2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

Flow-Decoupled Normalized Reprojection Error for Visual Odometry

Martin Buczko[1] and Volker Willert[1]

Abstract— **In this paper, we present an iterative two-stage scheme for precise and robust frame-to-frame feature-based ego-motion estimation using stereo cameras. We analyze the characteristics of the optical flows and reprojection errors that are independently induced by each of the decoupled six degrees of freedom motion. As we will show, the different characteristics of these induced optical flows lead to a reprojection error that depends on the coordinates of the features. When using a proper normalization of the reprojection error, this coordinate-dependency can be almost completely removed for decoupled motions. Furthermore, we present a way to use these results for automotive application where rotation and forward motion are coupled. This is done by compensating for the flow that is induced by the rotation, which decouples the translation flow from the overall flow. The resulting method generalizes the ROCC approach [4], where a robust outlier criterion was introduced and proved to increase robustness and quality for large forward translation motions. Therewith the proposed method generalizes ROCC to almost all possible automotive motions. The performance of the method is evaluated on Kitti benchmark and currently[2] reaches the best translation error of all camera-based methods.**

I. Introduction

For autonomous driving, robust and precise self-localization of vehicles is one of the main challenges. Hence, great effort is put into the improvement of visual odometry methods to obtain additional localization measurements for automotive applications, as can be seen in numerous publications e.g. in the odometry section of Kitti benchmark [7]. Visual odometry approaches estimate the ego-motion of a vehicle. This ego-motion consists of a rotation component **R** and a translation component **T**, whereas each induces a specific optical flow pattern. Comparing the best performing visual odometry methods, one similarity turns out: Rotation and translation of the ego-motion are calculated in two separate processes, as in [4], [5]. This is intuitive, since there are fundamental differences between rotation and translation estimation, which are exploited by this separation: The reconstruction of the translation is dependent on a depth estimate and therefore sensitive to multiple error sources: Ambiguous correspondences in the stereo matching can lead to wrong disparities, respectively depths. Furthermore, the resolution of the reconstructed depth of each feature reduces quadratically with disparity. Additionally, the optical flow from translation decreases proportionally to the feature's inverse depth (see Sec.IV-A). By contrast, the flow from rotation is not dependent on depth and is therefore not susceptible to these effects. This means, that ideal features for rotation estimation are not necessarily good choices when estimating translation. In [4], [5] this is taken care of by setting up a dedicated process for each of these estimations. This enables these methods to achieve mean rotation errors of 0.002 to 0.003 °/m without reoptimizing the structure (bundle adjustment). Splitting the overall motion estimate into independent processes for the estimation of rotation and translation obviously is a helpful first step.

In contrast to handling rotation and translation estimation independently, we propose to use the rotation estimate within the translation estimate to improve outlier detection by decoupling the optical flow components, which are induced by the rotation and translation parts of the ego-motion. Before discussing the details of the proposed method, Fig.1 shows the result of the improved outlier detection for a turning maneuver. In the upper image, the reprojection errors of three ideal measurements are plotted as green bars and

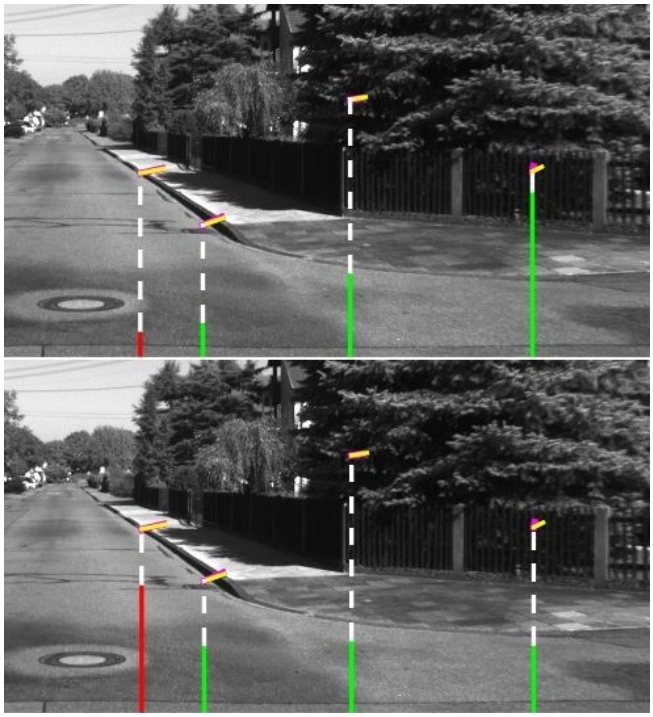

Fig. 1. In each image, the measured optical flow (yellow) and estimated flow according to an error-prone motion hypothesis (magenta, top of dashed white lines) of a right-turn are shown. The reprojection error (top) is visualized by the red and green bars. Our proposed decoupled normalized reprojection error is shown in the lower image. The errors of ideal measurements are marked green and the left feature with erroneous optical flow is marked red. As one can see, the reprojection error judges the real outlier with a low error and some of the inliers with a high value. By contrast, our proposed decoupled normalized reprojection error (bottom) shows an almost constant error for all features due to the error-prone motion hypothesis and a further increased value for the erroneous correspondence.

[1]Control Methods and Robotics Lab,
TU Darmstadt, Germany.

[2]At time of paper submission September 13, 2016.

