control loop of the stabilizing actuator.

Therefore in more recent works visual information is incorporated in the stabilization process. Either information from inertial sensors and vision is fused as e.g. in [3], or the inertial sensors are calibrated based on visual information [4]-[6]. The latter approach is similar to the Vestibulo-Ocular Reflex (VOR), that many vertebrates use for gaze stabilization. The VOR not only stabilizes the image on the retina, but also continuously adapts to erroneous measurements of the vestibular organ located in the inner ear [6]. This adaptation process is based on visual feedback.

In [4]-[6] sensor and actuator errors are compensated by means of an adaptive FIR (finite impulse response) filter as suggested in [7]. The filter adaptation is driven by optical flow. The camera platform in [4] executes only rotational motions in an indoor environment. Thus optical flow results solely from rotational motion. In [5] and [6] the camera platform is mounted behind the windshield of a car. Optical flow computation is based on distant features. This way translational platform motion has little impact on the resulting optical flow. But calculating the optical flow by utilizing distant features only is not always possible, consider e.g. a car driving on a forest track or in an innercity environment.

The active inertial gaze stabilization, that we propose, is inspired by the Vestibulo-Ocular Reflex, too. In contrast to the aforementioned works however we model sensor and actuator errors explicitly. From those error models we derive a filter transfer-function suitable to modify the sensor signal in a way, that the overall stabilization error is compensated. The adaptation of the unknown filter parameters is driven by the residual camera motion.

The estimation of the unknown residual camera motion is based on image data. This is commonly refered to as *visual odometry*. Recently the estimation of non-holonomic vehicle motion in real time was presented in [8] and [9]. An overview on the relevant literature is also given therein. *Visual odometry* relies on the calculation of image correspondences between consecutive image frames. A method to calculate these correspondences in real time is given in [10].

Most works on *visual odometry* aim at estimating camera motion from visual data only. We instead are rather interested in the estimated camera motion than in *visual odometry* itself. To enhance robustness we incorporate the vehicle velocity and the steering angle in the estimation process. These measurements are available on most ground moving vehicles.

One fundamental problem in *visual odometry* is the determination of the depth of image keypoints. To avoid this, we assume that the drivable area in front of the vehicle is an ideal ground plane and search only there for image correspondences. Then the depth of the image keypoints is determined implicitly by the camera position relative to the ground plane.

Please note that the gaze stabilization is independent of the camera motion estimation. Only the calibration of the angular rate sensor is based on the estimated camera motion. But the calibration can be switched off, whenever the estimated camera motion seems to be unreliable. This may be the case e.g. if too little image correspondences are available or if the ground plane assumption is invalid.

The outline of the paper is as follows: In section II we give an overview of the proposed online-calibration process. Sections III and IV detail the two major components of our approach, namely the adaptive filter, that compensates sensor and actuator errors, and the vision-based camera motion

All authors are with the Institute for Autonomous Systems Technology (TAS), University of the Bundeswehr Munich, 85577 Neubiberg, Germany. Contact author email: alois.unterholzner@unibw.de

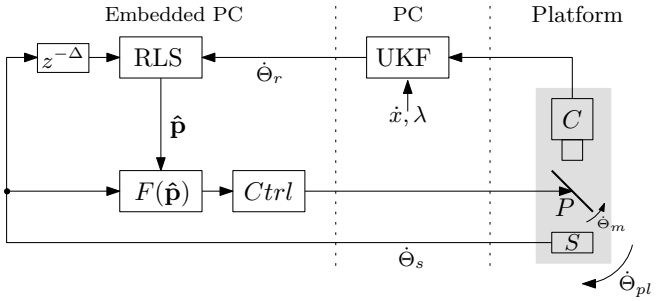


Fig. 1. Block diagram of the inertial gaze stabilization and its calibration

estimation. Experimental results derived during test drives on a bumpy dirt road are given in section V. The paper concludes with section VI.

II. Calibrating the Inertial Gaze Stabilization

The proposed gaze stabilization comprises the open loop gaze stabilization based on an MEMS angular rate sensor and the calibration of this sensor driven by visual feedback (Fig. 1).

The angular rate sensor S measures the pitch velocity $\dot{\Theta}_{pl}$ of the camera platform. After passing the filter $F(\hat{\mathbf{p}})$ the sensor signal $\dot{\Theta}_s$ is the reference value for the control ($Ctrl$) of the gaze stabilizing plant P. A mirror together with an actuator is used for gaze stabilization. In fact the reference value for the control of P is half the size of the filtered sensor signal, since for a mirror the incident angle equals the emergent angle.

