

Evaluation of Visual Inertial Navigation System for Autonomous Robot Tours on Campus

Subhobrata Chakraborty
Department of Computer Science
California State University
Northridge, CA, USA
subhobrata.chakraborty.026@my.csun.edu

Abhishek Verma
Department of Computer Science
California State University
Northridge, CA, USA
abhishek.verma@csun.edu

Amiel Hartman
Department of Mechanical Engineering
California State University
Northridge, CA, USA
amiel.hartman@csun.edu

Abstract— Autonomous mobile robot navigation is a challenging area of research due to various physical, hardware, and software issues. In this research, an autonomous robot system has been developed, which incorporates a visual inertial navigation system (VINS) with the goal that the robot could conduct automated university campus tours. Mapping and state estimation rely on the accuracy acquired from fusing the data from the cameras and inertial measurement units (IMU). The fusion of these two sensors makes VINS systems more accurate and robust.

We created a custom stereo inertial system and performed extensive evaluation to mitigate calibration issues such as noise and bias for accurate state estimation. The custom sensor infrastructure can be mounted on any mobile system. Furthermore, we tested our evaluation methodology on two challenging benchmark datasets, namely, EuRoC and TUM to determine the precision of the state estimator. Experiments show that VINS achieved highly accurate results in terms of calibration, reprojection, and trajectory estimation.

Keywords— visual inertial navigation system, MSCKF, calibration, state estimation, EKF, UKF.

I. INTRODUCTION

Autonomous ground based mobile robotic systems and unmanned aerial vehicles (UAVs) are increasingly widespread. Their extensive use can be attributed, in part, to a significant surge in computational capabilities coupled with a simultaneous reduction in the cost and power consumption. Empowering such robotic systems with the capability to comprehend and interpret their positions within local surroundings is essential for various applications ranging from augmented reality and virtual reality to self-directed navigation that frequently incorporates the utilization of visual inertial navigation systems (VINS). Those systems leverage information from onboard cameras and inertial measurement sensors, which are subsequently fused together to provide accurate state estimation for the robotic system.

Creating an operational visual inertial navigation system algorithm from the ground up is challenging. There is a lack of available open-source codebases with comprehensive documentation and derivations. Such a situation has caused the progress to be slow in this domain and it also inhibits the scope of extended research. There are various open-source visual inertial codebases [1, 2, 3] but they do not prioritize extensibility and suffer from inadequate documentation and evaluation tools. However, open VINS [4] has been essential in bridging that gap.

We evaluate a visual inertial navigation system pipeline for autonomous robots with the goal that the robot can conduct automated university campus tours. Our research is focused on extending the application of such VINS systems to a more modular setup, so that it could be deployed on any robotic system. Our research also focuses on creating custom stereo inertial sensor configuration to facilitate the evaluation of the simultaneous localization and mapping (SLAM) system in real time. We focus on calibration of the sensors to mitigate noise and bias while also comparing the results on two challenging benchmark datasets to assess the competence of the system.

The subsequent sections of the paper are organized in the following manner. Section II focuses on the prior relevant works related to our research. Section III provides details of the benchmark datasets used for conducting the experiments. Section IV covers details of the evaluation methodology of this research, while section V outlines the experimental findings, results, and discussions. Section VI gives the conclusion and the probable areas to extend this research.

II. RELATED WORK

Over the preceding few decades, state estimation has been the center of immense attention in research. Numerous methods have been proposed to address the accurate estimation of 6-DoF (Degrees of Freedom) poses. Several methods have been proposed focusing on different types of sensors such as visual-based [5, 6, 7], LiDAR-based [8], RGB-D-based [9] and event-based [10]. Achieving 6-DoF pose estimation with a monocular camera could be challenging due to the inherent inability to recover absolute scale from a single camera. To enhance the dependability of the system, it is a common practice to integrate multiple sensors.

There are two prevalent trends that exist in multi-sensor fusion approaches and VINS systems, these are optimization based and filter-based methods. Filter-based methods are known to utilize the extended Kalman filter (EKF). In such methods, inertial and visual measurements are often filtered together to estimate the state of the robotic system. High-rate inertial sensors are responsible for facilitating state propagation, while visual measurements provide updates [11, 12]. The multi-state constraint Kalman filter (MSCKF) [13, 14], maintains various poses of the camera and utilizes various camera viewpoints for a multi-constraint update. However, methods relying on filter-based approaches often come across challenges due to early linearization of states, which leads to errors caused by imprecise linear points.

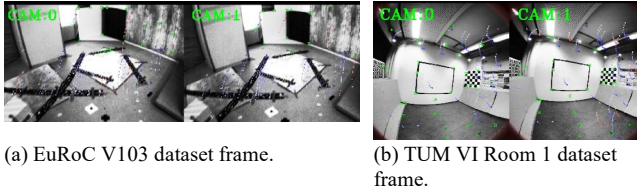


Fig. 1. Examples of EuRoC [17] and TUM VI [18] dataset frames used for evaluation of VINS.

