

Video-Based Roll Angle Estimation for Two-Wheeled Vehicles

Marc Schlipfing*, Jakob Schepanek*, and Jan Salmen*

Abstract— Video-based driver assistance systems are a key component for intelligent vehicles today. Applications for lane detection, traffic sign recognition, and collision avoidance have been successfully deployed in cars and trucks. State-of-the-art algorithms rely on machine learning and therefore depend on invariance conditions, e.g. a fixed image perspective. In order to apply current modules in two-wheeled vehicles one needs to determine the roll angle, i.e. the angle between the road plane and the slanted vehicle. It can either be used for parametrisation of the algorithms or for rotation of the video image back to a horizontal alignment. Using an inertial measurement unit to acquire this data is unreasonably expensive. We propose a video-based module that estimates the current roll angle based on gradient orientation histograms to overcome this flaw. Due to the visual structure of a traffic scene we are able to derive possible roll angles from the gradient statistics by correlation with learnt data. Analogously, we estimate the roll rate by correlating subsequent image statistics and stabilise both measures within a linear Kalman filter. Experiments on real image data from various test scenarios show high accuracy of the proposed approach. Thus, estimating the roll angle/rate from video only, enables us to employ established video-based assistance modules for two-wheeled vehicles without any additional hardware expense.

I. INTRODUCTION

In the last decades, more and more technical systems were used in mass-production vehicles in order to increase safety for all road users and comfort for the driver. Such systems make use of different types of sensors, e.g., mechanical, ultra sound, or infrared. All these sensors have typical restrictions introducing limitations for the following applications.

For current and future developments, video-based systems are the most popular and most promising approach. Cameras mounted on vehicles provide all information needed for safe, comfortable, and economic driving: Human drivers rely almost exclusively on visual information.

Based on camera data, very different applications can be realized, e.g., detection, classification, and tracking of relevant objects, estimation of optical flow, and stereo vision.

However, many systems and algorithms show limited robustness against varying external conditions. Especially when machine learning techniques are involved, methods rely on an either constant or known scene perspective. For object detection tasks with car-mounted systems the perspective changes are usually small enough to be neglected. The large interval of possible roll angles on two-wheeled vehicles enforces alternatives:

- Make the module invariant to rotations within the image plane. This will require at least a retraining or even a redesign of the module and will generally lead to a decrease in performance.
- Try several discrete rotation angles to pre-process the input. This will increase the runtime by a multiple and will nevertheless be of insufficient precision.

In a multi-module system those procedures would apply to each application.

To apply the same algorithms to a motorcycle-mounted system one will need to compensate for the roll angle by rotating the acquired images back to horizontal alignment. There are approaches to estimate vehicle states by integration of several sensor responses, i.e. velocity, acceleration, roll/pitch/yaw rate, and distance to ground plane [1], [2], [3], [4], [5], [6]. Their set-ups include inertial measurement units (IMU) in order to analyse vehicle dynamics or rely on stereo vision information [7]. But applying an IMU will – in the given scenario – cause unreasonable high costs and/or yet not deliver accurate results on motorcycles.

We propose the approach of estimating the roll angle/rate from gradient information of grey-value images only. Looking at road scenes one can find typical compositions and shapes that produce major gradients in the recorded images, e.g. horizon, vehicles, housing, lane borders and markings. Fig. 1 illustrates gradient orientations from a motorway scene. Their orientations result in a characteristic gradient angle histogram, which is similar for most images from a vehicle-mounted camera.

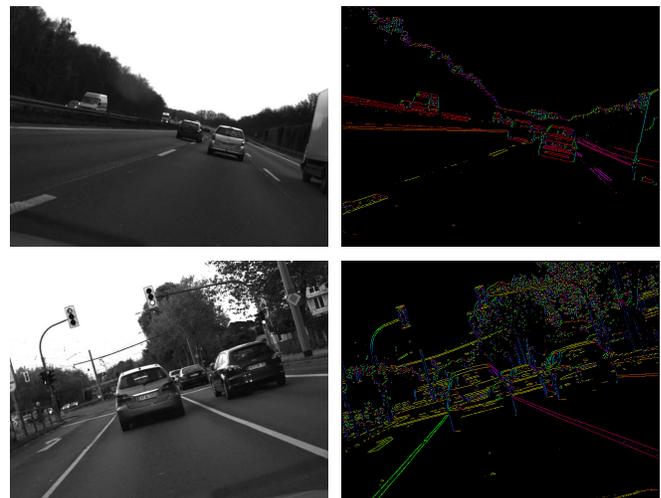


Fig. 1. Example images of different scenes and their colour-coded (thresholded) gradient orientations.

*Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany.

{marc.schlipfing, jakob.schepanek, jan.salmen}@ini.rub.de

Learning these statistics from numerous training images enables us to correlate the histogram of a single test image with several translations of the learnt statistics. The translations cover the interval of reasonable roll angles. The current estimate is the one maximising the correlation measure.

In a second step we independently estimate the roll rate, i.e. the roll angle's change per second, by maximising the correlation of orientation histograms from two subsequent images. Finally both angle and rate are fed into a linear Kalman filter for robust integration over time. The filter is initialised with physically reasonable dynamics and noise covariances.

We performed experiments on real-world video data covering different scenarios. The roll angle was simulated in order to allow for a systematic evaluation with ground-truth data. The experimental results substantiate robustness, satisfactory precision and real-time capability of the presented approach.

The upcoming section will introduce the types of histograms all estimation is based on. In Sec. III and IV we present the estimation process which is followed by a filtering step in Sec. V. Finally, experimental results (VI) and the conclusion (VII) of the proposed approach are stated.

II. ORIENTATION HISTOGRAMS

Image processing on grey-scale images is – due to missing colour information – mainly driven by the analysis of structure. Many object recognition or scene categorisation algorithms make use of gradients, their orientation and energy. Examples of popular image features are *histogram of oriented gradients* (HOG) [8] and *Haar-like* features [9]. They are able to describe local properties for objects or parts.

Our approach uses the idea of HOG-features in a global manner by building only a single histogram for the entire image. Realistic roll angles for two-wheeled vehicles lie in $[-35^\circ, 35^\circ]$ for normal traffic situations. Thus, we cannot and do not want to make assumptions concerning local image properties.

Nevertheless, all road scenarios contain main characteristics featuring strong gradient responses. Many objects captured from vehicle-mounted cameras are aligned with respect to the ground plain and/or road, e.g. other vehicles, buildings, walls and fences. Therefore, we assume the distribution of their orientation to code for the current roll angle of the camera with respect to the horizon. An orientation histogram is used to indicate that distribution – one histogram bin for each discrete angle in $[0^\circ, 180^\circ)$.

From the image gradients g_x and g_y (cf. (2) and (3)) we derive the orientation α . For robustness one wants to prohibit small gradients to have large impact on the calculated histogram h . There are two ways to deal with the issue:

- Only consider image positions that exceed a certain energy threshold or
- all measured angles contribute with their corresponding energy E (cf. (6)).

$$f = \begin{pmatrix} -1 & -2 & 0 & 2 & 1 \\ -2 & -4 & 0 & 4 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{pmatrix} \quad (1)$$

$$g_x = \text{img} * f \quad (2)$$

$$g_y = \text{img} * f^T \quad (3)$$

$$E(i) = \sqrt{g_x^2(i)g_y^2(i)} \quad (4)$$

$$\angle(i) = \angle \begin{pmatrix} g_x(i) \\ g_y(i) \end{pmatrix} \quad \angle \in [0^\circ, 180^\circ) \quad (5)$$

$$h(\alpha) = \frac{\sum_i \begin{cases} E(i), [\angle(i)] = \alpha \\ 0, \text{otherwise} \end{cases}}{\sum_i E(i)} \quad \alpha \in \mathbb{N}_0 \quad (6)$$

The convolution mask f in (1) is a combination of a gradient filter and an orthogonal smoothing filter in order to avoid the extraction of artefacts which emerge when using smaller filter masks.

Given this normalised histogram (cf. Fig. 2) as image feature we are able to correlate either a test image with a learnt histogram distribution (roll angle) or two subsequent images (roll rate). Those approaches are exposed in the following two sections.

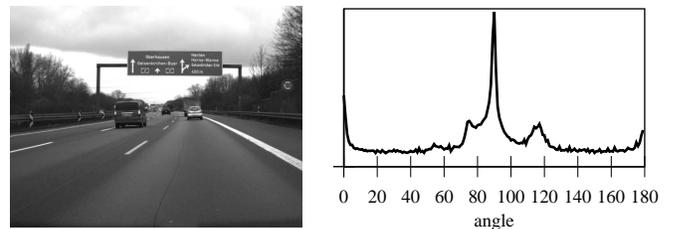


Fig. 2. A horizontally aligned input image and its corresponding orientation histogram. Gradient orientation of 90° correspond to horizontal edges.

