2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Self-Validation for Automotive Visual Odometry

Martin Buczko[1], Volker Willert[1], Julian Schwehr[1] and Jürgen Adamy[1]

***Abstract*— Automotive visual odometry has become a highly researched topic. The published works have turned cameras into a very precise source of ego motion estimation. However, for automotive application, visual odometry has to be integrated into a sensor fusion to simultaneously obtain global localization, maximum availability as well as highest precision. Since any visual odometry remains sensitive to its environment, good motion estimation cannot always be guaranteed. Consequently, a self-validation scheme is one of the barriers towards application in a sensor fusion system. In order to solve this problem, we first formulate an Ackermann vehicle's motion as a function of its forward speed and yaw rate. Secondly, we present a data-driven model, which achieves to reconstruct the sideward speed as a function of the yaw rate only. As we show, both models reach the quality of the estimated sideward motion from visual odometry. The therewith achieved redundancy can be used for different tasks. Of course, the estimation of the sideward motion can be excluded from the visual odometry scheme to save computation time. This is of special interest for monocular systems, where yet no absolute scale of a translation motion could be directly calculated. Instead, we propose to maintain the estimation of the sideward motion in the visual odometry and to compare the result to the modeled motion. As we show, the emerging deviation is a very good metric for self-validation of the overall visual odometry estimate. Integrating the resulting method into our visual odometry system, we currently achieve the best frame-to-frame result in the KITTI benchmark.**

I. Vehicle Localization

In current automotive applications, self-localization is mostly used for navigation systems that increase the driver's comfort. These systems require precise information on the ego-position. Consequently, they are not available if localization information is absent, which leads to a decreased level of comfort and increases the attention requirement of the driver. By contrast, a much higher degree of availability is demanded for future autonomous tasks: Cars will be able to perform driving in certain scenarios as well as specialized tasks like valet parking without needing a driver. Here, the absence of localization information renders the vehicle unable to navigate to its target and therefore forces the system to stop its current task. This may lead to disturbing and even dangerous situations and may lower the acceptance of autonomous systems. Consequently, the steady availability of localization information is crucial for the success of autonomous driving. Because of this, physically independent localization principles are highly relevant. Due to their growing use for advanced driver assistance systems, cameras are more and more present in newly built cars. Consequently, gaining motion information from cameras – the so called visual odometry – currently raises large interest and is highly researched, as the numerous publications in this field (e.g. [5], [6], [13], [16]) show. Its advantages are the ability to reconstruct the vehicle's 3D rotation vector $\mathbf{R}_{VO} = [\alpha, \beta, \gamma]$ as well as the 3D translation vector $\mathbf{T}_{VO} = [t_x, t_y, t_z]$ from the data of two calibrated cameras. Here, α, β and γ are the pitch, yaw and roll between consecutive frames at time $t-1$ and time t, while t_x, t_y and t_z are translation to the right, downwards and forwards for the same frames. The respective angular rates and translational speeds can be obtained by multiplication with the frame rate f. Since the purpose of its use in automotive context is to increase the availability of an overall system, the visual motion estimate is fused with the motion estimates from others sensors, as depicted in Fig. 1. In contrast to many other sensors, visual odometry is highly affected by the current environment of the vehicle. Therefore, no scenario-invariant noise model can be derived and furthermore no output is available that gives information on the reliability of the current estimate in state of the art visual odometry. Since the reliability is a crucial demand for the sensor fusion, this work deals with the development of a self-validation signal for visual odometry. The remainder of this paper is organized as follows: In Sec. II, we discuss different published approaches to increase the robustness of visual odometry. Subsequent Sec. III deals with the most prominent class of metrics to detect the failure of visual

