Real-time 3D Mapping using a 2D Laser Scanner and IMU-aided Visual SLAM

Shaowu Yang
National University of Defense Technology

Real-time 3D Mapping using a 2D Laser Scanner and IMU-aided Visual SLAM

Mengxiao Chen, Shaowu Yang, Xiaodong Yi and Dan Wu

Abstract— **In this paper, we present a solution to 3D mapping using a 2D laser scanner with pose estimates from an IMU-aided visual SLAM system. Accurate motion estimation of a robot is achieved by visual-inertial fusion based on an extended Kalman filter (EKF). Range measurements scanned on the vertical plane are received constantly by a 2D laser scanner mounted on the robot, which are re-organized as point clouds in Cartesian space. With the pose estimates of the robot, the point clouds can be transformed into the world frame in real time. Furthermore, these point clouds received between two consecutive keyframes of the visual SLAM system are accumulated together into a unit relative to a keyframe, which can be corrected later by loop closing in visual SLAM. The 3D globally consistent map is built simultaneously by gathering these units of point clouds. The proposed approach and its performance are demonstrated and evaluated by our indoor experiments using a Turtlebot mounted with a Kinect camera, an IMU and a 2D laser scanner.**

I. INTRODUCTION

Real-time and accurate three-dimensional (3D) mapping is of great importance to autonomous navigation of mobile robots, especially in environment without prior knowledge. Laser scanners which can provide accurate range measurements are widely used in environmental mapping. However, 3D laser scanners are with high prices, and would require extensive time cost for environmental mapping. Thus, some methods have been proposed to achieve 3D mapping with 2D scanners such as combining multiple laser scanners [1], [2], rotating a scanner on a moving robot [3], [4], fusing information from other sensors [5], [6], etc. When a laser scanner is moving, creating a 3D map requires accurate pose estimation for the scans to register the points into a fixed coordinate system. Otherwise, the point clouds would be severely distorted [4]. The Iterative Closest Point (ICP) algorithm [7] and its derivatives are widely used in registration of point clouds. However, it is computationally expensive due to iterative calculations. How to obtain robust and accurate pose estimation and build map in real time is still an open issue in 3D mapping using 2D laser scans.

The work presented in this paper attempts to achieve real-time 3D mapping with a 2D laser scanner, the poses are estimated based on IMU-aided visual SLAM. A 2D laser scanner mounted vertically on a robot is used to gather scans constantly. Measurements from visual SLAM and inertial sensors have been fused based on Extended Kalman Filter (EKF) to provide robust and accurate 6 degree-of-freedom (DOF) pose estimation. With the pose estimation, the laser scans can be registered into a fixed coordinate system to incrementally generate a consistent 3D map. Furthermore, the point clouds received between two consecutive keyframes are accumulated together into a unit, which can be corrected later by loop closing in visual SLAM.

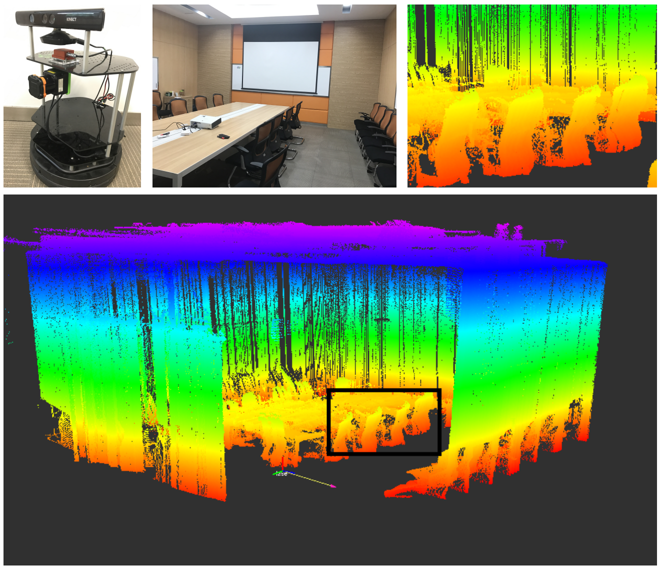


Fig. 1: 3D Maping using a 2D laser scanner with state estimation based on visual-inertial fusion. Top left: The Turtlebot robot mounted with a Kinect, an IMU and a upward-pointed laser scanner. Top middle: The indoor environment of our experiments. Top right: A part of point clouds in the bottom map. Bottom: The full point cloud of the resulting 3D map, colored according to height.

