E3S Web of Conferences **94**, 02005 (2019) https://doi.org/10.1051/e3sconf/20199402005
ISGNSS 2018

Stereo visual-inertial odometry with an online calibration and its field testing

Jae Hyung Jung[1], *Sejong Heo*[2], and *Chan Gook Park*[1*]

[1]Dept. of Mechanical & Aerospace Engineering / Automation and Systems Research Institute, Seoul National University, Republic of Korea
[2]Hanwha Corporation/Defense, Republic of Korea

Abstract. In this paper, we present a visual-inertial odometry (VIO) with an online calibration using a stereo camera in planetary rover localization. We augment the state vector with extrinsic (rigid body transformation) and temporal (time-offset) parameters of a camera-IMU system in a framework of an extended Kalman filter. This is motivated by the fact that when fusing independent systems, it is practically crucial to obtain precise extrinsic and temporal parameters. Unlike the conventional calibration procedures, this method estimates both navigation and calibration states from naturally occurred visual point features during operation. We describe mathematical formulations of the proposed method, and it is evaluated through the author-collected dataset which is recorded by the commercially available visual-inertial sensor installed on the testing rover in the environment lack of vegetation and artificial objects. Our experimental results showed that 3D return position error as 1.54m of total 173m traveled and 10ms of time-offset with the online calibration, while 6.52m of return position error without the online calibration.

1 Introduction

Ego-motion estimation is one of the most crucial tasks for unmanned vehicles such as planetary rovers or autonomous driving cars to successfully carry out their missions. However, to deal with the absence or outage of GNSS signals, alternative navigation algorithms should be considered. For instance, NASA's Martian rovers are equipped with stereo cameras and localize itself by the vision-based navigation called visual odometry (VO) [1]. While VO suffers from the well-known error accumulation, visual-inertial odometry (VIO) decreases its rate by filling a gap between small baselined images using IMU readings [2]. A fusion of a camera and IMU is an attractive solution due to their complementary features.

Most of the visual-inertial fusion algorithms assume that output data from a camera and IMU is timely synchronized and the sensors are spatially well aligned. However, this causes significant estimation errors when time-delay of a camera is not negligible or a camera-IMU system is not well calibrated since the measurement model is linearized around the currently available estimate referenced at the camera frame. Even if a camera-IMU system is calibrated in advance, this cannot reflect uncertainties on calibration parameters to an estimator. In the worst case, calibration parameters could be changed due to external shocks.

Many efforts to deal with the above issue has been made. The authors of [3] showed that the cam-IMU extrinsic parameter, the scale factor, and the global gravity is observable with the global pose measurements.

Fig. 1. Testing rover mounted the visual-inertial sensor

However, the measurement model which assumes that images output global poses was somewhat unrealistic. Guo et al. in [4] proved that cam-IMU extrinsic parameter is observable using the proposed basis functions under the known depth (feature point) assumption. The work of [5] focused on the temporal calibration of a cam-IMU system. They theoretically showed that time-offset between cam-IMU system can be recovered, while practically implemented the online calibration algorithm in the extended Kalman filter (EKF) framework. Also in [6], camera intrinsics, as well as IMU intrinsics (misalignment, g-sensitivity) was modeled in the estimator.

In this paper, we exploit the theoretic results of [4,5] and formulate EKF-based VIO algorithm using feature point measurements obtained from the stereo camera. Specifically, we augment the state vector with the time-

* Corresponding author: chanpark@snu.ac.kr

offset and extrinsic parameter. Fig. 1 shows the testing rover equipped with the visual-inertial sensor to record the dataset which lacks artificial object and vegetation.