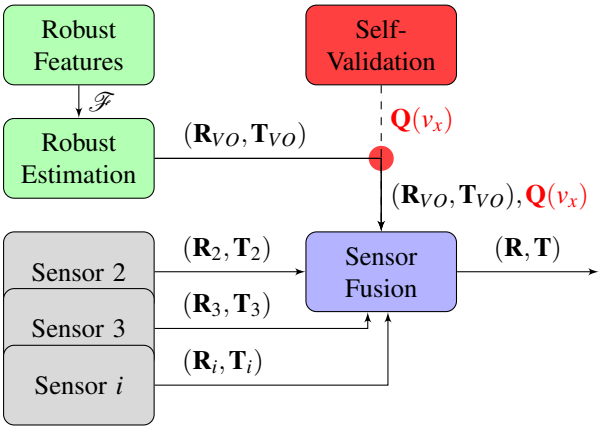

Fig. 1. The schematic visualizes the process of fusing a motion hypothesis from visual odometry to a set of other sensors (gray). Current approaches mostly focus on improving the robustness of the feature extraction and motion estimation process (green). Despite all effort, the estimate from visual odometry remains highly dependent on the vehicle's environment. Consequently, to achieve the availability constraints of automotive application it has to be embedded in a sensor fusion system. Unfortunately, a false visual odometry estimate $(\mathbf{R}_{VO}, \mathbf{T}_{VO})$ induces a bias to the result of a filter-based sensor fusion (blue). In order to ensure a high quality of the overall motion estimation process, it therewith becomes inevitable to develop a self-validation scheme (red) which judges the quality of the visual odometry estimate.

[1]All authors are with Control Methods & Robotics Lab, TU Darmstadt, Germany. www.rmr.tu-darmstadt.de/{surname}

odometry systems, which is the number of available features. Based on this, we present our approach in Sec. IV. Here, two different models are derived, that achieve redundancy between different estimated degrees of freedom. In the subsequent Sec V, we prove that these models can be used to self-validate the motion hypothesis from visual odometry. A conclusion and outlook are given in Sec. VI.

II. Robust Visual Odometry

Visual Odometry can be structured into two main processes: The search for **feature correspondences** aims at finding temporal correspondences of a landmark as well as spatial correspondences between the left and right camera. This results in the feature set $\mathscr{F}$. The **motion estimation** uses this feature set $\mathscr{F}$ to generate a motion hypothesis $(\mathbf{R}_{VO}, \mathbf{T}_{VO})$, that describes the camera's motion. Because of the high relevance for the automotive application of visual odometry, numerous developments to increase the robustness against critical scenarios and systematic weaknesses were made, which we will address in the following section.

A. Robust Feature Correspondences

Many steps to increase the robustness of visual odometry algorithms were developed, evaluated and have become common today. Already at the search for corresponding features, sophisticated pipelines evolved over time. Motion model based prediction of the optical flow as e.g. applied in [1] in conjunction with forward/backward matching of the optical flow vectors from $t-1$ to t and from t to $t-1$ as proposed in [6], [10] or even in a full loop search over the left and right images at $t-1$ and those at time t as in [1], [4] are commonly applied in state of the art methods. With the use of suitable motion models and redundant calculation circles, the quality of the resulting features can be drastically improved. The authors of [4] furthermore propose the selection of features according to their structure and additionally track the features over multiple frames to increase consistence. The same idea of preferring features that are trackable over longer periods of time is used in [16].

B. Robust Motion Estimation

Also when calculating the motion estimate based on the resulting feature set $\mathscr{F}$, numerous developments were achieved during the last years. The most common probably is the application of random sample consensus (RANSAC) for visual odometry as proposed in [4], [11]. Here, a minimal sample of features that theoretically suffices to calculate a motion hypothesis is randomly chosen from the overall feature set. The resulting motion hypothesis is used as a reference hypothesis with respect to which an error of all remaining features is calculated. The features with an error below a predefined threshold are considered as inliers for the current motion hypothesis. After a predefined number of repetitions of this draw-and-evaluate-run, the motion hypothesis with the highest number of inliers is selected as the best hypothesis and the set of inliers for this hypothesis is chosen as the final inlier set on the base of which the final motion estimation is carried out. A concurrent approach is the maximum subset outlier rejection (MASOR) as presented in [1], [2]. Here, the quality of the correspondences is assumed to be high enough to allow a good initial motion estimate based on the full feature set. Next, the error for each feature with respect to this motion hypothesis is calculated and the features with the highest errors are removed from the inlier set. In the next iteration, a new motion hypothesis based on the reduced feature set is calculated and the evaluation and outlier rejection is repeated. This procedure is continued until a termination criterion is met as e.g. a maximum number of iterations. Besides different schemes for outlier detection, also advanced measures were introduced as e.g. in [1], [2], where the coordinate-variance could be approximately eliminated from the commonly used reprojection error, respectively the directional error in [3]. This results in superior detection ability of measurement errors. In addition to that, current approaches more and more exploit long-term considerations on the level of frame-to-frame estimates. The authors of [13], [16] rely on bundle adjustment on top of their already robust motion estimation pipeline to keep their motion estimates consistent. This method refines a set of frame-to-frame motion estimates in a windowed optimization scheme over time. Also the approach from [4] benefits from the idea of smoothing the single motion estimates by averaging over different estimates and by keyframe based optimization over multiple frames.

