2016 IEEE Intelligent Vehicles Symposium (IV)
Gothenburg, Sweden, June 19-22, 2016

How to Distinguish Inliers from Outliers in Visual Odometry for High-speed Automotive Applications

Martin Buczko[1] and Volker Willert[1]

***Abstract*— In this paper, we present an outlier removal scheme for stereo-based visual odometry which is especially suited for improving high-speed pose change estimations in large-scale depth environments. First we investigate the variance of the reprojection error on the 3D position of a feature given a fixed error in pose change to conclude that a detection of outliers based on a fixed threshold on the reprojection error is inappropriate. Then we propose an optical flow dependent feature-adaptive scaling of the reprojection error to reach almost invariance to the 3D position of each feature. This feature-adaptive scaling is derived from an approximation showing the relation between longitudinal pose change of the camera, absolute value of the optical flow, and distance of the feature. Using this scaling, we develop an iterative alternating scheme to guide the separation of inliers from outliers. It optimizes the tradeoff between finding a good criterion to remove outliers based on a given pose change and improving the pose change hypothesis based on the current set of inliers. Including the new outlier removal scheme into a pure two-frame stereo-based visual odometry pipeline without applying bundle adjustment or SLAM-filtering we are currently ranked amongst the top camera-based algorithms and furthermore outperform camera and laser scanner methods in Kitti benchmark's high-speed scenarios.**

I. Introduction

Stereo-based visual odometry estimates the motion of a camera given by its three-dimensional pose change between temporal consecutive images from a sequence of stereo image pairs. This information can be used as an estimate of the current driving states for different driver assistance systems such as anti-lock braking system (ABS), electronic stability control (ESC), cruise control or roll stability control (RSC) and many more. Further on, it can be used as input for any dead reckoning approach to estimate the trajectory and location of the vehicle. Challenging situations for camera-based systems include an insufficient number of correspondences of static scene points, heavily changing illumination and low brightness, an unstructured environment with homogeneous, non-textured surfaces, or an improperly low frame rate. High-speed scenarios along motorways combine several of these problems, making them one of the most challenging situations. Especially, the loss of suitable near features complicates the estimation. In this paper, we show why the classical approach of an outlier detection based on reprojection error fails in such situations and how to circumvent this problem. We analyze the variance of the reprojection error with regard to the 3D position of a feature for a fixed estimation error shown in Fig.1 and explained in Sec. III-A. This uncovers that outlier removal based on the reprojection error often leads to problems when estimating the translation in scenarios with large differences in the features' 3D positions. Solving this problem, we derive a new measure to valuate the quality of feature correspondences. Fig.1 shows the comparison between the classical and our approach to rate features for outlier removal. In Sec. IV we present an iterative optimization scheme alternating between outlier rejection and pose change estimation refinement, which is based on the new criterion. An evaluation based on the Kitti benchmark [8], which provides city, overland and freeway scenarios and comparisons to state-of-the-art methods are given in Sec. V. Our system currently[2] ranks first place amongst camera-based algorithms in the Kitti benchmark. Furthermore, at speeds higher than 70 km/h it achieves better results than the best camera and laser-scanner based methods in the benchmark.

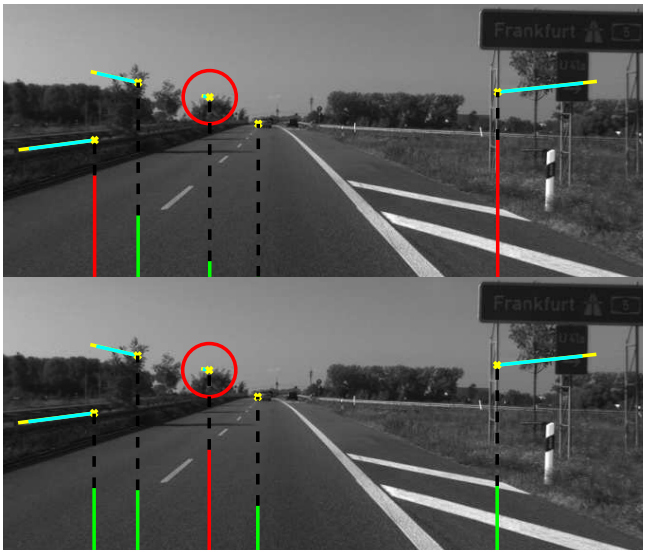

Fig. 1. The estimated motion from visual odometry leads to the cyan optical flows instead of the yellow real flow induced by the motion of the ego-vehicle. All correspondences but the one marked by the red circle are error-free. This true outlier has an additive error in the depth estimate. The reprojection error (proportional to the length of the red and green bars in the upper image) shows a low error for the distant outlier, thus selected as an inlier (green bar) but high values for close true inliers, which are marked as outliers (red bars). Our proposed normalized reprojection error (proportional to the length of the red and green bars in the lower image) shows an almost constant offset for all features due to the estimation error and an increased error for the true outlier, now also marked as an outlier (red bar). The true inliers are now selected as inliers (green bars).