2 The filter description

The error state vector of the presented algorithm consists of the 15th order of IMU state, the calibration parameters: cam-IMU time-offset, extrinsic parameter and the sliding window pose/velocity as in Eq. (1).

$$\begin{aligned}\tilde{\mathbf{x}} &= \begin{bmatrix}\tilde{\mathbf{x}}_I^{\mathrm{T}} & \tilde{\mathbf{x}}_C^{\mathrm{T}} & \tilde{\mathbf{x}}_S^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \\ \tilde{\mathbf{x}}_I &= \begin{bmatrix}\tilde{\boldsymbol{\theta}}_{GB}^{\mathrm{T}} & {}^{G}\tilde{\mathbf{p}}_B^{\mathrm{T}} & {}^{G}\tilde{\mathbf{v}}_B^{\mathrm{T}} & \tilde{\mathbf{b}}_a^{\mathrm{T}} & \tilde{\mathbf{b}}_g^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{15} \\ \tilde{\mathbf{x}}_C &= \begin{bmatrix}\tilde{\boldsymbol{\theta}}_{CB}^{\mathrm{T}} & {}^{C}\tilde{\mathbf{p}}_B^{\mathrm{T}} & t_d\end{bmatrix} \in \mathbb{R}^{7} \\ \tilde{\mathbf{x}}_S &= \begin{bmatrix}\tilde{\boldsymbol{\theta}}_{GB_i}^{\mathrm{T}} & {}^{G}\tilde{\mathbf{p}}_{B_i}^{\mathrm{T}} & {}^{G}\tilde{\mathbf{v}}_{B_i}^{\mathrm{T}}\end{bmatrix} \in \mathbb{R}^{9N}\end{aligned} \quad (1)$$

In this expression, we denote the global frame as $\{G\}$, the camera frame as $\{C\}$, the body (IMU) frame as $\{B\}$, and the number of sliding window as N. Also, we define the error state as $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ where $\hat{\mathbf{x}}$ means estimated value. For instance, $\tilde{\boldsymbol{\theta}}_{GB}$ is an attitude error expressed in a 3-dimensional vector, ${}^{G}\tilde{\mathbf{p}}_B$ and ${}^{G}\tilde{\mathbf{v}}_B$ are a position and velocity error expressed in $\{G\}$. Also, $\tilde{\boldsymbol{\theta}}_{CB}$ and ${}^{C}\tilde{\mathbf{p}}_B$ are extrinsic parameter error in a cam-IMU system, t_d is a cam-IMU time-offset due to latency in sensors defined as in [5]. Note that we include the sliding window velocity in the state vector since the measurement Jacobian matrix requires the current estimate of the velocity.

1.1 Prediction step

The IMU measurements are modeled as Eq. (3) with the zero-mean white Gaussian noise process (**n**), and the random walk process (**b**).

$$\begin{aligned}\mathbf{a}_m(t) &= \mathbf{R}_B^G(t)\left({}^{G}\mathbf{a}(t) - {}^{G}\mathbf{g}\right) + \mathbf{b}_a(t) + \mathbf{n}_a(t) \\ \mathbf{w}_m(t) &= \mathbf{w}_t(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t)\end{aligned} \quad (3)$$

Where ${}^{G}\mathbf{a}(t)$ is the true acceleration of the sensing platform, ${}^{G}\mathbf{g}$ is the global gravity that is approximately $\begin{bmatrix}0 & 0 & 9.81\end{bmatrix}^{\mathrm{T}}\ \mathrm{m/s^2}$. The error-state continuous time propagation is as follow, while calibration-related states are assumed to be constant.

$$\begin{aligned}\dot{\tilde{\mathbf{x}}}_I(t) &= \mathbf{F}_I(t)\tilde{\mathbf{x}}_I(t) + \mathbf{G}_I(t)\mathbf{n}_I(t) \\ \mathbf{n}_I(t) &= \begin{bmatrix}\mathbf{n}_a^{\mathrm{T}}(t) & \mathbf{n}_g^{\mathrm{T}}(t) & \mathbf{n}_{wa}^{\mathrm{T}}(t) & \mathbf{n}_{wg}^{\mathrm{T}}(t)\end{bmatrix}^{\mathrm{T}} \\ \dot{\mathbf{b}}_a(t) &= \mathbf{n}_{wa} \\ \dot{\mathbf{b}}_g(t) &= \mathbf{n}_{wg}\end{aligned} \quad (4)$$

In Eq. (4), $\mathbf{n}_{wa}$ and $\mathbf{n}_{wg}$ are zero-mean white Gaussian noise processes. The nominal IMU state is integrated through the closed-form state-transition matrix derived in [7].