C. Visual Odometry Self Validation

The already presented class of methods aims at improving the robustness of different parts of the visual odometry pipeline. Here, the aim was to avoid false motion estimates. By contrast, the following approaches try to define indicators that can be used to mark the outcome of the visual odometry scheme as valid or invalid. The authors of [9] state that for the automotive application, a failure of visual odometry is usually acceptable if this failure is detected and the estimate can be marked as incorrect. By contrast, unrecognized false estimates may lead to drastic consequences. Hence, they include detection schemes in the algorithm to classify false estimates as incorrect. They discard their estimate if the features are not sufficiently spread over the image. Furthermore, they demand at least ten features to be available and the mean reprojection error of the used feature set with respect to the final motion estimate to lay below a predefined threshold. In [13] at least 50 features are demanded for the visual odometry. Otherwise, the authors propose to use a motion model as fallback strategy. A more dynamic-based method is used in [12]. Here, the authors use a constant velocity and constant angular velocity model that runs in parallel to their visual odometry pipeline. In order to judge the consistency between the estimate and the motion model, they evaluate the Kalman innovations and reject the visual odometry result, if it is too large.

III. Number of Feature Correspondences

Isolated wrong motion estimates of a visual odometry system mainly originate in two different classes of reasons: If the vast majority of measured features is incorrect, the robust motion estimation process is bound to fail. However, if the vast majority of measured features is correctly measured, the motion estimation process still may fail, which is mostly due to the convergence towards distant points as discussed e.g. in [2]. This incorrect convergence can for example be taken care of by applying the decoupled normalized reprojection error, that is presented in [1]. Since our considerations base on the implemented visual odometry from this publication, we here assume the main reason for failure of the system to be erroneous correspondences. This motivates to investigate the number of available features as an indicator for the quality of the resulting motion estimate.

A. Absolute Number of Feature Correspondences

As pointed out in Sec. II, one of the few yet presented measures to flag a visual odometry result as invalid is the number of correctly matched features that are available for the motion estimation. This seems to be intuitive, since most approaches base on internal motion models for the prediction of the optical flow. Consequently, a high number of successfully tracked features indicates that large parts of the observed scene fit to the motion model and seem to be plausible measurements. By contrast, a low number of corresponding features might indicate a discrepancy between the motion model and the optical flow from the images. For judging the number of corresponding features as a quality-metric, we consider the scenario from KITTI track 07 that is depicted in Fig. 2. During the crossing of the truck in the top image, only very few optical flow measurements remain. This is because of the fact, that the crossing motion of the truck stands in contrast to the optical flow that is used as a prediction for the optical flow estimation, as described in [2]. As a result, only few optical flow vectors are successfully calculated. These focus on the right part of the camera image. Consequently, the number of matched features drops to about 10 % of the maximum available number around frame 640 in the center graph. Despite this heavy drop, no visible error can be seen in the right image: As an example for all degrees of freedom we here consider the forward speed v_z in the bottom image: The visual odometry estimate is plotted in red on top of the green reference signal. This consideration proves, that the absolute number of features is not a suitable indicator of the estimate's quality. If there is a low number of good features available, the estimate may meet or even exceed the quality of an estimate based on a larger feature set, if the remaining features are of good quality. As a consequence, we suggest to not discard an estimate only based on the absolute number of features in the feature set.