III. ROLL ANGLE ESTIMATION

The approach of roll angle estimation on unseen image data is based on learning the distribution of the discrete training histograms. The image sequences for training and test data were generated from a car-mounted camera. In order to simulate possible roll angles the images were rotated within the image plane – obtaining the ground truth angles for the learning and evaluation phase at the same time. In a later project stage we will be able to record data from an experimental motor-cycle equipped with front view camera and IMU. The temporal rolling behaviour was simulated by a sine-wave with varying amplitude α_{\max} and wave-length ω .

$$\alpha(t) = \alpha_{\max} \sin 2\pi\omega t \quad \alpha_{\max} \leq 35^\circ, \omega \in \left[\frac{1}{20s}, \frac{1}{2s} \right] \quad (7)$$

In order to handle different scene properties we used recordings of the following road scenarios for training and test images:

- motorway
- country road
- urban

Details of the evaluation procedure will be documented in Sec. VI.

Given the ground-truth angle one is able to rotate the measured gradient vectors back to horizontal alignment which will be the reference for training. We then learn the histogram distribution for each bin (angle) individually by computing mean μ and variance σ^2 over all training examples' histograms h_n .

$$\mu(\alpha) = \frac{1}{N} \sum_{n=1}^N h_n(\alpha) \quad (8)$$

$$\sigma^2(\alpha) = \frac{1}{N-1} \sum_{n=1}^N (h_n(\alpha) - \mu(\alpha))^2 \quad (9)$$

Fig. 3 shows the learnt statistics from three *country road* sequences where significant maxima arise around 0° and 90° .

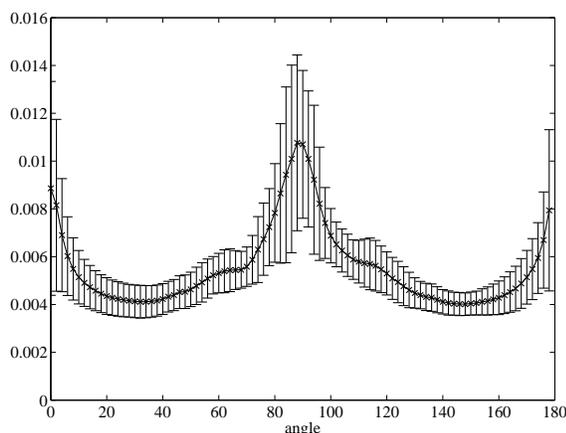


Fig. 3. Learnt mean and standard deviation for each orientation histogram bin (one degree).

For a given test image we translate its histogram h_{test} by a range of reasonable roll angles γ and determine maximal correlation with the learnt statistics.

$$\gamma_{\text{estim}} = \arg \max_{\gamma} \sum_{\alpha} \frac{h_{\text{test}}(\alpha + \gamma) - \mu(\alpha)}{\sigma(\alpha)}, \quad \gamma \in [-35^\circ, 35^\circ] \quad (10)$$

This *sum-of-weighted-differences* (SWD) procedure assumes the histogram bins to be independent and normally distributed which should be an appropriate approximation.

$$h(\alpha) \sim \mathcal{N}(\mu(\alpha), \sigma^2(\alpha)) \quad (11)$$

Alternatively, we evaluated *sum-of-absolute-differences* (SAD) and *normalised cross-correlation* (NCC) as similarity measure (cf. Sec. VI).

To receive a robust estimate one should make use of temporal integration. This will be done by a Kalman filter. Thereby, we are able to integrate the roll rate and account for their physical dependence.

IV. ROLL RATE ESTIMATION

As a second measure – mainly with the goal to stabilise the angle estimate – we look at the roll rate. Being the derivative of the roll angle the rate is measured in degrees per second. In contrast to the angle we will estimate the rate from the orientation histograms h_t and $h_{t-\Delta t}$ of two subsequent frames. Therefore, we now do not rely on the learnt distributions but correlate images with mainly the same scene content which facilitates the task.

$$\gamma'_t = \frac{1}{\Delta t} \arg \max_{\beta} r(h_t(\alpha + \beta), h_{t-\Delta t}(\alpha)) \quad (12)$$

where r is a similarity measure, e.g. normalised cross-correlation:

$$r(f, g) = \frac{1}{\alpha_{\text{max}} - 1} \sum_{\alpha} \frac{(f(\alpha) - \bar{f})(g(\alpha) - \bar{g})}{\sigma_f \sigma_g} \quad (13)$$

The choice of Δt is a trade-off:

- With respect to the limited accuracy of the gradient orientation and resolution of the histograms it is advisable to choose Δt large enough to encounter an angle difference $\gamma \neq 0$.
- On the other hand, for γ'_t to be a good estimate of the rate at time t one should choose Δt small.