1.2 Calibration parameter error modeling

To deal with the calibration parameters of a cam-IMU system, their error state should be modeled in the measurement model. Note that the extrinsic parameter is observable under the point features [4], and the time-offset is also observable up to the time referenced at an IMU [5]. These motivate us to jointly estimate the parameters along with the navigation solution in a stereo vision scenario which provides reliable depth information.

In order to build constraints among multiple views for point features, sliding window poses/velocity should be augmented to the state vector. Propagating the current IMU state up to the time-offset (t_d), the sliding window state is as follow,

$$\tilde{\mathbf{x}}_{S_i} = \begin{bmatrix}\tilde{\boldsymbol{\theta}}_{GB_i}^{\mathrm{T}}(t+\hat{t}_d) & {}^{G}\tilde{\mathbf{p}}_{B_i}^{\mathrm{T}}(t+\hat{t}_d) & {}^{G}\tilde{\mathbf{v}}_{B_i}^{\mathrm{T}}(t+\hat{t}_d)\end{bmatrix}^{\mathrm{T}} \quad (5)$$

Accordingly, the Jacobian matrix with regard to the IMU state is

$$\tilde{\mathbf{x}}_{S_i} \approx \begin{bmatrix}\mathbf{I}_9 & \mathbf{0}_{9\times 6} & \mathbf{0}_{9\times 6} & \mathbf{J}_{t_d} & \mathbf{0}_9\end{bmatrix}\tilde{\mathbf{x}}_I(t+t_d) \quad (6)$$

Where $\mathbf{J}_{t_d}$ is the Jacobian related to the time-offset, and can be derived using 1st order approximation,

$$\mathbf{J}_{t_d} = \begin{bmatrix}\hat{\mathbf{R}}_B^G(t+\hat{t}_d)\left\{\mathbf{w}_m - \hat{\mathbf{b}}_g(t+\hat{t}_d)\right\} \\ {}^{G}\hat{\mathbf{v}}_B(t+\hat{t}_d) \\ \hat{\mathbf{R}}_B^G(t+\hat{t}_d)\left\{\mathbf{a}_m - \hat{\mathbf{b}}_a(t+\hat{t}_d)\right\} + {}^{G}\mathbf{g}\end{bmatrix} \in \mathbb{R}^9 \quad (7)$$

Assuming that a stereo camera is well calibrated in advance, the point feature measurement model is

$$\begin{aligned}\mathbf{z}(t) &= \begin{bmatrix}1/Z_L\,\mathbf{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & 1/Z_R\,\mathbf{I}_2\end{bmatrix}\begin{bmatrix}{}^{C_L}\mathbf{p}_f(1:2) \\ {}^{C_R}\mathbf{p}_f(1:2)\end{bmatrix} + \mathbf{n}_z \\ {}^{C_L}\mathbf{p}_f(t+t_d) &= \begin{bmatrix}X_L & Y_L & Z_L\end{bmatrix}^{\mathrm{T}} \\ {}^{C_R}\mathbf{p}_f(t+t_d) &= \begin{bmatrix}X_R & Y_R & Z_R\end{bmatrix}^{\mathrm{T}}\end{aligned} \quad (8)$$

with zero-mean white Gaussian noise process, $\mathbf{n}_z$. The linearized measurement model is given by,

$$\mathbf{r} = \mathbf{z}(t) - \hat{\mathbf{z}}(t+\hat{t}_d) \approx \mathbf{H}(t+\hat{t}_d)\tilde{\mathbf{x}}(t+\hat{t}_d) + \mathbf{n}_z \quad (9)$$