II. RELATED WORK

Laser scanning sensors, which can provide high frequency and accurate range measurements, are widely used in on-board environment sensing and navigation of mobile robots, including 3D laser scanners [8], [9]. However, scans can be easily distorted when a 3D scanner is in fast motion. Furthermore, 3D scanners are not suitable for mobile robots in application for high price and time cost.

Recently, many researchers have attempted to construct maps with 2D scanners, considering that they are light-weighted, relatively inexpensive and can operate in a wide range of environmental conditions to provide 2D range data [6]. To gather 3D scans, a typical method is mounting more

M. Chen, Dr. S. Yang, Asso.Prof. Dr. X. Yi, and Asso.Prof. Dr. D. Wu are with the State Key Laboratory of High Performance Computing (HPCL) and College of Computer, National University of Defense Technology, Changsha, China {chenmengxiao11, shaowu.yang, yixiaodong, daisydanwu}@nudt.edu.cn
Project 91648204 supported by NSFC, project ZDYYJCYJ20140601 supported by NUDT, and project 201602-01 supported by HPCL.

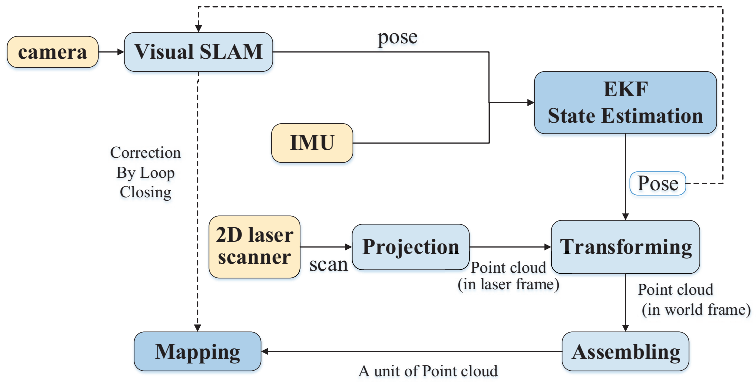


Fig. 2: Overview of our approach: real-time 3D mapping using 2D laser scans with pose estimation based on sensor fusion

scanners on a moving platform. Two 2D scanners are used in [1], while the horizontally mounted one builds 2D map to recover pose and the upward pointed one grabs vertical scan lines which are transformed into point clouds. Since this method require very precise motion estimation and can not be applied to the full three-dimensional case, researchers attempt to use a rotating 2D scanner driven by a motor in [3], [4]. Continuous rotation enables the 2D scanner to have large fields of view, resulting in relatively complex mechanical configuration, however.

Fusing information from other sensors can also help 2D laser scanners in 3D mapping. In [5], visual odometry is implemented to estimate the ego-motion and register point clouds from a lidar, while scan matching based lidar odometry is used to refine the estimation and registration. The device called *Zebedee* proposed in [6] consists of a 2D laser scanner and a rigidly connected inertial measurement unit (IMU), both of which are mounted on a spring fixed to a moving platform. Carefully modeling the mechanical property of the spring is required for this passively actuated laser scanner.

As range measurements are received at different timestamps, accurate 6DOF poses should be estimated to register the points into a fixed coordinate system. Otherwise, it would involve distortion in the resulting point cloud. Thus, efficient solutions are in great need to estimate the pose precisely and construct the environmental map accurately.