B. Relative Number of Feature Correspondences

As discussed in Sec. III-A, the absolute number of features is not a suitable criterion to evaluate the quality of the motion estimation. We therefore conclude with the change of the

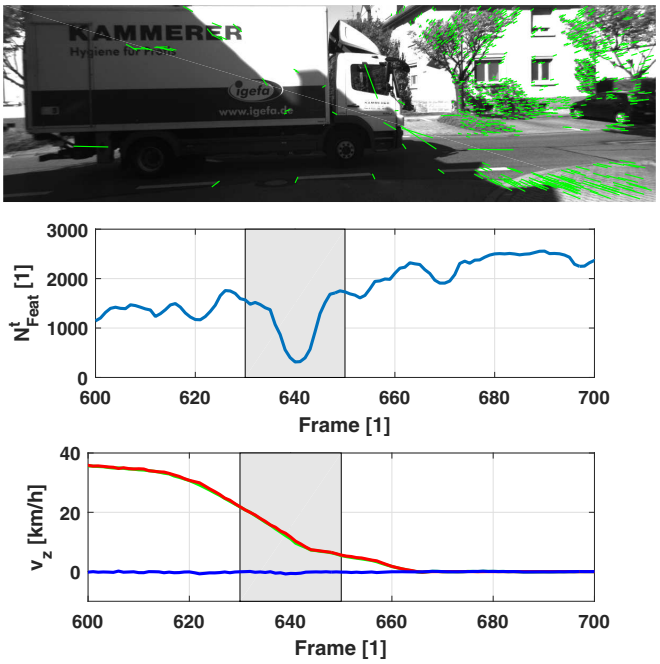

Fig. 2. The top image shows a truck, crossing the ego vehicle from left to the right. As can be seen, only few optical flow vectors are found on the truck. Nevertheless, the remaining parts of the image suffice to reconstruct the ego-motion with very high quality. The center image shows the number of matched features, that are available for the motion estimation. During the crossing of the truck (gray window), this number decreases to about 300 of the possible 3000 features, which could be assumed to indicate a bad motion estimate. Comparing the green reference data of the forward speed v_z in the bottom image with the red visual estimate, drawn onto the green line, no relevant error can be observed during the crossing. Also the other five degrees of freedom which are not shown here, do not show any relevant deviation. This means that even a rate of only 10 % of the possible features is not necessarily indicating a failed estimate.

number of available features: We assume that the frame rate of the cameras f is high enough to consider the perception of the environment of the vehicle as continuous. Therewith, the number of matched features N^t_{Feat} can be assumed to behave approximately steady in the case of correct matching and therefore $N^{t-1}_{\text{Feat}} \approx N^t_{\text{Feat}}, \forall t \in \mathbb{N}^+$. In this case, the change of available features $\Delta_{\text{Feat}} = N^t_{\text{Feat}} - N^{t-1}_{\text{Feat}}$ should be small, if the visual odometry works correctly. However, in reality the discrete imaging times and the effect of uneven feature distribution have to be taken into account. This is especially required due to the fact, that features are only found on structured areas of the image and therefore tend to cluster in distinct areas. In order to evaluate the change of the amount of features Δ_{Feat} over time as a validation signal, Fig. 3 shows an example from KITTI track 02. As can be seen, despite a change Δ_{Feat} of more than 500 features around frame 280, no extraordinary error in the forward speed v_z can be observed. Also none of the remaining five degrees of freedom motions shows a deviation that goes beyond that of times with minor deviation of the number of features. This shows, that also the change of features is not suitable to reliably determine, whether the current estimate is successful or not.

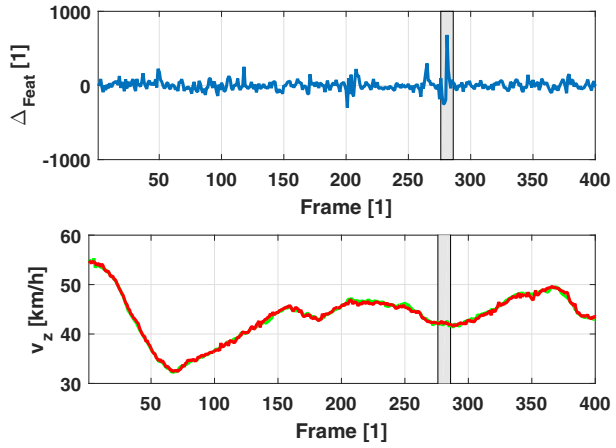

Fig. 3. Despite a change of more than 500 features at the marked time in the top graph, only little error at the forward speed v_z occurs in the bottom graph. Also none of the remaining five degrees of freedom shows a deviation that goes beyond the normal level, which is not shown here.