where **H** matrix is the measurement Jacobian matrix. Specifically, this matrix is computed from the left and right measurements,

$$\mathbf{H}(t+\hat{t}_d) = \left(\frac{\partial \mathbf{z}}{\partial {}^{C_L}\mathbf{p}_f}\frac{\partial {}^{C_L}\mathbf{p}_f}{\partial \mathbf{x}} + \frac{\partial \mathbf{z}}{\partial {}^{C_R}\mathbf{p}_f}\frac{\partial {}^{C_R}\mathbf{p}_f}{\partial \mathbf{x}}\right)_{t+\hat{t}_d} \quad (10)$$

To compute Eq. (10), the global pose is perturbed up to the 1st order Taylor series expansion with respect to the time-offset,

$$\begin{aligned} \mathbf{R}_B^G(t+t_d) &\approx \mathbf{R}_B^G(t+\hat{t}_d) + \dot{\mathbf{R}}_B^G(t+\hat{t}_d)\tilde{t}_d \\ {}^G\mathbf{p}_B(t+t_d) &\approx {}^G\mathbf{p}_B(t+\hat{t}_d) + {}^G\mathbf{v}_B(t+\hat{t}_d)\tilde{t}_d \end{aligned} \quad (11)$$

The linearization in Eq. (11) enables us to model the time-offset in the measurement model.

3 Experimental results

The testing rover in Fig. 1 consists of Pioneer3-AT (rover platform), Xsens MTi-300 (IMU), ZED stereo camera, and the on-board computer for the purpose of data recording. While IMU outputs its data at 200Hz, the stereo camera gives 1280x720 grey images at 15Hz. Although both sensors are timestamped under the ROS environment, the nature of the separate system motivates us to estimate the time-offset. The initial guess of the extrinsic parameter was computed using Kalibr toolbox [8]. Also, a human pilot drove the testing rover returning to the starting point to quantify return position error. The typical environment of the site is shown in Fig. 2 which lacks artificial object and vegetation.

In what follows, we describe details of the vision front-end implementation and field testing results.

3.1 Vision front-end design

Features are provided to the estimator when either tracking fails or the number of tracks exceeds a user-defined maximum sliding window. To obtain reliable sets of feature tracks, we design a stereo feature tracker shown in Fig. 2 as similar to [9]. Assuming that the feature correspondence at t_{k-1} is obtained, features on the left image are tracked to the next time step t_k, then 8-point RANSAC eliminates outlier sets. Survived inliers on the left image are kept tracked to the right image. Again, 8-point RANSAC detects outliers between temporal right images at t_{k-1} and t_k. In the only case when the feature is successfully tracked both the temporal and static tracking, the feature is fed to the estimator. In contrast to a monocular case, the stereo features give scale information due to the baseline. Specifically, we triangulate feature points from the farthest two-view; for instance, the oldest frame in the left and the latest frame in the right before the multi-view triangulation. This strategy enables us to compute the feature depth, even the sensors are in static.

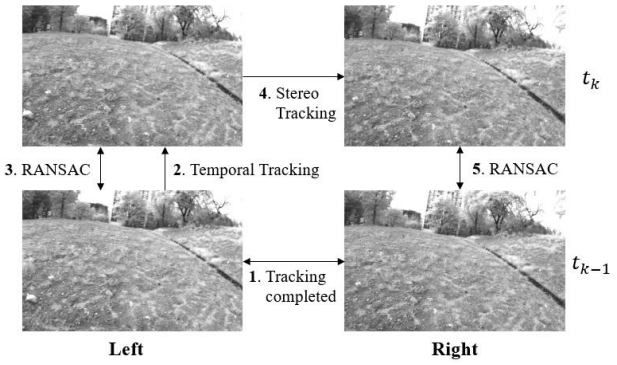

Fig. 2. Feature tracking strategy using stereo images

3.2 Field testing results

To test the presented VIO, the testing rover shown in Fig. 1 traveled the total distance of 173m for 224 seconds commanded by the human pilot. The testing site mainly consisted of soil with small rocks where typical images are shown in Fig .2. To compute an initial attitude with respect to the navigation frame (NED-frame), outputs of the accelerometer at the first 2 seconds were used. Also, the testing rover started from the static state; the initial velocity was set to zero.