[1]Control Methods & Robotics Lab,
Technical University of Darmstadt, Germany

[2]At the time of paper submission January 11, 2016

II. Related Work

The essential part of any visual odometry system is the detection of outliers. Therefor, a broad variety of methods has been introduced: Purely **flow-based** approaches can be found in [1], [10], [16]. All of them are based on the assumption, that the flow follows patterns which are induced by the egomotion of the car. Next, **motion model-based** approaches for outlier detection exist, that explicitly constrain the flow using a certain motion model as in [18]. The majority of existing systems use **reprojection error-based** approaches. Here mainly two different ways for finding a proper inlier set are used. The first one is **RANSAC** [7], which is based on the following principle: In each iteration, a minimum number of random samples is taken from the correspondences to create a motion hypothesis. Then, a score for each feature is calculated that describes whether it supports the hypothesis. If the motion estimate reaches a predefined support of the features, the non-supporting features are marked as outliers. Otherwise, a new random sample is drawn and the next iteration starts. In order to define the support of a feature in this RANSAC-scheme, the authors of [3], [11], [12], [17] calculate the reprojection error for each feature and compare it to a constant threshold. Trying to optimize the random process of finding the right hypothesis to separate the features into inliers and outliers, numerous extensions were created. A comparison between the most prominent ones can be found in [15].

Due to the random selection of correspondences one can not expect a steady improvement of the resulting motion estimation during the iterations. Coping with this problem, an alternative method was applied in [2], [13], [23]. Following the naming that was used for RANSAC we unite this class of methods under the notation **MAximum Subset Outlier Removal** (MASOR). Here, the maximum number of features instead of a minimum random sample is taken to calculate a motion hypothesis. This motion estimation and a subsequent outlier rejection step are repeated in an iterative scheme. Then a support score is calculated for every feature. Instead of judging the hypothesis, the score is interpreted as a measure for the quality of each feature, as the hypothesis is considered to be a good estimate. Non-supporting features are rejected and the next iteration starts with the remaining features. The process is repeated until a termination criterion is met. This approach is a good alternative to RANSAC if the number of inliers is sufficient enough to create a hypothesis that is good enough to separate the outliers, which is fulfilled in scenarios that we tested. Due to the broad application of the reprojection error for outlier detection, our proposed transformation is of interest for a wide class of outlier rejection schemes such as RANSAC and MASOR.

III. How to define an outlier?

We start with the classical least squares estimator

$$(\hat{\mathbf{R}},\hat{\mathbf{T}}) = \operatorname{argmin}_{\mathbf{R},\mathbf{T}} \sum_{i=1}^{N} \left(\varepsilon_i^t\right)^2 , \tag{1}$$

$$\varepsilon_i^t = \|\mathbf{x}_i^{t-1} - \pi(\mathbf{R}\lambda_i^t\mathbf{x}_i^t + \mathbf{T})\|_2 , \tag{2}$$

where the norms ε_i^t are called the reprojection errors for each feature f_i^t indexed by i at time t within the feature set $\mathscr{F}^t = \{f_i^t\}_{i=1}^N$. Here, $\{\mathbf{x}_i^{t-1},\mathbf{x}_i^t\} \in \mathbb{R}^3$ is the correspondent pair of coordinates denoted in homogeneous normalized[3] image coordinates $\mathbf{x}_i^t = [x_i^t, y_i^t, 1]^T$ for all 3D points $p_i \in \mathbb{R}^3$ with camera coordinates $\mathbf{X}_i^t = [X_i^t, Y_i^t, Z_i^t]^T = \lambda_i^t\mathbf{x}_i^t$. The pose change of the camera from time $t-1$ to time t is given by the 3D translation vector[4] $\mathbf{T} = [t_x, t_y, t_z]^T \in \mathbb{R}^3$ and the rotation matrix[5] $\mathbf{R} \in SO(3)$ and π denotes the standard planar projection $[X,Y,Z]^T \longmapsto [X/Z, Y/Z, 1]^T$ with lateral coordinate X, transversal coordinate Y and forward coordinate Z.

Following the classical visual odometry pipeline, we assume that for each point p_i the depth $\lambda_i^t \in \mathbb{R}$ is measured by some stereo vision algorithm, the image coordinates $\mathbf{x}_i^{t-1}$ are extracted by some feature detector and the correspondent image coordinates in the next frame $\mathbf{x}_i^t$ are measured by some optical flow algorithm. To find the optimal estimate of the pose change $(\hat{\mathbf{R}},\hat{\mathbf{T}})$ via minimizing the objective (1) with an iterative gradient descent method[6] some initial guess for the pose change has to be given.