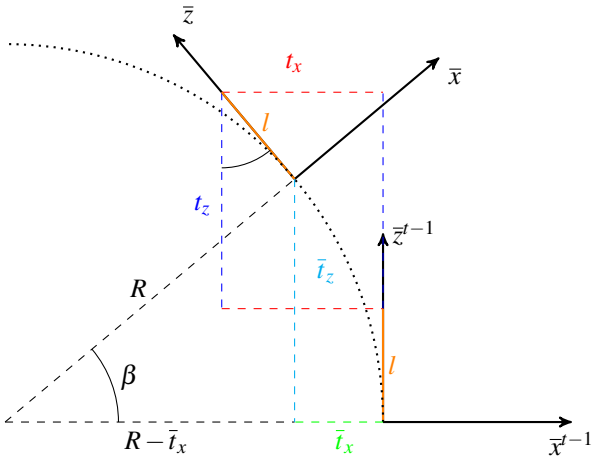

Fig. 4. Under the assumption of a planar circular vehicle motion with radius R, an analytic relation between the camera's sideward translation t_x, forward translation t_z and yaw angle β can be formulated. The required items for this are visualized in this figure. Here, x^{t-1}/z^{t-1} stand for the planar camera coordinate system at $t-1$, while x/z represent the camera coordinate system after the vehicle motion $(\mathbf{R},\mathbf{T})$ from time $t-1$ to time t.

IV. Sideward Motion Modeling

In this section, we show that the six degrees of freedom of an automotive motion are not independent of each other. Assuming planar motion, the sideward speed v_x can be modeled as a function of the yaw rate $\dot{\beta}$ and the forward speed v_z. As turned out, an even simpler model can be used, that describes the sideward speed v_x as a function of the yaw rate $\dot{\beta}$. In the following, we present the **two parameter model** $v_x(\dot{\beta},v_z)$ and the **one parameter model** $v_x(\dot{\beta})$.

A. Two Parameter Model

As stated in [15], Ackerman vehicles always have an instant center of rotation around which the vehicle rotates. In theory, it lies on the axis through the rear axle of the vehicle. Because of this, we assume a planar circular vehicle motion as done in [14] for outlier detection. In the following, the overline indicates coordinates and motions with respect to the center of the rear axle. We model the camera to be mounted in front of the vehicle's center of the rear axle with distance l. This leads to the schematic in Fig. 4.

The right triangle on the left side of Fig. 4 with hypotenuse R and legs $\bar{t}_z$ respectively $(R-\bar{t}_x)$ is used to formulate the following equation:

$$(R-\bar{t}_x)^2+\bar{t}_z^2=R^2 \tag{1}$$

$$\Rightarrow \bar{t}_x^2-2R\bar{t}_x+\bar{t}_z^2=0. \tag{2}$$

Solving (2) for the rear axle's sideward translation $\bar{t}_x$ yields:

$$\bar{t}_{x,1/2}=R\pm\sqrt{R^2-\bar{t}_z^2}. \tag{3}$$

For the relation between the radius R and the rear axle forward translation $\bar{t}_z$ and the yaw angle β, it holds that

$$R\sin\beta=\bar{t}_z. \tag{4}$$

Plugging this back into (3) is used to eliminate the unknown radius R and thereby leads to the following expression for the rear axle sideward translation $\bar{t}_{x,1/2}$:

$$\bar{t}_{x,1/2}=\bar{t}_z\frac{(1\pm\cos\beta)}{\sin\beta}. \tag{5}$$

For realistic vehicle motions with $|\bar{t}_z|\gg|\bar{t}_x|$ and $|\beta|\ll\frac{\pi}{2}$, only one solution remains for the rear axle's sideward translation $\bar{t}_x(\beta,\bar{t}_z)$:

$$\bar{t}_x(\beta,\bar{t}_z)=\frac{\bar{t}_z}{\sin\beta}(1-\cos\beta). \tag{6}$$

Based on the trigonometric relations in Fig. 4, two relations between camera motion and rear axle motion can be derived to take the camera mounting position into account:

$$\bar{t}_x=t_x-l\sin\beta, \tag{7}$$

$$\bar{t}_z=t_z+l-l\cos\beta. \tag{8}$$

Integrating these into (6) and exploting $v_x=ft_x$, $v_z=ft_z$ and $\dot{\beta}=f\beta$, we get the model for the sideward speed v_x:

$$v_x(\dot{\beta},v_z)=f\frac{\frac{v_z}{f}+l(1-\cos\frac{\dot{\beta}}{f})}{\sin\frac{\dot{\beta}}{f}}(1-\cos\frac{\dot{\beta}}{f})+l\sin\frac{\dot{\beta}}{f}. \tag{9}$$

According to the KITTI benchmark documentation the camera is mounted 1.03 m in front of the rear axle, see [7], [8] for details. The resulting modeled sideward speed $v_x(\dot{\beta},v_z)$ is visualized as a yellow graph in Fig. 5, which shows 1000 frames from KITTI track 02. The result for assuming the camera to be mounted at the center of the rear axle is shown in cyan.

