Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning

Joern Rehder[1,2], Paul Beardsley[2], Roland Siegwart[1] and Paul Furgale[1]

***Abstract*— This work presents a novel approach to spatio-temporal calibration of a laser range finder (LRF) with respect to a combination of a stereo camera and an inertial measurement unit (IMU). Spatial calibration between an LRF and a camera has been extensively studied, but so far the temporal relationship between the two has largely been neglected. While this may be sufficient for applications where the setup is mounted on a vehicle, which imposes bounds on the dynamics, we aim for employment on a hand-held scanning device, where angular velocities can easily exceed hundreds of degrees per second. Employing a continuous-time batch estimation framework, this work demonstrates that the transformation between the LRF and the visual/inertial setup—but also its temporal relationship—can be estimated accurately. In contrast to the majority of established calibration approaches, our approach does not require an overlap in the field of view of the LRF and camera, allowing for previously infeasible sensor configurations to be calibrated. Preliminary results for a novel hand-held scanning device suggest improvements in 3D reconstructions and image based point cloud coloring, especially for highly dynamic motions.**

I. Introduction

Time-of-flight laser scanning for 3D reconstruction is a mature technology with applications in fields ranging from reverse engineering of industrial plants, to architecture, to archaeology [1]. However, the vast majority of commercially available scanners operates stationarily, and in order to completely capture more complex environments where occlusions are present, the device has to be repositioned multiple times. On the other hand, triangulating, hand-held 3D scanners for small scale objects are widely applied in industry due to their easy deployment [2].

This work is motivated by the goal of developing a system that can be used to scan large scale structures, but that provides the same ease of deployment of a hand-held sensor, like the one depicted in Fig.1, comprised of a Hokuyo UTM-30LX laser range finder (LRF), rigidly connected to a visual/inertial sensor [3]. Arising from the extended range of dynamic motions of a human operator as compared to a ground vehicle, we employ a novel calibration approach that extends current state of the art by estimating the spatial transformation between the sensors as well as the time offset at which measurements are recorded. Additionally, our framework allows for the calibration of a broader variety of sensor configurations by discarding of the requirement of an overlapping field of view between camera and LRF.

[1]The authors are with the Autonomous Systems Lab (ASL), ETH Zurich {joern.rehder, paul.furgale}@mavt.ethz.ch, rsiegwart@ethz.ch

[2]The authors are with Disney Research Zurich pab@disneyresearch.com

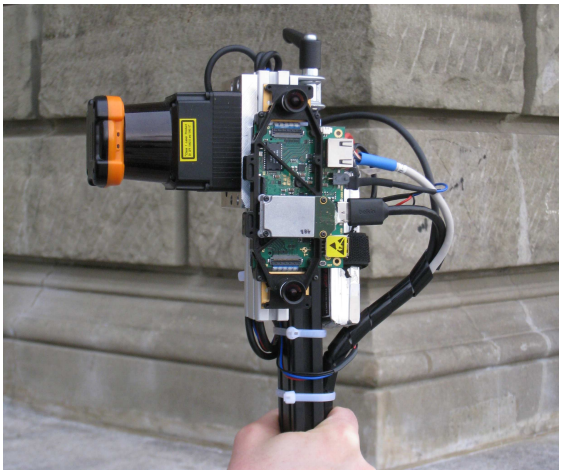


Fig. 1: The handheld scanning device comprised of a Hokuyo UTM-30LX laser range finder and a visual/inertial sensor [3].

To this end, a continuous-time batch estimation framework [4] is employed, extending our previous work on camera/IMU calibration [5] to integrate laser range measurements. As we perceive the requirement of an overlap in the field of view between the camera and the laser as unnecessarily constraining, ubiquitous planes are identified in the scan data and exploited for modelling the range measurements.

Section IV presents quantitative results for the calibration, demonstrating that the transformation can be estimated to millimeter accuracy and the time delay is determined up to about 5 ms precision. Furthermore, point cloud reconstructions obtained with our system prove the accuracy of the overall system and are well comparable with results reported for similar hand-held scanners [6], [7].

II. Related Work

Intensive research has gone into calibrating the transformation between laser scanners and cameras. Many approaches are designed for setups where the scanning plane is rotated [8], [9], [10], or for multi-beam systems [11], [12], [13], and are hence not applicable to our setup. For calibrating a setup of a rigidly connected camera and single-beam laser scanner, Zhang et al. [14] proposed an approach, where a set of simultaneously acquired images and static scans of a planar calibration target is used to establish the transformation between the two sensors. Other groups improved upon this algorithm [15], [16], while maintaining the same fundamental principle. Similarly, Núñez et al. [17] use simultaneous observations of a planar pattern, but additionally employ an inertial measurement unit to further

constrain the problem. Mei et al. [18] present an algorithm that makes use of the laser trace being visible in the image, which, while not relaxing the requirement of an overlapping field of view, constitutes a different approach to calibration. Bok et al. [6] generally follow the calibration procedure of [14], but additionally extract measurements of the edges of the calibration pattern from laser data to improve the results. Finally, [19] proposes a calibration method that matches edges detected in the image to plane intersections and boundaries in point clouds recorded with their Zebedee system [20]. The approach is capable of calibrating for devices, where the field of view of the camera does not overlap with the field of view of the range sensor. However, it requires the setup to be able to generate an accurate point cloud irrespectively of the transformation that is calibrated for, and hence is not applicable to our case.