Now, we are faced with the main problem of visual odometry: Given the set of all extracted features, we need to find suitable features – the inliers – and reject all other features from the set – the outliers. This is usually done by selecting only features with reliable measurements $\{\lambda_i^t, \mathbf{x}_i^{t-1}, \mathbf{x}_i^t\}$ and defining some criterion to evaluate how well these measurements fit to some hypothesis of the estimate $(\tilde{\mathbf{R}},\tilde{\mathbf{T}})$.

The reliability of a measurement has two aspects. First, since in stereo vision depth $\lambda_i^t = b/d_i^t$ is reconstructed from disparity d_i^t using a stereo rig with a fixed known baseline b and both the disparity d_i^t and the pairs $\{\mathbf{x}_i^{t-1},\mathbf{x}_i^t\}$ are based on a correspondence search, only unambiguous correspondences, e.g. not facing the aperture problem, should be taken into account. Second, the accuracy of these correspondences are limited by the resolution of the images. So even if the correspondences are unambiguous the smaller their distances in image space $\|\mathbf{x}_i^t - \mathbf{x}_i^{t-1}\|$ and d_i^t, the less accurate the pose change can be estimated. This is because the ratios $\|\mathbf{x}_i^t - \mathbf{x}_i^{t-1}\|/\Delta p$ and $d_i^t/\Delta p$ between distances $\|\mathbf{x}_i^t - \mathbf{x}_i^{t-1}\|$, d_i^t and the limited image resolution Δp are getting smaller with smaller image-distances and thus the signal-to-resolution-ratio decreases. Especially for the accuracy of the reconstructed depth $\lambda_i^t = b/d_i^t$, this is crucial because the resolution of depth $\partial\lambda_i^t \propto \partial d_i^t (\lambda_i^t)^2$ reduces quadratically with disparity.

Considering these facts, it seems to be easy to figure out good features. Choose near features with large optical flow that are based on highly confident correspondence estimates. Additionally, each correspondence has to fulfill the epipolar constraint for one optimal estimate $(\hat{\mathbf{R}},\hat{\mathbf{T}})$, thus the features have to be projections of static points in the scene only. Since we cannot guarantee that the measurements are all confident

[3] Known intrinsic camera parameters are assumed.

[4] The time index for $\mathbf{T}$ and $\mathbf{R}$ is neglected for convenience.

[5] The space of rotation matrices is denoted by $SO(3) := \{\mathbf{R} \in \mathbb{R}^{3\times3} | \mathbf{R}^T\mathbf{R} = \mathbf{I}, det(\mathbf{R}) = 1\}$.

[6] For example the Gauss-Newton or Levenberg-Marquardt method.

and we do not have the optimal pose change estimate at hand, we need to find a good hypothesis $(\tilde{\mathbf{R}},\tilde{\mathbf{T}})$ and a proper criterion to keep as much suitable features as possible.
To resolve this very task, we investigate the reprojection error (2) in two ways: On the one hand, it should be used as the criterion to remove outliers based on a threshold given a pose change hypothesis and on the other hand, it should improve the hypothesis of the pose change given the inliers. In order to combine both subproblems in an alternating scheme, we figure out how to use the reprojection error for both subtasks such that as many inliers as possible are kept which also leads to a more accurate estimate of the pose change.
To find a good criterion for the outlier removal, we examine the variance of the reprojection error on the values of the measurements assuming error-free measurements $\{\hat{\lambda}_i^t,\hat{\mathbf{x}}_i^{t-1},\hat{\mathbf{x}}_i^t\}$ and an imprecise pose change hypothesis $(\tilde{\mathbf{R}},\tilde{\mathbf{T}})$. For improving the pose change estimation, we assume error-prone measurements $\{\tilde{\lambda}_i^t,\tilde{\mathbf{x}}_i^{t-1},\tilde{\mathbf{x}}_i^t\}$ and try to maximize the improvement of the iterative pose change estimation $(\hat{\mathbf{R}},\hat{\mathbf{T}})$ by using the knowledge about the variance of the reprojection error on the values of the measurements again.

As stated in [22], high translational errors occur at large longitudinal pose changes along the optical axis. The translation estimates get especially poor for long distance features [14]. To receive a first impression on the consequences for the sensitivity of the reprojection error in such driving scenarios Fig.2 shows the dependency of the reprojection error ε_i^t for an increase in longitudinal translation error estimates Δt_z for varying feature depths λ_i^t. It clearly illustrates that the reprojection error (RE) linearly increases with increasing translation error but the sensitivity of the reprojection error (the slope of the lines) decreases with increasing distance of the features.