Fig. 2. Comparison between high-speed scenario (left column) and low-speed turning maneuver (right column). The top line shows the rotation-only flow (green). The translation-only flow is shown in the bottom line (yellow). The high-speed motion shows only minor or no rotation, whereas the turning maneuver shows vast rotation-induced flow components. By decoupling the rotation and translation motion induced flows, also turning maneuvers can be transformed into a pure translation, quasi-high-speed scenario. This decoupling transformation allows the application of an almost coordinate-invariant outlier detection scheme, which we show in this paper.

that of a falsified correspondence as a red bar, for an erroneous translation estimate and ideal rotation estimate. The inliers are judged with high reprojection errors whereas the outlier receives a small error. The reason for this is that the reprojection error is dependent on the image coordinates of the features and thus is not a proper criterion to differ between errors that stem from wrong correspondences and error-prone motion hypotheses, as will be derived in Sec.IV. By contrast, our proposed approach in the bottom of Fig.1, which is presented in Sec.V, identifies the outlier with a high error and sets an almost constant offset for the inliers. As mentioned, the idea here is to not estimate rotation and translation in isolated processes: First, we estimate the rotation only. Next, we use this estimated rotation to transform the measured correspondences into a pure translation scenario. This transformation is described in Sec.IV-C and visualized in Fig.2: The high-speed scenario in the left column shows no rotation flow component (green), but forward translation flow (yellow). By contrast, the low-speed scenario in the right column shows vast rotation components (green). After compensating for these in the overall flow, the resulting translation flow (yellow) shows the characteristics of the high-speed flow. This decoupled translation flow allows the application of an almost feature-coordinate independent outlier criterion, leading to a much more precise outlier detection.

Before going into detail, we present the relevant literature in Sec.II. After defining our notation and the basic pipeline in Sec.III, we investigate outlier measures for independent and joint ego-motion flows in Sec.IV. Our proposed outlier criterion, which decouples translation flow from rotation flow, making use of the results in Sec.IV, is explained in Sec.V. The evaluation of the resulting outlier scheme is done via simulation in Sec.IV and with real data in Sec.VI.

II. Related Work

One possibility to divide and characterize outlier detection methods for visual odometry are the different motion models, that are assumed to describe the vehicle's pose changes.

The first category uses a full six degree of freedom approach, which is the most general way to find a motion hypothesis. In [1], [9], this is done by minimizing the reprojection error in a RANSAC framework. Here in each iteration, a full motion hypothesis is created, based on a minimum number of random samples from the correspondences. This hypothesis is considered as reference motion and gets valued by the reprojection error of the remaining features. This is iteratively repeated until a termination criterion is met. The best hypothesis is then used to finally divide all features into inliers and outliers and the resulting hypothesis is calculated based on these features.

The second category of visual odometry methods uses a restrictive motion model and performs outlier rejection within this subspace. An example for this restrictive motion assumption can be found in [14], where the vehicle's motion is limited to forward translation, pitch and yaw. In [11], an even more restrictive model is used. Here, a locally planar and circular vehicle motion is assumed. This allows to apply Ackermann's steering principle and therewith to describe the motion with one rotation parameter plus the unknown scale. Through this restrictive motion hypothesis the number of necessary correspondences is reduced to just one, which leads to a very fast outlier removal and motion estimation scheme. To summarize, if the vehicle's motion is covered by the assumed model, these approaches are very effective and lead to fast and accurate results. However, if the vehicle's motion does not fit to the assumed model the outlier rejection stage looses good correspondences, causing the estimate to describe a false motion within the assumed subspace.

Our proposed method retains the advantages of a full motion estimate, while preserving the advantages of a restrictive motion model. For this, we separate the estimation of rotation and translation into two dedicated processes, which are coupled by a transformation. This transformation compensates for the rotation within the measurement and therefore allows to apply a restrictive motion model that assumes pure translation without reducing the degrees of freedom, which are taken into account. Fig.3 illustrates the concept briefly.

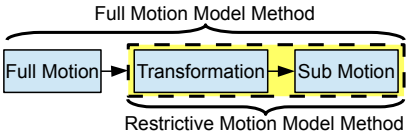

Fig. 3. Using a transformation, allows the application of a restrictive motion method while estimating the full six degrees of freedom motion.

III. Definition of the Overall Problem

The backbone of our motion estimation is the classical least squares estimator for pose change $(\hat{\mathbf{R}}^t, \hat{\mathbf{T}}^t)$

$$(\hat{\mathbf{R}}^t, \hat{\mathbf{T}}^t) = \text{argmin}_{\mathbf{R},\mathbf{T}} \sum_{n=1}^{N^t} \left(\varepsilon_n^t\right)^2, \tag{1}$$

$$\varepsilon_n^t = \|\mathbf{x}_n^{t-1} - \pi(\mathbf{R}\lambda_n^t \hat{\mathbf{x}}_n^t + \mathbf{T})\|_2, \tag{2}$$

where the norms ε_n^t are called the reprojection errors. These are calculated for each feature f_n^t indexed by n at time t within the feature set $\mathscr{F}^t = \{f_n^t\}_{n=1}^{N^t}$. The standard planar projection $[X,Y,Z]^T \mapsto [X/Z, Y/Z, 1]^T$ is denoted as π, with lateral coordinate X, transversal coordinate Y and forward coordinate Z. Here, $\{\mathbf{x}_n^{t-1}, \mathbf{x}_n^t\} \in \mathbb{R}^3$ is the correspondent pair of pixel coordinates denoted in homogeneous coordinates $\mathbf{x}_n^t = [x_n^t, y_n^t, 1]^T$ with depth $\lambda_n \in \mathbb{R}$. The respective image coordinates[3] at time t can be written as

$$\hat{\mathbf{x}}_n^t = \left[\hat{x}_n^t, \hat{y}_n^t, 1\right] = \left[\frac{x_n^t - o_x}{f}, \frac{y_n^t - o_y}{f}, 1\right]^T. \tag{3}$$