With the exception of [19], these approaches have in common that they calibrate the setup based on static scans and completely neglect the temporal relationship between camera and laser scanner. While this might be sufficient for platforms with slow dynamics, it may result in significantly distorted reconstructions for hand-held systems, where angular velocities can reach hundreds of degrees per second. In contrast to these stationary calibration approaches, our calibration is based on continuous-time batch estimation [4], which allows for a seamless integration of time delays into the calibration framework, and it is an extension of [5]. The Zebedee system [20] continuously estimates the delay between different sensors in operation. While we can see the beauty in this system, our work takes a different approach: In order to increase robustness and decrease the size of the state estimated online, we try to accurately calibrate such quantities beforehand in an offline procedure in a lab environment.

III. Methodology

A. Experimental Setup

The scanning device is based on the visual/inertial sensor detailed in [3]. This sensor combines two global shutter MT9V034 WVGA image sensors in a plane-parallel stereo setup with an ADIS16448 inertial measurement unit. The integration of a XILINX Zynq, a combination of a dual core ARM processor with FPGA fabric, allows for accurate, exposure-compensated triggering of the cameras as well as synchronized polling of inertial data. The sensor has been augmented with a Hokuyo UTM-30LX, which has been rigidly mounted to the visual/inertial sensor.

B. Calibration

Our calibration is based on the continuous-time batch estimation framework proposed in [4]. In order to estimate the transformation between the laser range finder and the visual/inertial sensor and the inter-sensor delays, we extend our previous work on visual/inertial calibration presented in [5]. In the following, a brief recapitulation of the visual/inertial calibration framework will be provided, before the contribution to the objective function arising from laser measurements is derived in detail. With this, the description of the algorithm closely follows its processing procedure, as a two step approach is employed, where a smooth sensor path is estimated in a first step, followed by a step that adds laser terms to the estimation. We follow this two step approach, since a sufficiently accurate sensor trajectory is a prerequisite for obtaining an initial point cloud, which in turn is used to obtain a model for the laser measurements. The calibration procedure itself is similar—the setup is waved in front of a checkerboard in a way that excites all degrees of freedom sufficiently to render the calibration parameters observable—but we additionally require the sequence to be recorded in an environment, where a subset of the laser measurements are induced by at least one plane.

Recapitulation Camera/IMU Calibration: We employ B-splines to represent time-varying states and parametrize the time-varying transformation from the inertial coordinate frame into the world frame as a 6×1 spline, applying a Euclidean parametrization to translations and an angle/axis representation to orientations. With $\mathbf{C}(\cdot)$ being a function that constructs a rotation matrix from our orientation parametrization $\boldsymbol{\varphi}$ and $\mathbf{t}$ being the translation, the transformation from the body reference frame defined to coincide with the one of the IMU into the world reference frame $\mathbf{T}_{w,i}$ at time t may be expressed as

$$\mathbf{T}_{w,i}(t) = \begin{bmatrix} \mathbf{C}(\boldsymbol{\varphi}(t)) & \mathbf{t}(t) \\ \mathbf{0}^T & 1 \end{bmatrix} . \tag{1}$$

Given translations represented as composition of continuously differentiable basis functions, velocities $\mathbf{v}(t)$ and accelerations $\mathbf{a}(t)$ can be obtained by derivation. Angular velocities $\boldsymbol{\omega}(t)$ are obtained similarly with an additional transformation $\mathbf{S}(\cdot)$ relating parameter rates to angular velocities.

With this, the contributions from visual and inertial measurements to the objective function are

$$\mathbf{e}_{y_{mj}} := \mathbf{y}_{mj} - \mathbf{h}\left(\mathbf{T}_{c,i}\mathbf{T}_{w,i}(t_j)^{-1}\mathbf{p}_w^m\right) \tag{2a}$$

$$J_y := \frac{1}{2}\sum_{j=1}^{J}\sum_{m=1}^{M}\mathbf{e}_{y_{mj}}^T\mathbf{R}_{y_{mj}}^{-1}\mathbf{e}_{y_{mj}} \tag{2b}$$

$$\mathbf{e}_{\alpha_k} := \boldsymbol{\alpha}_k - \mathbf{C}\left(\boldsymbol{\varphi}(t_k)\right)^T\left(\mathbf{a}(t_k) - \mathbf{g}_w\right) + \mathbf{b}_a(t_k) \tag{2c}$$