In order to address the inconsistencies resulting from the linearized errors, the observability constrained EKF [15] was suggested to enhance precision and stability. An alternative approach using the unscented Kalman filter (UKF) [16] was introduced, which integrated LiDAR, visual and GPS. The UKF is an extension of the EKF without the analytic Jacobians and it aims to mitigate sensitivity to time synchronization inherent in filter-based methods. Delayed measurements in those methods can disrupt the filtering procedure as states cannot be retroactively propagated. Therefore, a specific sequencing mechanism is essential to ensure that data from multiple sensors are sequenced in the proper manner.

III. DATASET DESCRIPTION

A. EuRoC

EuRoC [17] is a publicly available dataset used for research and is primarily designed to assess and benchmark the effectiveness of various methods within the context of micro aerial vehicles and their sensor infrastructure. The data was generated from several sensors including stereo cameras and inertial measurement units. It was collected using micro aerial vehicles in diverse indoor and outdoor settings. It is a benchmark dataset for conducting experiments in visual odometry, SLAM, sensor fusion, and other applications in computer vision and robotics.

B. TUM VI

Fusing sensor and inertial data improves the precision and resilience in visual inertial odometry methods. TUM visual inertial dataset [18] consists of a varied range of data sequences captured in distinct scenarios for evaluation of visual inertial odometry algorithms. It has stereo pairs of images with a resolution of 1,024 x 1,024. The images are captured at 20 frames per second with HDR and photometric calibration. The inertial measurement unit (IMU) calculates the acceleration and angular velocities at 200 frames per second along the three axes. The IMU along with the camera data are synchronized in time within the hardware. To assess the precision of the trajectory, they provide the ground truth data that is collected while utilizing a motion capture system operating at a frequency of 120 Hz. It also provides color and depth data including the ground truth. It contains the color and the depth images captured at a frequency of 30 frames per second (FPS) and a sensor resolution of 640 x 480. The accelerometer data was collected from the Kinect sensor. Fig. 1 illustrates sample frames for the EuRoC V103 and TUM VI room 1 datasets.

IV. RESEARCH METHODOLOGY

We created custom stereo inertial sensor configuration to facilitate the evaluation of VINS in real time for autonomous

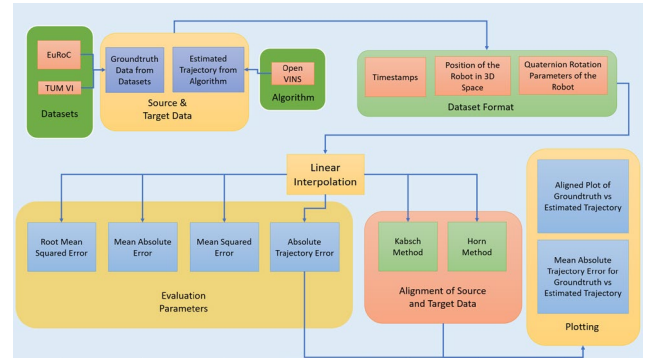


Fig. 2. Evaluation methodology of the visual inertial navigation system.

robots with the goal that the robot could conduct automated university campus tours. Furthermore, we tested our evaluation methodology on two challenging benchmark datasets such as EuRoC and TUM to determine the precision of the state estimator. Fig. 2 presents evaluation methodology of the VINS. One of the goals of our research is to implement camera IMU calibration and improve the time synchronization amongst the cameras and IMU. This helps in real time accurate state estimation and localization. The rest of this section discusses the implementation details of the evaluation methodology to ascertain the accuracy of a SLAM system or any state estimator.

A. Source Data

The source data consists of the ground truth data from the open-source benchmark datasets such as EuRoC and TUM VI along with the estimated trajectory of the camera generated from the state estimator. The evaluation methodology implemented in our research can incorporate open-source methods such as open VINS. The dataset format is in the following format: timestamps, position of the robot in 3D space, and quaternion rotation of the robot.

B. Evaluation Parameters

The evaluation parameters in the implemented methodology comprise of the mean absolute error (MAE), root mean squared error (RMSE), absolute trajectory error (ATE), and mean squared error (MSE). RMSE is a widely used metric used for evaluating the accuracy of a predictive model or the quality of estimation or reconstruction. MAE exhibits lower sensitivity to outliers compared to RMSE and MSE. This is a suitable choice when the data contains extreme values capable of disproportionately influencing other error metrics. ATE is a metric used to assess the precision of predicted trajectory when compared to the ground truth data in SLAM systems. Given a ground truth and a predicted path, the ATE is calculated as an average of the Euclidean measures amongst the corresponding poses along the trajectories.



Fig. 3. Custom stereo inertial sensor configuration using Intel RealSense D455 and MicroStrain IMU.