Image correspondences computed from already stabilized images from camera C together with the vehicle velocity $\dot{x}$ and the vehicle steering angle λ serve as measurements for an Unscented Kalman Filter (UKF) [11]. The UKF is used to estimate the residual pitch velocity $\dot{\Theta}_r$, which is the difference between the real platform velocity $\dot{\Theta}_{pl}$ and two times the mirror velocity $\dot{\Theta}_m$. $\dot{\Theta}_r$ again is used as an error signal for a Recursive Least Squares (RLS) parameter estimator, that estimates the parameters $\hat{\mathbf{p}}$ of the filter $F(\hat{\mathbf{p}})$ in a way, that the residual image motion is minimized. To consider the latencies Δ resulting from image processing in the parameter estimation process, the sensor signal $\dot{\Theta}_s$ is artificially delayed by Δ before entering the RLS parameter estimator.

III. Compensating the Open Loop Errors

To derive an adaptive filter that compensates the open loop errors of the inertial gaze stabilization, we model the sensor and actuator characteristics.

The MEMS angular rate sensor signal

$$\dot{\Theta}_s = k_s \dot{\Theta}_{pl} + \delta_s \tag{1}$$

is considered to be biased resulting from an offset δ_s. Moreover the sensor is thought to have a sensitivity k_s differing from the nominal sensitivity. Please note that the sensitivity k_s may also be influenced by the mounting position of the sensor, e.g. when the sensor is not exactly aligned with the axis of rotation of the mirror.

The stabilizing mirror and its control are modelled as a first-order transfer-function in the Laplace domain

$$\dot{\Theta}_m = \frac{0.5 k_a}{\tau_a s + 1} \dot{\Theta}_s \tag{2}$$

with the time constant τ_a describing the bandwith of the control loop and the gain k_a. The gain k_a describes the steady state deviation of the mirror velocity $\dot{\Theta}_m$ from half of the measured sensor value $\dot{\Theta}_s$.

The goal of the adaptive filter design is to reconstruct the real platform pitch velocity $\dot{\Theta}_{pl}$ from the sensor signal $\dot{\Theta}_s$ and to compensate the influence of the actuator control loop on the gaze stabilization. We choose the filter equation as follows:

$$\dot{\Theta}_F = F(\hat{\mathbf{p}}) = \frac{\hat{\tau}_a s + 1}{\hat{k}_a \hat{k}_s} (\dot{\Theta}_s - \hat{\delta}_s) \tag{3}$$

Using the output of filter (3) as input to the mirror control loop (2) and inserting (1) in (3) we derive

$$\dot{\Theta}_m = \frac{0.5 k_a}{\tau_a s + 1} \frac{\hat{\tau}_a s + 1}{\hat{k}_a \hat{k}_s} (k_s \dot{\Theta}_{pl} + \delta_s - \hat{\delta}_s) \tag{4}$$

If the estimated parameters $\hat{\mathbf{p}} = [\hat{\tau}_a,\ \hat{k}_a,\ \hat{k}_s,\ \hat{\delta}_s]^T$ approach their true values $\mathbf{p} = [\tau_a,\ k_a,\ k_s,\ \delta_s]^T$ equation (4) reduces to

$$\dot{\Theta}_m = 0.5 \dot{\Theta}_{pl} \tag{5}$$

which is the desired optimal gaze stabilization.

To get estimates $\hat{\mathbf{p}}$ of the unknown parameters $\mathbf{p}$ with a discrete time RLS parameter estimator (Fig. 1), the filter equation (3) has to be discretized. Since the sampling interval T is small compared to the vehicle's pitch dynamics, we can use backward differences $\dot{x}(k) = \frac{x(k) - x(k-1)}{T}$ for discretization:

$$\dot{\Theta}_F(k) = \frac{\hat{\tau}_a + T}{\hat{k}_a \hat{k}_s T} \dot{\Theta}_s(k) - \frac{\hat{\tau}_a}{\hat{k}_a \hat{k}_s T} \dot{\Theta}_s(k-1) - \frac{\hat{\delta}_s}{\hat{k}_a \hat{k}_s} \tag{6}$$

The resulting discretized filter equation (6) is linear in its parameters and can be written as a linear regression

$$\begin{aligned} \dot{\Theta}_F(k) &= \left[\dot{\Theta}_s(k),\ \dot{\Theta}_s(k-1),\ 1\right]^T \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{d} \end{bmatrix} \\ &=: \phi^T(k-1)\hat{\theta} \end{aligned} \tag{7}$$

which is suitable for RLS parameter estimation [12].