For reference, the sideward speed v_x, which we want to model is shown in red. As can be seen, assuming the camera to be at $l=0$ as well as modeling it according to the data sheet $l=1.03$ leads to a consistent misestimation of the actual sideward speed v_x. Because of this, we optimize the shift l to the rear axle in a least squares approach to

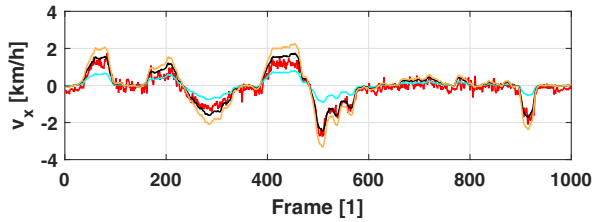

Fig. 5. Optimizing the camera shift of the camera leads to a significantly improved model reconstruction (black), compared to the non-shifted model (cyan) and the shift according to the data sheet (yellow). The optimized model is close to the measurement at all considered amplitudes of the sideward speed v_x estimate from the visual odometry system (red).

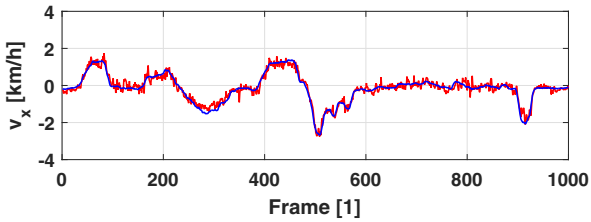

Fig. 6. Comparison between the visual odometry estimate (red) and the linear least squares model (blue). The model succeeds to describe the characteristics of the system very well, while only requiring two constants to be determined.

minimize the error between motion and model. This results in a mounting position $l \approx 0.66\,\mathrm{m}$ in front of the rear axle. The emerging graph is shown in black in the figure. As can be seen, the modeling accuracy is much higher than with the position from the data sheet. We therefore use the optimized parameterization for the evaluation in Sec. V.

B. One Parameter Model

After the analytical derivation of a model with good precision, we proceed with modeling the sideward speed v_x solely from the yaw rate $\dot{\beta}$. Current visual odometry approaches as those presented in [1], [5] propose separate estimation processes for the rotation and translation to increase the overall motion hypothesis quality. A model $v_x(\dot{\beta})$ therefore bears the advantage of being able to reconstruct one degree of the translation estimation process from one degree of the rotation estimation process. Therewith, the results of the two separate estimation processes can be cross-evaluated by using the model $v_x(\dot{\beta})$. Because of this, the in practice much more robust rotation estimation can be used to evaluate the much more sensitive translation estimation. As turned out, a very simple model suffices to achieve this. This model consists of two setup-related constants c_1 and c_2 and is formulated as

$$v_x(\dot{\beta}) = c_1\dot{\beta} + c_2. \tag{10}$$

These constants were determined by a data-driven least squares optimization of the deviation between the output of the model from (10) and the measured visual odometry data. For the evaluation of our model, Fig. 6 shows the comparison between the estimate of the sideward speed v_x from the visual odometry system in red and from the one parameter model in blue. As can be seen in the example from KITTI track 02, the estimates and the model show consistent behavior with only minor deviations. Despite requiring only the yaw rate $\dot{\beta}$ as an input, the sideward speed v_x can be modeled very precisely with the one parameter model.