$$J_\alpha := \frac{1}{2}\sum_{k=1}^{K}\mathbf{e}_{\alpha_k}^T\mathbf{R}_{\alpha_k}^{-1}\mathbf{e}_{\alpha_k} \tag{2d}$$

$$\mathbf{e}_{\omega_k} := \boldsymbol{\varpi}_k - \mathbf{C}\left(\boldsymbol{\varphi}(t_k)\right)^T\boldsymbol{\omega}(t_k) + \mathbf{b}_\omega(t_k) \tag{2e}$$

$$J_\omega := \frac{1}{2}\sum_{k=1}^{K}\mathbf{e}_{\omega_k}^T\mathbf{R}_{\omega_k}^{-1}\mathbf{e}_{\omega_k} \tag{2f}$$

$$\mathbf{e}_{b_a}(t) := \dot{\mathbf{b}}_a(t) \tag{2g}$$

$$J_{b_a} := \frac{1}{2}\int_{t_1}^{t_K}\mathbf{e}_{b_a}(\tau)^T\mathbf{Q}_a^{-1}\mathbf{e}_{b_a}(\tau)\,d\tau \tag{2h}$$

$$\mathbf{e}_{b_\omega}(t) := \dot{\mathbf{b}}_\omega(t) \tag{2i}$$

$$J_{b_\omega} := \frac{1}{2}\int_{t_1}^{t_K}\mathbf{e}_{b_\omega}(\tau)^T\mathbf{Q}_\omega^{-1}\mathbf{e}_{b_\omega}(\tau)\,d\tau \tag{2j}$$

where $\mathbf{h}(\cdot)$ is an arbitrary projection model, accepting checkerboard corners $\mathbf{p}_w^m$ transformed from the world frame

into the camera frame via the transformation $\mathbf{T}_{w,i}^{-1}$ at time t_j and the rigid transformation between inertial and camera frame $\mathbf{T}_{c,i}$. In contrast to [5], a time delay between camera and IMU is not considered, since it is compensated for in hardware [3]. Inertial measurements at times t_k contribute accordingly, with $\mathbf{g}_w$ being the estimated gravity and $\mathbf{b}$ denoting inertial sensor biases, parametrized as B-splines and modelled as driven by zero-mean white Gaussian processes, governed by covariance $\mathbf{Q}$. All measurements are weighted according to the inverse covariances, $\mathbf{R}^{-1}$, of the additive, zero-mean Gaussian distributed perturbation assumed to corrupt the measurement.

With these terms, the initial objective function J for estimating the sensor trajectory is composed as $J := J_y + J_\alpha + J_\omega + J_{b_a} + J_{b_w}$, which is minimized using the Levenberg-Marquardt (LM) algorithm [21]. Note that with the error term formulation provided above, this constitutes the maximum-likelihood estimation assuming that the perturbation model is sufficiently accurate.

Incorporating Laser Range Measurements: In the following section, we will derive the approach to modelling laser range measurements for a seamless integration into the objective function J. While there exist calibration approaches based on minimizing point cloud entropy [22] that omit assumptions about the scanned structure, we decided to explicitly model the laser range measurements as induced by a structure. For this, some knowledge about the structure is indispensable. Due to their ubiquity, we decided to identify laser measurements induced by planes and model them accordingly. Not relying on a calibration pattern for modelling laser measurements has distinct advantages: Intuitively, the observability of the transformation improves with target size, and manufacturing large targets can quickly become impractical. In contrast to this approach, most calibration approaches rely on an overlapping field of view between camera and laser [14], [15], [16], [17], which limits the applicable sensor configurations. For modelling the laser measurement, we make the following assumptions.

- The visual/inertial sensor is capable of estimating a sufficiently accurate trajectory.
- The initial guess for the transformation between laser and sensor is sufficiently accurate.
- A subset of all individual distances measured by the LRF is induced by planar structures.
- The range measurements are corrupted by additive zero-mean Gaussian distributed noise.
- There is zero error on the beam directions reported by the range finder.

Of these assumptions, the accuracy of the initial estimate of the transformation is the most constraining, as it may be hard to obtain for some setups. In this context, accuracy is sufficient if planes can reliably be identified in the point cloud obtained from the sensor trajectory and with the initial transformation, which is the case when a majority of laser measurements induced by a plane falls within an envelop defined by the chosen RANSAC [23] threshold.

