A Classification and Temporal Filtering based System for Online Extrinsic Camera Calibration

Jens Westerhoff\textsuperscript{1}, Stephanie Lessmann\textsuperscript{2}, Mirko Meuter\textsuperscript{3} and Anton Kummert\textsuperscript{4}

Abstract—This paper proposes an online extrinsic camera calibration system able to determine the pitch and roll angle of a moving, forward-looking camera relative to the road surface. Some road surfaces do not show enough texture to allow reliable optical flow calculation and accurate per-frame pitch and roll estimations, respectively. If inaccurate per-frame estimations are regarded in the final calibration, no exact and stable determination of calibration is possible.

Therefore, the main contribution of this paper is a classifier system which enables us to distinguish between accurate and inaccurate per-frame angle estimations. To achieve a final calibration only the accurate (classified as positive) per-frame estimations are incorporated into a temporal filter. In our research we found that a fixed amount of positive classified per-frame estimations (a fixed temporal filter length) can be specified which guarantees an exact final calibration. With that, our system is able to autonomously signalizes the moment at which an accurate calibration is achieved while driving. With our developed system, road surfaces with no texture only delay the moment at which a final calibration is achieved but do not affect the exactness of calibration.

I. INTRODUCTION

Different camera-based applications in the field of “Advanced Driver Assistance Systems” (ADAS) require information about the extrinsic calibration (mounting) of the camera relative to the vehicle. For instance, a “Forward Collision Warning” (FCW) system needs extrinsic calibration information in order to calculate distances of objects relative to the camera and to the vehicle, respectively. During the life cycle of a vehicle, repairs as well as additional load may affect the camera’s extrinsic calibration and therefore, a re-calibration becomes mandatory. For online calibration it is still an open question at which moment an exact and stable calibration is achieved while the car is moving.

In this paper we focus on the calibration of the pitch and roll angle of a forward-looking camera relative to a road surface under the assumption of planar roads. The determination of pitch and roll angle in this paper relies on optical flow vectors calculated on the road surface. Roads, showing weak road texture (e.g. due to unstructured asphalt or snow cover) often cause inaccurate optical flow vectors and therefore inaccurate per-frame angle estimations which should not be used for the determination of the final calibration. For this reason, we developed a classification system able to identify the inaccurate per-frame angle estimations. To determine the final calibration, the accurate (classified as positive) per-frame estimations are incorporated into temporal filtering. As our research shows, it is possible to specify a fixed temporal filter size (meaning a fixed amount of positive classified per-frame estimations) which ensures a stable and exact pitch and roll angle calibration. For example, the specified amount of positive classified estimations can be reached after 30 seconds, 2 minutes or 10 minutes, depending on the available texture on the road surface. The advantage of this approach is that due to specified filter length the system is able to signalize when a stable and exact calibration is achieved, independent of the variation of road structure.

A prerequisite of our calibration is the ego-motion (rotation, translation) of the camera between two frames which is provided by an algorithm developed in [1]. In addition, overall metric scale is obtained from a velocity sensor and from the camera height which is assumed to be constant. The intrinsic camera parameters are determined via standard calibration procedure.

In the following the final pitch and roll angle result of our system is named “calibration”, whereas single per-frame pitch and roll angle estimations are referred as “estimation”. The reminder of this paper is divided into seven main sections. Section II introduces the literature related to online extrinsic camera calibration. An overview of our system pipeline is given in III and explains the main parts of the system introduced in IV - VI. In section VII experimental results are evaluated while section VIII concludes this paper with a summary and an outlook.

II. RELATED WORK

Previous developments in the area of online extrinsic calibration of cameras differ from each other related to available input data, motion model and the estimated output parameters. Our paper deals with a mono camera and therefore the literature review does as well. However, there are approaches for stereo- and multi-camera rig calibration [2]–[5] and offline approaches based on calibration objects [6]–[8], too.

