2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Visual Odometry Driven Online Calibration for Monocular LiDAR-Camera Systems

Hsiang-Jen Chien and Reinhard Klette
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: jchien@aut.ac.nz

Nick Schneider and Uwe Franke
Environment Perception, Daimler AG
Sindelfingen, Germany
Email: nick.schneider@daimler.com

***Abstract*—Recently LiDAR-camera systems have rapidly emerged in many applications. The integration of laser range-finding technologies into existing vision systems enables a more comprehensive understanding of 3D structure of the environment. The advantage, however, relies on a good geometrical calibration between the LiDAR and the image sensors. In this paper we consider visual odometry, a discipline in computer vision and robotics, in the context of recently emerging online sensory calibration studies. By embedding the online calibration problem into a LiDAR-monocular visual odometry technique, the temporal change of extrinsic parameters can be tracked and compensated effectively.**

I. Introduction

Nowadays LiDAR-integrated vision systems are pervasive in a wide range of real world applications. Advanced driver assistance systems [1], autonomous driving, navigation [2], [3], and scene modeling are just a few examples among them. A good integration of a LiDAR scanner into a vision system heavily relies on a set of accurately calibrated extrinsic parameters. These parameters specify how laser measurements are cast to the camera's coordinate system, in a way that the vision system can benefit from the highly accurate depth provided by the laser scanning technology.

A geometric calibration procedure can either be offline (e.g. [4], [5]) or online (e.g. [6], [7], [8], [9]), depending on the available time and setup. Offline calibration of LiDAR-camera systems have been well studied in the last decade. Such a calibration procedure is deployed in a controlled environment with the use of one or many referenced targets with known geometry for determining the relative pose of the LiDAR sensor with respect to a vision system. Issues such as mechanical vibration and temperature change may decrease the performance of the calibrated extrinsic parameters over time, after the system goes online and starts to operate. In this case, an online calibration strategy needs to be deployed to constantly verify the parameters and make adjustments accordingly. The process is also known as recalibration, and is usually done automatically on-site with no user intervention.

Existing online calibration techniques are based, for example, on using *mutual information* (MI) maximisation together with LiDAR scans. The underlying idea is that, given a set of optimal extrinsic parameters, a LiDAR scan, once projected onto an image, must have some of its properties best correlated (e.g. using MI maximisation) to the imagery information. Figure 1 shows an example of maximised MI, which maps high intensity LiDAR points onto bright pixels Such an idea might fail when 3D structure of the scene is not taken into account when a laser point is associated with a pixel. As shown in the same figure, some occluded points are wrongly mapped to image pixels. This suggests that a recalibration algorithm *corrects* the extrinsics in a way that a better correlation is achieved.

Visual odometry, on the other hand, provides promising results for the recovery of camera trajectories from video sequences [10]. The development of visual odometry traces back to the early 1980's. Nowadays we have maturity of a variety of computer vision disciplines including stereo matching, optical flow computation, and feature tracking. This delivers a highly accurate estimation of a camera's ego-motion. In general, a state-of-the-art algorithm is capable to achieve a drift error within 1% in real-time along a long travel distance up to a few kilometres [11] .

In this paper we first propose a LiDAR-engaged visual odometry framework. Based on the framework, we also develop an online strategy to automatically track and correct the change of extrinsic parameters. The strategy is novel and different to any previous work; it takes image-structure consistency into account. Inheriting from the underlying visual odometry framework, our online calibrator evaluates the accuracy of the parameters.

The rest of this paper is organised as follows. Section II reviews some recently proposed online calibration techniques.

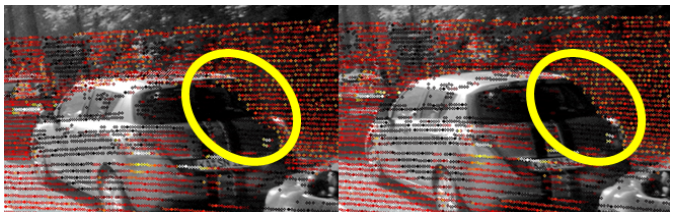

Fig. 1. Example of maximised MI. The projection of LiDAR points before (left) and after (right) the adjustment of extrinsic parameters is colour-coded by their intensity readings. The optimisation wrongly associates occluded points behind the car's back window.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

Section III explains the basics of our LiDAR driven visual odometry. Section IV proposes a novel visual odometry-driven solution to the online calibration problem. Experimental results and studies are given in Section V. Section VI concludes.