Now, we are faced with the main problem of visual odometry: Given the set of all extracted features, we need to find suitable features (inliers) and reject all other features (outliers) from the set. The standard measure to judge the features when given a motion hypothesis, is the reprojection error from Eq.(2). In the following, we analyze it with respect to each degree of freedom motion independently and compare it to a coupled motion afterwards.

IV. Characteristics of Independently and Jointly Induced Motion Components

We start with the pixel coordinate of a feature, being induced by a generic motion with pitch α, yaw β, roll γ and translations to the side t_x, downwards t_y, and forwards t_z. Justified by the data from Kitti benchmark and reasonable camera frame rates of at least 10 fps, we assume limited pitch, yaw and roll of $\alpha, \beta, \gamma < \frac{5^\circ \pi}{180^\circ}$ per frame and therefore approximate the trigonometric functions with their first order Taylor series at operating point 0, leading to rotation matrix $\mathbf{R_z R_y R_x} = \mathbf{R}$:

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \approx \begin{pmatrix} 1 & \alpha\beta - \gamma & \alpha\gamma + \beta \\ \gamma & 1 + \alpha\beta\gamma & \beta\gamma - \alpha \\ -\beta & \alpha & 1 \end{pmatrix}. \tag{4}$$

We then obtain the general induced coordinate by full motion

$$\begin{pmatrix} x_n^{t-1} \\ y_n^{t-1} \end{pmatrix} = \begin{pmatrix} f\frac{\lambda_n(r_{11}\hat{x}_n^t + r_{12}\hat{y}_n^t + r_{13}) + t_x}{\lambda_n(r_{31}\hat{x}_n^t + r_{32}\hat{y}_n^t + r_{33}) + t_z} + o_x \\ f\frac{\lambda_n(r_{21}\hat{x}_n^t + r_{22}\hat{y}_n^t + r_{23}) + t_y}{\lambda_n(r_{31}\hat{x}_n^t + r_{32}\hat{y}_n^t + r_{33}) + t_z} + o_y \end{pmatrix}. \tag{5}$$

From this starting point, we now derive the reprojection error for independent motions and compare it to the normalized error which was presented for forward-only motion in [4].

A. Optical Flow from Independent Rotations

For pitch motion with angle α and an erroneous estimated pitch $\tilde{\alpha}$ of $\tilde{\alpha} = E\alpha$, Eq.(5) with $\alpha^2 E\hat{y}_n^t \ll 1$ and $\alpha\hat{y}_n^t \ll 1$ leads to an approximated reprojection error ε_n^t of

$$\varepsilon_n^t \approx f|\alpha(E-1)| \left\| \begin{pmatrix} \hat{x}_n^t \hat{y}_n^t \\ \hat{y}_n^t \hat{y}_n^t + 1 \end{pmatrix} \right\|_2, \tag{6}$$

which is dependent on the pixel coordinates of the feature. The same dependency occurs at the approximated error-free absolute value of the optical flow $\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2$ from the pitch motion, with $\alpha\hat{y}_n^t \ll 1$, using Eq.(3):

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 \approx f|\alpha| \left\| \begin{pmatrix} \hat{x}_n^t \hat{y}_n^t \\ \hat{y}_n^t \hat{y}_n^t + 1 \end{pmatrix} \right\|_2. \tag{7}$$

When dividing the reprojection error by the optical flow, the resulting normalized reprojection error (NRE) ν_n^t becomes independent of the feature coordinate and is characterized by the error of the motion hypothesis $|E-1|$ only:

$$\nu_n^t = \frac{\varepsilon_n^t}{\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2} \approx |E-1|. \tag{8}$$

In a similar way, reprojection error and optical flow measurement for an erroneous motion with yaw β and estimate $\tilde{\beta} = E\beta$ can be used to eliminate the reprojection error's dependency on the feature coordinates, with $\beta\hat{x}_n^t \ll 1$, $E\beta\hat{x}_n^t \ll 1$ and $E\beta^2\hat{x}_n^t\hat{x}_n^t \ll 1$:

$$\varepsilon_n^t \approx f|\beta(E-1)| \left\| \begin{pmatrix} \hat{x}_n^t \hat{x}_n^t + 1 \\ \hat{x}_n^t \hat{y}_n^t \end{pmatrix} \right\|_2, \tag{9}$$

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 \approx f|\beta| \left\| \begin{pmatrix} \hat{x}_n^t \hat{x}_n^t + 1 \\ \hat{x}_n^t \hat{y}_n^t \end{pmatrix} \right\|_2, \tag{10}$$

$$\nu_n^t \approx |E-1|. \tag{11}$$

Also for roll motion with incorrect estimation $\tilde{\gamma}$ with $\tilde{\gamma} = E\gamma$ the reprojection error ε_n^t is biased by the feature's position, which can be eliminated by normalizing with the optical flow $\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2$, leading to the normalized reprojection error ν_n^t:

$$\varepsilon_n^t = f|\gamma(E-1)| \left\| \begin{pmatrix} \hat{x}_n^t \\ \hat{y}_n^t \end{pmatrix} \right\|_2, \tag{12}$$

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 = f|\gamma| \left\| \begin{pmatrix} \hat{x}_n^t \\ \hat{y}_n^t \end{pmatrix} \right\|_2, \tag{13}$$

$$\nu_n^t = |E-1|. \tag{14}$$

[3] Focal length f and principle point $\mathbf{o} = [o_x, o_y]$ are assumed to be known.