Usually the RLS parameter update is driven by the difference of a measured signal y and a prediction $\hat{y}$ of y based on the estimated parameters $\hat{\theta}$. The difference $y - \hat{y}$ is a measure for the RLS parameter estimation error. The residual camera pitch velocity $\dot{\Theta}_r$ can also be considered to be a measure for the parameter estimation error. Unless $\dot{\Theta}_r = 0$ the estimated parameters $\hat{\mathbf{p}}$ of filter $F(\hat{\mathbf{p}})$ are not equal to their true values $\mathbf{p}$. Therefore we use

$$\hat{\theta}(k+1) = \hat{\theta}(k) + K(k)\dot{\Theta}_r \tag{8}$$

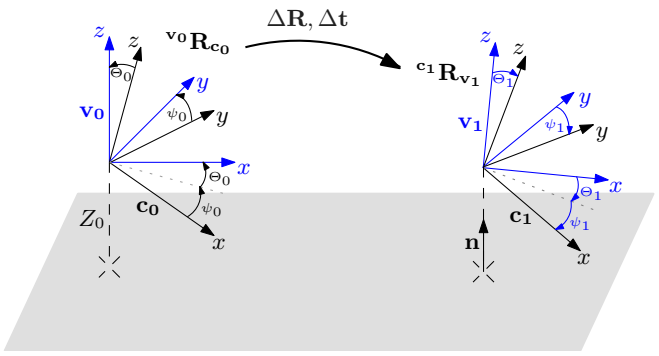


Fig. 2. Camera motion model

as parameter update equation, with $K(k)$ being the parameter update gain of the RLS parameter estimator.

Simulations indicate that the proposed filter $F(\hat{\mathbf{p}})$ together with the parameter estimator is adequate for compensating the modelled sensor errors.

IV. Camera Motion Estimation

The camera motion is estimated within an UKF-framework that comprises measurement and process model as well as a preselection of valid image correspondences. As a tribute to computation speed, the calculated image correspondences include some *false positives* [10]. Thus the preselection of valid correspondences is necessary; otherwise the estimation performance may be corrupted. The following subsections present the UKF measurement and process model together with a method for selecting valid image correspondences.

A. Measurement Model

The measurement model incorporates a mapping of a homogeneous image point $\tilde{\mathbf{m}}_\mathbf{0}$ to its corresponding homogeneous image point $\tilde{\mathbf{m}}_\mathbf{1}$ in a subsequent frame (cf. [13]).

A 3D-point $\mathbf{M}_0 = (X_0,\ Y_0,\ Z_0)^T$ undergoes the coordinate transformation

$$\mathbf{M_1} = \mathbf{R}\mathbf{M_0} + \mathbf{t} \tag{9}$$

if the coordinate system $\mathbf{c_0}$ is rotated and translated to the new coordinate system $\mathbf{c_1}$.

Combining the coordinate transformation (9) with the mapping

$$X\tilde{\mathbf{m}} = \mathbf{A}\mathbf{M} \Longleftrightarrow \mathbf{M} = X\mathbf{A^{-1}}\tilde{\mathbf{m}} \tag{10}$$

of a 3D-point $\mathbf{M}$ to a homogenous 2D-image point $\tilde{\mathbf{m}} = (1,\ \tilde{y}/X,\ \tilde{z}/X)^T$ with depth $X = \tilde{x}$ we derive:

$$\tilde{\mathbf{m}}_\mathbf{1} = \mathbf{A}\mathbf{R}\mathbf{A^{-1}}\tilde{\mathbf{m}}_\mathbf{0} + X_0^{-1}\mathbf{A}\mathbf{t} \tag{11}$$

This equation can be used to compute the homogenous 2D-point $\tilde{\mathbf{m}}_\mathbf{1} = (\tilde{x}_1,\ \tilde{y}_1,\ \tilde{z}_1)^T$ from a given image point $\tilde{\mathbf{m}}_\mathbf{0} = (1,\ \tilde{y}_0/\tilde{x}_0,\ \tilde{z}_0/\tilde{x}_0)^T$, in case of a known camera motion $[\mathbf{R}|\mathbf{t}]$ from $\mathbf{c_0}$ to $\mathbf{c_1}$. $\mathbf{A} = \mathbf{diag}(1,\ \alpha,\ \alpha)$ is the camera calibration matrix, if the camera distortion is neglected. α represents the focal length of the camera in terms of (quadratic) pixel dimensions.

In general the depth X_0 of the image point $\tilde{\mathbf{m}}_\mathbf{0}$ in equation (11) is unknown. For this reason we use point correspondences from the drivable area in front of the vehicle only. Then we can assume, that all the 3D-points $\mathbf{M}$ corresponding to image points $\tilde{\mathbf{m}}$ lie on a ground plane. If the camera coordinate system $\mathbf{c_0}$ has the distance Z_0 to this plane, the *Hessian Normal Form* of this ground plane with respect to $\mathbf{c_0}$ is

$$\mathbf{n^T}\mathbf{M} - Z_0 = 0 \Longleftrightarrow Z_0^{-1}\mathbf{n^T}\mathbf{M} = 1 \tag{12}$$

Inserting (10) in (12) results in combination with (11) in a mapping

$$\tilde{\mathbf{m}}_\mathbf{1} = \mathbf{A}(\mathbf{R} + Z_0^{-1}\mathbf{t}\,\mathbf{n^T})\mathbf{A^{-1}}\tilde{\mathbf{m}}_\mathbf{0} \tag{13}$$

from the image point $\tilde{\mathbf{m}}_\mathbf{0}$ to the homogenous point $\tilde{\mathbf{m}}_\mathbf{1}$, that is independent of the unknown depth X_0.