The authors of [9] aim for the extrinsic camera orientation of a mono side-looking camera under the assumption of planar motion. Known vehicle velocity and camera height serve as input data minimizing a one-dimensional cost function based on homography. Finally, the estimated orientation parameters are accumulated over time in a histogram. As shown in

\textsuperscript{1}School of Electrical, Information and Media Engineering, University of Wuppertal, D-42119 Wuppertal, Germany jens.westerhoff@uni-wuppertal.de

\textsuperscript{2}Intelligent Systems Group, University of Duisburg-Essen, D-47057 Duisburg, Germany stephanie.lessmann@uni-due.de

\textsuperscript{3}Delphi Electronics & Safety Advanced Engineering, D-42119 Wuppertal, Germany mirko.meuter@delphi.com

\textsuperscript{4}School of Electrical, Information and Media Engineering, University of Wuppertal, D-42119 Wuppertal, Germany kummert@uni-wuppertal.de
the experimental section (VII-B) of our paper the histogram based approach can benefit from our classification based approach.

A further online calibration system has been introduced by [10] for a rear-parking camera directly estimating height, pitch and roll angle based on planar motion model. In comparison to [9] and [10] our approach focuses on the pitch and roll angle estimation of a forward-looking camera and is not restricted by the planar motion model.

Another work assumes that the vehicle travels on a straight line [11] in order to obtain not only rotation but position parameters as well. A different method is applied by the authors of [12]. They use triangulated 3D points to estimate the extrinsic camera calibration of a forward-looking camera conditioned by straight forward motion of the vehicle as [11]. In our paper the pitch and roll angle of the camera to the road is determined. Therefore, the applicability of our calibration system is not restricted to straight ahead movement.

Orientation parameter optimization by particle filtering is done in [13] with straight ahead movement restriction and additional vanishing point estimation based on lane markers. Publications [14]–[16] require lane markers as well to determine camera orientation of a mono camera. In comparison to these publications our system is not limited by the need for road markings or any special environmental markers.

In conclusion to the best of our knowledge, no investigations related to the question at which moment a stable and exact extrinsic camera calibration is achieved have been done so far.

III. SYSTEM OVERVIEW

This section introduces the pipeline of our calibration system depicted in Fig. 1 which enables us to determine the moment at which an exact extrinsic calibration is achieved. The orientation (angle) estimation is the basis of the calibration system and therefore discussed in the next chapter (see IV). As can be seen in Fig. 1 orientation angles are only estimated for frames passing the pre-classifier at the beginning of the pipeline. The decision of the pre-classifier is based on thresholds explained in V-A. For frames for which the angle estimation is done, features related to the estimation are calculated. These features serve as input data for the supervised classifier (see V-B). With the supervised classifier we are able to decide whether the angle estimation results are accurate or not. If the supervised classifier takes a positive decision (estimation is accurate) the angle estimation results are accurate or not. If the supervised classifier takes a positive decision the angle estimation is obtained to provide a final accurate extrinsic calibration.

IV. ANGLE ESTIMATION

In general, camera orientation relative to the road surface can be derived from the normal vector of the road. The transformation of the normal vector in the world coordinate system \( \mathbf{w}_n \) to the normal vector in camera coordinates (see Fig. 2 and Fig. 3) is defined by

\[
\mathbf{c}_n = \mathbf{C}_w \mathbf{R}_\gamma \mathbf{R}_\theta \mathbf{R}_\phi \mathbf{w}_n_0 \quad (1)
\]

where \( \mathbf{w}_n_0 = [0, 0, -1]^T \). The rotation matrices representation related to the world coordinate system is given by \( \mathbf{R}_\gamma \) for the roll rotation around x-axis, \( \mathbf{R}_\theta \) for the pitch rotation around y-axis and \( \mathbf{R}_\phi \) for the yaw rotation around z-axis. Furthermore, matrix

\[
\mathbf{C}_w = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad (2)
\]

transforms the axes from the world coordinate system to the camera coordinate system. For abbreviation purpose we set \( \mathbf{w}_n = \mathbf{C}_w^T \mathbf{c}_n \). With the assumption of \( \mathbf{R}_\gamma = \mathbf{I}_{3 \times 3} \) we can simplify (1) to

\[
\mathbf{w}_n = \mathbf{R}_\phi \mathbf{R}_\theta \mathbf{w}_n_0. \quad (3)
\]