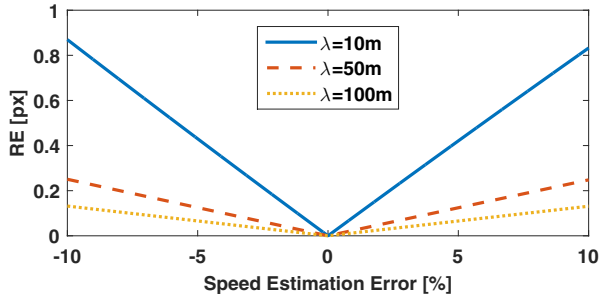

Fig. 2. Reprojection error (RE) ε_i^t for an error-prone translation $\tilde{t}_z=\hat{t}_z+\Delta t_z$, with $\hat{t}_z=100\,\mathrm{km/h}$ and a range of $\Delta t_z=[-10\%;10\%]\hat{t}_z$ for varying error-free feature depths $\hat{\lambda}_i^t$ of 10, 50 and 100 m. The decrease of the sensitivity (slope of the lines) with increasing depth can clearly be seen.

A. The Reprojection Error in High-Speed Scenarios

For error-free measurements $\{\hat{\lambda}_i^t,\hat{\mathbf{x}}_i^{t-1},\hat{\mathbf{x}}_i^t\}$ and the optimal motion estimate $(\hat{\mathbf{R}},\hat{\mathbf{T}})$, the reprojection error (2) becomes zero because it holds

$$\hat{\mathbf{x}}_i^{t-1}=\pi(\hat{\mathbf{R}}\hat{\lambda}_i^t\hat{\mathbf{x}}_i^t+\hat{\mathbf{T}})\,,\quad\forall i,t\,. \tag{3}$$

In order to find a proper threshold on the reprojection error to reject outliers, we have to define some motion error range $(\Delta\mathbf{R},\Delta\mathbf{T})$ on the optimal estimate $(\tilde{\mathbf{R}},\tilde{\mathbf{T}})=(\hat{\mathbf{R}}\Delta\mathbf{R},\hat{\mathbf{T}}+\Delta\mathbf{T})$ to find the reprojection error range given the motion error range and error-free measurements. This results in the sensitivity of the reprojection error

$$\Delta\varepsilon_i^t=\|\hat{\mathbf{x}}_i^{t-1}-\pi(\tilde{\mathbf{R}}\hat{\lambda}_i^t\hat{\mathbf{x}}_i^t+\tilde{\mathbf{T}})\|_2\,,\quad\forall i,t\,. \tag{4}$$

Now, considering high-speed scenarios, we can assume very small rotations

$$\text{1. high-speed approximation:}\quad\mathbf{R}\approx\mathbf{I}\,, \tag{5}$$

thus the rotation matrix approximately equals the identity $\mathbf{I}$ and much larger longitudinal than horizontal and vertical movements $t_z\gg t_x,t_y$, thus the lateral and transversal components of the translation are approximately equal to zero

$$\text{2. high-speed approximation:}\quad t_x,t_y\approx 0\,. \tag{6}$$

Applying approximation (5) and (6) we get an approximation of the sensitivity of the reprojection error (4) under high-speed for an error-prone motion hypothesis $\tilde{t}_z=\hat{t}_z+\Delta t_z$ which reads

$$\Delta\varepsilon_i^t\approx\left\|\begin{pmatrix}\frac{\hat{\lambda}_i^t\hat{x}_i^t}{\hat{\lambda}_i^t+\hat{t}_z}-\frac{\hat{\lambda}_i^t\hat{x}_i^t}{\hat{\lambda}_i^t+\hat{t}_z+\Delta t_z}\\ \frac{\hat{\lambda}_i^t\hat{y}_i^t}{\hat{\lambda}_i^t+\hat{t}_z}-\frac{\hat{\lambda}_i^t\hat{y}_i^t}{\hat{\lambda}_i^t+\hat{t}_z+\Delta t_z}\end{pmatrix}\right\|_2 \tag{7}$$

$$=\left|\frac{\hat{\lambda}_i^t\Delta t_z}{(\hat{\lambda}_i^t+\hat{t}_z)(\hat{\lambda}_i^t+\hat{t}_z+\Delta t_z)}\right|\left\|\hat{\mathbf{x}}_i^t\right\|_2\,. \tag{8}$$

The sensitivity of the reprojection error is *scaled by the absolute value of the image coordinate* $\|\hat{\mathbf{x}}_i^t\|_2$ and *damped by the feature's depth* $\hat{\lambda}_i^t$. This means, an incorrect motion hypothesis $\tilde{t}_z=\hat{t}_z+\Delta t_z$ with a fixed error range Δt_z produces a variant sensitivity $\Delta\varepsilon_i^t$ dependent on the feature's position, as illustrated in Fig. 3. Thus, methods that base the outlier removal on a constant threshold on the reprojection error remove close features, although the measurements are error-free (or error-prone in the same range as for distant features).

This leads to the breakdown of outlier removal in high-speed scenarios for methods based on a fixed threshold on the reprojection error. As close features with high absolute values of their correspondences are lost during the outlier rejection process, the sensitivity of the reprojection error against forward translation gets lost, as can be seen in Fig. 2. In turn, this results in worse estimates of the translation because the signal-to-resolution ratio is getting small and cannot be exploited anymore. To conclude, a reasonable threshold to judge the feature's quality can not be a constant value but must incorporate the depth as much as the length of the image coordinate in order to make a meaningful statement on the feature's quality.