II. Related Work

The scope of the proposed method crosses two domains, namely visual odometry and online LiDAR-camera calibration. As the former has already been extensively studied for decades, in this section we discuss only online calibration techniques. For a comprehensive review of the development on visual odometry, readers are referred to [10].

For one of the early attempts to automatically find extrinsic parameters between a LiDAR and a camera system, see [6]. The work aligns LiDAR frames to camera images by contour matching. The edge points in each laser scan are identified and projected onto the image. The extrinsic parameters are then adjusted accordingly to improve the alignment of projected edge points to object contours detected in images.

A similar idea is proposed and verified in later work in [7]. An objective function is defined to capture the correlation of edge points and image edges. To instantiate such an objective function, edges are detected in an image. An *inverse distance transform* (IDT) is then applied on the edge image to produce an energy map. In the LiDAR domain, depth gradients are calculated and associated to each 3D measurement. Those measurements are then projected onto the IDT to find the associated energy, taking image and depth gradients into account. The summation of all the energy is then used to evaluate the fitness of hypothetic extrinsic parameters. Optimal parameters are determined by means of energy minimisation. To avoid associating occluded points to wrong pixels, [7] uses the 2D topology of the multi-layer LiDAR systems to filter out non-convex points. In some scenarios, we found that edges, which were due to image textures or shadows, were wrongly used as targets. This is not desired as these edges do not reflect depth discontinuities of scene structures.

In more recent work, [9] demonstrates how the change of extrinsics can be successfully identified and compensated by means of MI maximisation. The approach projects laser points into an intensity image, then uses these points' intensity readings from the LiDAR system plus the associated image intensities to compute the MI, which represents the dependency of data from different sensors. A high MI value is believed to indicate that the sensors are well modeled by the extrinsics. The authors also propose a sampling strategy to give a probability estimation on the extrinsic parameters' optimality. Occlusions, however, are not addressed in their paper. In our experiments we also found that image shadows may provide incorrect hints that guide the extrinsics' adjustment into a wrong way.

To deal with occlusions and shadows a stereo camera setup could be leveraged to acquire pixel-wise depth estimates which could then be used to identify inconsistent LiDAR projections. In this work we address such issues in a different way, without the requirement of an additional camera. The proposed method is inspired by the principle of visual odometry, which relies on good correspondences between scene structures and imagery data to deliver accurate motion estimates. The key concept is to use the ego-motion estimation as a guide to evaluate the correctness of LiDAR-image correspondences, and in turn the hypothetic extrinsic parameters. To our best knowledge, using visual odometry for online sensory calibration has never been reported before elsewhere.

III. Visual Odometry and LiDAR Systems

Recent research identified the potentials of LiDAR systems to dramatically boost the performance of odometry under challenging conditions over conventional vision techniques. In this section we first illustrate principles of visual odometry techniques. This is followed by a demonstration of a LiDAR-enhanced strategy.

Theory of Visual Odometry. The motion of a camera can be recovered given a set of correspondences from two consecutive frames. Let χ be an image feature, $\rho_k(\chi) \in \mathbb{R}^2$ its image coordinates, and $g_k(\chi) \in \mathbb{R}^3$ its 3D coordinates in frame k.

The motion $(\mathbf{R}_k, \mathbf{t}_k)$ of the camera, where $\mathbf{R}_k \in SO(3)$ is the rotation matrix and $\mathbf{t}_k \in \mathbb{R}^3$ the translation vector from frame k to $k+1$, can be determined in different ways, depending on the choice of correspondence types. In the 3D-to-3D case, $(\mathbf{R}_k, \mathbf{t}_k)$ is solved immediately from a set of Euclidean correspondences $g_k(\chi_i) \leftrightarrow g_{k+1}(\chi_i)$ using Horn's analytical solution [15]. This case is not preferable for the computer vision community as the 3D coordinates are usually measured from a disparity map, which scales errors in a nonlinearly anisotropic way once converted to Euclidean space.

For a stereo setup, the state-of-the-art approach uses 3D-to-2D correspondences to estimate camera motion by solving the *perspective-n-point* (PnP) problem. Given a set of correspondences $g_k(\chi) \leftrightarrow \rho_{k+1}(\chi)$, it finds the optimal parameters $(\mathbf{R}, \mathbf{t})$ that minimise the sum-of-square errors between each projected feature and its observation. This is modeled by the re-projection error function

$$\phi_{\text{RPE}}(\mathbf{R}, \mathbf{t}) = \sum_i \|\rho_{k+1}(\chi_i) - \pi(\mathbf{R} g_k(\chi_i) + \mathbf{t})\|^2_{\mathbf{\Omega}_i} \qquad (1)$$

where $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is the camera projection function and $\mathbf{\Omega}_i \in \mathbb{R}^{2\times 2}$ is the error covariance matrix of feature χ_i. The optimal solution of Eq. (1) can be approached by any nonlinear least-square solver; the Levenberg-Marquardt algorithm is one of the most popular choices [18].