C. Trajectory Alignment

We use two alignment methods, the Kabsch method and the Horn method. The Kabsch method calculates the optimized rotation matrix that minimizes the root mean squared deviation between a pair of points. It also requires the translation vector to align two sets of points in three-dimensional space. The steps include centering where the sets of points are centered around the centroids and covariance matrix computation where the covariance matrix between the two centered points are calculated. Application of singular value decomposition to the covariance matrix allows for the extraction of the rotation matrix. The determination of the translation vector is based on the centroids of the point sets.

The Horn method extracts the optimal rotation and translation amongst a group of points in three-dimensional space. It relies on the usage of unit quaternions for representations of rotations and minimization of the sum of squared differences between corresponding points. The Horn method does not account for the scale factor because of which the Kabsch method is preferred in certain scenarios. The results are plotted for a better visual representation of the aligned and unaligned data for comparison between the two methods.

D. Camera IMU Calibration

For conducting the experiments in this research, we used a custom stereo-inertial configuration comprising of two Intel RealSense D455 cameras and a MicroStrain IMU as illustrated in Fig. 3. The monocular RGB module of each RealSense camera was used to create a stereo configuration. The open-source toolbox Kalibr [19] was used for calibrating the cameras. Kalibr supports multiple camera calibration, camera-IMU calibration, multi IMU calibration and rolling shutter camera calibration [20]. The purpose of performing camera calibration is to extract the intrinsic and the extrinsic parameters of the camera along with the lens distortions, which are crucial to accurate state estimation. The transformation between each sensor is also determined from the calibration process.

The calibration tool for the camera IMU system establishes the temporal and spatial attributes of a camera system in relation to the IMU, which is intrinsically calibrated. The IMU data and the image must be supplied in the rosbag format. The calibration attributes are measured through a comprehensive batch optimization process, utilizing splines to represent the system's poses.

Calibration of intrinsic attributes of the IMU is essential and the correction measurements are implemented on the raw measurements. The noise density and the bias random walk for the gyroscope and accelerometer of the IMU need to be calculated. This was done using the Allan variance ROS [21] package. The Allan variance serves as a statistical metric employed to describe the stability and noise properties of a time series signal. It is a measure of how the variance of a signal changes as the average time increases. We collected more than 22 hours of IMU data to compute the Allan deviation.

A rosbag was created to collect the direct image streams from the sensors. The calibration target was fixed while the camera IMU configuration was being shifted in front of the calibration target to stimulate every axis of the IMU. During this

process, the calibration target was evenly illuminated, and the camera shutter durations were minimized to prevent motion blur.

E. Open VINS System

The state vector in VINS encompasses the present inertial navigation state, a group of past IMU poses, a group of intrinsic and extrinsic parameters of cameras along with a set of landmarks in the environment. The representation of the environment landmarks simplifies things because it exclusively contains global 3D positions. The open VINS framework accommodates various representations such as inverse MSCKF [22, 23], complete inverse depth [24] as well as anchored 3D positions [25]. The calibration vector encompasses the intrinsic attributes of the camera that includes focal length, optical center, lens distortions as well as camera IMU extrinsic parameters, which is also referred to as the relative orientation or the spatial transformation from the IMU towards each camera. To take into account the synchronization of the camera clocks, a single time offset is incorporated between the camera clock and the IMU into the calibration vector.

$$\mathbf{x}_k = [\mathbf{x}_I^T \ \mathbf{x}_C^T \ \mathbf{x}_M^T \ \mathbf{x}_W^T \ c_{t_1}]^T \quad (1)$$

$$\mathbf{x}_I = [{}^I_k \bar{q}^T \ G_{pI_k}^T \ G_{vI_k}^T \ \mathbf{b}_{\omega_k}^T \ \mathbf{b}_{a_k}^T]^T \quad (2)$$

$$\mathbf{x}_C = [{}^I_k \bar{q}^T \ G_{pI_{k-1}}^T \ \dots \ {}^{I_k-c} \bar{q}^T \ G_{pI_{k-c}}^T]^T \quad (3)$$

$$\mathbf{x}_M = [G_{pI_1}^T \ \dots \ G_{pI_m}^T]^T \quad (4)$$

$$\mathbf{x}_W = [{}^I_c \bar{q}^T \ c_{1pI}^T \ \zeta_0^T \ \dots \ c_w^I \bar{q}^T \ c_{wpI}^T \ \zeta_w^T]^T \quad (5)$$

Equations 1 to 5 represent the state of VINS. It consists of the present state estimation with a set of past IMU poses denoted by c and landmarks denoted by m . ${}^I_k \bar{q}$ determines the quaternion unit that parameterizes the rotation with respect to the global frame G and the local frame of the IMU denoted by I_k where k is time. \mathbf{x}_I denotes the inertial state. \mathbf{b}_{ω} and \mathbf{b}_a represent bias of the gyroscope and accelerometers respectively. \mathbf{x}_C denotes the historic IMU poses and \mathbf{x}_M denotes the landmarks in the environment. \mathbf{x}_W represents the calibration attributes.