B. Optical Flow from Independent Translations

The reprojection error and optical flow measurement for pure forward motion at hypothesis $\tilde{t}_z$ with $\tilde{t}_z = Et_z$ can be calculated as

$$\varepsilon_n^t = f \left| \frac{\lambda_n t_z (E-1)}{(\lambda_n + t_z)(\lambda_n + Et_z)} \right| \left\| \begin{pmatrix} \hat{x}_n^t \\ \hat{y}_n^t \end{pmatrix} \right\|_2, \quad (15)$$

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 = f \left| \frac{t_z}{(\lambda_n + t_z)} \right| \left\| \begin{pmatrix} \hat{x}_n^t \\ \hat{y}_n^t \end{pmatrix} \right\|_2. \quad (16)$$

Here the reprojection error is depending on the feature coordinate and the feature depth. With $\lambda_n \gg Et_z$, this approximately reduces to $|E-1|$ when normalizing:

$$\nu_n^t = \left| \frac{\lambda_n (E-1)}{\lambda_n + Et_z} \right| \approx |E-1|. \quad (17)$$

The dependency on the depth can also be eliminated by proceeding analogously with the reprojection error and optical flow for sideways motion with hypothesis $\tilde{t}_x = Et_x$,

$$\varepsilon_n^t = f \left| \frac{t_x (E-1)}{\lambda_n} \right|, \quad (18)$$

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 = f \left| \frac{t_x}{\lambda_n} \right|, \quad (19)$$

$$\nu_n^t = |E-1|. \quad (20)$$

For upward motion $\tilde{t}_y = Et_y$, as well the dependencies disappear for the normalized reprojection error

$$\varepsilon_n^t = f \left| \frac{t_y (E-1)}{\lambda_n} \right|, \quad (21)$$

$$\|\mathbf{x}_n^{t-1} - \mathbf{x}_n^t\|_2 = f \left| \frac{t_y}{\lambda_n} \right|, \quad (22)$$

$$\nu_n^t = |E-1|. \quad (23)$$

These results are visualized in Fig.4. Here, only ideal measurements are considered. Each line shows a one degree of freedom motion. The first three lines show rotations with pitch, yaw and roll of 3 °. The last three lines show translations with t_x, t_y and t_z of 1 m. Each feature is evaluated with reprojection error and normalized reprojection error for a non-ideal motion hypothesis with an error of 10 % leading to an error of 0.3 ° respectively 0.1 m.
For this setup, the reprojection error is heavily depending on the position of each feature for all motions. By contrast, the normalization of the NRE compensates for the flow characteristics of each motion and judges error-free correspondences with almost the same value for the performed isolated motions. This value is the relative motion hypothesis error $|E-1| = 0.1$. Minor deviations at pitch, yaw and forward translation come from the approximations, which were made in Sec.IV-A and IV-B. This is a highly relevant result, when decoupled motions can actively be performed, like at calibration. This prerequisite is usually not met in automotive localization, because the motion of the car is induced by the driver and restricted by the dynamics of the vehicles. How these results can still be applied in automotive context, is the scope of the next section.