Recently, integrating multi-modal sensors for 6DOF state estimation and omnidirectional environment perception has received much attention. A multi-modal perception system consisting of vision, laser scanning, inertial sensing and GPS for the exploration and mapping of a river is presented in [10]. A dual 3D laser scanner, three stereo camera pairs, an IMU and an onboard computer are integrated to achieve state estimation and mapping in [11]. In [12], with the inertial guidance system which computes the pose from differential GPS systems and an IMU, the data from a single line scanner is integrated into point clouds.

Related work listed above has shown that the performance of using a 2D laser scanner to construct 3D maps accurately in real time can be improved with sensor fusion. As for sensor fusion, loosely coupled methods based on extended Kalman filter (EKF) from visual and inertial measurements have been shown in [13], [14], [15] to achieve state estimation in real time. Similar method is used to fuse the measurements from camera and IMU in this work. With the pose estimate by sensor fusion, range measurements in the vertical plane re-organized as point clouds are transformed into the world frame in real time. Specifically, with the recorded relation between these units of point cloud and keyframes, point clouds can be optimized according to the correction of keyframes in loop closing of visual SLAM.

III. 3D MAPPING USING 2D LASER SCANS

A. *Overview of the approach*

In this work, we achieve real-time 3D mapping using a 2D laser scanner mounted on a robot with the assistance of state estimation by sensor fusion, as shown in Fig. 2. The laser scanner mounted on the mobile platform is upward pointed to rapidly acquire range data comprising the distance and angular values measured within a vertical 2D plane. The method of pose estimation is based on online extended Kalman filter (EKF), while the state is predicted by IMU measurements and updated with 6DOF pose estimates from a visual SLAM system. With the accurate and high frequency pose estimation from EKF, point clouds corresponding to the current laser scan can be registered into a fixed coordinate system. Since the method of visual SLAM we use is keyframe-based, these point clouds obtained between two consecutive keyframes can be assembled into a unit. Such units obtained constantly are combined to incrementally build a consistent global 3D map.

B. *State estimation based on sensor fusion*

In order to achieve state estimation robustly and accurately, we have implemented an IMU-aided visual SLAM system based on extended Kalman filter (EKF). Fig. 3 shows the situation of coordinate frames in our system. Specifically, the state prediction is driven by an IMU, while the camera position measurements obtained from the visual SLAM system are used for the filter updates.

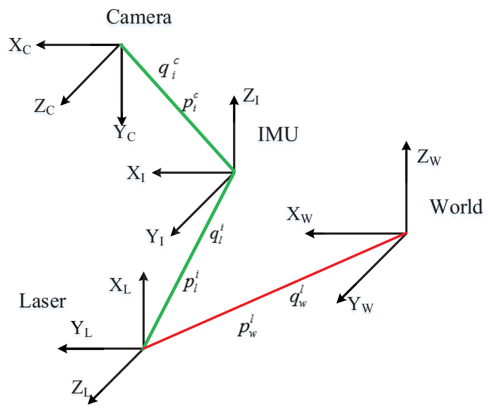


Fig. 3: The different coordinate frames. The IMU-camera transformation and IMU-laser transformation colored in green are regarded as constant values. Red line indicates the pose of the laser scanner in the world frame.

The IMU gyroscopes and accelerometers provide measurements of the rotational velocity ω_m and acceleration a_m with a certain bias b and zero-mean white Gaussian noise n. The real angular velocities ω and accelerations a in the IMU frame are as follows:

$$\omega = \omega_m - b_\omega - n_\omega, \qquad a = a_m - b_a - n_a. \tag{1}$$

The state can be represented as:

$$X = \{p_w^{i\,T}\ v_w^{i\,T}\ q_w^{i\,T}\ b_\omega^T\ b_a^T\ \lambda\ p_i^c\ q_i^c\}. \tag{2}$$

The state consists of the position of IMU in the world frame (W) p_w^i, the attitude quaternion q_w^i (IMU frame w.r.t. the world frame, expressed in the world frame), the velocity v_w^i, the gyroscope and acceleration biases b_ω and b_a, the visual scale factor λ, the rotation from the IMU frame into the camera frame q_i^c and the position of the camera center in the IMU frame p_i^c, while q_i^c and p_i^c are regarded as constant values and have been calibrated already.