Therefore, we can calculated the roll angle by

\[
\phi = - \arctan \frac{w_{ny}}{w_{nz}} \quad (4)
\]
and the pitch angle by reverting the roll angle as shown by

\[ \theta = \arctan \left( \frac{R^T_{\phi}(\phi)W_n}{x} \right). \]  \hspace{1cm} (5)

In general, the normal vector of a plane (here: road surface) in the camera coordinate system is part of a perspective transformation called homography between two different views (captured from two different positions) of this particular plane. Put simply, the homography matrix \( H \) is able to transform a point belonging to the plane in the first view \( p \) to the respective point \( q \) in the second view in homogenous coordinates by

\[ q = Hp. \]  \hspace{1cm} (6)

To obtain the image coordinates \((u, v)\) from the transformed homogenous vector \( q \) we calculate

\[ u = \frac{q_1}{q_3}, \quad \text{and} \quad v = \frac{q_2}{q_3}. \]  \hspace{1cm} (7)

The homography matrix

\[ H = K \left( R - \frac{t}{d} C_n^T \right) K^{-1} \]  \hspace{1cm} (8)

is built from the intrinsic camera calibration matrix \( K \), the relative rotation \( R \) and translation \( t \) between the two views given in camera coordinates and the camera height above ground \( d \).

The components \( R \) and \( t \) between the two views are obtained via ego-motion estimation by [1]. As already mentioned in the introduction, the overall metric scale is given by car’s velocity sensor. The camera height is measured manually and the intrinsic camera matrix \( K \) is known.

The components of the normal vector \( C_n \) which are needed to calculate the pitch and roll angle of the camera are obtained by creating the functional dependencies between the image point observations of the plane in both images and the normal vector components [17]. The point observations are provided by optical flow vectors which are calculated using a Kanade-Lucas-Tomasi feature tracker (KLT, [18]). After disassembling and rearranging the equations we obtain a non-homogeneous linear equation system from which the normal vector components can be determined [17].

Evenly distributed optical flow vectors at fixed starting positions are calculated in a limited area in the lower half of a 1024 \( \times \) 512 pixel image. In general, there are two reasons for incorporating optical flow vectors only in front of the vehicle, to be more precise within the ego-lane.

Firstly, different lanes of the same road can have different horizontal slopes (e.g. roof profile) in order to drain the road which may lead to wrong roll angle estimations if the vector field exceeds the ego-lane. Secondly, inexact intrinsic camera calibration or temperature influences cause significant offset errors related to the roll angle estimations increasing towards image borders. The start positions of the optical flow vectors are calculated in a pre-defined intersection area of reasonable parameter limitations, like \(+/-5^\circ\) pitch and roll angle.

The described angle estimation algorithm is embedded in a RANSAC procedure used to discard optical flow vectors which do not match the current angle estimation. The inlier/outlier decision of the RANSAC procedure is done by thresholding the error between the actual endpoint of the optical flow vector and the endpoint of the vector estimated by the angle estimation, the homography transformation, respectively.

At this point an estimation of pitch and roll angle of a camera relative to the road surface can be performed for every frame based on the described algorithm. Unfortunately, the results of the per-frame estimations show high variability. The following sections introduce a system able to deal with this high variability resulting in an accurate extrinsic calibration with additional information about the moment at which an accurate calibration is achieved.

V. Classification

In this section we introduce the core components of our online extrinsic camera calibration system: The pre- and supervised classifier. These classifiers take the task to find frames which do provide reliable pitch and roll angle estimations.

A. Pre-Classifier

In general, the pre-classifier needs not be trained in terms of supervised classification and takes the task of rejecting frames from which no accurate angle estimations can be expected. In the following two cases are described which need special treatment.

In common driving scenarios small curve radii increase the risk that the optical flow field in front of the car includes environmental geometry (curbstones, embankment, ...) which do not correspond to the road plane geometry. We observe that these cases provide accurate angle estimations in terms of the features of the supervised classifier described in V-B and therefore cannot be filtered out by the supervised classifier. Obviously, in these cases the angle estimations do not reflect the camera orientation relative to the road but for instance the camera orientation relative to a curbstone in front of the vehicle. For that reason, frames exceeding an empirically determined curve radius (calculated from vehicle’s speed and yaw rate) are treated as negative classification results related to the pre-classifier.