The camera motion from $\mathbf{c_0}$ to $\mathbf{c_1}$ comprises a composition of rotations and translations (Fig. 2).

At first the camera coordinate system undergoes a rotation ${}^{\mathbf{v_0}}\mathbf{R}_{\mathbf{c_0}}$ from coordinate system $\mathbf{c_0}$ to a coordinate system $\mathbf{v_0}$. The distance of $\mathbf{v_0}$ to the ground plane is Z_0 and the axes of $\mathbf{v_0}$ are aligned to the axes of the vehicle coordinate system. The rotation angles are the negative values of the camera yaw- and pitch-angle. This coordinate transformation corresponds to the image frame 0.

Secondly the incremental vehicle motion between two consecutive image frames 0 and 1 is incorporated in the calculation of the camera motion with the rotation $\Delta\mathbf{R}$ and the translation $\Delta\mathbf{t}$. The rotation $\Delta\mathbf{R}$ determines the incremental yaw- and pitch-movement of the vehicle. $\Delta\mathbf{t}$ is the incremental vehicle translation along its x-axis. The resulting coordinate system is $\mathbf{v_1}$. Analog to $\mathbf{v_0}$ the axes of $\mathbf{v_1}$ are aligned to the axes of the vehicle coordinate system that corresponds to frame 1.

At last ${}^{\mathbf{c_1}}\mathbf{R}_{\mathbf{v_1}}$ rotates the coordinate system $\mathbf{v_1}$ towards the camera coordinate system $\mathbf{c_1}$. The rotation angles are the camera yaw- and pitch-angle corresponding to frame 1.

Inserting ${}^{\mathbf{v_0}}\mathbf{R}_{\mathbf{c_0}}$, $\Delta\mathbf{R}$, $\Delta\mathbf{t}$ and ${}^{\mathbf{c_1}}\mathbf{R}_{\mathbf{v_1}}$ in (13) we finally derive the measurement equation for the image correspondences

$$\tilde{\mathbf{m}}_\mathbf{1} = \mathbf{A}\,{}^{\mathbf{c_1}}\mathbf{R}_{\mathbf{v_1}}\,(\Delta\mathbf{R} + Z_0^{-1}\Delta\mathbf{t}\,\mathbf{n^T})\,{}^{\mathbf{v_0}}\mathbf{R}_{\mathbf{c_0}}\,\mathbf{A^{-1}}\tilde{\mathbf{m}}_\mathbf{0} \tag{14}$$

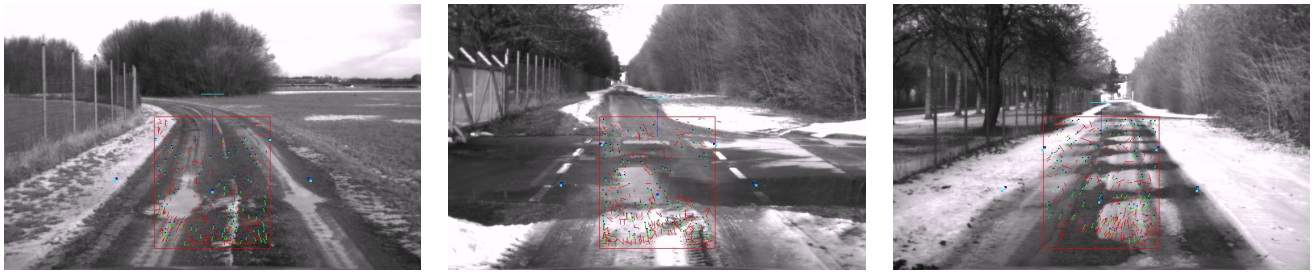
To increase the robustness of the camera motion estimation we include vehicle velocity $\dot{x}$ and steering angle λ in the measurement model. The vehicle velocity is directly used as measured. The steering angle λ is used to calculate the vehicle's incremental yaw angle

$$\Delta\psi = \lambda T_f \dot{x}/l_w \tag{15}$$

according to the so called Ackermann steering-model with vehicle wheelbase l_w and frame time T_f [1]. The equations (14) and (15) together with $\dot{x}$ build the measurement model. The resulting measurement vector is:

$$\mathbf{y} = (\dot{x},\ \Delta\psi,\ \tilde{\mathbf{m}}_\mathbf{1}(1)\ \ldots\ \tilde{\mathbf{m}}_\mathbf{1}(N))^T \tag{16}$$

N is the maximum number of image correspondences that can be incorporated in the estimation process.