The forward propagation of the inertial state involves the utilization of incoming IMU measurements in the form of linear accelerations and angular velocities. This propagation is achieved through a nonlinear general IMU kinematics framework, which advances the state from the previous timestamp to the next. The covariance matrix of the state is generally advanced by linearizing the nonlinear model based on the current estimation. The pose clones, environmental landmarks, and calibration states remain constant over time. Consequently, the associated state Jacobian entries maintain an identity status with no propagation noise. This helps in the utilization of the sparsity for efficient computation.

The fundamental element of this visual inertial navigation system is the indexing method, which is based on types. Drawing inspiration from a popular graph-based optimization framework like GTSAM [26], the users don't need to manipulate the covariance directly but are instead presented with tools, which are capable of handling the state and its covariance autonomously. Adopting such an approach result in a substantial decrease of implementation time and a diminished susceptibility to developmental errors arising from explicit

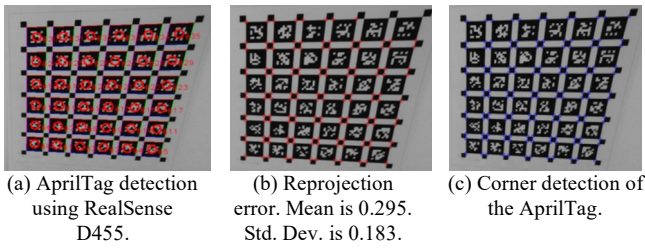


Fig. 4. AprilTag detection and corner detection for the calibration process using the RealSense D455 camera.

access to covariance and state. Each internal state variable type contains details about the position in the error state, automatically adjusting during operations such as initialization, cloning, marginalization, which impact the ordering of variables. A type is characterized by its size of state, position in covariance and prevailing estimations. The present value might not exclusively be a vector and can also be represented as a matrix. The error state remains consistently represented as a vector for all types, which prompts each type to establish the boxplus mapping that connects its error state and representation of manifold, essentially serving as a function for update. A primary benefit of this system is its capability to ease the incorporation of new features while creating sparse Jacobians. Rather than generating a Jacobian encompassing every state element, the sparse Jacobian is utilized to incorporate every state element specific to the function of measurement. This not only conserves computational resources in instances where a measurement is linked to a select few state elements but also permits measurement functions to remain indifferent to the overall state as long as the requisite state variables are present.

The objective is to efficiently calculate the primary estimation of a fresh state variable along with its associations and covariance with state variables that are already in place. To illustrate, the process of initializing a new SLAM landmark is described thereby providing a generalized approach to any new state variable. The initial step involves conducting QR decomposition [27].

The landmark measurement model accommodates diverse feature parameterizations such as 3D positions, inverse depth, and others. The measurement functions pertain to the inherent projection, distortion, and transformation operations and the associated measurement Jacobians can be determined using a straightforward application of the chain rule. The errors are

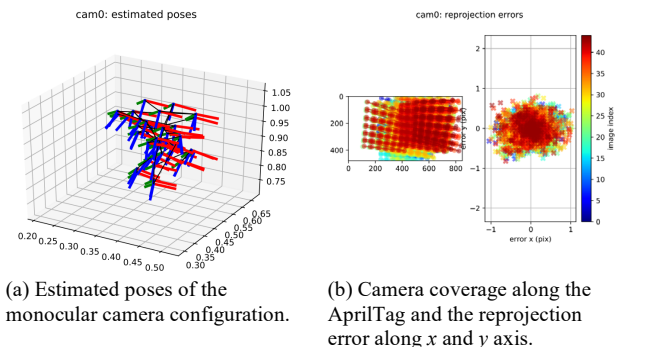


Fig. 5. Estimated camera poses of the monocular camera while capturing calibration data along with the resultant reprojection errors after calibration.

calculated on the original uv pixels enabling the adjustments of camera intrinsics through calibration. The function can also be modified to accommodate other camera models such as equidistant and radial tangential.

V. EXPERIMENT SETUP, RESULTS, AND DISCUSSION

The hardware setup includes Intel RealSense D455 camera, Microstrain IMU and the integrated inertial measurement unit (IMU) of the RealSense camera. The software was set up on Ubuntu 20.04 and the libraries used were Matplotlib, NumPy, SciPy and Math. The calibration of the monocular and the stereo pair of cameras was done using Kalibr, which is an open-source calibration tool developed for the purpose of multi camera calibration, visual-inertial calibration, multi-inertial calibration, and rolling shutter camera calibration. The Allan variance ROS package was employed to process an extensive IMU data sequence, determining the bias instability, gyro random walk, angle random walk for the gyroscope, bias instability, velocity random walk, and accelerometer random walk. This package is compatible with Kalibr and ROS Noetic, which was also used for conducting the experiments.