V. Evaluation

For the judgment of the two parameter model and the one parameter model, we evaluate the mean localization error over all tracks of the KITTI benchmark. The baseline is the result of the **RotROCC** method, that was published in [1] and achieves a mean error of 0.74 % on the test data. Setting the sideward translation estimate t_x to zero after the actual estimation process leads to a significant drop of the performance to 0.90 %. This shows the importance of the sideward motion for the overall trajectory. In contrast to this result, using the modeled sideward translation t_x instead of the actual estimate, the two parameter model achieves an error of 0.75 %, while the one parameter model even leads to a result of 0.73 %. This proves, that the two parameter model as well as the one parameter model lead to a very good estimate of the sideward translation, which can compete with the output from a state of the art visual odometry system.
For detecting the failure of a visual odometry estimate, we use a two model scheme. This scheme consists of the one parameter model to cross-validate the forward speed v_x and the yaw rate $\dot{\beta}$ and of the two parameter model, which additionally considers the forward speed v_z. If the one parameter model shows vast deviations, the estimated sideward speed v_x does not fit to the estimated yaw rate $\dot{\beta}$. If the second model shows vast deviations despite the first model surveillance not facing errors, the forward speed v_z is assumed to be wrong. Therefore, we consider the visual odometry system as failing if at least one surveillance stage shows large deviations between measured and modeled sideward speed v_x. We define the self-validation signal $\mathbf{Q}(v_x)$ as a vector, consisting of the error of the two parameter model and the error of the one parameter model:

$$\mathbf{Q}(v_x) = \begin{pmatrix} v_x - v_x(\dot{\beta}, v_z) \\ v_x - v_x(\dot{\beta}) \end{pmatrix}. \tag{11}$$

For the detection of system failure, we analyze the frames 800 to 1000 of KITTI track 01 that contain a clear estimation error. The result is depicted in Fig. 7. The top graph shows the first dimension of the self-validation signal $\mathbf{Q}(v_x)$ in blue and the second in black. In this case, low quality is reported close to the end, when the blue and black lines reach the highest amplitudes. Taking a look at the red visual estimate of the sideward speed v_x in the center graph, it turns out that the detection is correct: The behavior of the estimated v_x, that is described by the red graph is not realistic and in this case caused by very few structure in the image during a harsh curve. By contrast, the blue/black output from the two/one parameter model shows smooth and realistic behavior. The failure can also be seen in the bottom graph, which shows the estimated forward speed v_z in red and the reference data in green. When the self-validation signal indicates low quality,

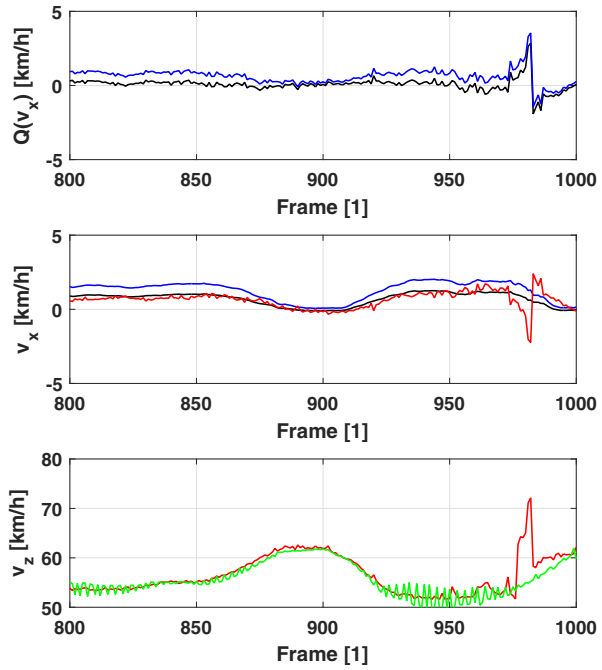

Fig. 7. The top graph shows the self-validation signal $\mathbf{Q}(v_x)$ for the two parameter model (blue) and one parameter model (black) that are used as self-validation signal for the visual odometry motion estimation over time. Throughout the track, only minor deviations between the modeled and estimated sideward speed v_x occur. As can be seen, the one parameter model (black) again shows higher reconstruction quality than the two parameter model. Towards the end, both show a drastically increased error and therewith state low estimation quality. As can be seen by the impulse-like deviation between the plausible shape of the two motion model (blue/black) and the visual estimate (red) in the center graph, this detection is correct. The bottom graph shows the estimate (red) and reference data (green) of the forward speed v_z. Towards the end, highly volatile behavior occurs, where the self-validation signal $\mathbf{Q}(v_x)$ reports low quality. This proves the incorrect estimate and correct self-validation.