Following the method presented in [16], we propagate the state variables according to their differential equations. With the differential equations for the continuous time error state and the continuous time system noise covariance matrix, the discretized error state propagation and error process noise covariance matrices can be calculated. And the propagated state covariance matrix can be computed according to the filter equation in [16].

As for visual SLAM, we use the popular keyframe-based ORB-SLAM2 framework [17] with a RGB-D camera. By using the ORB features [18] for tracking and mapping tasks and inserting keyframes frequently, it can provide pose estimation, which can be further optimized with loop closing and full bundle adjustment.

With the position measurements p_w^c from visual SLAM, the measurement model is represented as:

$$z_p = p_w^c = (p_w^i + C_{(q_w^i)}^T p_i^c)\lambda + n_p. \tag{3}$$

with $C_{(q_w^i)}$ as the rotation matrix corresponding to the the attitude of IMU and n_p as zero-mean, white and Gaussian noise of the visual measurement.

The position error is defined as:

$$\tilde{z}_p = z_p - \hat{z}_p, \tag{4}$$

which can be linearized to

$$\tilde{z}_{pli} = H_p \tilde{x}. \tag{5}$$

For the rotation measurement, we define the error as:

$$\tilde{z}_q = z_q - \hat{z}_q, \tag{6}$$

Then the measurements can be stacked together as:

$$\tilde{\mathbf{z}} = \mathbf{H}\tilde{\mathbf{x}}. \tag{7}$$

With the measurement matrix $\mathbf{H}$, the following calculations can be carried out to update the estimates according to the well known Kalman filter procedure:

1) residual: $\tilde{z} = z - \hat{z}$
2) innovation: $S = HPH^T + R$
3) Kalman gain: $K = PH^T S^{-1}$
4) correction: $\hat{\tilde{x}} = K\tilde{z}$

With the correction $\hat{\tilde{x}}$, the updated state variables can be computed. Following the equations in [16], the error quaternion is calculated and the error state covariance can be updated.

Based on the EKF algorithm above, the IMU-aided visual SLAM is capable of getting robust, accurate and high-frequency pose estimation of the inertial sensor. With the calibrated transformation between the IMU frame and the laser-scanner frame, we can obtain precise pose estimation while the scanner receiving range measurements. Specifically, the rotation and translation of the world frame to the laser frame expressed in the world frame are denoted as q_k and x_k, while the orientation q_k can be expressed as a rotation matrix R_k with the time index k.

C. 3D Mapping

As shown in Fig. 4, with the robust and accurate pose estimation based on sensor fusion, the point clouds corresponding to the 2D range measurements can be transformed into a fixed coordinate system in real time. These point clouds which fall between two consecutive keyframes are accumulated together into a unit of point cloud. A global consistent map is built incrementally by adding these units together.

The 2D laser scanner (HOKUYO UTM-30LX) mounted on a Turtlebot is upward pointed to obtain range measurements in the vertical plane. It has an viewing angle of 270° with a resolution of 0.25°.