The depth parameter of the RealSense camera was set to true with a resolution of 840 x 480. The gyroscope and accelerometer frames per second (FPS) were set to 400 and 250. The calibration was done for both monocular and stereo configuration. The custom stereo inertial system was used for the stereo calibration process. The AprilTag used for the calibration process had the following configuration: size of each tag is 0.088 m, number of AprilTags in each row and column is six, space between each tag is 0.3 m.

A. Monocular Calibration

For the monocular calibration, the integrated IMU of the RealSense D455 camera was used, which published the data at 400 FPS and the rosbag color image data was recorded at 30 FPS. The monocular RGB color module was utilized and the pinhole camera model was used for the calibration. Fig. 4 illustrates the calibration of the monocular camera configuration along with the reprojection error. The figure illustrates the April Tag detection along with the corner detection, which helps in

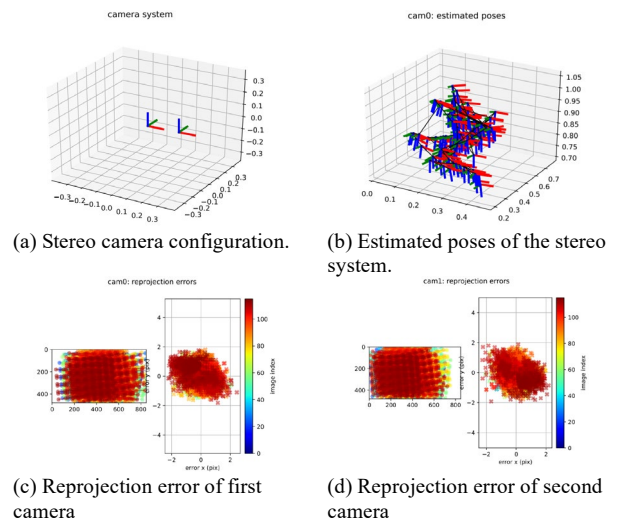


Fig. 6. Stereo camera system reprojection error and estimated poses.

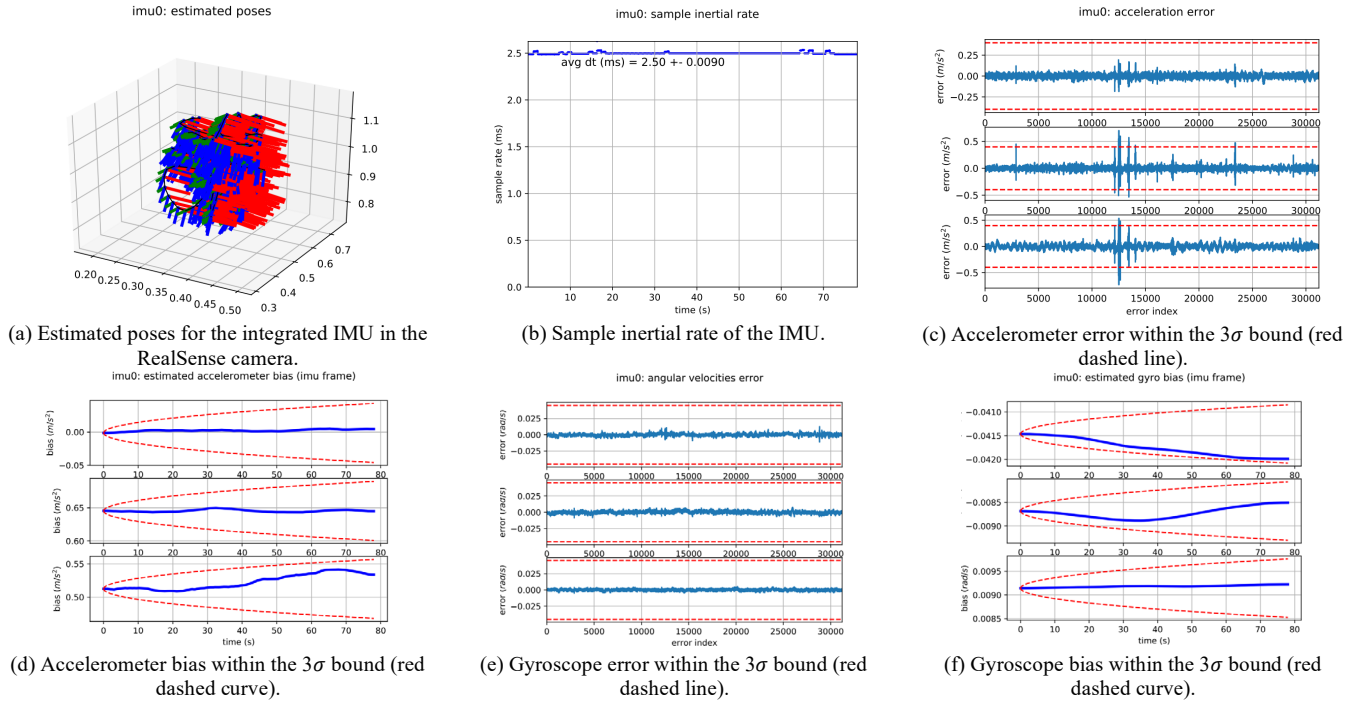


Fig. 7. IMU estimated poses along with the sample inertial rate of the IMU demonstrating steady flow of data with the corresponding accelerometer and gyroscope errors and biases.

identifying the intrinsic and extrinsic attributes of the camera along with the coefficients of distortion. However, since the pinhole camera is used for the experiments, distortion coefficients are almost equivalent to zero.