B. Almost Invariant Criterion for Outlier Removal

To reduce the variance of the reprojection error on the feature position we can either apply a position adaptive threshold for outlier removal or normalize the reprojection error for coordinate $\|\hat{\mathbf{x}}_i^t\|_2$ scaling and depth $\hat{\lambda}_i^t$ damping.

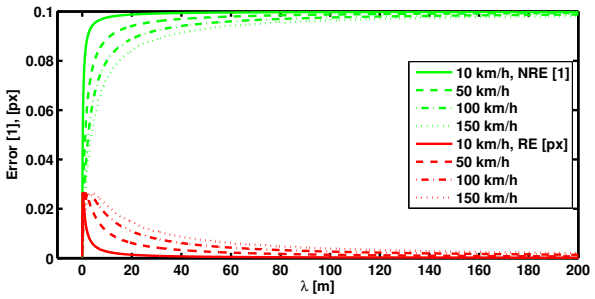

Fig. 3. Comparison between sensitivity of reprojection error $\Delta\varepsilon_i^t$ (RE, red) and sensitivity of normalized reprojection error $\Delta\overline{\varepsilon}_i^t$ (NRE, green) over depth λ_i^t. The reprojection error is dampened by distant and slow features and a fixed threshold outlier criterion tends to lose close and fast features. By contrast, the normalized reprojection error amplifies distant and slow features up to some saturation, thus a fixed threshold outlier criterion tends to keep close and fast features.

Since the resolution of the measured depth values decreases with distance and depth measurements are error-prone in general, we do not want to incorporate them to compensate the depth damping of the reprojection error. Instead, we use the dependency of the absolute value of the optical flow on the depth and use the optical flow measurements to normalize the reprojection error as follows:

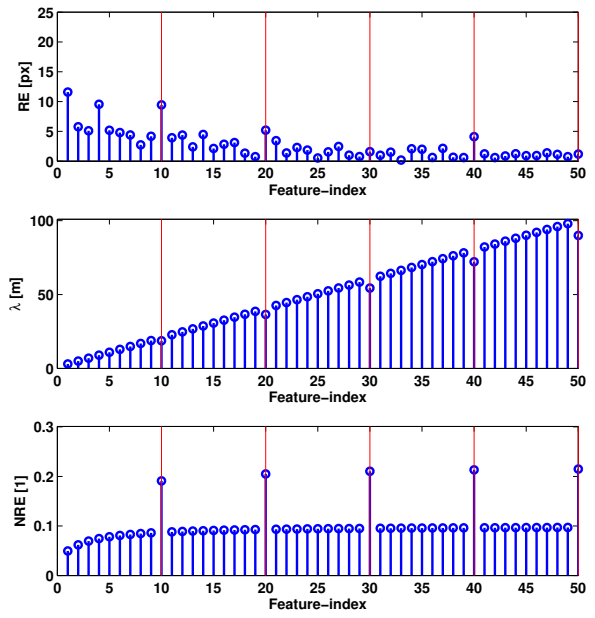

Fig. 4. The reprojection error ε_i^t (RE, top) in comparison to the normalized reprojection error $\overline{\varepsilon}_i^t$ (NRE, bottom) for a scenario with a simulated forward motion t_z of 100 km/h in which every tenth feature f_i^t has an error of 10 % in the estimated depth λ_i^t (middle). The depths start at 3 m for feature f_1^t and end at 100 m for feature f_{50}^t.

The error-free absolute value of a feature's optical flow induced by an error-free straight forward motion $\hat{t}_z$ again assuming (5) and (6) reads

$$\left\|\hat{\mathbf{x}}_i^{t-1} - \hat{\mathbf{x}}_i^t\right\|_2 = \left\|\pi(\hat{\mathbf{R}}\hat{\lambda}_i^t\hat{\mathbf{x}}_i^t + \hat{\mathbf{T}}) - \hat{\mathbf{x}}_i^t\right\|_2 \tag{9}$$

$$\approx \left|\frac{\hat{t}_z}{\hat{\lambda}_i^t + \hat{t}_z}\right| \left\|\hat{\mathbf{x}}_i^t\right\|_2 . \tag{10}$$

Using the absolute value of the current flow as a normalization to the sensitivity of the reprojection error, we get

$$\Delta\overline{\varepsilon}_i^t = \frac{\Delta\varepsilon_i^t}{\left\|\hat{\mathbf{x}}_i^{t-1} - \hat{\mathbf{x}}_i^t\right\|_2} \approx \left|\frac{\hat{\lambda}_i^t \Delta t_z}{\hat{t}_z(\hat{\lambda}_i^t + \hat{t}_z + \Delta t_z)}\right| \approx \left|\frac{\Delta t_z}{\hat{t}_z}\right| . \tag{11}$$

Here, the second approximation assumes the depth being much larger than the longitudinal motion $\hat{\lambda}_i^t \gg \hat{t}_z + \Delta t_z$. For this reason, the normalized reprojection error is not scaled by the absolute value of the image coordinate anymore and almost not dependent on the distance of long distant features. This can also be seen in Fig. 3.