As already mentioned we use ego-motion information as input data for our angle estimations. The used ego-motion algorithm [1] provides the additional information whether the estimated ego-motion is reliable or not. Because of the dependency of our angle estimation from ego-motion we
treat frames as negative classification results for which no reliable ego-estimation is available.

B. Supervised Classifier

Features related to the optical flow field and the angle estimation itself enable us to separate accurate angle estimations from inaccurate ones. These features can be used to train a supervised classifier. Here, we choose a SVM classifier [19] because of its generally known ability to achieve good generalization with limited training data and avoidance of local minima during optimization. We use a soft-margin SVM consisting of a linear kernel.

A compromise related to good generalization property and overfitting of the SVM is found by a heuristic search regarding the soft-margin parameter of the SVM. More information about training and testing is given in section VII. For the SVM training and testing we obtain the ground truth related to the pitch and roll angle by applying an offline external camera calibration method based on [8]. This ground truth information is not needed for the final application of our angle estimation pipeline but only for the training and testing of the classifier. In particular, pitch angle estimations which do not extend an error of $+/-1^\circ$ relative to ground truth are labeled as positive classification results, roll angle estimations $+/-2^\circ$, respectively. In the following a description of the used features for the supervised classification will be given.

As mentioned, the angle estimation uses optical flow vectors as input data. It depends on the availability of texture on the road whether an optical flow estimation is feasible or not. Therefore, a different amount of flow vectors is available for the angle estimation algorithm depending on the appearance properties of the road in the current frame. The angle estimation is embedded in a RANSAC procedure. At the end, the RANSAC procedure outputs flow vectors (homography transformation inliers) which contributes to the angle estimation for every frame. In experiments we observed that the accuracy of angle estimation benefits from a well distributed homography inlier vector field. Therefore, we divide the optical flow field in six distinct subregions (see Fig. 4 and Fig. 5). The amount of inlier vectors $N$ in each subregion is taken as a classifier feature. By taking these six features $(N_1, \ldots, N_6)$ into account the classifier can learn how the amount of inliers and its distribution contributes to the accuracy of angle estimation. In conclusion, our feature vector is defined by

$$f = (N_1, N_2, N_3, N_4, N_5, N_6)^T.$$  \hspace{1cm} (9)

VI. TEMPORAL FILTERING AND VALIDITY DECISION

The classification system makes sure (up to a certain probability) that angle estimations which are further processed are located inside an accuracy interval of $+/-1^\circ$ for pitch and $+/-2^\circ$ for roll angle. To obtain an accurate final calibration we use recursive moving average filters to smooth the angle estimations inside the accuracy intervals. In the following we determine the amount of positive classified estimations required to obtain a final accurate extrinsic calibration and show the related results.

VII. EXPERIMENTS

The dataset we use in our experiments consists of city and highway scenarios with different camera roll angles, namely $0^\circ$, $3^\circ$ and $5^\circ$ (40,000 frames). In all scenarios the pitch angle of the camera relative to the road is fixed to $-2.1^\circ$. The $0^\circ$ roll angle dataset serves as training and validation data in section VII-A for the supervised classifier. To analyze if the calibration system shows good performance related to other roll angles as well, sections VII-B and VII-C evaluate overall system performance on a $3^\circ$ and $5^\circ$ roll angle test set, respectively. In addition, we compare our method to two other methods which do not use our classification system.

A. Training, Operational Point and Filter Specification

The $0^\circ$ roll angle dataset is divided into training and validation data. After SVM training we use the validation data set to choose an operational point for the final extrinsic calibration system on the classifier’s ROC curve in Fig. 6. By empirical experiments based on the validation data we found that a false positive rate (FPR) in the range of $6\%$ (see cross in Fig. 6) in combination with a moving average filter length of $500$ positive classified frames gives promising overall calibration results. The desired low FPR is at the expense of a FNR (false negative rate) of $54\%$ (see TABLE I). However, this FNR is not an issue for our online extrinsic calibration because false negatives only delay the moment at which an accurate calibration is obtained. We will show the impact of these specifications related to the test data in VII-B and VII-C.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
 & 1 & 0 \\
\hline
1 & 0.46 & 0.06 \\
0 & 0.54 & 0.94 \\
\hline
\end{tabular}
\caption{Confusion matrix of the selected classifier}
\end{table}