Fig. 5 (a) shows the estimated poses of the monocular camera configuration and the reprojection error along the x and y axes is represented in (b). It also represents the coverage area of the camera while collecting the calibration data. Once the monocular camera is calibrated separately, the IMU needs to be calibrated as well. A 22 hour 59 seconds long static IMU data was collected to estimate the transformation between the camera and the integrated IMU of the RealSense camera. The gyroscope and the accelerometer bias of the IMU was calculated using the Allan variance ROS package. Fig. 7 (a) represents the estimated

poses of the IMU while collecting the rosbag data. The same rosbag data was used for both the camera and camera IMU calibration. Fig. 7 (b) represents the sample inertial rate of the IMU, which is a measure of the angular velocity of the IMU at a specific sampling rate. During calibration, the IMU is often subjected to static positions and the inertial rate measurements are collected at regular intervals. These measurements help in characterizing the performance by compensating for the errors and biases. The accelerometer and gyroscope errors and biases are represented in (c), (d), (e), and (f) respectively. The fact that all of these are within the 3σ bound, which is marked with the red-dashed line suggest accurate calibration results. Fig 8 (a) and (b) represents the Allan standard deviation of the accelerometer and gyroscope respectively.

B. Stereo Calibration

To calibrate the stereo pair of cameras, the same intel RealSense D455 cameras were used but an external MicroStrain IMU was used for the visual inertial system calibration. The

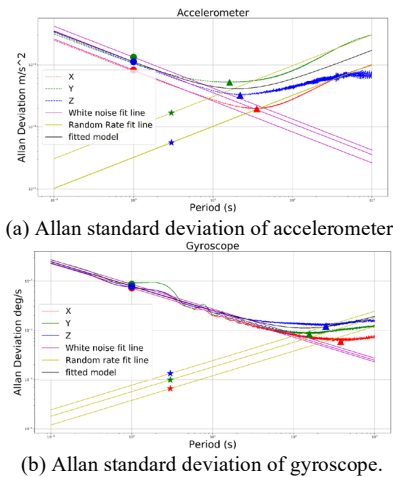


Fig. 8. Allan standard deviation of accelerometer and gyroscope with manually identified noise processes.

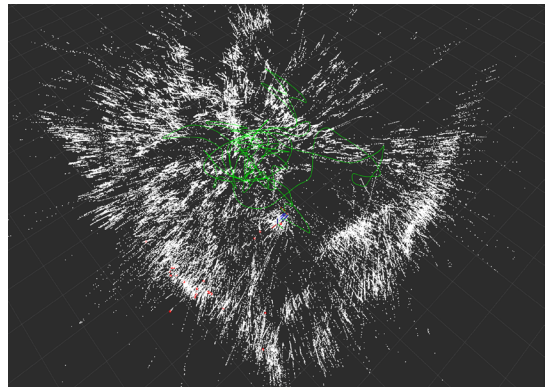


Fig. 9. Sparse map generation of the EuRoC V103 dataset.

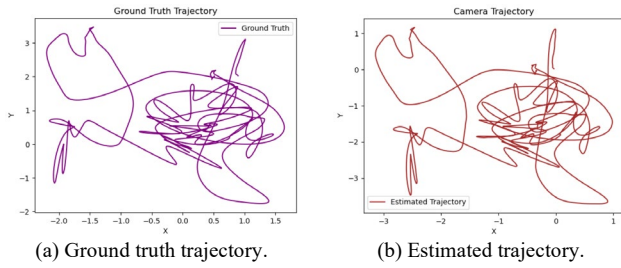


Fig. 10. Ground truth and estimated trajectory of EuRoC V103 dataset.

same calibration process was followed as done for the monocular calibration, but both cameras had to be calibrated separately. One camera was considered to be the global camera coordinate frame and the transformation of the other camera was determined with respect to the first camera. The IMU transformations were determined with respect to both cameras to determine the exact orientation and position of the IMU in the sensor configuration. This is very important for accurate state estimation.

The IMU data was published at 100 Hz whereas the stereo camera pair published the color image data at 30 frames per second (FPS). The scale misalignment model was used for calibrating the IMU for both the stereo and monocular sensor configurations. Fig. 6 (a) shows the stereo camera sensor configuration along with the estimated poses of the stereo system in (b), while the reprojection errors of the first and second cameras are represented in (c) and (d) respectively.