(a) Pothole (b) Dip (c) Washboard

Fig. 3. Selected images from the test track

B. Process Model

The process model describes the motion and the position of the coordinate system $\mathbf{v}$ (Fig. 2) relative to the ground plane. The state vector

$$\mathbf{x} = (\Theta,\ \dot{\Theta},\ \Delta x,\ \dot{x},\ \Delta\psi,\ z)^T \tag{17}$$

represents the system state of $\mathbf{v}$. The states Θ and $\dot{\Theta}$ are the pitch-angle and the pitch-velocity of $\mathbf{v}$ with respect to the ground plane. The state Δx is the incremental translation of $\mathbf{v}$ along its x-axis. $\dot{x}$ is the velocity in x-axis direction of $\mathbf{v}$. This is also the vehicle velocity, since the axes of $\mathbf{v}$ and the vehicle coordinate system are aligned. $\Delta\psi$ is the incremental yaw rotation of $\mathbf{v}$ and z is the distance of $\mathbf{v}$ to the ground plane ($z = Z_0$).

Since we are mainly interested in the pitch-angle and pitch velocity of $\mathbf{v}$ it is sufficient to estimate only the incremental changes of x and ψ. We know from simulations, that the roll-angle of $\mathbf{v}$ and the translation of $\mathbf{v}$ along its y-axis can be neglected. Therefore the resulting process model is

$$\begin{aligned}
\Theta_{k+1} &= \Theta_k + T_f\dot{\Theta}_k \\
\dot{\Theta}_{k+1} &= \dot{\Theta}_k \\
\Delta x_{k+1} &= T_f\dot{x}_k \\
\dot{x}_{k+1} &= \dot{x}_k \\
\Delta\psi_{k+1} &= \Delta\psi_k \\
z_{k+1} &= z_k
\end{aligned} \tag{18}$$

In our experiments we used a constant acceleration model and the harmonic oscillator model from [14] alternatively as process model for the pitch-angle Θ. However, these models had almost no influence on the estimation results and therefore we stick to the constant velocity model for Θ.

Finally, after having received an estimate of $\mathbf{x}$, the residual camera pitch velocity $\dot{\Theta}_r$ can be derived from the pitch component of the rotation matrix ${}^{\mathbf{c_1}}\mathbf{R}_{\mathbf{c_0}} = {}^{\mathbf{c_1}}\mathbf{R}_{\mathbf{v_1}}\ \boldsymbol{\Delta}\mathbf{R}\ {}^{\mathbf{v_0}}\mathbf{R}_{\mathbf{c_0}}$ divided by the frame time T_f.

C. Selection of Correspondences

For the selection of valid correspondences we assume, that the camera translation between two consecutive image frames is known exactly. As a consequence the uncertainty in the prediction of the keypoints $\tilde{\mathbf{m}}_\mathbf{1}$ results from the uncertainty in the camera rotation only. If we neglect the roll component, the uncertainties $\delta\Theta$ and $\delta\psi$ in the pitch and yaw component remain. Applying the small angle approximation the resulting pixel uncertainties δz and δy in the image line and column are

$$\delta z = \alpha \cdot \delta\Theta \quad \text{and} \quad \delta y = \alpha \cdot \delta\psi \tag{19}$$

Only those image keypoints $\tilde{\mathbf{m}}_\mathbf{1}$ are selcted as measurements, whose coordinates $z_1 = \tilde{z}_1/\tilde{x}_1$ and $y_1 = \tilde{y}_1/\tilde{x}_1$ satisfy

$$\left(\frac{z_1}{\delta z}\right)^2 + \left(\frac{y_1}{\delta y}\right)^2 \leq 1 \tag{20}$$

Hardly anything is known in advance about external disturbances causing undesired pitch rotation of the camera platform. On the contrary the yaw rotation is less uncertain, due to the measured steering angle λ. Therefore $\delta\psi$ can be chosen considerably smaller than $\delta\Theta$.

V. Experimental Results

The proposed gaze stabilization was evaluated during several test runs on a bumpy dirt road. To achieve comparable results we repeatedly drove along one section of this road. Figure 3 gives some impressions of the test track. A video is available on YouTube[1].