Given a continuous trajectory of the sensor estimated by minimizing J as defined in Section III-B and an initial guess of the transformation $\mathbf{T}_{i,l}$ between the inertial sensor and the laser, an initial point cloud can be obtained by transforming the laser measurements $\mathbf{m}_k = [\alpha_k, l_k]^T$ of a single beam, cast at angle α_k and measuring range l_k into a common coordinate frame:

$$[\mathbf{p}_k, 1]^T = \mathbf{T}_{w,i}(t_k + \delta t)\mathbf{T}_{i,l}\,[l_k cos(\alpha_k), l_k sin(\alpha_k), 0, 1]^T\,, \tag{3}$$

where $\mathbf{T}_{w,i}$ denotes the transformation from the inertial into the world coordinate frame, t_k is the timestamp of a laser range measurement l_k with corresponding beam angle α_k, and δt the unknown inter-sensor delay. Fig. 2a shows an initial point cloud acquired from a calibration sequence. While some planes are visible, the entire point cloud appears cluttered, with many measurements not stemming from planar structures. In order to identify measurements induced by planes, a RANSAC scheme [23] is applied to the point cloud, with plane hypotheses generated from a minimal set of three non-collinear points and model support being evaluated according to threshold t

$$|\mathbf{n}_i^T \mathbf{p}_k - d_i| < t \quad , \tag{4}$$

with $\mathbf{n}_i$ being the normal of plane i, d_i the distance to the origin, and $\mathbf{p}_k$ being a point evaluated for support. Note that not only measurements induced by this plane satisfy this condition, but points on any structure within the threshold t from the intersection with the plane defined by $\mathbf{n}_i$ and d_i. To avoid outliers, the points on each plane are clustered by evenly discretizing them into spatial bins and, starting from the most populated bin, invoking adjacent bins into the plane until the ratio of points in adjacent bins falls below a threshold. As the laser itself scans in a plane, resting the sensor during calibration may result in an accumulation of points that may be detected as a plane in the RANSAC step. To avoid picking such accumulations as planes, we further require the number of points that fall into bins invoked into the plane to represent a certain percentage of all potential plane points and discard of the plane hypothesis otherwise. Having identified the measurements induced by plane i in the environment, and assuming zero error on the beam angle, a prediction of the range $\hat{l}_k$ measured by the LRF can be modelled as

$$\hat{l}_k = \left| \frac{\mathbf{n}_i^T \mathbf{t}_{w,l}(t_k + \delta t) - d_i}{\mathbf{n}_i^T \mathbf{r}_k(t_k + \delta t)} \right| \quad , \tag{5}$$

where $\mathbf{t}_{w,l}(t_k + \delta t) = \mathbf{C}_{w,i}(t_k + \delta t)\mathbf{t}_{i,l} + \mathbf{t}_{w,i}(t_k + \delta t)$ and $\mathbf{r}_k(t_k + \delta t) = \mathbf{C}_{w,i}(t_k + \delta t)\mathbf{C}_{i,l}\,[cos(\alpha_k), sin(\alpha_k), 0]^T$. With this, the contribution of the laser range measurements to the objective function J can be determined as

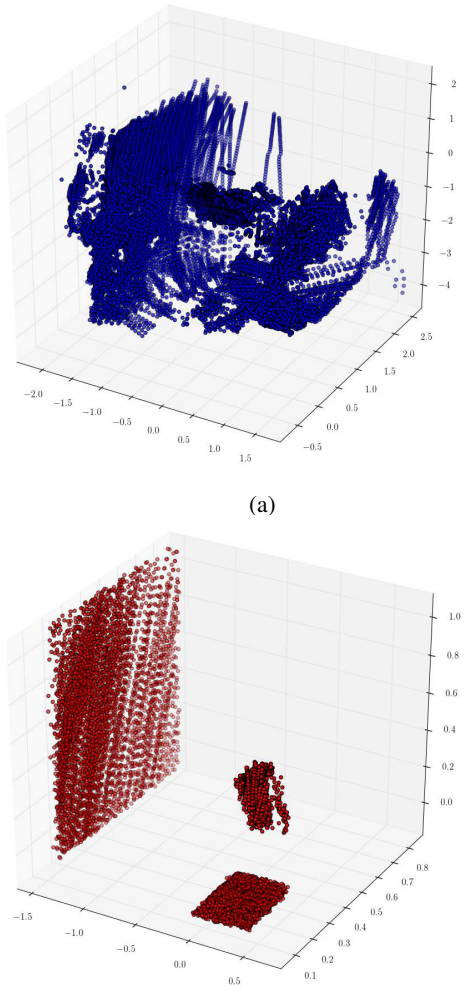
$$\mathbf{e}_l := \hat{l}_k - l_k \tag{6a}$$

$$J_l := \sum_{i=1}^{I} \sum_{k=1}^{K} \mathbf{e}_l^T \mathbf{R}^{-1} \mathbf{e}_l \tag{6b}$$

where $\mathbf{R}$ denotes covariance on the range measurements. The laser calibration quantities $\mathbf{T}_{i,l}$ and δt—along with the plane parameters $\mathbf{n}_i$ and d_i—are then estimated by minimizing the augmented objective function $J := J_y + J_\alpha + J_\omega + J_{b_a} + J_{b_w} + J_l$ analogously to Section III-B, using the result of the

trajectory generation as initial guess. To improve robustness to outliers, we further employ the Huber cost function [24] for the LRF error terms.

Note that the number of planes identified in the RANSAC step is a design parameter that can be adapted to the number of dominant planes in the environment scanned during calibration. Also note that we chose to implement the two different error metrics mostly for convenience: While the distance to the plane in the RANSAC step can be evaluated rapidly with minimal data association, the second metric models the plane induced distance measurement more accurately.