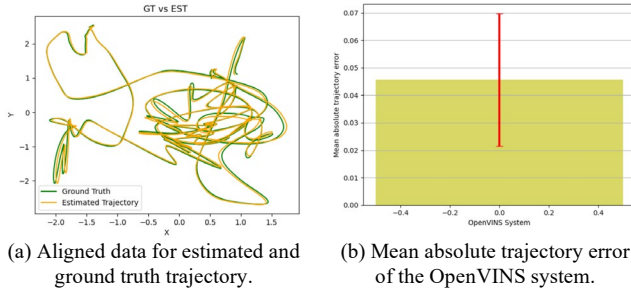


Fig. 11. Aligned trajectory and the mean absolute trajectory error of the EuRoC V103 dataset.

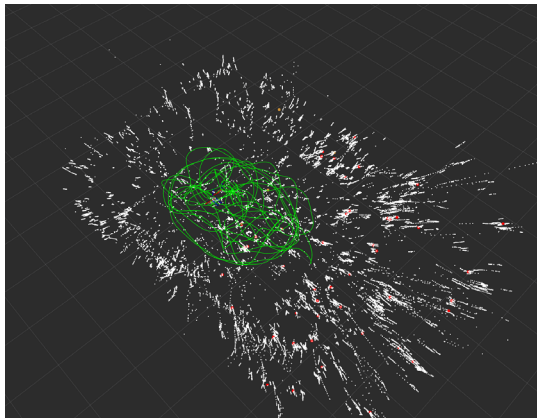


Fig. 12. Sparse map generation of the TUM VI room 1 dataset.

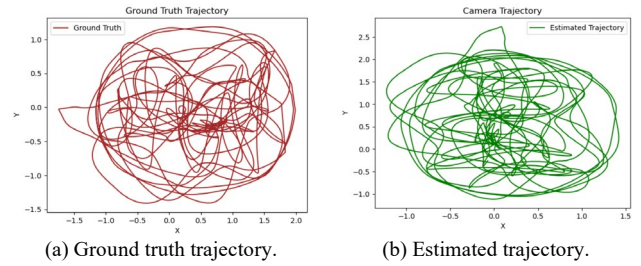


Fig. 13. Ground truth and estimated trajectory of TUM VI room 1 dataset.

TABLE 1. COMPARISON OF TRAJECTORY ALIGNMENT ERROR BETWEEN KABSCH AND HORN METHODS ON EUROC AND TUM VI DATASETS.

	Kabsch (RMSE)	Horn (RMSE)
EuRoC	0.052	0.051
TUM VI	0.057	0.009

C. Evaluation on EuRoC V103 Dataset

The visual inertial navigation system pipeline was evaluated on the V103 dataset from EuRoC. A similar pinhole radtan model was used for the calibration of the stereo camera system. A sparse map was also generated along with the state estimation, but the map is not stored. The sparse map was a result of delaying the feature decay value to an extended period of time. Fig. 9 shows the sparse map generated from the V103 dataset from EuRoC. Once the entire rosbag file was executed we acquired the corresponding timestamp values with the position and orientation of the robot in three-dimensional space. Leveraging that data, it was possible to generate the estimated trajectory and evaluate the results with respect to the ground truth trajectory. Fig. 10 shows the ground truth trajectory and the estimated trajectory generated from the open VINS system.

The estimated trajectory and the ground truth trajectory are then subsequently aligned using the Kabsch method and the root mean squared error and the mean absolute trajectory error is calculated. The RMS error is calculated to be approximately 0.052. Fig. 11 (a) shows the aligned trajectory of the ground truth and estimated trajectory and the mean absolute trajectory error plot in (b).

D. Evaluation on TUM VI Room 1 Dataset

The visual inertial navigation system was also evaluated on the TUM VI room 1 dataset where a similar set of experiments were conducted. A sparse map was generated with a 45 second decay time for the features. Fig. 12 illustrates the sparse map generated by the open VINS system. The sparse map is the projection of features on the three-dimensional plane, but the map is not stored in this scenario. The features will disappear

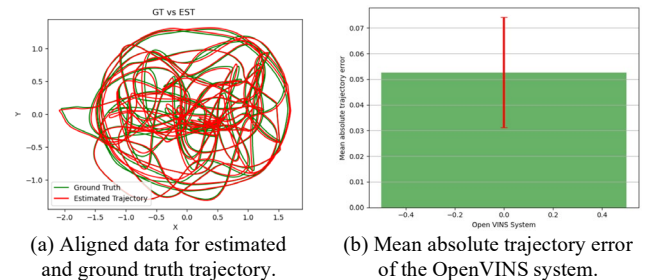


Fig. 14. Aligned trajectory and the mean absolute trajectory error of the TUM VI room 1 dataset.

with time increments. After the entire rosbag data was executed successfully, the estimated trajectory was obtained. Fig. 13 shows the plot of the estimated trajectory and the ground truth trajectory. The estimated trajectory and the ground truth trajectory were aligned using the Kabsch method, which is represented in Fig. 14 (a) along with the mean absolute trajectory error. The RMS error was evaluated to be 0.057 and the mean absolute trajectory error plot is also represented in (b).