Thus, using a threshold $\varepsilon_{\text{thresh}}$ on the normalized reprojection error $\overline{\varepsilon}_i^t$ to mark each feature f_i^t as a member of the current feature set $\mathscr{F}^t$, we apply

$$f_i^t \begin{cases} \in \mathscr{F}^t, & \text{if} \quad \overline{\varepsilon}_i^t = \frac{\varepsilon_i^t}{\left\|\mathbf{x}_i^{t-1} - \mathbf{x}_i^t\right\|_2} < \varepsilon_{\text{thresh}} , \\ \notin \mathscr{F}^t, & \text{else} . \end{cases} \tag{12}$$

This criterion (as part of an outlier removal scheme explained in Sec. IV) turns out to be very suitable for outlier removal, especially in high-speed scenarios, because it is almost invariant to the features' 3D position. Fig. 4 shows a comparison between the reprojection error ε_i^t (top) and the normalized reprojection error $\overline{\varepsilon}_i^t$ (bottom) for some error-prone depth estimates and a forward motion t_z of 100 km/h. The reprojection error does not allow a separation between inliers and outliers because ε_i^t scales with the absolute value of the coordinate of the features. By contrast, $\overline{\varepsilon}_i^t$ (bottom) leads to a clear separability.

C. Hypothesis Refinement on the Inlier Set

One question was not yet addressed: Since the normalized reprojection error (11) improves the outlier rejection, is it also suitable to get better hypothesis refinements for the least squares problem formulated in (1)? Now, assuming error-prone measurements $\{\tilde{\lambda}_i^t, \tilde{\mathbf{x}}_i^{t-1}, \tilde{\mathbf{x}}_i^t\}$ and trying to minimize (1) to get a better pose change estimate $(\hat{\mathbf{R}}, \hat{\mathbf{T}})$ the features with measurements that possess a high resolution should contribute more to the estimate than measurements with low resolution. This can be realized with an extension of (1) to a weighted least squares problem that realizes a decrease of the weights for distant features and an increase of the weights for larger optical flow amplitudes. Looking at the approximation of the reprojection error (8) for high longitudinal speeds, this weighting is intrinsically done by the reprojection error itself, whereas the normalized reprojection error would treat the features more or less equally weighted. Thus, for refinement of the motion hypothesis based on the current set of inliers, the (un-normalized) reprojection error is already most suitable.

IV. Incremental Alternating Outlier Removal and Pose Refinement Scheme

To realize an iterative optimization scheme that carefully alternates between incremental outlier rejection and pose change refinement, we need a suitable set of a reasonable number of features to start with. A suitable feature has unambiguous temporal as well as stereoscopic correspondence measurements to get as much reliable optical flow and depth estimates as possible.

A. Per Frame Initialization

Our initial feature set for every stereo-frame-pair is created as follows applying only standard functions of the OpenCV library [5]: We start with the feature-selection using the Shi and Tomasi method [21]. For each feature the disparity at time $t-1$ is calculated using SAD-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 [14]). After that the optical flow for the left and right image between time $t-1$ and t is refined with the Lucas-Kanade method [4]. The final feature set $\mathscr{F}_0^t = \{\mathbf{x}_i^{t-1}, \mathbf{x}_i^t, \lambda_i^t\}_{i=1}^{N_0}$ with a starting number N_0 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 [9]).

B. Alternating Iteration Based on MASOR

We iterate over p alternating between a) pose refinement keeping the current inlier set F_{p-1}^t fixed and b) outlier removal keeping the current pose $(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p)$ fixed:

a) Pose refinement starts with $\mathscr{F}_0^t$ at first iteration $p=0$. The pose estimate is initialized with the estimate of the last frame $\hat{R}_0^t = \hat{R}^{t-1}$ and $\hat{T}_0^t = \hat{T}^{t-1}$. In the following we omit time index t for simplicity.

$$(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p) = \text{argmin}_{\mathbf{R},\mathbf{T}} \sum_i \left(\varepsilon_i^t\right)^2, \forall f_i^t \in \mathscr{F}_{p-1}^t \qquad (13)$$

b) Outlier removal applying our combined criterion, which we call *Robust Outlier Criterion for Camera-based Odometry* (ROCC):

$$f_i^t \begin{cases} \in \mathscr{F}_p^t, & \text{if} \quad \bar{\varepsilon}_i^t(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p) < \bar{\varepsilon}_p^{\text{thresh}} \\ & \text{and} \quad \varepsilon_i^t(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p) < \varepsilon_p^{\text{thresh}}, \\ \notin \mathscr{F}_p^t, & \text{else}. \end{cases} \qquad (14)$$