To quantify the performance of the algorithm, we compare return position errors among three cases: "full calibration (td + extrinsic)", "partial calibration (only td)" and "no calibration". Table. 1 shows 3D return position of 3 cases in the Cartesian coordinate in which the starting point was $[0 \quad 0 \quad 0]^T m$. As expected the full calibration yields the best performance (2-norm) that is 76.4% error decrease when compared to the no calibration. It is interesting to note that the z-axis position of the partial calibration drifted up to -4.07m. We argue that this is due to the inaccurate extrinsic parameter that is computed beforehand. Also, Fig. 3 plots the entire estimated 2D trajectories of all cases. It is clearly seen in Fig. 3 that the no calibration largely drifts after the first 180 deg turning when compared the others.

Fig. 4 plots the estimated time-offset with its 3-sigma envelopes in the full calibration scenario. After quick convergence at the beginning, it converges to -10.3ms. Remind that the sampling time of images is 66.7ms (15Hz), thus the time-offset is not negligible.

Table 1. 3D return position errors for 3 cases

	No Calibration	Partial Calibration	Full Calibration
xyz Return position [m]	-0.8229; 3.3352; -5.5402	0.3627; 0.1956; -4.0725	-1.1922; 0.5711; -0.7856
2-norm [m]	6.52	4.09	1.54