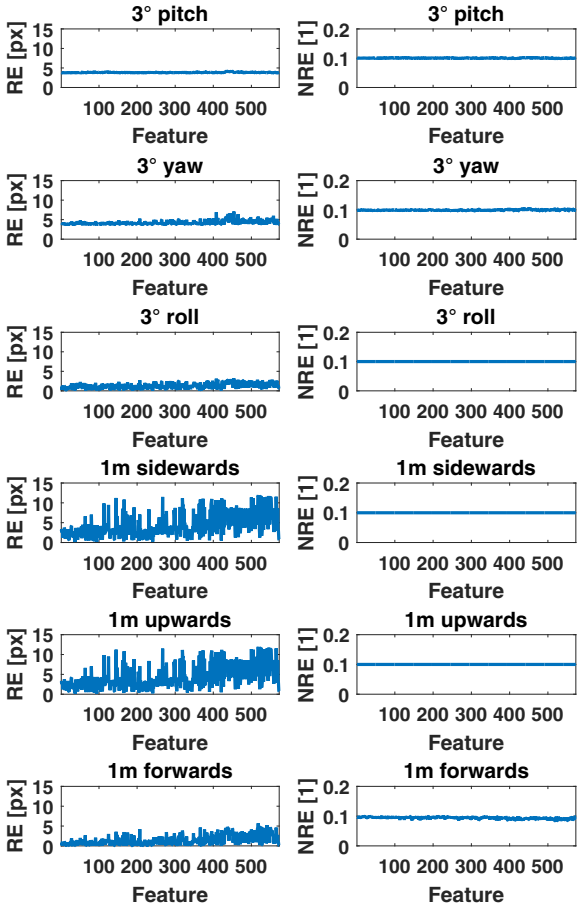

Fig. 4. Reprojection error (RE, left column) and normalized reprojection error (NRE, right column) for ideally measured flow vectors and motion hypothesis error of 10 %. In the first line, a pitch motion of 3 ° is shown. Here the reprojection error shows an almost constant error for all features. The normalized reprojection error also shows only minor variations due to the approximations from Sec.IV-A. In the second and third line, a yaw and a roll motion of 3 ° is compared. The reprojection error shows a heavy dependency on the feature position. By contrast, the normalized reprojection error shows an almost position-independent result for yaw and roll, with small deviations caused by the approximations, made in Sec.IV-A. The same pertains for the translation with 1 m along all three axes. Here, only the NRE for forward motion shows slight deviations, which come from the approximations, made in Sec.IV-A and are less than 0.0154. In all cases, the NRE is approximately the relative motion hypothesis error $|E-1| = 0.1$.

C. Optical Flow from Combined Motions

While independent motions allow a coordinate and depth independent feature judging scheme by normalizing the reprojection error with the optical flow, coupling the individual motions induces the dependencies again: An exemplary vehicle motion which shows this dependency is presented in Fig.5. Granting realistic conditions, we use a real-world feature distribution. The vehicle's virtual pose change at a framerate of 10 Hz consists of a turning motion of 50 °/s with simultaneous pitch of 10 °/s and roll of 5 °/s at a speed of 50 km/h. Here, we assume an error-free rotation estimate and error-prone translation estimate with $|E-1| = 0.05$. An ideal

outlier-detection scheme would lead to a constant offset, when regarding ideal measurements, as the hypothesis is the same for all features. By contrast, the reprojection error and the normalized reprojection error are highly dependent on the feature-position. This leads to errors between 0 and 5 pixels for the reprojection error and between 0 and 0.2 for the normalized reprojection error. This high variance masks the simulated depth errors of 5 % in the measurement for the features, which are marked by the red vertical lines.
This problem is solved by the **decoupled normalized reprojection error (DNRE),**

$$\delta_n^t = \frac{\varepsilon_n^t(\hat{\mathbf{R}}^t, \hat{\mathbf{T}}^t)}{\|\mathbf{x}_n^{t-1} - \pi\left(\hat{\mathbf{R}}^t \mathbf{x}_n^t\right)\|_2}. \tag{24}$$

Here, the flow from the vehicle's rotation is compensated by the estimated rotation with $\hat{\mathbf{R}}^t$. Therefore $\|\mathbf{x}_n^{t-1} - \pi\left(\hat{\mathbf{R}}^t \mathbf{x}_n^t\right)\|_2$ is the pure translation flow. With this, translation induced optical flow is decoupled from rotation induced flow. For an ideal estimate of the rotation $\mathbf{R}^t = \hat{\mathbf{R}}^t$, the reprojection error represents the error from translation only. This normalization transforms the measurement into the scenario of forward-only motion, which was described in Sec.IV-B. By doing so, the dependency on the feature coordinate is eliminated and outliers can be identified. The resulting errors of the decoupled normalized reprojection error are visualized in Fig.5, where an almost constant offset due to the motion hypothesis error of $|E-1| = 0.05$ affects every feature. A higher value indicates the evaluation correspondence errors, which here come from the added depth error of 5 %.

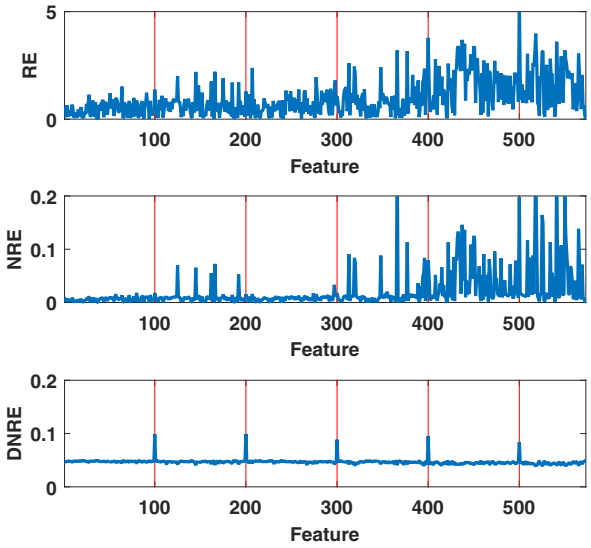

Fig. 5. Simulated comparison between reprojection error (RE, top), normalized reprojection error as proposed in [4] (NRE, middle) and decoupled normalized reprojection error (bottom, DNRE). Using a real-data feature distribution, the vehicle performs a turning motion of 50 °/s with simultaneous pitch of 10 °/s and roll of 5 °/s at a speed of 50 km/h. Features 100, 200, 300, 400 and 500 (marked with red lines) have an depth error of 5 %, while the other features are measured ideally. Reprojection error as well as the normalized reprojection error lead to heavily varying values due to the error that is induced by a 5 % deviation of the translation estimate. By contrast, the decoupled normalized reprojection error shows an almost constant offset of $|E-1| = 5\%$ for all features and additionally an increased value for the erroneous depth estimates.

V. Flow Decoupling Outlier Removal and Pose Refinement Scheme (RotROCC)

After having derived the criterion for improved outlier detection in Eq.(24), we now go through our full method in detail. Before estimating a motion hypothesis and identifying outliers, an initial set consisting of a reasonable number of suitable features is required. A suitable feature combines unambiguous temporal as well as stereoscopic correspondence measurements and is vital for reliable optical flow and depth estimates. This initial feature set is detected as follows in the next section.