As the range measurements are in polar coordinates of the sensor plane, straightforward projection into Cartesian space is taken firstly. Each ray in a scan is simply projected according to the appropriate angle:

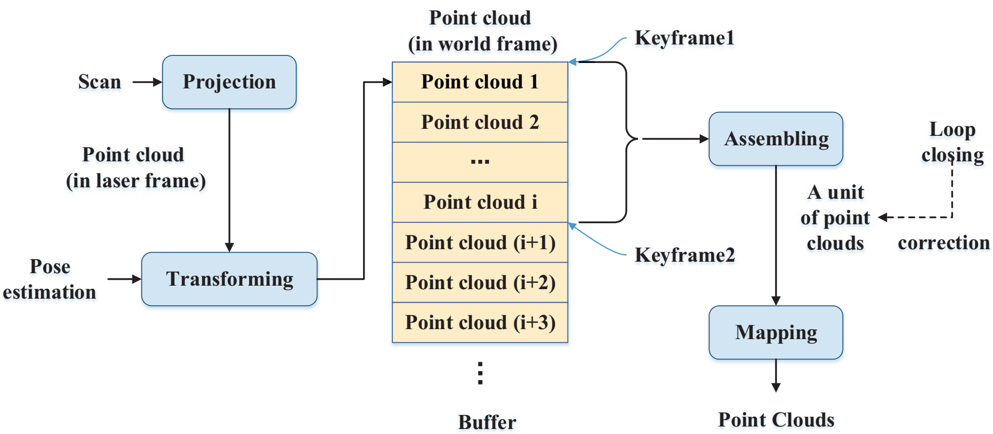


Fig. 4: Procedure of the 3D mapping process: Projecting the 2D range measurements into a point cloud in the laser frame in Cartesian space; Transforming the point cloud into the world frame with the pose estimation; Assembling these point clouds between two consecutive keyframes into a unit of point cloud which may be corrected later by loop closing; Adding these units of point clouds together to build a global consistent map.

$$\begin{pmatrix} x_k^i \\ y_k^i \\ z_k^i \end{pmatrix} = r_k^i \begin{pmatrix} \cos \alpha_k^i \\ \sin \alpha_k^i \\ 0 \end{pmatrix}, \tag{8}$$

with α_k^i being the angle of the i-th element in the range vector of the scan and the r_k^i being the measured range value.

Since these scan measurements are received at different time-stamps during continuous motion, in order to build a global consistent map, the primary point clouds obtained in the scan frame need to be transformed into the world frame W. The pose esitimation of the scanner, represented by a translation x_k and a rotation R_k, are used as follows:

$$\begin{pmatrix} X_k^i \\ Y_k^i \\ Z_k^i \end{pmatrix} = R_k \begin{pmatrix} x_k^i \\ y_k^i \\ z_k^i \end{pmatrix} + x_k. \tag{9}$$

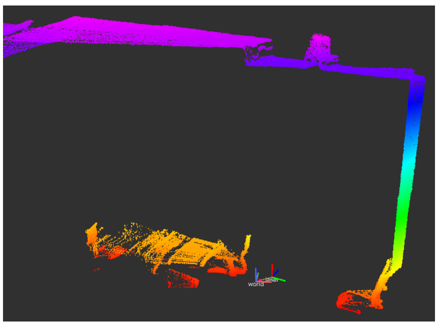


Fig. 5: A unit of primary point clouds which is assembled using these point clouds fall between two consecutive keyframes.

With the transformation above, the single scan received has been projected and registered into the world frame W in real time. Considering that keyframes are frequently inserted, these point clouds stored in a buffer that fall between two consecutive keyframes are accumulated into a unit as shown in Fig. 5. As a result, a unit of point cloud and a corresponding keyframe are interrelated. The whole 3D point cloud map will be generated by adding such units incrementally.