In the monocular case where the Euclidean measurement $g(\chi)$ is not available, the epipolar condition is adopted instead. It is possible to algebraically recover camera motion purely from 2D correspondences $\rho_k(\chi) \leftrightarrow \rho_{k+1}(\chi)$, up to four candidate solutions along with scale ambiguity (i.e. the norm of $\mathbf{t}$ remains unknown). Ego-motion estimation is based on minimisation of the sum-of-squares epipolar error

$$\phi_{\text{EPI}}(\mathbf{R}, \mathbf{t}) = \sum_i \left\| \mathbf{y}^\top_{i,k+1} \cdot [\mathbf{t}]_\times \mathbf{R} \cdot \mathbf{y}_{i,k} \right\|^2 \qquad (2)$$

where $[\cdot]_\times$ is the skew-symmetric 3×3 matrix form of a normalised vector, and $\mathbf{y}_{i,k}$ are the homogeneous coordinates of the i-th feature's canonical coordinates in frame k, with

$$\mathbf{y}_{i,k} = \mathbf{K}^{-1} \begin{bmatrix} \rho_k(\chi_i) \\ 1 \end{bmatrix} \tag{3}$$

where $\mathbf{K}$ is the 3×3 camera matrix.

Existing visual odometry frameworks deploy a variety of disciplines, including feature matching and optical flow computation, for updating function ρ and to maintain good tracking of features; stereo matching or triangulation techniques are used to implement function g. The next subsection discusses some alternative solutions which incorporate LiDAR range data.

LiDAR-enabled Visual Odometry. Few LiDAR-enabled visual odometry systems were proposed in recent years. These systems take the advantage of accurate laser scanning and outperform conventional vision-based odometry methods. By early 2016, top-ranked odometry on the KITTI benchmark website [11] is dominated by LiDAR-based odometry (e.g. [12], [13], [14]).

A straightforward implementation of a LiDAR-engaged visual odometry system is to use laser-rendered depth maps for replacing those computed from stereo images. Laser points are projected on the image plane and triangulated to produce an up-sampled dense depth map (see Fig. 2 for an example). To enhance the resolution of the rendered depth map, multiple scans are accumulated and aligned using the estimated ego-motion [13].

An alternative strategy is to use the projections of laser points to establish initial features and start to track these features to assemble inter-frame 3D-to-2D correspondences. Let $\mathcal{L}_k(\chi)$ be the laser measured 3D coordinates of feature χ in the k-th LiDAR frame. We redefine the Euclidean measurement function

$$g_k(\chi; \mathbf{\Gamma}, \tau) = \mathbf{\Gamma}\mathcal{L}_k(\chi) + \tau \tag{4}$$

where g is now parametrised over extrinsic parameters $\mathbf{\Gamma} \in SO(3)$ and $\tau \in \mathbb{R}^3$. The image coordinates of χ in frames k and $k+1$ are then, respectively, decided by

$$\rho_k(\chi) = \pi \left[g_k \left(\chi; \mathbf{\Gamma}, \tau \right) \right] \tag{5}$$

and

$$\rho_{k+1}(\chi; \mathbf{\Gamma}, \tau) = \nu_k(\rho_k) \tag{6}$$

where an image feature transfer function $\nu_k : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is introduced. The function can be implemented using an optical flow technique or a point-tracking algorithm (e.g. the KLT algorithm [16]). Following these formulations, Eqs. (1) and (2) are now parametrised by the extrinsic parameters as well. This defines an embedded optimisation problem when parameters $(\mathbf{\Gamma}, \tau)$ are the considered variables. The framework is depicted by Fig. 3. In Section IV we study how such an embedded structure can be used to solve the online calibration problem.

Implementation Details. LiDAR-aided PnP ego-motion estimation, described previously, needs to be enhanced when applied to real-world sequences where non-stationary features and noisy tracking are present. In our implementation, image feature detection and extraction are performed independently from the LiDAR-engaged framework. For distinguishing the features inducted either from a projection of LiDAR points or directly detected in the image, we use $\hat{\chi}$ for denoting the latter ones. These image features are then temporally tracked in feature space.