A. Framework Outline

Our initial feature set for every stereo-frame-pair is created as follows, applying only standard functions of the OpenCV library [3]: We start with the feature-selection using the Shi and Tomasi method [13]. For each feature the disparity at time $t-1$ is calculated using sum-of-absolute-differences-based block matching. For optical flow initialization, we triangulate each feature's position in 3D space at time $t-1$ and reproject the features to the current frame at time t using a modified constant turn rate and velocity model based on the last estimated pose change (which is a variant of *motion model predicted tracking by matching* proposed in [10]). After that, the optical flow for the left and right image between time $t-1$ and t is refined with the Lucas-Kanade method [2]. The final feature set $\mathscr{F}_0^t = \{\mathbf{x}_n^{t-1}, \mathbf{x}_n^t, \lambda_n^t\}_{n=1}^{N_0^t}$ with a starting number N_0^t for initialization is reached via a left-right consistency check at time t for all remaining optical flow estimates (which is a variant of *circular matching* proposed in [8]).

B. Flow Decoupling Motion Estimation

The motion estimation is performed in two dedicated parts, which are coupled by a transformation of the measurement. First, the best feature set for rotation estimation $\mathscr{F}_i^{t,R}$ is searched for in the iterative rotation optimization, using a classical reprojection error based outlier rejection criterion. Superscript ^{R} is used to denote the affiliation to the rotation estimation process.
In the subsequent transformation, we rotate the points in camera coordinates and pixel coordinates by the reconstructed rotation. The distance between the rotated pixel coordinates and the measured feature position at time $t-1$ within the image is the new measurement flow for the resulting translation-dedicated estimation part, where superscript ^{T} denotes the affiliation to this sub-process. The emerging scheme can be formulated as:

1) **Iteratively optimize rotation** at time t and iteration i until the feature set does not change $\mathscr{F}_i^{t,R} = \mathscr{F}_{i-1}^{t,R}$ or the maximum number of iterations is reached $i = i_{\max}$. For the first iteration, the feature set $\mathscr{F}_0^{t,R}$ is set to the full set $\mathscr{F}_0^t$ and the motion estimate is initialized with last frame's results $\hat{\mathbf{R}}_i^{t,R} = \hat{\mathbf{R}}^{t-1}$ and $\hat{\mathbf{T}}_i^{t,R}$ to $\hat{\mathbf{T}}^{t-1}$.

a) Estimate motion of iteration i with current inlier set $\mathscr{F}_{i-1}^{t,R}$:

$$(\hat{\mathbf{R}}_i^{t,R}, \hat{\mathbf{T}}_i^{t,R}) = \text{argmin}_{\mathbf{R},\mathbf{T}} \sum_{n=1}^{N_i^{t,R}} \left(\varepsilon_n^t\right)^2, \forall f_n^t \in \mathscr{F}_{i-1}^{t,R} \tag{25}$$

b) Remove outliers by evaluating the standard reprojection error:

$$f_n^t \begin{cases} \in \mathscr{F}_i^{t,R}, & \text{if} \quad \varepsilon_n^t(\hat{\mathbf{R}}_i^{t,R}, \hat{\mathbf{T}}_i^{t,R}) < \varepsilon_i^{R,\text{thresh}} \\ \notin \mathscr{F}_i^{t,R}, & \text{else}, \end{cases}, \tag{26}$$

with $\varepsilon_i^{R,\text{thresh}}$ being the maximum between the bth highest reprojection error and a predefined fixed termination limit $\varepsilon^{R,\text{thresh}}$.

2) **Transform measurements** to compensate for rotation. This is done by rotating the 3D points, which is the same as keeping a fixed $\mathbf{R}^t$ for the optimization:

$$\mathbf{X}_n^{t,T} = \lambda_n^t \mathbf{x}_n^{t,T} = \lambda_n^t \hat{\mathbf{R}}_i^{t,R} \mathbf{x}_n^t. \tag{27}$$

and calculating the rotation-compensated optical flow once for each feature:

$$\|\mathbf{x}_n^{t-1} - \pi\left(\hat{\mathbf{R}}_i^{t,R} \mathbf{x}_n^t\right)\|_2 \tag{28}$$

3) **Iteratively optimize translation** at time t and iteration j until the feature set does not change $\mathscr{F}_j^{t,T} = \mathscr{F}_{j-1}^{t,T}$ or the maximum number of iterations is reached $j = j_{\max}$. Before the first iteration, $\mathscr{F}_0^{t,T}$ is set to the full set $\mathscr{F}_0^t$ and the translation is initialized with last frame's results $\hat{\mathbf{T}}_1^{t,T} = \hat{\mathbf{T}}^{t-1}$. At this point, the transformed measurements contain translation only.

a) Estimate motion of iteration j with current inlier set $\mathscr{F}_{j-1}^{t,T}$:

$$\hat{\mathbf{T}}_j^{t,T} = \text{argmin}_{\mathbf{T}} \sum_{n=1}^{N_j^{t,T}} \left(\varepsilon_n^t\right)^2, \forall f_n^t \in \mathscr{F}_{j-1}^{t,T} \tag{29}$$

b) Remove outliers by evaluating the decoupled normalized reprojection error:

$$f_n^t \begin{cases} \in \mathscr{F}_j^{t,T}, & \text{if} \quad \frac{\varepsilon_n^t(\hat{\mathbf{T}}_j^{t,T})}{\|\mathbf{x}_n^{t-1} - \pi\left(\hat{\mathbf{R}}_i^{t,R} \mathbf{x}_n^t\right)\|_2} < \varepsilon_j^{T,\text{thresh}} \\ \notin \mathscr{F}_j^{t,T}, & \text{else}. \end{cases}, \tag{30}$$

With $\varepsilon_j^{T,\text{thresh}}$ being the maximum between the bth highest normalized reprojection error and a predefined fixed termination limit $\varepsilon^{T,\text{thresh}}$ that describes the expected final error $|E-1|$.