An extra measure has been taken to improve the accuracy of the map. During the process of visual SLAM, once a loop has been detected, the loop would be corrected by pose-graph optimization and the pose of a certain keyframe would be corrected at the same time. With the transformation T_{KF_k} between the poses of this keyframe obtained before and after the loop closing, the pose of the unit of point clouds X_w related to this keyframe can be corrected to $X_{w(cor)}$ as

$$X_{w(cor)} = T_{KF_k} X_w. \tag{10}$$

IV. EXPERIMENTS AND RESULTS

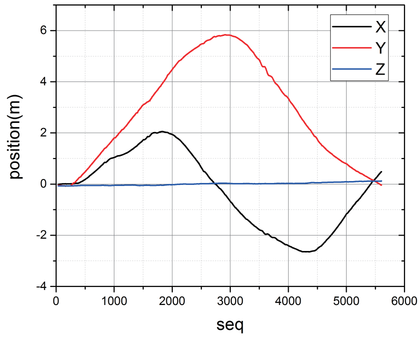
A. *Experiment setup*

In the experiments, we use a Turtlebot platform mounted with a Kinect sensor, an inertial sensor IMU (Xsens MTi-30) and a upward-pointed Hokuyo UTM-30LX laser scanner. The transformations among sensors are regarded as constant values which are calibrated offline. The proposed 3D mapping system runs on a laptop equipped with an Intel i5 core 1.70 GHz CPU and a 4 GB RAM. It is implemented in the Robot Operating System (ROS) [19] in Ubuntu operating system.

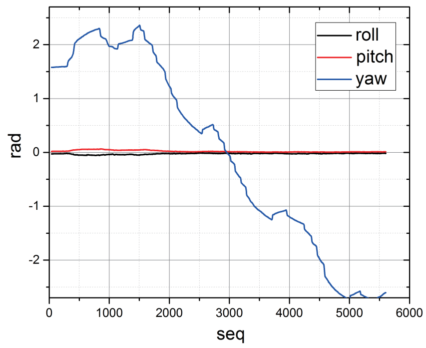
B. *Localization and mapping results*

As shown in Fig. 1, the Turtlebot was driven in an office and a logfile of sensor data from Kinect, IMU and the laser scanner was taken as a ROS bag file [1]. Then we replayed the bag at full speed to evaluate our proposed 3D mapping system. The pose estimates of the IMU sensor based on visual-inertial fusion are shown in Fig. 6, with which we can obtain the pose estimation of the laser scanner. Since the Turtlebot is moving on the floor, the pose estimates on the z-axis mainly remain to be constant values, and only rotation in yaw angle changes obviously. Primary point clouds in the

[1]http://wiki.ros.org/rosbag



(a)



(b)

Fig. 6: The pose estimation of the IMU based on sensor fusion. (a) The position estimation of the IMU frame w.r.t to the world frame. (b) The orientation estimation of the IMU frame w.r.t to the world frame.

scanner frame are transformed into the world frame using pose estimation. And these point clouds which fall between two consecutive keyframes are accumulated into a unit of point cloud as shown in Fig. 5. The global 3D map as shown in Fig. 1 can be built incrementally by joining such units together, which can provide rich environmental information in our experiment. Moreover, once there is a loop detected in visual SLAM, the correlative units of point cloud could be optimized by the correction of keyframes.

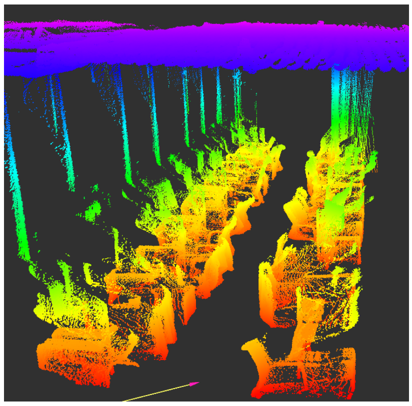
In addition, the 3D maps of other two rooms in our laboratory have been constructed in similar experiments as shown in Fig. 7.

C. Evaluation of the performance

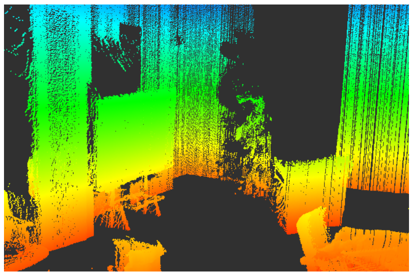
Firstly, we can evaluate the performance of pose estimation based on sensor fusion by comparing with the visual SLAM in the experiment as shown in Fig. 1. The full trajectory of the Turtlebot on the x - y plane in the experiment is shown in Fig. 8.