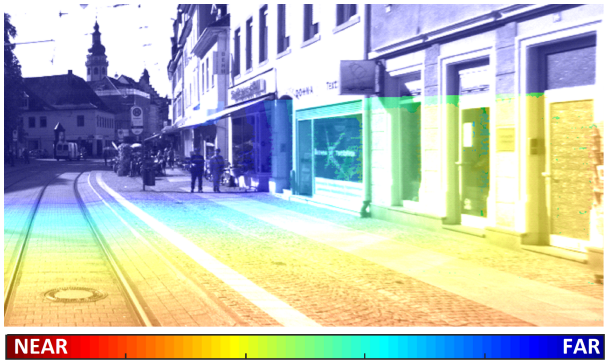

Fig. 2. Example of dense depth map from up-sampled LiDAR data.

To reject non-stationary image features, a robust *least-median-of-squares* (LMedS) estimator is adopted to compute the fundamental matrix from tracked feature correspondences $\hat{\chi}_k \leftrightarrow \hat{\chi}_{k+1}$. The fundamental matrix is then used to identify outliers from the correspondences $\hat{\chi}_k \leftrightarrow \hat{\chi}_{k+1}$, which are yielded by ν, by means of the Sampson error [18]

$$\varepsilon(\mathbf{x}, \mathbf{x}'; \mathbf{F}) = \frac{(\mathbf{x}'\mathbf{F}\mathbf{x})^2}{(\mathbf{F}\mathbf{x})_1^2 + (\mathbf{F}\mathbf{x})_2^2 + (\mathbf{F}^\top\mathbf{x}')_1^2 + (\mathbf{F}^\top\mathbf{x}')_2^2} \tag{7}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are the homogeneous coordinates of two tracked features, $\mathbf{F}$ is the estimated fundamental matrix, and $(\mathbf{F}\mathbf{x})_i^2$ is the square of the i-th entry of $\mathbf{F}\mathbf{x}$.

The correspondences $\hat{\chi}_k \leftrightarrow \hat{\chi}_{k+1}$ are also used to better condition the objective function ϕ_{RPE}. The epipolar geometry-derived regularisation term is defined as

$$\hat{\phi}_{\text{EPI}}(\mathbf{R}, \mathbf{t}) = \sum_j \left\| \varepsilon(\hat{\chi}_{j,k}, \hat{\chi}_{j,k+1}; \mathbf{K}^{-\top}[\mathbf{t}]_\times \mathbf{R}\mathbf{K}^{-1}) \right\|^2 \tag{8}$$

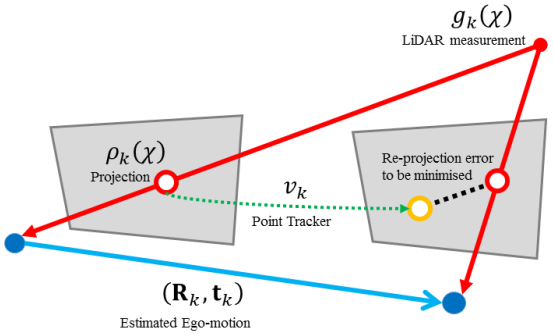

Fig. 3. Illustration of an alternative LiDAR-VO strategy considered in this paper. The framework is based on the PnP ego-motion estimation technique and robust feature tracking algorithms.

The term is then combined with the objective ϕ_{RPE} to yield a regularised objective function

$$\Phi(\mathbf{R},\mathbf{t}) = \phi_{\text{RPE}}(\mathbf{R},\mathbf{t}) + \lambda\ \hat{\phi}_{\text{EPI}}(\mathbf{R},\mathbf{t}) \tag{9}$$

where λ controls the importance of a penalty for a case that $(\mathbf{R},\mathbf{t})$ violates the epipolar condition constrained by $\hat{\chi}_k \leftrightarrow \hat{\chi}_{k+1}$.

Our implementation of the described monocular LiDAR-enabled visual odometry technique is tested on sequences from the KITTI dataset [11]. We choose SURF features to actualise the introduced outlier rejection and regularisation techniques which together achieve an average drift error within 2%. Figure 4 shows experimental results of a street-side sequence. By minimising only ϕ_{RPE} without the use of $\hat{\chi}$, the estimated trajectory deviates from the GPS/IMU data 4.5% after the vehicle traveled 200 m. With the outlier rejection using fundamental matrices from $\hat{\chi}_k \leftrightarrow \hat{\chi}_{k+1}$, the drift error reduced to 3.3%. By adding the epipolar regularisation term, the error further reduced to 1.6%, with λ set to 0.5. The threshold of the Sampson error is set to 0.05 pixel, above which a correspondence is considered to be an outlier.