Experiments show that distances of about five video frames ($\Rightarrow \Delta t \simeq 200 \text{ ms}$) work well.

To overcome the lack of accuracy we approximated r in the neighbourhood of the maximum by a quadratic function and chose its maximum. Fig. 4 compares histograms of the same scene with slightly different rotations.

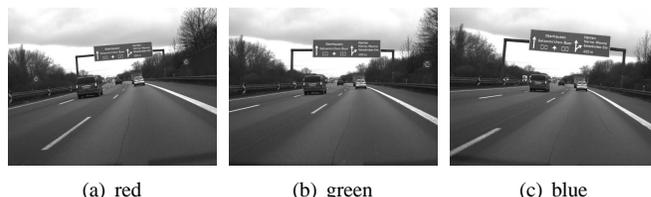
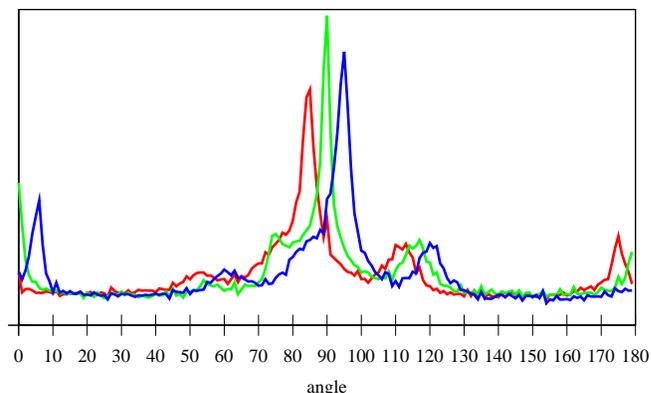


Fig. 4. Normalised histograms of three consecutive images with a distance of ten frames ($\simeq 400 \text{ ms}$).

V. KALMAN FILTERING

In contrast to conventional time series filters the Kalman filter supports an explicit separation of the system dynamics and the process of measurement. Therefore, it is a suited tool for many sensor fusion problems that incorporate a physical model [10].

The process of Kalman filtering applied to the given problem can be outlined by the following initialisation:

Estimated angle and rate define the observation vector $\mathbf{z}_t = (\gamma_t \ \gamma_t')^\top$, where $\gamma_t \equiv \gamma_{\text{estim}}$ from (10). Both are uncertain observations of the state \mathbf{x}_t containing noise, e.g., due to sensor noise and the estimation procedure (histogram resolution).

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \quad (14)$$

where $\mathbf{H}_t = \mathbf{I}$ is the observation model which maps state space to observed space (identical here). \mathbf{v}_t is the observation noise which is modelled as zero-mean Gaussian noise with covariance \mathbf{R} .

Given the internal state at $t-1$ the filter dynamics assume the following true state \mathbf{x}_t to emerge according to

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t \quad (15)$$

where the state transition model \mathbf{F}_t describes the physical behaviour of our system and process noise $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q})$ that is caused by acceleration a_t with standard deviation σ_a .

$$\mathbf{F}_t = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \quad (16)$$

$$\mathbf{Q} = \begin{pmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{pmatrix} \begin{pmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{pmatrix}^\top \sigma_a^2 = \begin{pmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 \end{pmatrix} \sigma_a^2 \quad (17)$$

A control input is not used in our scenario.

After choosing physically reasonable noise parameters \mathbf{R} and σ_a the filter is able to produce a smooth and robust estimate for the roll angle taking into account the independently observed roll rate and their physical relationship. The following section will illustrate this with several experimental results.