The red rectangles in the Figures 3(a) - 3(c) mark the Region of Interest (ROI) of the SidCell algorithm. Feature correspondences are only computed within this region. According to our ground plane assumption the ROI corresponds to the drivable area in front of the vehicle. The green lines in the Figures 3(a) - 3(c) mark correspondences which satisfy equation (20). Therefore they were accepted as measurements for the camera motion estimation. The red lines depict correspondences which were rejected as measurements since they failed to satisfy (20). Please note that large pixel uncertainties δz are required to detect sudden fast vehicle pitch rotations. However, large pixel uncertainties increase the propability of selecting invalid correspondences.

The calibration of the angular rate sensor S is driven by the residual camera motion $\dot{\Theta}_r$, which is computed from the estimated vehicle pitch velocity $\dot{\Theta}$. Therefore we compared – in a preliminary step – the estimated vehicle pitch velocity $\dot{\Theta}$ with the vehicle pitch velocity $\dot{\Theta}_{INS}$, measured with a high precision Inertial Navigation System (INS). Since

[1] http://www.youtube.com/watch?v=xl2I6h0RGZc

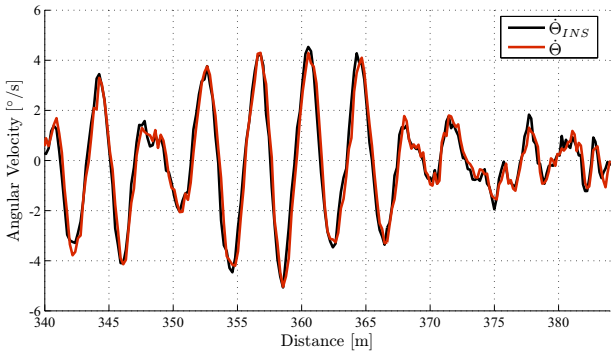


Fig. 4. Measured and estimated vehicle pitch velocity ($\dot{\Theta}_{INS}$ and $\dot{\Theta}$)

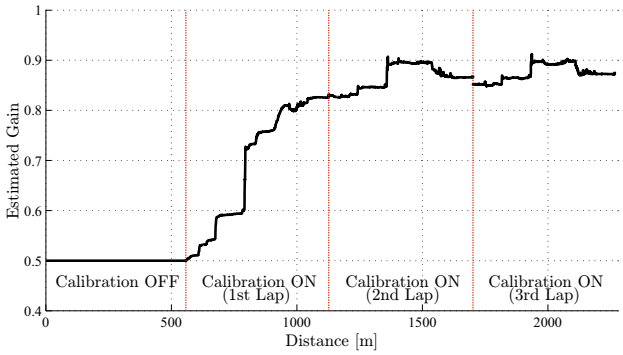


Fig. 5. Estimated gain of the filter $F(\hat{\mathbf{p}})$

the measured and the estimated pitch velocity are in good accordance with one another (Fig. 4), the residual camera pitch velocity $\dot{\Theta}_r$ can be used as an error signal for the RLS parameter estimator. The plot in Figure 4 corresponds to the so called "washboard" in our test track (Fig. 3(c)).

Although the estimated vehicle pitch velocity fits very well to the INS measured pitch velocity, it is not completely unbiased. To avoid a drifting estimated vehicle pitch angle, a decay factor $c < 1$ was introduced in the process model. Hence the first equation in (18) changed to

$$\Theta_{k+1} = c\,\Theta_k + T_f \dot{\Theta}_k$$

As a result Θ gradually tends to zero. This is a feasible modification, as the mean vehicle pitch angle relative to the ground plane is zero.

However, the decay factor does not eliminate a potential offset in the estimated vehicle pitch velocity $\dot{\Theta}$. Consequently the residual camera pitch velocity $\dot{\Theta}_r$ is also biased. Therefore we use a high-pass filter to derive an unbiased error signal for the parameter estimation. Incorporating the long term correspondences provided by the SidCell algorithm [10] in the estimation process may improve estimation results with respect to slow vehicle pitch motions. Obviously high-pass filtering of the error signal $\dot{\Theta}_r$ prevents the estimation of the sensor offset δ_s. This implies that the sensor signal $\dot{\Theta}_s$ also has to be high-pass filtered. Otherwise the mirror would drift away.

To test the proposed vision-based online-calibration of the MEMS angular rate sensor, we drove along the test track three times with active sensor calibration and active gaze stabilization. Figure 5 depicts the estimated gain $\hat{k}_f = \frac{1}{\hat{k}_a \hat{k}_s}$ of the filter $F(\hat{\mathbf{p}})$. For evaluation purposes the filter gain was artificially reduced from 1 to 0.5. During parameter estimation the gain rose from $\hat{k}_f = 0.5$ to $\hat{k}_f \approx 0.88$, which is beneath the nominal gain 1.