IV. Visual-Odometry Driven Online Calibration

The proposed LiDAR-enabled visual odometry framework is not only an accurate ego-motion estimator but also a good tool for finding optimal extrinsic parameters. This is based on a straightforward idea that, if the LiDAR-camera extrinsics are far from the true values, the performance of the estimated ego-motion must be negatively affected. The idea is supported by our experimental findings. Figure 5 shows average drift errors over one of the test sequences with respect to shifted ground truth extrinsics along $x-$ and $y-$axes. With a clear convexity, the minimum is achieved when the calibrated extrinsics are not contaminated (i.e. $\Delta\tau_x = 0$ and $\Delta\tau_y = 0$). Such linkage allows us to embed the ego-motion estimation problem into online calibration which in turn leads to a bilevel optimisation structure. A similar solution to the structure-from-motion problem is discussed in the early development of a bundle adjustment technique [17], where the optimal scene structure recovery is embedded in the ego-motion optimisation process, or alternatively in reverse order.

Bilevel optimisation has been studied since the pioneering efforts on game theory by the German economist Heinrich Freiherr von Stackelberg in 1934. Unfortunately, due to its nested nonlinear structure, solving a bilevel optimisation problem still remains hard today. Next we propose a technique to ease the difficulty of solving such a problem in the context of an online calibration problem by introducing data constraints.

Embedded Optimisation Problem. The regularised objective function Eq. (9) is parameterized over the extrinsics (Γ,τ), which are now considered adjustable. We formulate the embedded optimisation problem as

$$\min_{\mathbf{\Gamma},\tau} \Psi(\mathbf{\Gamma},\tau;\mathbf{R},\mathbf{t}) \text{ s.t. } (\mathbf{R},\mathbf{t}) \in \arg\min \Phi(\mathbf{R},\mathbf{t};\mathbf{\Gamma},\tau) \tag{10}$$

where Ψ is an objective function that evaluates the fitness of hypothetic extrinsics, given an optimal ego-motion estimate.

A good choice of Ψ is to use the drift of the estimated ego-motion. However, such metric is not directly measurable without knowing the ground truth of camera motion. Alternatively, we perform a backward consistency check by evaluating the inverse re-projection error function

$$\bar{\phi}_{\text{RPE}}(\mathbf{R},\mathbf{t}) = \sum_i \left\| \rho_k(\chi_i) - \pi(\mathbf{R}^\top g_{k+1}(\chi_i) - \mathbf{R}^\top \mathbf{t}) \right\|^2_{\Omega_i} \tag{11}$$

which projects LiDAR measurements from frame $k+1$ as features and compares their back-traced image points with their projections in frame k using the inverse of $(\mathbf{R},\mathbf{t})$.

The effectiveness of such a measurement is verified by experiments which showed a significantly positive connection between Eq. (11) and the drift error by correlation analysis. Note that we avoid using a forward re-projection error as

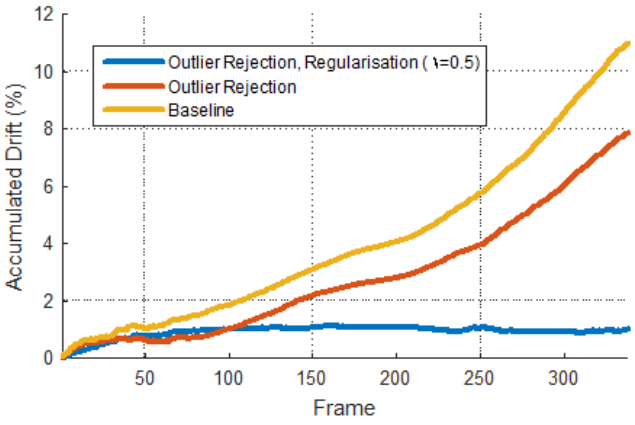

Fig. 4. LiDAR-VO framework results. The epipolar constrained objective function with outlier rejection strategy achieved significantly better accuracy in ego-motion estimation.