(a)

(b)

Fig. 2: Different stages of point cloud processing during calibration. Fig. 2a shows the initial point cloud generated from a calibration sequence. While planes are visible, the point cloud is cluttered with measurements stemming from non planar objects. Fig. 2b depicts the point cloud after identifying the three most dominant planes and clustering.

C. Model Generation

After calibration, the rig can be used to generate 3D models that consist of points from the LRF colored by intensity obtained with the cameras. To estimate the ego-motion of the sensor, the approach presented in [25] is employed. This algorithm administers a non-linear optimization over a sliding window of key-frames and corresponding inertial measurements. In order to keep computational complexity at bay and allow for real-time operation, entries that correspond to past velocities and inertial sensor biases are continuously marginalized from the estimated state. A detailed description of this framework is outside of the scope of this work, and the interested reader is kindly referred to the original publication.

We use the ROS framework [26] for transforming the laser range measurements from the coordinate frame of the scanner into the global coordinate frame given the static transformation between the laser and the visual/inertial sensor and state estimates from our visual/inertial framework. State estimates are published at the rate of the inertial measurements and interpolated to accommodate for the fact that the range measurements of a single scan are not acquired instantaneously but consecutively over a period of multiple milliseconds.

The laser scan data is augmented with image intensity values using the static transform between the two sensors. To this end, the range measurements are transformed into the camera coordinate frame and projected into the camera using its previously calibrated intrinsics and distortion model. Intensities are sampled at the projections of the laser range respective measurements. This approach constitutes a rather ad-hoc method to laser point cloud coloring, as it neglects the consecutive nature of laser range scans and does not take varying exposure times and gains in successive images into account. There are other approaches that employ more sophisticated coloring schemes, e.g. [27]. However in that approach, global exposure and gain equalization is performed in post-processing, which makes the method less well suited for immediate model feedback. Furthermore, by considering a single image and scan, the necessity for sophisticated occlusion reasoning is reduced. Although Fig. 5 seems to provide a counter argument for that, given that the method would sample incorrect intensity values for the cupboard occluded by the upper left corner of the checkerboard, one has to consider that this problem is induced by the baseline of the laser with respect to the camera, which is mitigated by the distance to the object, and thus less apparent for the majority of scanning use cases, as obvious from the correctly textured building shown in Fig. 7.

IV. Results

All following experiments employed the sensor device detailed on in Section III-A, hand guided by an operator. To this end, stereo pairs were recorded on a laptop computer at a rate of 20 Hz, inertial measurements at 200 Hz, and laser range scans at 40 Hz.

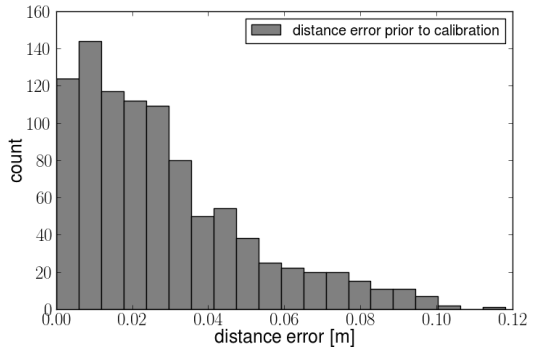
Calibration Results: This section details on results for the spatial and temporal calibration of the LRF with respect to the visual/inertial sensor. Apart from measurement covariances of all sensors and the parameters for the Gaussian processes modelling inertial biases, an initial guess for the transformation between LRF and sensor is required along

with a set of free parameters, that can be roughly categorized into parameters inherent to continuous-time batch estimation and parameters governing plane identification and robust estimation. For the continuous time estimation, we employed 6th order B-splines with 120 support points per second. We chose to identify three planes and picked a RANSAC threshold t of 50 mm, allowing for some envelope around the plane to account for errors due to time-delay and the inaccurate initial guess for the inter-sensor calibration. We stopped invoking bins into the plane when the neighbouring bin contained less then half as many points as the current, and discarded a plane hypotheses as potentially having been induced by the scanning plane rather than a structure when the number of clustered points represented less than half of all plane inlier. For down-weighting potential outliers via the Huber cost function, an outlier threshold of 20 mm was chosen in accordance with the measurement accuracies of the Hokuyo as reported by the manufacturer. For all experiments, we assumed zero delay and zero displacement of the laser with respect to the visual/inertial sensor. The relative orientation between the two in Euler angles in XZY convention was approximated by $180°, 0°$ and $90°$ Note that the correct selection of these parameters is crucial for achieving accurate calibration results, and we noticed that the approach is sensitive to a correct choice of the RANSAC threshold, which we found to be best set within the range of the accuracy of the LRF.