For future enhancement one could work with dynamic uncertainty for each measurement. Translating the confidence of the angle and rate estimates to noise covariance \mathbf{R}_t in each time step will support the filtering process. However, the mapping from correlation results to an angle covariance is not straight forward and needs further investigation. For now we choose

$$\mathbf{R}_t = \mathbf{R} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad (18)$$

VI. EXPERIMENTS

As stated in Sec. III the experimental data was acquired from rotated image sequences recorded from a car-mounted camera. The ground-truth rotation angle was defined by a sinusoidal wave

$$\alpha(t) = \alpha_{\text{max}} \sin 2\pi\omega t \quad (19)$$

with combinations of

$$\alpha_{\text{max}} \in \{20^\circ, 25^\circ, 30^\circ, 35^\circ\} \quad (20)$$

and

$$\omega \in \left\{ \frac{1}{12s}, \frac{1}{10s}, \frac{1}{8s}, \frac{1}{6s} \right\}. \quad (21)$$

We generated 15 sequences with a length of at least one wave up to one minute. This resulted in more than 6,000 images. The sequences are a mixture of the three different scenarios – *motorway*, *country road*, and *urban* (see Fig. 5). Moreover, images were recorded under various weather and lighting conditions.



Fig. 5. Images of the three scene categories.

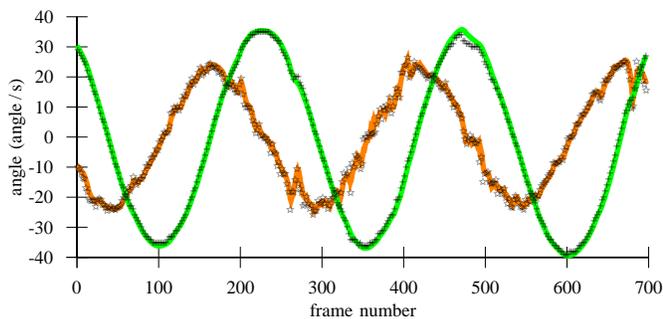
a) Training: Three of the *country road* sequences were exclusively used for the training process where we estimated the distribution of the gradient orientation by calculating mean and standard deviation of all histogram bins independently. The ground truth was employed to produce the histograms with the horizontal reference alignment (see Sec. III). Note that this procedure can be applied in the same way for recordings from a motor-cycle.

b) Testing: The estimation was applied to the rest of the sequences including all three scenarios. We observed the raw estimates for the roll angle, the rate, and the filter behaviour for both. The following plots (Fig. 6 and 7) show results of the estimation process in comparison to the ground-truth data.

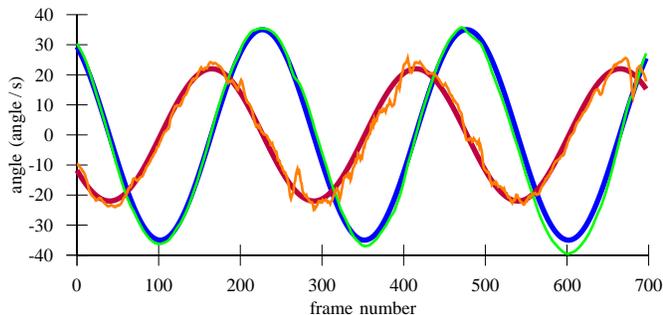
c) Performance: In order to measure the performance for each sequence individually the mean-squared-error of the filtered roll angle was used. Table I states the performance of the angle estimation grouped by scenarios and applied correlation methods. We are able to record that no correlation method dominates any other – however its choice does not influence the performance significantly. A mean error of

TABLE I
MEAN SQUARED ERRORS BY ROAD SCENARIOS AND CORRELATION METHODS (NORMALISED CROSS-CORRELATION, SUM-OF-ABSOLUTE-DIFFERENCES, SUM-OF-WEIGHTED-DIFFERENCES).

Scenario	NCC	SAD	SWD
Motor way	1.82	1.96	2.47
Country road	2.40	2.47	2.43
Urban	2.12	2.04	2.41
All	2.09	2.15	2.44



(a) Raw and filtered estimates of angle (cross/green) and rate (star/orange).



(b) Estimated vs. ground-truth angle (green/blue) and rate (orange/purple).

Fig. 6. Experimental results from a country road sequence using *normalised cross-correlation*.

about 2 degrees can be regarded as low and especially the fact that only sequences from *country roads* were used for training documents good generalisation of our approach.

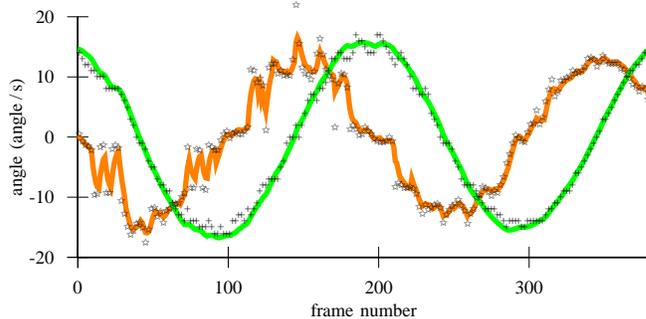
Finally the computational complexity of the applied methods is small. After the determination of gradient and orientation image all further computations are based on a low dimensional histogram which makes it easily applicable in the context of real-time driver assistance systems where limited hardware resources are an issue.