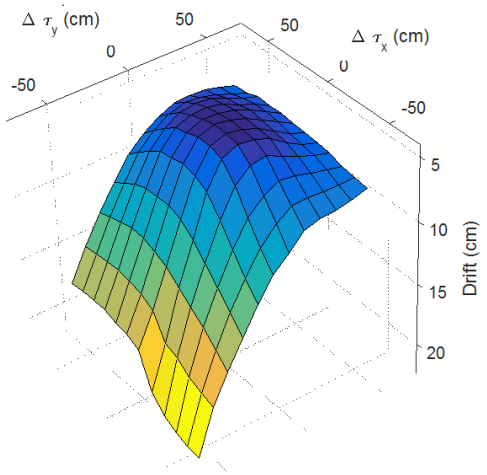

Fig. 5. Visualisation of ego-motion drift errors with respect to the shifted extrinsic parameters along $x-$ and $y-$axis. We applied the proposed LiDAR-VO strategy with modified extrinsics $\tau'_x = \tau_x + \Delta\tau_x$ and $\tau'_y = \tau_y + \Delta\tau_y$.

defined in Eq. (1), since it is already optimised by solving the inner problem. Figure 6 shows correlation coefficients of the forward and backward re-projection errors with respect to the drift errors of a test sequence. We used 144 sets of extrinsics to calculate the errors for each frame. The results are then accumulated over time to yield a better estimate of the Pearson product-moment correlation coefficients. The experiments suggest that the backward re-projection error, which is not directly optimised in the inner problem, presents a stronger connection to motion drift compared to the optimised forward re-projection error.

Despite being effective, the inverse re-projection function is ill-posed, as multiple local optima are found. In the following subsections we introduce data constraints to improve the condition of Ψ, which in turn allows us to deploy computational inexpensive solvers to solve the embedded problem.

Intensity Constraint. Commercial LiDAR systems return not only range data but also the strength of sensed laser pulses, which can be converted into intensity readings. To take into account photometric constraints, we use intensity data from the camera and the LiDAR system. We adopted the MI technique proposed in [9]. The MI of two signals is defined as

$$MI(\mathbf{a},\mathbf{b}) = \sum_i \sum_j p(\mathbf{a}_i,\mathbf{b}_j) \log\left(\frac{p(\mathbf{a}_i,\mathbf{b}_j)}{p(\mathbf{a}_i)p(\mathbf{b}_j)}\right) \quad (12)$$

where $p(\mathbf{a_i})$ and $p(\mathbf{b_i})$ are marginal distribution functions of $\mathbf{a}$ and $\mathbf{b}$ respectively, and $p(\mathbf{a}_i,\mathbf{b}_j)$ is the joint probability function.

In the context of a LiDAR-camera setup, we have $\mathcal{R}_k$ for the intensity readings and $\mathcal{I}_k$ for the intensity image of frame k. Applying Eq. (12) to associate the intensity reading of a feature χ and the image intensity value at the location of its projection, it yields

$$\phi_{\text{MI}}(\mathbf{\Gamma},\tau) = \frac{1}{MI\left\{\mathcal{R}_k(\chi_i),\mathcal{I}_k[\rho_k(\chi_i)]\right\}} \quad (13)$$

Our implementation uses a discretized *kernel method* to estimate the probability distribution functions. The estimation is efficiently done by convolving 1D and 2D histograms built from $\mathbf{a}$ and $\mathbf{b}$ with a chosen kernel function which is Gaussian in our work.

Discontinuity Constraint. An accurate set of extrinsic parameters is also supported by a good alignment of edge points (i.e. from the LiDAR point cloud to the image edges). We follow an approach similar to [7]. For each LiDAR point, we calculate its edginess by comparing its depth to the two nearest neighbors. Considering a feature χ_i and its nearest neighbors χ_i' and χ_i'', it defines

$$\delta_k(\chi_i) = \max\{z_k(\chi_i') - z_k(\chi_i), z_k(\chi_i'') - z_k(\chi_i), 0\}^\gamma \quad (14)$$

where z is the function to retrieve a feature's depth in frame k, and $\gamma > 0$ is the smoothness factor of the edginess transform. In this paper we set γ to 0.5, as suggested in [7]. The *alignedness* of the edge points to the image edges is then denoted by

$$\phi_{\text{EG}}(\mathbf{\Gamma},\tau) = \frac{1}{\sum_i \delta_k(\chi_i)\mathcal{J}_k[\rho_k(\chi_i)]} \quad (15)$$

where $\mathcal{J}_k$ is the inverse distance transform of the binary edge image of $\mathcal{I}_k$. Note that the formula listed here is the inverted version of its original definition, as authors in [7] solved the extrinsics by energy maximisation while in this work we consider minimisation problems.

Cost Aggregation and Optimal Solutions. For improving the smoothness and convexity of the backward re-projection error, we add data constraints. The aggregated objective function is then defined as

$$\Psi(\mathbf{\Gamma},\tau;\mathbf{R},\mathbf{t}) = \bar{\phi}_{\text{RPE}}(\mathbf{R},\mathbf{t};\mathbf{\Gamma},\tau) + \alpha\phi_{\text{MI}}(\mathbf{\Gamma},\tau) + \beta\phi_{\text{EG}}(\mathbf{\Gamma},\tau) \quad (16)$$

where α and β, respectively, control the penalties for violating the intensity and the edge-alignedness constraints.