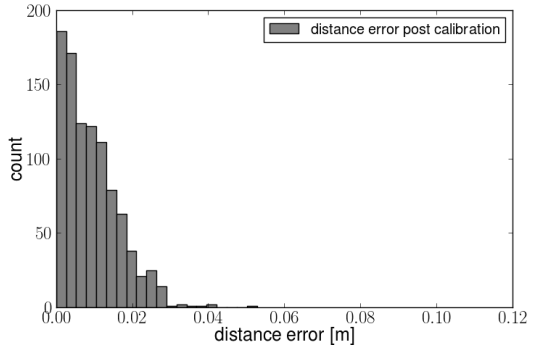
Fig. 3 displays the distribution in errors of the modelled range readings with respect to measured distance for a single plane in one dataset of about 10 s. After initializing the point cloud, but prior to performing the estimation (Fig. 3a) errors span an interval of up to 12 cm distance. Optimizing over the augmented cost function and including the spatio-temporal system parameters as well as plane parameters drives down the error to within the tolerance reported by the datasheet of the LRF, with few outliers (Fig. 3b). Neglecting the temporal relationship between the sensors results in a broader distribution of errors, with many modelled measurements falling outside of the range specified by the manufacturer (Fig. 3c). This suggests that neglecting the temporal relationship between the LRF and other sensors results in impaired reconstruction results.

Fig. 4 visualizes an experiment for evaluating the accuracy of the temporal calibration. As we lack information on the true delay between the sensors, we simulated different delays by altering the timestamps of the range measurements by -20 ms, 0 ms and 20 ms respectively. We then estimated each time-offset on ten datasets of 4 seconds of data taken from a longer calibration sequence. Of these ten estimates, the mean delays are 47.5 ms, 64.5 ms and 82.2 ms with standard deviations σ of 4.78 ms, 5.49 ms and 6.26 ms respectively. These results show that the approach is capable of estimating the simulated additions, which suggests that it can measure the absolute offset with similar accuracy.

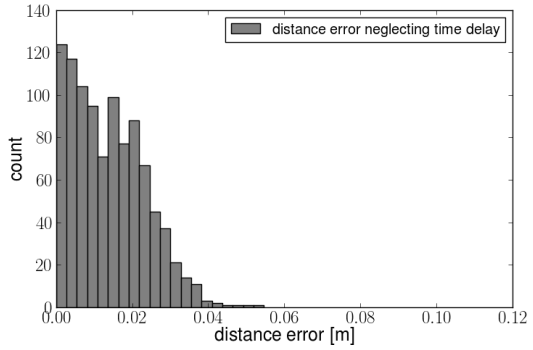
In order to evaluate the spatial calibration, a total of 20 datasets, each of about 40 seconds length, was recorded. The mean and standard deviation of all translation estimates expressed in the coordinate frame of the LRF is $6.34\ cm \pm 0.84\ cm$, $-1.59\ cm \pm 0.64\ cm$ and $-7.97\ cm \pm 2.75\ cm$, as compared to approximate hand measurements of $6.2\ cm, -1.4\ cm$ and $-9.4\ cm$. The large uncertainty in estimating the z component of the translation may be explained with a lack of rotational excitement in one axis in the datasets, likely caused both by a conservative operation by the user and by the narrow field of view of the image sensors that causes the sensor to lose track of the calibration pattern more easily when rotating in a certain axis. Orientation in Euler angles in XZY convention was estimated to be $179.73° \pm 0.10°, -1.47° \pm 0.47°$ and $90.78° \pm 0.46°$.



(a)



(b)



(c)

Fig. 3: Histogram of laser distance errors prior to (Fig. 3a) and after (Fig. 3b) estimation of the inter-sensor transformation and delay. Fig. 3c depicts results when the time delay between sensors is neglected, resulting in larger overall errors.

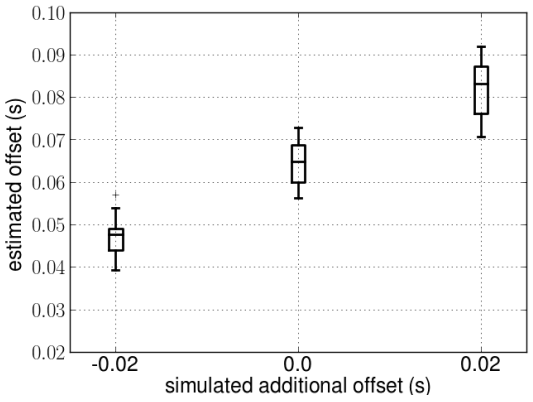


Fig. 4: Distribution of the intersensor delay. As ground truth was not available for the delay, the accuracy was evaluated by adding a simulated delay of -20 ms, 0 ms and 20 ms respectively to sets of each ten calibration sequences. Results suggest that the proposed approach is capable of accurately estimating the time delay for this sensor combination.