Here we increase the thresholds $\bar{\varepsilon}_p^{\text{thresh}}$ and $\varepsilon_p^{\text{thresh}}$ in a coarse to fine manner during the iterations. If the number N_p of the feature set does not change any more, a minimum number of features $N_{\min}$ is reached or a maximum number of iterations p_{max} is reached, we terminate our robust pose estimation scheme and perform one last refining optimization run with the remaining features. This run is initialized with the rotation and direction of the translation estimate from openCV's standard least median of squares 2D-2D five point method.

In order to evaluate our robustified criterion for outlier detection, we compare our results with MASOR approaches that use the reprojection error. In 2005, the authors of [23] applied the following criterion to classify outliers:

$$f_i^t \begin{cases} \in \mathscr{F}_p^t, & \text{if} \quad \varepsilon_i^t(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p) - \mu_p < 1.5\sigma_p, \\ \notin \mathscr{F}_p^t, & \text{else}. \end{cases} \qquad (15)$$

With mean error $\mu_p = \sum_i^{N_p} \varepsilon_i^t\left(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p\right)/N_p$ and squared standard deviation $\sigma_p^2 = \sum_i^{N_p} \left(\varepsilon_i^t\left(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p\right) - \mu_p\right)^2/(N_p-1)$. The total number of iterations was set to a fixed value.

In 2011 the authors of [2] changed the criterion slightly:

$$f_i^t \begin{cases} \in \mathscr{F}_p^t, & \text{if} \quad \varepsilon_i^t(\hat{\mathbf{R}}_p, \hat{\mathbf{T}}_p) < 3^2\mu_p, \\ \notin \mathscr{F}_p^t, & \text{else}. \end{cases} \qquad (16)$$

The estimate of the forward translation t_z in Fig.5 shows massive breakdowns when applying the methods from [2] (Mean-based) and [23] (Std-based). As we derived in Sec. III-A, this is due to the use of the reprojection error. By contrast, our new ROCC shows only minor errors and leads to a robust estimation.

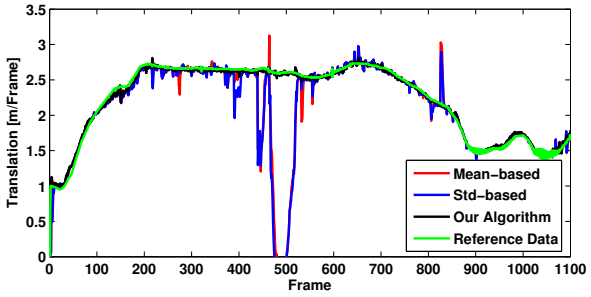

Fig. 5. Comparison of the MASOR-methods in [2] (Mean-based), [23] (Std-based) and our new approach in the freeway-scenario of track 01.

V. Evaluation

In order to evaluate the performance of our method in high-speed scenarios, we first compare it with state-of-the art visual odometry algorithms without additional sensor-data: The method from [6] achieves the second best overall result in the Kitti benchmark with a translation error of 1.03 %. The authors use feature tracking on the base of many images. With an error of 1.09 %, the algorithm from [14] shows a slightly worse quality. Here, the authors apply bundle adjustment to improve the motion estimation. By contrast to these three methods, we do not use the feature's history. As depicted in Fig.6, this leads to an almost constant additional error in comparison to the two other methods. Despite the loss of precision due to not using the feature's history, our new outlier rejection scheme leads to a lowered error from a speed of approximately 65 km/h on. This shows, that the application of our new measure enables even a comparative imprecise system to outperform state-of-the-art methods. In order to underline the performance in high-speed scenarios, we also compare our results to methods that incorporate

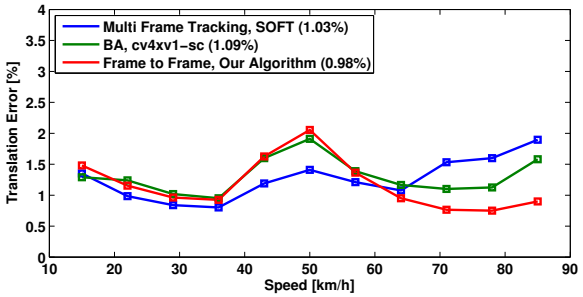

Fig. 6. Comparison between our method and the two best camera-based algorithms in the Kitti benchmark. From a speed of 65 km/h on, our system achieves the best reconstruction quality amongst the top three ranked methods.

the information from a high-precision laser scanner: With an error of 0.88 % and 1.14 %, the methods from [27] and [26] reach a very high reconstruction quality. The top-ranked method from [28] even achieves an overall-error of 0.75 %. Fig.7 illustrates the comparison between the laser scanner extended systems and our system. Despite showing an inferior overall reconstruction quality due to the inferior sensor-setup, our careful outlier rejection again leads to the best reconstruction quality at speeds higher than 70 km/h.