The inner problem of optimising Φ as defined by Eq. (9) is essentially a least-squares minimisation problem which is efficiently solved using the Levenberg-Marquardt algorithm in our work. The outer problem, after being better conditioned, is solved by a gradient-descent search technique. The process is depicted by Fig. 7.

V. Experiments

The proposed online calibration technique has been tested using street-side sequences from the KITTI dataset [11]. The extrinsics are initially contaminated by a synthetic rigid transformation. The magnitude of the rotational disturbance is set to 2°, and the norm of the translational noise is set to 10 cm. The proposed method is then applied to adjust

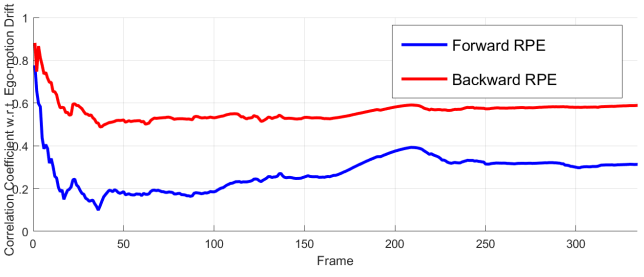

Fig. 6. The plot illustrates the correlation between forward and backward re-projection errors with respect to the drift of ego-motion.

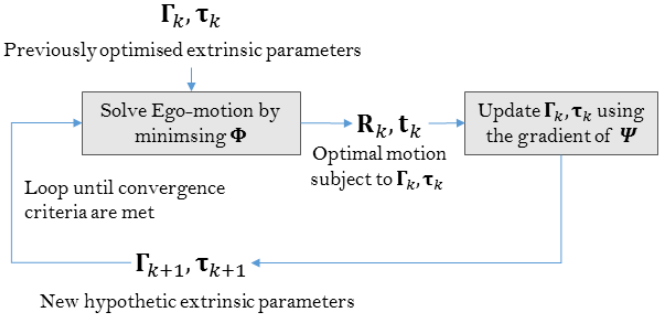

Fig. 7. Process of our online parameter adjustment strategy.

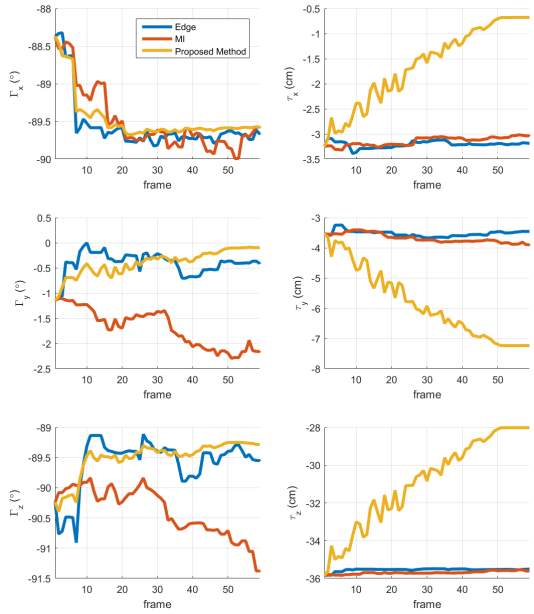

Fig. 8. Adjustment of extrinsics through first 60 frames.

the contaminated extrinsics through the sequence, using only the left greyscale camera. For the aggregation of Ψ, we used $\alpha = 0.5$ and $\beta = 0.5$. To evaluate the performance of the LiDAR-VO approach, we also applied online calibration using only the edge alignment objective ϕ_{EG} (based on [7]) and the intensity MI objective ϕ_{MI} (based on [9]), respectively. The adjusted parameters are compared with ground truth, which is obtained using offline calibration as proposed in [5]. The differential motion is calculated to evaluate the performance of each technique, and the Euler angles [i.e. pitch ($\mathbf{\Gamma}_x$), yaw ($\mathbf{\Gamma}_y$), and roll ($\mathbf{\Gamma}_z$)] are derived from the rotation part.