Table I shows the comparison between the Kabsch method and the Horn method used for the evaluation of the accuracy of the algorithm. For the EuRoC dataset, the Kabsch and the Horn method showed similar performance but for the TUM VI dataset, the Horn method performed better than the Kabsch method. Both methods have achieved highly accurate results as is evident from Fig. 11 (a) and Fig. 14 (a). The ground truth and the estimated trajectory are well aligned thus representing the accuracy of the algorithm.

VI. CONCLUSION AND FUTURE WORK

The real time camera and IMU calibration was performed and evaluation methodology was used for the alignment of the estimated and ground truth trajectories. The evaluation methodology was created to account for the RMS, rotational error, timebound error, and mean absolute trajectory error. This methodology provides an extensive evaluation toolkit for any SLAM system provided the source data format is preserved. Furthermore, two challenging benchmark datasets, EuRoC and TUM VI were evaluated. The alignment of the data proved to be very accurate. This state estimation process was used for both indoor and outdoor robot navigation. The custom sensor configuration shown in Fig. 3 can be mounted on any robot for accurate state estimation. The open VINS system can be further extended to generate full scale dense maps of the environment to make it a complete SLAM system.

REFERENCES

- [1] M. Bloesch, M. Burri, S. Omari, M. Hutter and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," in *Int. J. of Robotics Research* 36, no. 10 (2017): 1053-1072.
- [2] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," in *the Int. J. of Robotics Research* 34, no. 3 (2015): 314-334.
- [3] T. Qin, J. Pan, S. Cao and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," in *arXiv preprint arXiv:1901.03638* (2019).
- [4] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4666-4672. IEEE, 2020.
- [5] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," in *IEEE Trans. on Robotics* 37, no. 6 (2021): 1874-1890.
- [6] J. Engel, T. Schöps and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *European Conf. on Computer Vision*, pp. 834-849. Cham: Springer International Publishing, 2014.
- [7] C. Forster, M. Pizzoli and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *2014 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 15-22. IEEE, 2014.
- [8] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems*, vol. 2, no. 9, pp. 1-9. 2014.
- [9] C. Kerl, J. Sturm and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2100-2106. IEEE, 2013.
- [10] H. Rebecq, T. Horstschäfer, G. Gallego and D. Scaramuzza, "Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time," in *IEEE Robotics and Automation Letters* 2, no. 2 (2016): 593-600.
- [11] M. Bloesch, S. Omari, M. Hutter and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 298-304. IEEE, 2015.
- [12] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to mav navigation," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 3923-3929. IEEE, 2013.
- [13] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 3565-3572. IEEE, 2007.
- [14] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," in *the Int. J. of Robotics Research* 32, no. 6 (2013): 690-711.
- [15] G. P. Huang, A. I. Mourikis and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," in *the Int. J. of Robotics Research* 29, no. 5 (2010): 502-528.
- [16] S. Shen, Y. Mulgaonkar, N. Michael and V. Kumar., "Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV," in *2014 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4974-4981. IEEE, 2014.
- [17] M. J. N. P. G. Burri, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik and R. Siegwart, "The EuRoC micro aerial vehicle datasets," in *Int. J. of Robotics Research* 35, no. 10 (2016): 1157-1163.
- [18] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler and D. Cremers., "The TUM VI benchmark for evaluating visual-inertial odometry," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1680-1687. IEEE, 2018.
- [19] "Kalibr," in <https://github.com/ethz-asl/kalibr/wiki/installation>.
- [20] L. Oth, P. Furgale, L. Kneip and R. Siegwart, "Rolling shutter camera calibration," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1360-1367. 2013.
- [21] "Allan Variance ROS," in <https://github.com/ethz-asl/kalibr/wiki/IMU-Noise-Model>.
- [22] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. 2007 IEEE Int. Conf. on Robotics and Automation*, pp. 3565-3572. IEEE, 2007.
- [23] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," in *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep 2* (2005): 2005.
- [24] J. Civera, A. J. Davison and J. M. Montiel, "Inverse depth parametrization for monocular SLAM," in *IEEE Trans. on Robotics* 24, no. 5 (2008): 932-945.
- [25] M. K. Paul, K. Wu, J. A. Hesch, E. D. Nerurkar and S. I. Roumeliotis, "A comparative analysis of tightly-coupled monocular, binocular, and stereo VINS," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 165-172. IEEE, 2017.
- [26] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," in *Georgia Institute of Technology, Tech. Rep 2* (2012): 4.
- [27] M. Li, "Visual-inertial odometry on resource-constrained systems," in *Dissertation. University of California, Riverside, 2014. Retrieved from https://escholarship.org/uc/item/4nn0j264*.