B. Test Set: $3^\circ$ Roll Angle

In this section we show the overall performance of the calibration system on a test set with $3^\circ$ roll angle. Furthermore, we compare our classification-based system with a standard
To show the superiority of our classification-based system over a standard statistical approach, we calculate a moving average considering the variances of the estimations. The final angle variance estimation is calculated by

\[ \sigma^{2}_{\phi} = J_{\theta,\phi}(A_{n,1}A^{T})^{-1}J_{\theta,\phi}^{T} \]

(10)

is obtained with the help of the Jacobian matrix of formula (4) and (5)

\[ J_{\theta,\phi} = \frac{\partial (\theta, \phi)}{\partial (W_{n,1}, W_{n,2})} = \begin{bmatrix} \frac{\partial \theta}{\partial W_{n,1}} & \frac{\partial \theta}{\partial W_{n,2}} \\ \frac{\partial \phi}{\partial W_{n,1}} & \frac{\partial \phi}{\partial W_{n,2}} \end{bmatrix} \]

(11)

for pitch and roll angle, respectively. In order to ensure a fair comparison we set \( L = 500 \) equally to the averaging of our classification-based system. This approach represents a maximum likelihood estimator of the mean assuming independent and normal distributed data. The right plot of Fig. 8 shows the results of this approach. In contrast to our system it is neither possible to obtain a stable calibration nor to determine the moment at which an accurate calibration is achieved, especially for the roll angle.

Another benefit of our technique is the ability of supporting extrinsic calibration techniques relying on histograms containing estimated angles like [9]. For comparison, Fig. 9 shows results of a pure histogram method where the peak of the distribution is selected as final estimation result. Due to the smeared histogram distribution it is hard to find this particular peak. Our classifier system can help to sharpen the distribution (see Fig. 10) and therefore to find a more reliable extrinsic calibration.
C. Test Set: 5° Roll Angle

At last, we present a sequence with 5° camera roll angle to illustrate the impact of a significant amount of false-negative classification events. The right plot of Fig. 11 shows many pre-classifier FN’s (red dots) due to motorway on-ramp with a small curve radius where no disturbing obstacles are present contrary to the assumption in section V-A. However, this is not an issue for our extrinsic calibration because the results for pitch and roll angle after $L = 500$ positive classified estimations (Fig. 11, left plot, green stars) are nonetheless exact (roll error: 0.10°, pitch error 0.14°) and only temporally delayed.

![Fig. 9: A smeared histogram distribution makes it difficult to rely on histogram based estimation methods.](image1)

![Fig. 10: Our system can be used to sharpen the histograms, particularly helpful for roll angle estimation.](image2)

![Fig. 11: False-negative classifications only extend the time by which 500 positive classified estimations are obtained (green stars in left plot) but does not affect the final calibration accuracy (left figure, roll error: 0.10°, pitch error 0.14°)](image3)

VIII. Conclusion and Future Work

In this paper we introduced a system able to determine an accurate extrinsic camera calibration (related to pitch and roll angle) relative to the road surface for a vehicle mounted camera. Furthermore, the system autonomously signalizes the moment at which this stable calibration is achieved. The angle estimations for every frame rely on a homography-based estimation using optical flow vectors on the road as input data. Based on a developed classification system using presented features we are able to decide whether the current angle estimations for a single frame is accurate or not. In experiments, we found that 500 positive classified estimations incorporated in a moving average filter achieve an accurate extrinsic camera calibration.

Furthermore, we presented the advantages of our system by showing the differences to a standard variance-weighted averaging. In addition, histogram-based methods may benefit from the developed classification system as shown in the previous section. For future research bigger datasets for classifier training would be desirable. Further work will focus on improvements of temporal filtering by using output probabilities generated by the classifier in order to provide stable and accurate angle estimation results even faster.

References


学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。
图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具