The adjustment of extrinsics is plotted in Fig. 8. The figure shows that all the tested techniques corrected the pitch angle to values close to the ground truth in the first 15 frames, while only the edge alignment approach and the proposed method converged to yaw and roll angles properly. We found that the high sensitivity of the vertical angular adjustment is due to the sampling of the Velodyne LiDAR system, which span the field of view of the camera very densely in horizontal direction but only narrowly in vertical direction. Under such circumstances, a little change in pitch greatly influences the objective functions.

The experiment also shows that the VO driven approach outperforms the other two in terms of translational parameter adjustment. We inspected the derivatives of ϕ_{EG} and ϕ_{MI} with respect to τ and found that they do not significantly influence a prominent update in the gradient descent iterations, as a change of several centimetres results in shifts under 1 pixel on the projection of the LiDAR points. When the VO technique is engaged, such small differences are captured in the ego-motion estimation stage and provide meaningful hints to guide the adjustment of parameters. The final extrinsic parameters are listed in Table I. The offline calibrated parameters are provided as the ground truth for comparison.

TABLE I
CALIBRATED EXTRINSIC PARAMETERS

	Γ_x (°)	Γ_y (°)	Γ_z (°)	τ_x (cm)	τ_y (cm)	τ_z (cm)
ϕ_{EG}	−89.7	−0.4	−89.6	−3.19	−3.46	−35.5
ϕ_{MI}	−89.6	−2.2	−91.4	−3.04	−3.89	−35.6
Ψ	−89.6	−0.1	−89.3	−2.79	−6.11	−28.3
Offline	−89.6	0.0	−89.1	−0.41	−7.63	−27.2

VI. CONCLUSION

In this paper we proposed a novel online calibration solver based on monocular LiDAR-enabled visual odometry. To ease the difficulty of solving the bilevel optimisation, as introduced by embedding ego-motion estimation into the online calibration problem, data constraints are added. Being better conditioned, the calibration can then be efficiently approached using gradient approaches. Experimental results show that the proposed method is able to re-estimate the extrinsic parameters more effectively than existing methods.

REFERENCES

[1] Quanergy, "Future of 3D Sensing and Perception: 3D LiDAR for ADAS, Autonomous Vehicles & 3D Mapping" www.quanergy.com, accessed on Mar 31, 2016.
[2] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," J. Field Robotics, 23:3–20, 2006.
[3] M. Maimone, Y. Cheng, and L. Matthies, "Two years of visual odometry on the mars exploration rovers," J. Field Robotics, 24:169–186, 2007.
[4] F. Mirzaei, D.G. Kottas, and S.I. Roumeliotis "3D LiDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," Int. J. Robotics Research, 31:452–467, 2012.
[5] A. Geiger, F. Moosmann, Ã. Car and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in Proc. IEEE Int. Conf. Robotics Automation, pp. 3936–3943, 2012.
[6] S. Bileschi, "Fully automatic calibration of LiDAR and video streams from a vehicle," In Proc. IEEE International Conference on Computer Vision Workshops (ICCV), pp. 1457-1464, 2009.
[7] J. Levinson and S. Thrun, "Automatic online calibration of cameras and lasers," In Proc. Robotics Science Systems, 2013.
[8] L. Tamas and Z. Kato, "Targetless calibration of a LiDAR - perspective camera pair," In Proc. IEEE Int. Conf. Computer Vision Workshops, pp. 668–675, 2013.
[9] G. Pandey, J.R. McBride, S. Savarese, and R.M. Eustice, "Automatic extrinsic calibration of vision and LiDAR by maximizing mutual information," J. Field Robotics, Special Issue: Calibration for Field Robotics, 32:696–722, 2015.
[10] D. Scaramuzza, and F. Fraundorfer, "Visual odometry: Part I - the first 30 years and fundamentals," IEEE Robotics Automation Magazine, 18:80–92, 2011.
[11] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," Int. J. Robotics Research, 32:1229–1235, 2013.
[12] J. Zhang and S. Singh, "LOAM: LiDAR odometry and mapping in real-time," In Proc. Robotics Science Systems, 2014.
[13] J. Zhang, M. Kaess and S. Singh, "Real-time depth enhanced monocular odometry," In Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems, pp. 4973–4980, 2014.
[14] J. Zhang and S. Singh. "Low-drift and real-time LiDAR odometry and mapping," In Proc. Autonomous Robots, 2016.
[15] B.K.P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Optical Society America, 4:629–642, 1987.
[16] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," In Proc. Imaging Understanding Workshop, pp. 121–130, 1981.
[17] Z. Zhang, Y. Shan, "Incremental motion estimation through local bundle adjustment." Technical Report MSR-TR-01-54, Microsoft, 2001.
[18] R. I Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", 2nd Ed., Cambridge University Press, 2004.