also the estimate deviates largely from the reference signal. In order to test our model-based self-validation, we extended the RotROCC algorithm from [1] with this detection stage. As no other sensor is available in this setup, we try to detect large errors. Consequently, we rely on a constant turn rate and velocity model, whenever the maximum absolute value of $\mathbf{Q}(v_x)$ exceeds a threshold. The resulting algorithm **RotROCC+** is able to outperform the basic method by more than 5.5 % and therewith achieves a localization error of 0.83 %, being the best frame-to-frame visual odometry in the KITTI benchmark. We assume the real-world impact to be even higher, since most scenarios in the KITTI benchmark are not very demanding e.g. in terms of illumination or dense traffic. Furthermore, we only use high amplitudes of $\mathbf{Q}(v_x)$ in this test, that state a vast motion estimation failure.

VI. Conclusion and Future Work

In this paper, we motivate the need for intrinsic surveillance of a visual odometry system's motion estimate. To solve this task, we present an analytically derived two parameter model to recover the sideward speed v_x from the forward speed v_z and the yaw rate $\dot{\beta}$. Subsequently, a data-driven one parameter model is shown, which models the sideward speed v_x from only the yaw rate $\dot{\beta}$. Despite its simplicity, this model achieves even better reconstruction quality than the two parameter model. Both models can be used for system self-validation by comparing the modeled sideward speed v_x to the estimate from the visual odometry. Because of the absence of further sensor data, we focus on detecting system failure here. Future work should evaluate the influence of $\mathbf{Q}(v_x)$ on the quality of a sensor fusion. Also, the substitution of the sideward motion in the translation estimation process to speed up visual odometry should be investigated. Finally, the one parameter model can potentially serve to extend the compensation of single degrees of freedom from [1] in the measured optical flows. The method that was proposed there shows superior outlier rejection capability under the assumption of one degree of freedom motions. After compensating the rotation, the sideward motion was assumed to be approximately zero, there. This could be assured, if the sideward motion is known from the here proposed model.

Acknowledgments

We kindly thank Continental AG for funding this project.

References

[1] M. Buczko and V. Willert. Flow-decoupled normalized reprojection error for visual odometry. In *19th International IEEE Conference on Intelligent Transportation Systems*, pages 1161–1167, 2016.

[2] M. Buczko and V. Willert. How to distinguish inliers from outliers in visual odometry high-speed automotive applications. In *IEEE Intelligent Vehicles Symposium*, pages 478–483, 2016.

[3] M. Buczko and V. Willert. Monocular outlier detection for visual odometry. In *IEEE Intelligent Vehicles Symposium*, pages 739–745, 2017.

[4] I. Cvišić, J. Ćesić, I. Marković, and I. Petrović. Soft-slam: Computationally efficient stereo visual slam for autonomous uavs. In *Journal of Field Robotics*, 2017.

[5] I. Cvišić and I. Petrović. Stereo odometry based on careful feature selection and tracking. In *European Conference on Mobile Robots*, pages 1–6, 2015.

[6] J. Deigmoeller and J. Eggert. Stereo visual odometry without temporal filtering. In *German Conference on Pattern Recognition*, pages 166–175. Springer International Publishing, 2016.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[9] A. Howard. Real-time stereo visual odometry for autonomous ground vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3946–3952, 2008.

[10] Z. Kalal, K. Mikolajczyk, and J. Matas. Forward-backward error: Automatic detection of tracking failures. In *20th IEEE International Conference on Pattern Recognition*, pages 2756–2759, 2010.

[11] B. Kitt, , A. Geiger, and H. Lategahn. Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In *IEEE Intelligent Vehicles Symposium*, pages 486–493, 2010.

[12] T. Oskiper, Z. Zhu, S. Samarasekera, and R. Kumar. Visual odometry system using multiple stereo cameras and inertial measurement unit. 120:1–8, 2007.

[13] M. Persson, T. Piccini, M. Felsberg, and R. Mester. Robust stereo visual odometry from monocular techniques. In *IEEE Intelligent Vehicles Symposium*, pages 686–691, 2015.

[14] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *International Journal of Computer Vision 95.1*, 95(1), 2011.

[15] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza. *Introduction to Autonomous Mobile Robots*. MIT Press, 2011.

[16] S. Song, C. Chandraker, and C. Guest. Parallel, real-time monocular visual odometry. In *IEEE International Conference on Robotics and Automation*, 2013.