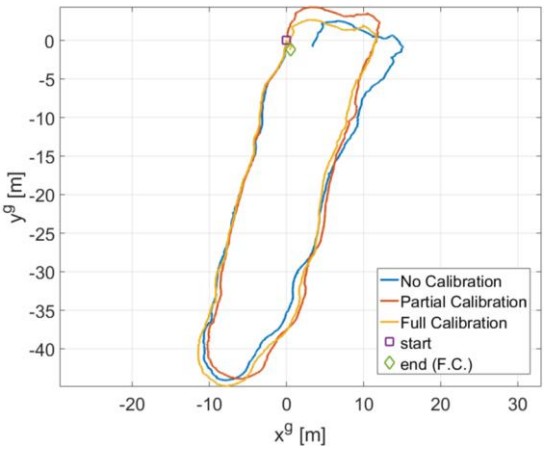

Fig. 3. Estimated 2D trajectory with the online calibration

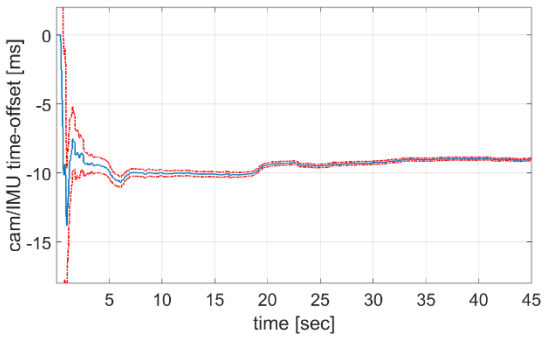

Fig. 4. Estimated cam/IMU time-offset with 3-sigma envelope

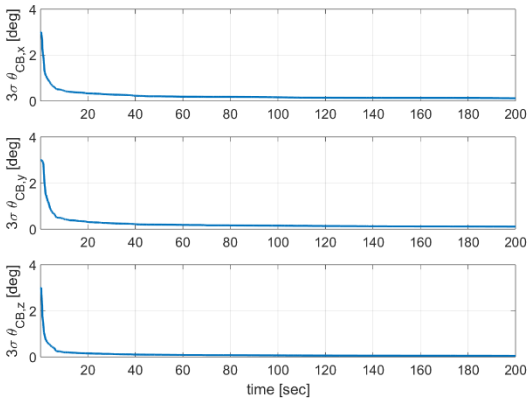

Fig. 5. Cam/IMU relative attitude 3-sigma envelopes

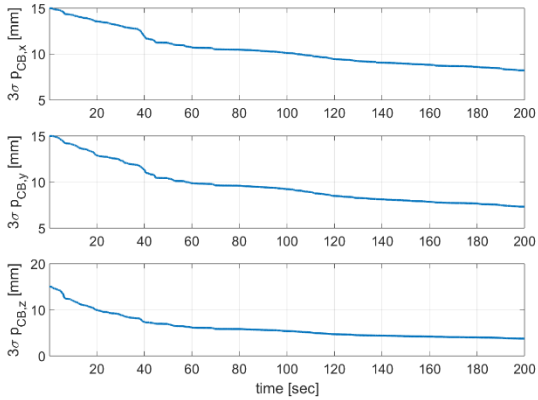

Fig. 6. Cam/IMU relative position 3-sigma envelopes

Fig. 5 and 6 show 3 standard deviations for the cam/IMU extrinsic parameter for each axis in the full calibration. We set the initial standard deviation as 1deg and 5mm respectively to cover the calibration uncertainty. Although we do not have the true value, the standard deviations are decreased as the filter is updated that is consistent results to the observability analysis in [4].

4 Conclusion

In this paper, we have presented the online calibration stereo VIO using naturally occurring point features in which the state vector is augmented by the time-offset and extrinsic parameters. To evaluate the presented VIO, the testing rover with the commercially available visual-inertial sensor recorded the dataset. Our experimental results have shown that when fusing independent sensors their extrinsic calibration is important; the online calibration method reduced the rover's return position error by 76.4% with respect to the no calibration method. Moreover, we experimentally showed that the time-offset and extrinsic parameter were observable under point features that is consistent with the observability analysis.

This work was supported by the Ministry of Science and ICT of the Republic of Korea through the Space Core Technology Development Program under Project NRF-2018M1A3A3A02 065722.

References

1. M. Maimone, Y. Cheng, L. Matthies, J. Field Roboticis, *Two years of visual odometry on the mars exploration rovers*, **24**, 3, 169-186 (2007)
2. D. Scaramuzza, F. Fraundorfer, IEEE robotics & automation magzine, *Visual odometry [tutorial]*, **18**, 4, 80-92 (2011)
3. J. Kelly, G. S. Sukhatme, Int. J. of Robotics Research, *Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration*, **30**, 1, 56-79 (2011)
4. C. X. Guo, S. I. Roumeliotis, IEEE Int. Conf. on Robotics and Automation, *IMU-RGBD camera extrinsic calibration: Observability analysis and consistency improvement*, Proc. IEEE Int. Conf. on Robotics and Automation (2013)
5. M. Li and A. I. Mourikis, Int. J. of Robotics Research, *Online temporal calibration for camera–IMU systems: Theory and algorithms*, **33**, 7, 947-964 (2014)
6. M. Li, H. Yu, X. Zheng, A. I. Mourikis, *High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation*, Proc. IEEE Int. Conf. on Robotics and Automation (2014)
7. M. Li, A. I. Mourikis, Int, J. of Robotics and Research, *High-precision, consistent EKF-based visual-inertial odometry*, **32**, 6, 690-711 (2013)
8. P. Furgale, J. Rehder, R. Siegwart, IEEE Int. Conf. On Intelligent Robots and Systems, *Unified temporal and spatial calibration for multi-sensor systems* (2013)
9. K. Sun, K. Mohtam B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, V. Kumar, IEEE Robotics and Automation Letters, *Robust stereo visual inertial odometry for fast autonomous flight*, **3**, 2, 965-972 (2018)