The estimated time constant $\hat{\tau}_a$ remained close to its initial value zero. This indicates that the bandwidth of the actuator control loop is sufficiently large and no phase-delay occurs during stabilization. The offset δ_s was estimated to be zero too, due to the high-pass filtered error signal $\dot{\Theta}_r$

To further evaluate the proposed gaze stabilization, we compared the estimated residual camera motion $\dot{\Theta}_r$ for the following three runs on the test track: One with deactivated gaze stabilization, another one with activated stabilization but uncalibrated sensor ($\hat{k}_f = 1$) and finally the one with activated gaze stabilization and calibrated sensor ($\hat{k}_f \approx 0.88$). During this last run the calibration was active, but the parameters had already converged.

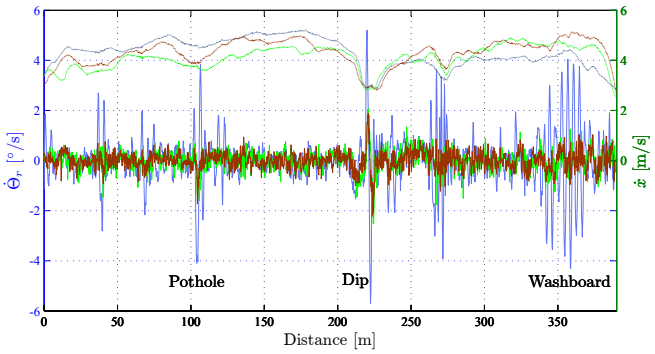
The stabilization results are plotted in Figure 6. Figure 6(a) gives an overview of the resulting estimated camera pitch velocities $\dot{\Theta}_r$ during the test runs. The labels "pothole", "dip" and "washboard" mark peaks in the pitch velocities $\dot{\Theta}_r$ that originate from passing the track sections depicted in the Figures 3(a) - 3(c). Figure 6(b) and Figure 6(c) give more detailed insights in the stabilization performance.

Driving through the dip (Fig. 3(b), 6(b)) not only induces a vehicle pitch rotation but also a changing vehicle height. However, the gaze stabilization can only compensate pitch rotations. Consequently a sudden change in the camera height results in a residual image motion. Since the estimated camera height z is considered to be slowly varying, the residual image motion due to a fast change in the camera height is likely to be interpreted as a pitch rotation by the camera motion estimation.

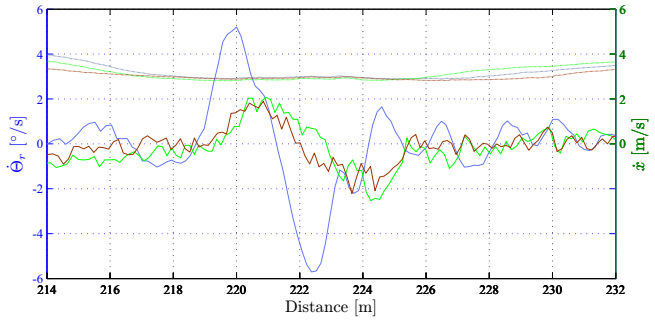
The gaze stabilization based on the uncalibrated MEMS sensor already showed good results. Especially for the washboard section 6(c) there is almost no difference between the stabilization results based on the calibrated or the uncalibrated angular rate sensor.

For better evaluation we calculated the rms-value of $\dot{\Theta}_r$ as a measure for the stabilization error (Fig. 7). The resulting rms-value was normalized with respect to the rms-value of $\dot{\Theta}_r$ in case of the unstabilized camera. The normalized rms-error for the unstabilized camera is then 100%. It turns out that the rms-error corresponding to the calibrated gaze stabilization is in any case smaller than the rms-error corresponding to the uncalibrated gaze stabilization.

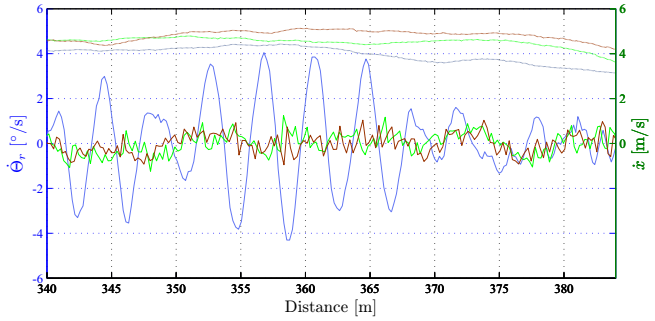
With respect to the whole test track (Fig. 7(a)), the rms-error for the calibrated stabilization is approximately 12% smaller than the rms-error resulting from the uncalibrated stabilization. This can be considered to be the average reduction of the stabilization error. It is remarkable, that the change in the filter gain $\hat{k}_f$ is similar to the reduction of the rms-error. Please note further, that it is the change of