VII. CONCLUSION

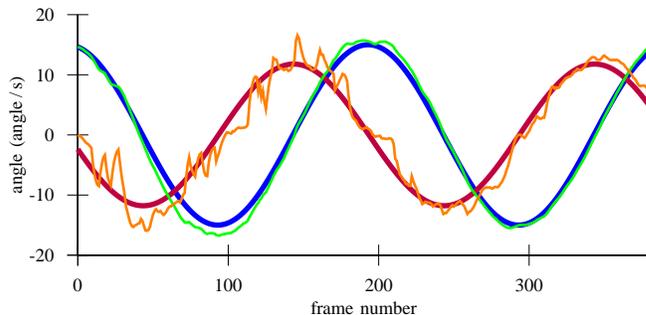
We proposed a new method of roll angle estimation for two-wheeled vehicles based on video images only. Its application is highly relevant to the progress of video-based driver assistance systems for motor-cycles. Knowing the current roll angle one is able to apply existing image processing modules that were developed for cars and trucks expecting horizontally aligned images.

Current alternatives implicate the usage of an inertial measurement unit, stereo vision information or the fusion of multiple sensory inputs which are very cost-intensive. Our approach is easily implemented as a preceding real-time image processing module that does not depend on additional sensors.

In a first project stage we developed a realistic evaluation procedure that allows for generating test sequences and ground-truth data recorded from a car-mounted camera. Sequences with various roll angle amplitudes from different



(a) Raw and filtered estimates of angle (cross/green) and rate (star/orange).



(b) Estimated vs. ground-truth angle (green/blue) and rate (orange/purple).

Fig. 7. Experimental results from a motor way sequence using *sum-of-absolute-differences*.

scene types were produced in order to learn the correlation of orientation statistics in an image and the present roll angle.

The evaluation of test sequences point out the robust behaviour of the estimation module at a mean precision of two degrees, which can be regarded as sufficient for the targeted field of application. Moreover, the developed learning module generalises very well with respect to varying external conditions (scenario, weather, lighting) without any parameter adjustment. The proposed procedures do not rely on camera calibration, but merely assume a constant camera pose with respect to the vehicle in training and test phase.

In further project stages we will apply the proposed method to sequences captured on a motor-cycle equipped with an IMU and extend the training procedure to all scenarios. In addition, we will further improve the Kalman filter process by an online adjustment of the observation noise – based on estimation confidences. Yet, mapping correlation confidences to noise covariance for the Kalman update is not straight forward.

REFERENCES

- [1] A. Rehm, “Estimation of vehicle roll angle,” *International Symposium on Communications, Control and Signal Processing*, pp. 1–4, 2010.
- [2] J. Ryu and J. C. Gerdes, “Estimation of vehicle roll and road bank angle,” in *Proceedings of the American Control Conference*, vol. 3, 2004, pp. 2110–2115.
- [3] I. Boniolo, M. Tanelli, and S. Savaresi, “Roll angle estimation in two-wheeled vehicles,” *International Conference on Control Applications*, pp. 31–36, 2008.

- [4] I. Boniolo, M. Norgia, M. Tanelli, C. Svelto, and S. Savaresi, "Performance analysis of an optical distance sensor for roll angle estimation in sport motorcycles," *IFAC World Congress*, pp. 135–140, 2008.
- [5] L. Gasbarro, A. Beghi, R. Frezza, F. Nori, and C. Spagnol, "Motorcycle trajectory reconstruction by integration of vision and mems accelerometers," in *Proceedings of the 43th Conference on Decision and Control*, 2004, pp. 779–783.
- [6] F. Nori and R. Frezza, "Accurate reconstruction of the path followed by a motorcycle from the on-board camera images," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2003, pp. 259–264.
- [7] R. Labayrade and D. Aubert, "A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision," in *Proceedings of the Intelligent Vehicles Symposium*, 2003, pp. 31–36.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [9] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2001.
- [10] A. Gelb, *Applied Optimal Estimation*, 1st ed. MIT Press, 1974.