TABLE I: TIME COSTS

Modules	Time costs(μs)
Projecting into Cartesian space	128
Transforming to the world frame	19
Assembling to a unit of point clouds	951



(a)



(b)

Fig. 7: The 3D maps built in other experiments. (a) The map of an office in our laboratory. (b) A part of the map of a meeting room.

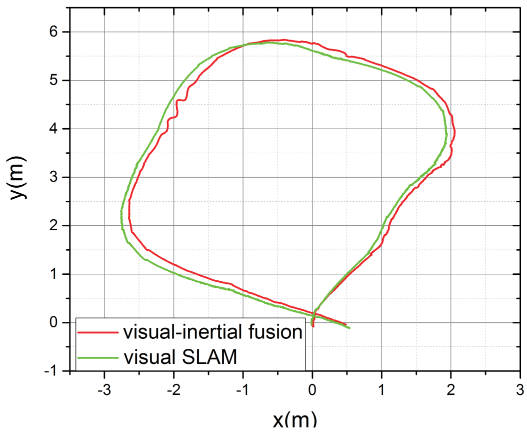
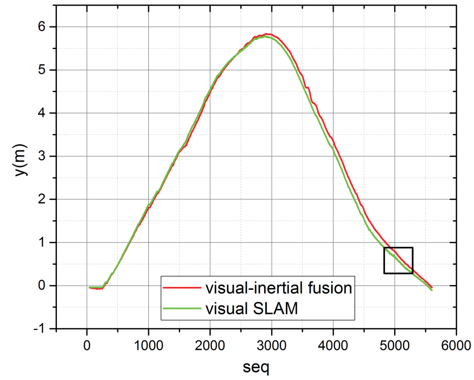


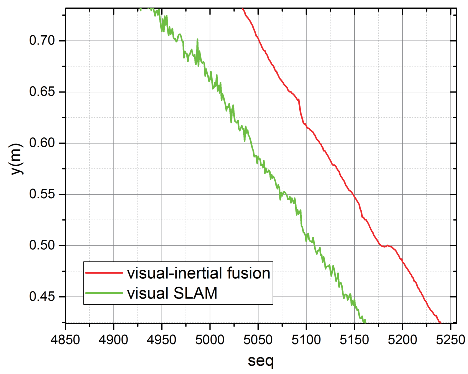
Fig. 8: The trajectory on the x - y plane of the Turtlebot in the experiment.

A more detailed comparison of pose estimation on the y-axis is shown in Fig. 9a, a part of which is enlarged and shown in Fig. 9b. We can easily find that there are much more vibrations in the estimation of visual SLAM, mainly because of image noises and limited number of good features being tracked for estimation. The pose estimates plotted in red show that there are less chances of such vibrations occur in the visual-inertial system. With the EKF-based fusion, the pose estimation can be more reliable and more accurate than using visual SLAM along.

With the pose estimation based on sensor fusion, the global 3D map can be built in real time after the range measure-



(a)



(b)

Fig. 9: The performance of the pose estimation. (a) Comparison of pose estimates on the y-axis between visual-inertial SLAM and visual SLAM. (b) A segment of estimation corresponding to the part of estimation marked in black rectangle in (a).

ments being projected into Cartesian space, transformed to the world frame and assembled into units of point cloud, as described in Sect. III-C. Table I shows the average time costs of these main modules per execution in our program.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, we propose using a 2D laser scanner, with the pose estimation from visual-inertial fusion, to achieve real-time 3D mapping. IMU-aided visual SLAM based on EKF is implemented to provide the required pose estimation. 2D laser scans during motion of a robot are accumulated to build a global consistent 3D map in real time. We demonstrate the theory presented in this work in the experiments by using a Turtlebot mounted with a Kinect, an IMU and a laser scanner to build 3D point cloud maps of several offices.