Fig. 5 provides an visual impression of the accuracy of the spatial calibration. As the image sensors on our device are susceptible to infra-red light, the scan line is visible as illuminated band on table and calibration pattern. The image is superimposed with the scan line modelled using the initial guess (blue) and the transformation as estimated by the framework (red). Although the intersection of the scan plane with two planes leaves unconstrained degrees of freedom, the resemblance of modelled and observed scan line in combination with the other results presented in this work suggest an accurate spatial calibration between the LRF and the visual/inertial sensor.

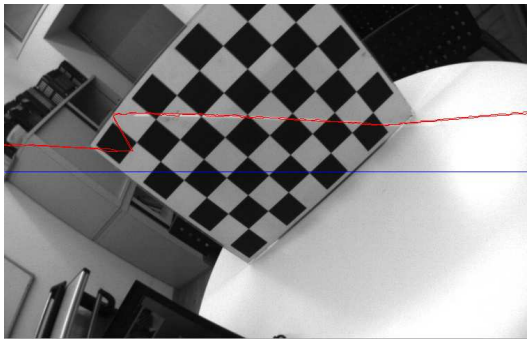
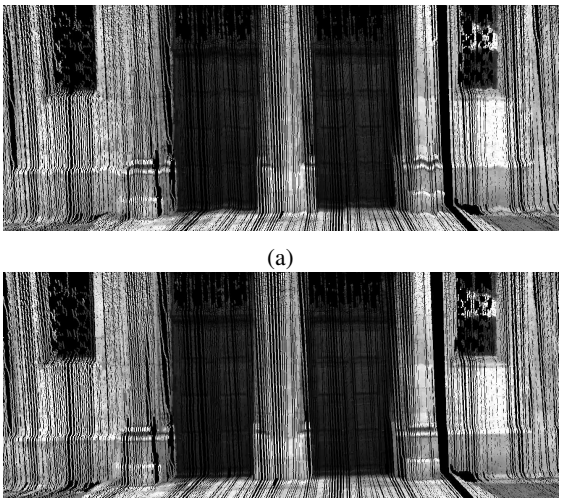


Fig. 5: Illustration of the calibration accuracy. The laser scanner emits infra-red light that the image sensors are susceptible to, rendering some of the points sampled by the laser range finder visible. The image is superimposed with a projection of the laser measurement into the camera using the initial guess (blue) and the calibrated transformation (red). The degree of coincidence of projected and visible scan line suggests an accurate calibration of the transformation between the sensors.

Preliminary Reconstruction Results: Fig. 6 provides the motivation for this work: In order to illustrate the necessity of a sufficient temporal calibration, the range scan measurements for generating the detail view in Fig. 6a were artificially delayed by 60 ms. The resulting point cloud exhibits noticeable distortions and incorrect coloring, most apparent for the right column, while correctly synchronized measurements allow for a crisp reconstruction(see Fig. 6b).



(a)

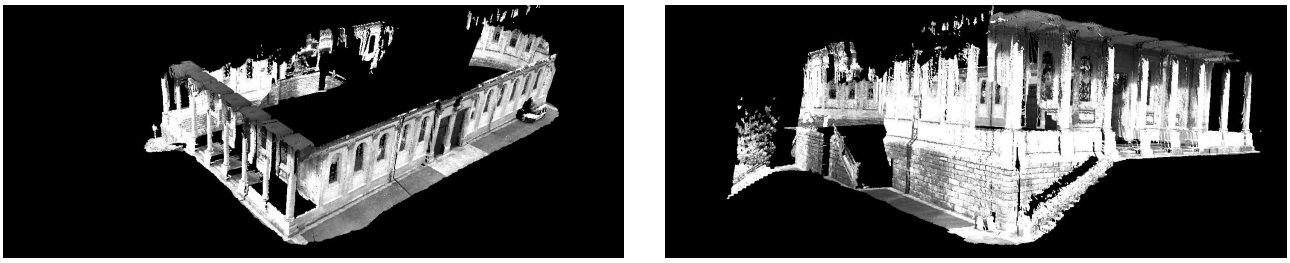
(b)

Fig. 6: Detail views of the rendered reconstructions. Laser range data in Fig. 6a exhibited a simulated delay of 60 ms with respect to the visual/inertial measurements, while in 6b, the sensors were correctly synchronized, resulting in an improved reconstruction of details, which is particularly apparent for the coloring of the column right of the door.

Fig. 7 depicts a sample reconstruction result obtained by scanning a church building. The dataset spans roughly 140 s with about 2800 image pairs, 5600 laser range scans and 28000 inertial measurements. The scanning path followed the contours of the façade. Please note the level of details both in reconstruction and coloring of the model, particularly in the details of the stone wall, which suggests that the calibration is reasonably accurate in estimating time delays as well as the transformation between the LRF and the camera sensor.