After completing this scheme, the final motion is put together from the rotation estimate $\hat{\mathbf{R}}_i^{t,R}$ based on feature set $\mathscr{F}_{j-1}^{t,R}$ and the translation estimate $\hat{\mathbf{T}}_j^{t,T}$ from feature set $\mathscr{F}_{j-1}^{t,T}$: $(\hat{\mathbf{R}}^t, \hat{\mathbf{T}}^t) = \left(\hat{\mathbf{R}}_i^{t,R}, \hat{\mathbf{T}}_j^{t,T}\right)$. If the vehicle is driving at very low speeds, phases two and three are skipped to avoid numerical inaccuracies due to the small optical flows. In this case, the final motion estimate is set to the result of phase one $(\hat{\mathbf{R}}^t, \hat{\mathbf{T}}^t) = \left(\hat{\mathbf{R}}_i^{t,R}, \hat{\mathbf{T}}_i^{t,R}\right)$.

VI. Evaluation

We base the evaluation of RotROCC on three major parts, using Kitti benchmark: First, we compare the results to implementations that are based on the reprojection error only. Secondly, we compare our results to top ranked state-of-the-art methods and finally we investigate the proposed method with regard to the performance of ROCC [4].

To give a first impression of the reconstruction quality, we evaluate test track 01. Here, we compare our results to two published approaches that base on the reprojection error in a similar way as we do in phase one of our scheme, presented in Sec.V-B: The authors of [15] suggested to classify outliers with reference to the mean reprojection error at iteration p:

$$f_i^t \begin{cases} \in \mathscr{F}_p^t, & \text{if} \quad \varepsilon_i^t(\hat{\mathbf{R}}_p^t, \hat{\mathbf{T}}_p^t) - \mu_p < 1.5\sigma_p, \\ \notin \mathscr{F}_p^t, & \text{else}, \end{cases} \tag{31}$$

with mean error $\mu_p = \sum_i^{N_p} \varepsilon_i^t\left(\hat{\mathbf{R}}_p^t, \hat{\mathbf{T}}_p^t\right)/N_p$ and squared standard deviation $\sigma_p^2 = \sum_i^{N_p} \left(\varepsilon_i^t\left(\hat{\mathbf{R}}_p^t, \hat{\mathbf{T}}_p^t\right) - \mu_p\right)^2/(N_p - 1)$ and a number of total iterations. The authors of [1] refer each feature's error to the standard deviation at iteration p:

$$f_i^t \begin{cases} \in \mathscr{F}_p^t, & \text{if} \quad \varepsilon_i^t(\hat{\mathbf{R}}_p^t, \hat{\mathbf{T}}_p^t) < 3^2\mu_p, \\ \notin \mathscr{F}_p^t, & \text{else}. \end{cases} \tag{32}$$

The comparison to our method is shown in Fig.6. Here, the estimate of the forward translation t_z shows massive breakdowns when applying the methods from [1] (Mean-based) and [15] (Std-based). By contrast, RotROCC achieves a much more robust and precise estimation.

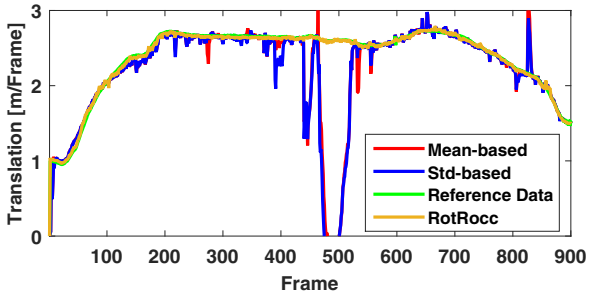

Fig. 6. Comparison of the methods in [1] (Mean-based), [15] (Std-based), and RotROCC in the freeway-scenario of track 01.

Giving a more general insight into RotROCC's results with regard to state of the art methods, we now compare it to the approach of rejecting features based on the reprojection error only, embedded in our framework. This means that we skip the process after the first phase. For this, we reconstructed the trajectories of tracks 00 to 10 with our new method and also with a simplified version that solely bases on the reprojection error. With a mean overall translation error of 0.70 %, our new method clearly outperforms the one-phase approach with a translation error of 0.80 %. Also when comparing low and middle speed scenarios only, our new method shows better results with 0.66 % against this approach with 0.72 %. In the benchmark, our approach currently ranks second place,

as can be seen in Fig.7. First place is SOFT [5], which is based on the tracking of features over time and achieves an overall translation error of 0,88 %. Third place is Svo2, which seems to be an extension of [6], relying on local bundle adjustment, and reaches an error of 0.94 %. These two methods work beyond a one time step temporal horizon to obtain their outstanding results. With an error of 0.88 %, RotROCC is the top ranked frame-to-frame method in the benchmark and the second ranked of all vision methods.