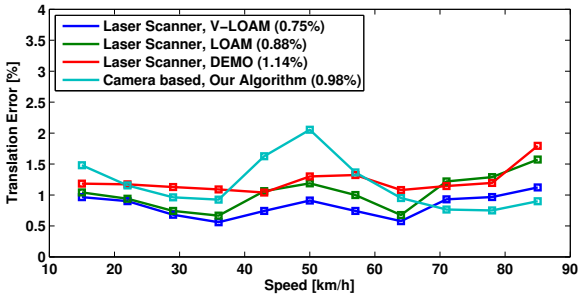

Fig. 7. Comparison between our camera based algorithm and the three best systems of the Kitti benchmark, based on stereo camera and laser scanner. From a speed of 70 km/h on, our system outperforms the top rated systems.

VI. Conclusion and Future Work

In our work, we motivated the need for a new error criterion in outlier detection schemes. After deriving a normalized reprojection error criterion from theoretical considerations, we applied it in an outlier detection within an iterative scheme for a frame-to-frame system. This leads to a careful rejection of outliers while simultaneously preserving as many close inliers as possible. Hereby we are able to drastically increase the robustness and reconstruction quality.
Next, we would like to investigate more accurate but fast and sparse optical flow algorithms [24], [25], the embedding into a local bundle adjustment framework [14] and the integration of additional environmental information [19], [20].

Acknowledgements

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

References

[1] A. Amit, E. Rivlin, and I. Shimshoni. Ror: Rejection of outliers by rotations. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.
[2] 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.
[3] F. Bellavia et al. Robust selective stereo slam without loop closure and bundle adjustment. In *International Conference on Image Analysis and Processing*. Springer Berlin Heidelberg, 2013.
[4] J.-Y. Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm, 2001. Intel Corp. 5.
[5] G. Bradski. Dr. dobb's journal of software tools, 2000.
[6] I. Cvisic and I. Petrovic. Stereo odometry based on careful feature selection and tracking. In *European Conference on Mobile Robots*, 2015.
[7] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, 1981.
[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*, 2012.
[9] A. Geiger, J. Ziegler, and C. Stiller. Stereoscan: Dense 3d reconstruction in real-time. In *Intelligent Vehicles Symposium*, 2011.
[10] B. Grinstead et al. Improving video-based robot self localization through outlier removal. In *Proceedings of the 1st Joint Emer. Prep. & Response/Robotic & Remote Sys. Top. Mtg*, 2006.
[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*, 2010.
[12] M. Maimone, C. Yang, and L. Matthies. Two years of visual odometry on the mars exploration rovers. In *Journal of Field Robotics*, 2007.
[13] C. Olson, L. Matthies, M. Schoppers, and M. Maimone. Robust stereo ego-motion for long distance navigation. In *IEEE Conference on Computer Vision and Pattern Recognition. Proceedings*, 2000.
[14] M. Persson, T. Piccini, M. Felsberg, and R. Mester. Robust stereo visual odometry from monocular techniques. In *IEEE Intelligent Vehicles Symposium*, 2015.
[15] R. Raguram, F. M. Frahm, and M. Pollefeys. A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In *European Conference on Computer Vision*, 2008.
[16] P. Santana and L. Correia. Improving visual odometry by removing outliers in optic flow, 2008.
[17] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. In *International Journal of Computer Vision 95.1*, 2011.
[18] 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.
[19] 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.
[20] M. Schreier, V. Willert, and J. Adamy. Compact representation of dynamic driving environments for adas by parametric free space and dynamic object maps. In *IEEE Transactions on Intelligent Transportation Systems*, 2016.
[21] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition. Proceedings*, 1994.
[22] S. Song and M. Chandraker. Robust scale estimation in real-time monocular sfm for autonomous driving. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
[23] N. Sünderhauf, K. Konolige, S. Lacroix, and P. Protzel. Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle. In *Autonome Mobile Systeme 2005*. Springer Berlin Heidelberg, 2006.
[24] V. Willert and J. Eggert. A stochastic dynamical system for optical flow estimation. In *IEEE 12th International Conference on Computer Vision Workshops*, 2009.
[25] V. Willert et al. Uncertainty optimization for robust dynamic optical flow estimation. In *IEEE 6th International Conference on Machine Learning and Applications*, 2007.
[26] J. Zhang, M. Kaess, and S. Singh. Real-time depth enhanced monocular odometry. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.
[27] J. Zhang and S. Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems Conference*, 2014.
[28] J. Zhang and S. Singh. V-loam: Visual-lidar odometry and mapping: Low-drift, robust, and fast. In *IEEE International Conference on Robotics and Automation*, 2015.