V. Conclusion and Future Work

This work proposed a spatio-temporal calibration for a combination of laser range finder, camera and inertial measurement unit. Furthermore, it presented preliminary results of colored point cloud reconstructions of buildings, acquired with a hand-held device. Its narrative follows previous work [5] in that it suggests an accurate and more complete offline-calibration of a multi-sensor setup, and part of its significance lies in the demonstration of unified spatio-temporal calibration applied to a novel combination of sensors. Another contribution lies in the demonstrated application to hand-held, large-scale scanning with results that compare well to other solutions [6], [7]. However, its broader applicability depends on the premise of a fixed time-delay between different sensors, which is not subject to changes on start-up or clock drift. In future work, it remains to be demonstrated that accurate timing can be reproduced over multiple start-ups of the system and that clock drift remains negligible within the time frame of a data collection campaign. Furthermore,



(a) (b)

Fig. 7: Two views of a 3D reconstruction of a church building. Only laser scan points are visualized for which an intensity value could be retrieved from the camera images. Please note the level of detail, particularly apparent on the stone wall, that suggest—besides accurate state estimation—a precise calibration of the laser with respect to the visual/inertial sensor.

by adding regularization terms to the objective function of the estimate, the approach could also be applied to a sensor suite without an IMU, and in the future, we would like to investigate this further.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude towards Pascal Gohl for his work on the experimental setup and Janosch Nikolic, Michael Burri and Stefan Leutenegger for their efforts invested into the visual/inertial sensor and the state estimation.

REFERENCES

[1] Leica Geosystems. (2014, Jan.) Leica ScanStation P20. [Online]. Available: http://www.leica-geosystems.com/en/Leica-ScanStation-P20_101869.htm

[2] Nikon. (2014, Jan.) ModelMaker MMDx digital laser scanner for portable 3D inspection and reverse engineering. [Online]. Available: http://www.nikonmetrology.com/en_EU/Products/Laser-Scanning/Handheld-scanning/ModelMaker-MMDx/

[3] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. T. Furgale, and R. Siegwart, "A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014.

[4] P. T. Furgale, T. D. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, St. Paul, MN, 14–18 May 2012, pp. 2088–2095.

[5] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, 3–7 November 2013, pp. 1280–1286.

[6] Y. Bok, Y. Jeong, D.-G. Choi, and I. S. Kweon, "Capturing village-level heritages with a hand-held camera-laser fusion sensor," *International Journal of Computer Vision*, vol. 94, no. 1, pp. 36–53, 2011.

[7] M. R. James and J. N. Quinton, "Ultra-rapid topographic surveying for complex environments: the hand-held mobile laser scanner (hmls)," *Earth Surface Processes and Landforms*, vol. 39, no. 1, pp. 138–142, 2014.

[8] H. Alismail, L. D. Baker, and B. Browning, "Automatic calibration of a range sensor and camera system," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, 2012, pp. 286–292.

[9] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 2007, pp. 4164–4169.

[10] R. Unnikrishnan and M. Hebert, "Fast extrinsic calibration of a laser rangefinder to a camera," Carnegie Mellon University, Robotics Institute, Tech. Rep. CMU-RI-TR-05-09, July 2005.

[11] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information," in *Proceedings of the AAAI National Conference on Artificial Intelligence*, Toronto, Canada, July 2012, pp. 2053–2059.

[12] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3d lidar–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 452–467, 2012.

[13] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3936–3943.

[14] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3. IEEE, 2004, pp. 2301–2306.

[15] F. Vasconcelos, J. P. Barreto, and U. Nunes, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2097–2107, 2012.

[16] G. Li, Y. Liu, L. Dong, X. Cai, and D. Zhou, "An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 2007, pp. 3854–3859.

[17] P. Núñez, P. Drews Jr, R. Rocha, and J. Dias, "Data fusion calibration for a 3d laser range finder and a camera using inertial data." in *ECMR*, 2009, pp. 31–36.

[18] C. Mei and P. Rives, "Calibration between a central catadioptric camera and a laser range finder for robotic applications," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 532–537.

[19] P. Moghadam, M. Bosse, and R. Zlot, "Line-based extrinsic calibration of range and image sensors," in *The 2013 IEEE International Conference on Robotics and Automation*, vol. 2, 2013, p. 4.

[20] M. Bosse, R. Zlot, and P. Flick, "Zebedee: Design of a spring-mounted 3-d range sensor with application to mobile mapping," *Transactions on Robotics*, vol. 28, pp. 1104–1119, October 2012.

[21] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial & Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[22] M. Sheehan, A. Harrison, and P. Newman, "Automatic self-calibration of a full field-of-view 3d n-laser scanner," in *Proceedings of the International Symposium on Experimental Robotics*, 2010, pp. 1–14.

[23] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[24] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Press, 2000, vol. 2.

[25] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart, "Keyframe-based visual-inertial slam using nonlinear optimization," in *Proceedings of Robotics: Science and Systems*, Berlin, Germany, 24–28 June 2013.

[26] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009.

[27] P. Chenga, M. Andersona, S. Heb, and A. Zakhor, "Texture mapping 3d models of indoor environments with noisy camera poses," in *SPIE Electronic Imaging Conference 9020, Computational Imaging XII*, San Francisco, CA, 2014.