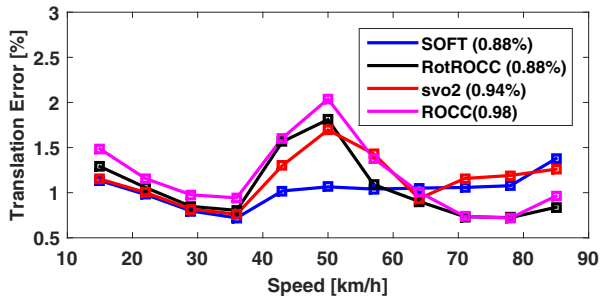

Fig. 7. Comparison of the top ranked methods. Though working on a frame-to-frame base only, RotROCC is competitive with the top ranked methods of Kitti benchmark. Integrating a time-horizon of more than one frame, the algorithms of first and third place achieve even better results at low and middle speeds. Despite using a less complex approach that is only based on frame-to-frame calculation, we come to a good result at these speeds and outperform all methods at speeds above 60 $\frac{km}{h}$.

For a better understanding of the results, Fig.8 shows the improvement of RotROCC, compared to ROCC. Despite introducing constant threshold within the robust estimation phases for rotation and translation and the elimination of backup schemes compared to ROCC, we were able to improve the reconstruction quality. At almost any speed, huge improvements are realized with this less restrictive approach.

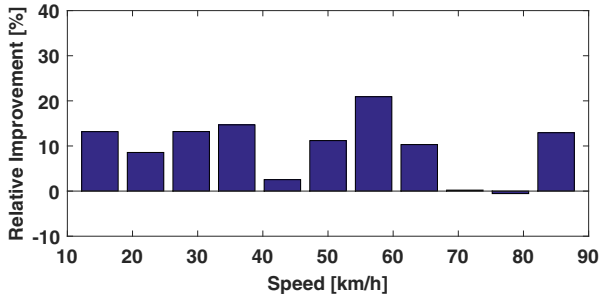

Fig. 8. Visualization of the relative improvement of the proposed RotROCC compared to ROCC. The new method shows an improved capability of estimating the vehicle's translation motion for a wide speed range.

RotROCC's robust and precise reconstruction of the vehicle's motion shows, that the decoupled normalized reprojection error (DNRE) is a more appropriate measure to evaluate correspondences for outlier detection than the commonly applied reprojection error.

VII. Conclusion and Future Work

As we showed, the reconstruction quality can be improved by decoupling the optical flows of rotation and translation and exploiting the resulting characteristics of the flow for outlier detection. We realized this concept by first estimating the rotation, using this estimate to compensate for the vehicle's rotation and afterwards using the normalized reprojection error to detect outliers in the measurement. The compensation allows to relax the restriction of rotation-free scenarios and to still apply an almost coordinate-independent error measure. By eliminating this restriction, the structure and parameters of our method could be notably simplified, which leads to an easier integrability into other implementations. Next, we would like to investigate more accurate but fast and sparse optical flow algorithms [16], [17], the gain of an additional local bundle adjustment as applied in [10] and the integration of additional environmental information as gained from the method which is described in [12].

Acknowledgements

We kindly thank Continental AG for funding this work within a cooperation.

References

[1] H. Badino and T. Kanade. A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion. In *IAPR Conference on Machine Vision Application*, 2011.
[2] J.-Y. Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm, 2001. Intel Corp. 5.
[3] G. Bradski. Dr. dobb's journal of software tools, 2000.
[4] M. Buczko and V. Willert. How to distinguish inliers from outliers in visual odometry high-speed automotive applications. In *IEEE Intelligent Vehicles Symposium*, 2016.
[5] I. Cvisic and I. Petrovic. Stereo odometry based on careful feature selection and tracking. In *European Conference on Mobile Robots*, 2015.
[6] C. Forster, M. Pizzoli, and D. Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *IEEE International Conference on Robotics and Automation*, 2014.
[7] 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*, 2012.
[8] A. Geiger, J. Ziegler, and C. Stiller. Stereoscan: Dense 3d reconstruction in real-time. In *Intelligent Vehicles Symposium*, 2011.
[9] 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*, 2010.
[10] M. Persson, T. Piccini, M. Felsberg, and R. Mester. Robust stereo visual odometry from monocular techniques. In *IEEE Intelligent Vehicles Symposium*, 2015.
[11] D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In *IEEE International Conference on Robotics and Automation*, 2009.
[12] M. Schreier, V. Willert, and J. Adamy. From grid maps to parametric free space maps : A highly compact, generic environment representation for adas. In *IEEE Intelligent Vehicles Symposium*, 2013.
[13] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition. Proceedings*, 1994.
[14] G. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *IEEE Intelligent Vehicles Symposium*, 2000.
[15] N. Sünderhauf, K. Konolige, S. Lacroix, and P. Protzel. Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle, 2006.
[16] V. Willert and J. Eggert. A stochastic dynamical system for optical flow estimation. In *IEEE 12th International Conference on Computer Vision Workshops*, 2009.
[17] V. Willert, M. Toussaint, J. Eggert, and E. Korner. Uncertainty optimization for robust dynamic optical flow estimation. In *IEEE 6th International Conference on Machine Learning and Applications*, 2007.