(a) Overview



(b) Dip



(c) Washboard

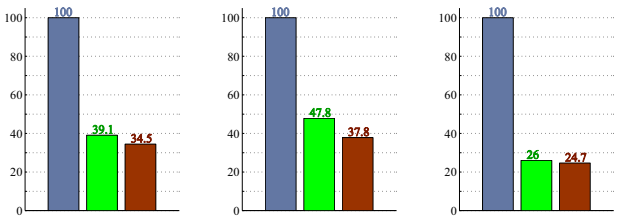
Fig. 6. Gaze stabilization results: The blue lines correspond to the deactivated stabilization, the green lines to the activated stabilization and the dark-red lines to the calibrated stabilization. Solid lines correspond to the residual camera motion $\dot{\Theta}_r$, dotted lines to the vehicle velocity $\dot{x}$

the gain which is of interest, since both an increasing and a decreasing gain can result in an error reduction.

The test runs were carried out at similar but not equal velocities (Fig. 6(a)). However, the velocity during the test run with calibrated stabilization (dark-red) was almost always higher than the velocity corresponding to the test run with active but uncalibrated stabilization (green). This fact rather supports than contradicts our results.

VI. Conclusion

Similar to the VOR, the gaze stabilization presented in this paper, continuously adapts to erroneous stabilization. The calibration of the inertial angular rate sensor and the compensation of actuator errors is based on visual feedback. An adaptive filter is used to compensate this errors. The



(a) Overview (b) Dip (c) Washboard

Fig. 7. The normalized rms error of the gaze stabilization (based on $\dot{\Theta}_r$). The blue bar corresponds to the deactivated stabilization, the green bar to the activated stabilization and the dark-red bar to the calibrated stabilization.

filter adaptation is driven by the estimated residual camera motion, which is based on the online computation of feature correspondences between consecutive image frames. Future work will incorporate long term feature correspondences in the estimation of the camera motion. This is expected to further improve the stabilization results.

References

[1] E. D. Dickmanns, *Dynamic Vision for Perception and Control of Motion*. Springer Verlag London Limited, 2007.

[2] J. Schiehlen, "Kameraplattformen für aktiv sehende Fahrzeuge," Dissertation, Universität der Bundeswehr München, 1995.

[3] J. Lobo, J. F. Ferreira, and J. Dias, "Robotic Implementation of Biological Bayesian Models Towards Visuo-Inertial Image Stabilization and Gaze Control," in *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, Bangkok, Thailand, February 2009.

[4] A. Lenz, T. Balakrishnan, A. G. Pipe, and C. Melhuish, "An Adaptive Gaze Stabilization Controller Inspired by the Vestibulo-Ocular Reflex," *Bioinspiration & Biomimetics*, vol. 3, 2008.

[5] W. Guenthner, S. Glasauer, P. Wagner, and H. Ulbrich, "Biomimetic Control for Adaptive Camera Stabilization in Driver-Assistance Systems," *Journal of Mechanical Science and Technology*, vol. 21, no. 6, pp. 930–934, June 2007.

[6] W. Guenthner, *Enhancing Cognitive Assistance Systems with Inertial Measurement Units*. Springer-Verlag Berlin Heidelberg, 2008.

[7] P. Dean, J. Porrill, and J. V. Stone, "Decorrelation Control by the Cerebellum Achieves Oculomotor Plant Compensation in Simulated Vestibulo-Ocular Reflex," *Proc. R. Soc. London: Biological Sciences*, vol. 269, no. 1503, pp. 1895–1904, August 2002.

[8] M. Schweitzer and H.-J. Wuensche, "Real-Time Visual Odometry for Ground Moving Robots using GPUs," in *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP 2010)*, Angers, France, May 2010.

[9] D. Scaramuzza, F. Fraundorfer, M. Pollefeys, and R. Siegwart, "Absolute Scale in Structure from Motion from a Single Vehicle Mounted Camera by Exploiting Nonholonomic Constraints," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2009)*, Kyoto, Japan, October 2009.

[10] M. Schweitzer and H.-J. Wuensche, "Efficient Keypoint Matching for Robot Vision using GPUs," in *Proceedings of the Fifth IEEE Workshop on Embedded Computer Vision at ICCV 2009*, Kyoto, Japan, October 2009.

[11] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 410–422, March 2004.

[12] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction And Control*, ser. Information and System Sciences, T. Kailath, Ed. Engelwood Cliffs, New Jersey 07632: Prentice-Hall, Inc, 1984.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry*, 2nd ed. Cambridge Univerity Press, 2006.

[14] R. Behringer and S. Hoetzl, "Simultaneous Estimation of Pitch Angle and Lane Width from the Video Image of a Marked Road," in *Proceedings of the IEEE International Conference on Robotics and Systems (IROS 1994)*, Munich, Germany, September 1994.