In the future work, we are interested in utilizing the laser scans in the SLAM optimization process, to achieve a complete multi-modal SLAM framework which fusing vision, laser, and inertial data.

REFERENCES

[1] Dirk Hahnel, Wolfram Burgard, and Sebastian Thrun. Learning compact 3d models of indoor and outdoor environments with a mobile robot. *Robotics Autonomous Systems*, 44(1):15–27, 2003.

[2] Huijing Zhao and Ryosuke Shibasaki. *Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras*. Springer Berlin Heidelberg, 2001.

[3] J. Zhang and S. Singh. Loam : Lidar odometry and mapping in real-time. 2014.

[4] Michael Bosse and Robert Zlot. Continuous 3d scan-matching with a spinning 2d laser. In *IEEE International Conference on Robotics and Automation*, pages 4244–4251, 2009.

[5] Ji Zhang and Sanjiv Singh. Visual-lidar odometry and mapping: low-drift, robust, and fast. In *IEEE International Conference on Robotics and Automation*, pages 2174–2181, 2015.

[6] Michael Bosse, Robert Zlot, and P Flick. Zebedee: Design of a spring-mounted 3-d range sensor with application to mobile mapping. *IEEE Transactions on Robotics*, 28(5):1104–1119, 2012.

[7] P. J Besl and Neil D Mckay. Method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 14(2):239–256, 1992.

[8] P. Allen, I. Stamos, A. Gueorguiev, E. Gold, and P. Blaer. Avenue: Automated site modeling in urban environments. In *International Conference on 3-D Digital Imaging and Modeling, 2001. Proceedings*, pages 357–364, 2001.

[9] Martial Hebert, Matthew Deans, Daniel Huber, Bart Nabbe, and Nicolas Vandapel. Progress in 3-d mapping and localization. *Surface Science*, 492(492):225–234, 2001.

[10] Sebastian Scherer, Joern Rehder, Supreeth Achar, Hugh Cover, Andrew Chambers, Stephen Nuske, and Sanjiv Singh. River mapping from a flying robot: state estimation, river detection, and obstacle mapping. *Autonomous Robots*, 33(1):189–214, 2012.

[11] M Beul, N Krombach, Yongfeng Zhong, and D Droeschel. A high-performance mav for autonomous navigation in complex 3d environments. In *International Conference on Unmanned Aircraft Systems*, pages 1241–1250, 2015.

[12] Sebastian Thrun, Michael Montemerlo, and Andrei Aron. Probabilistic terrain analysis for high-speed desert driving. In *Robotics: Science and Systems*, 2006.

[13] Anastasios I. Mourikis and Stergios I. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *IEEE International Conference on Robotics and Automation*, pages 3565–3572, 2007.

[14] Mingyang Li and Anastasios I. Mourikis. 3-d motion estimation and online temporal calibration for camera-imu systems. In *IEEE International Conference on Robotics and Automation*, pages 5709–5716, 2013.

[15] Simon Lynen, Markus W. Achtelik, Stephan Weiss, and Margarita Chli. A robust and modular multi-sensor fusion approach applied to mav navigation. In *Ieee/rsj International Conference on Intelligent Robots and Systems*, pages 3923–3929, 2013.

[16] S Weiss and R Siegwart. Real-time metric state estimation for modular vision-inertial systems. In *IEEE International Conference on Robotics and Automation*, pages 4531–4537, 2011.

[17] Raul Mur-Artal and Juan D Tardos. Orb-slam2: an open-source slam system for monocular, stereo and rgb-d cameras. *arXiv preprint arXiv:1610.06475*, 2016.

[18] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *IEEE International Conference on Computer Vision*, pages 2564–2571, 2011.

[19] Morgan Quigley, Ken Conley, Brian P Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.