IMAV2017-38
http://www.imavs.org/pdf/imav.2017.38

Development of Vision Based Navigation for Micro Aerial Vehicles in Harsh Environment

Hailong Qin*, Yingcai Bi, F. Lin and Ben M. Chen
National University of Singapore, Singapore

Abstract

Safe navigation and planning capabilities are of great importance to Micro Aerial Vehicles (MAVs) applications in different environments. MAVs are allowed to quickly access to the open sky or open area where far away from ground obstacles based on the onboard Global Positioning System (GPS) data. Except for these areas, MAVs become more demanding for missions in complex harsh environment, such as autonomous exploration, inspection and rescue tasks in disaster areas due to their maneuverability and low cost. To complete these missions, visual navigation for MAVs has been extensively studied because of the abundant 3D information and light-weight property.

In this article, a complete navigation system with a multi-heterogeneous sensor setup for visual sensing in harsh environment is proposed. Measurements of multiple RGBD visual sensors are utilized for localization in unknown harsh environment. After this, the measurement from visual sensors is fused with MAV onboard Inertial Measurement Unit (IMU) information through a preintegration approach to achieve autonomously takeoff, navigate and landing in unknown harsh environment. Extensive experiments have been conducted in both indoor and outdoor environments to evaluate the performance of the proposed system. Moreover, a preliminary fast navigation in challenge long corridor environment is also conducted to verify the robustness of the system.

1 Introduction

Due to small size and high manoeuvrability, MAVs have become a powerful and popular tool for rescue, surveillance and exploration. To perform such tasks, MAVs have to fly autonomously in unknown environments, which particularly depends on the results of localization and navigation. However, such operations are remarkably challenging in unstructured GPS-denied scenarios so that visual sensors, such as monocular and stereo cameras have become the most popular choices [1]. Furthermore, due to the stringent payload and power limit, using a high-power CPU or an advanced sensor set would not only significantly increases the power consumption and reduces the flight time, but also requires larger and more powerful motors and propellers, making the system more dangerous for civilian and defense applications. The characteristics of vision-based techniques has disadvantages of high sensitivity to illumination. Realizing localization and navigation for MAVs robustly and efficiently while employing auxiliary equipment as less as possible to save payload is still a critical problem and has been attracting increased attentions from researchers in control, robotics and vision communities.

To solve above issues, simultaneous localization and mapping (SLAM) has been popularly employed and comprehensively investigated for a robust navigation control loop of MAVs. Despite of its rapid progress and great achievements in recent years, SLAM still cannot meet some practical requirements. For instance, the classical SLAM is still a computational intensive process and hardly to be realized on portable computer to achieve real-time navigation [2] [3] [4]. In addition, due to the natural property of visual sensors, common vision-based SLAM can hardly work in low illumination environment.

Besides visual sensors, LIDAR is also a popular sensor used for SLAM problems and provides high frequency updating rate, wide-angle measurement and metric scale depth information [5]. Yet, LIDAR based localization algorithm is limited to the LIDAR sensing space constrain. For instance, using LIDAR to achieve 6 Degree of Freedom(DoF) pose estimation, it is usually required to employ a 3D LIDAR or a rotating 2D LIDAR. However, for both rotating 2D and 3D LIDAR, the point cloud distortion effect due to LIDAR motion still exists [6]. Thus additional sensors such as cameras are normally used for independent position estimation. Therefore, the LIDAR sensors are only largely deployed for dense mapping on Unmanned Ground Vehicles(UGVs) [7].

To operate MAVs successfully in harsh environment such as narrow featureless corridor, dark tunnel or firefighting scenario which filled with smoke, several challenges need to be addressed. First, the MAV should be highly manoeuvrable to fly in space limited environment with onboard sensors. Second, all the perception, planning and obstacle avoidance algorithms should run in real-time and be computational efficient for onboard processing. Third, multiple sensors in different directions should be carried and utilized so that the MAV can

*Email address(es): qinhailong722@gmail.com

even navigate and avoid obstacles in potential dark environment.

In this article, the developed MAV prototype with onboard sensors is firstly presented. The developed prototype is a light weight but efficient platform with onboard powerful computer and multiple visual sensors from Intel named RealSense. The aim of the developed platform is to achieve autonomously takeoff, navigate and land in general daylight indoor and outdoor environment. The developed onboard RealSense based localization and reconstruction algorithm will be presented in the following sections.

2 SYSTEM CONFIGURATION

Our designed platform is a customized small-size and light-weight quadrotor platform. The general properties of quadrotor are fast vertical takeoff, general in air fly and vertical landing. Besides these, Quadrotor platfrom has certain payload to carry onboard sensors. For our designed platform, the Pixhawk developed by ETH is utilized as flight controller [8]. To achieve a stable and robust performance, the robust perfect tracking (RPT) algorithm is developed and implemented on the flight controller. The Intel developed x86 computer, NUC is installed for onboard high level computing such as visual perception, obstacle avoidance and planning. For onboard perception, the depth camera developed by Intel named Realsense is installed in front of MAV platform for visual odometry. Another same type depth camera is downward facing the ground for velocity estimation and precise height measurement. Based on these design requirements, our designed platform has a tip to tip length of $108cm$. The designed platform is shown in Fig 1.

(a) MAV platform in front view (b) MAV platform in top down view

Figure 1: MAV platform in front view and top down view

The control system is divided into outer-loop using the RPT control concept and inner-loop with PID controller. The overall control structure can be described in following Fig. 2. Based on the generated position,velocity and acceleration reference from the mission control. The outer-loop controller generates the a_c. With a global-to-body transformation, the attitude references$(\phi_c, \theta_c, \psi_c)$ can be obtained. The detailed model formulation can be found in [9].

Since the visual odometry provides position measurement

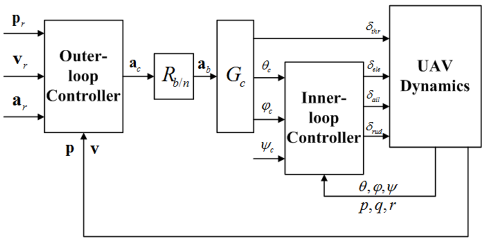

Figure 2: The dual loop control structure

with high frequency noise and low update rate, it can not be directly used as the input for control. Here, we adopt the Kalman filter in [9] to fuse the position measurement and provide a smooth MAV state estimation with a high update rate.

The Kalman filter framework can be described as a discrete-time linear system as following equation

$$\begin{aligned} x(k+1) &= Ax(k) + B(u(k) + w(k)) \\ y(k) &= Cx(k) + v(k) \end{aligned} \tag{1}$$

where x represent the state information. The input vector u equals to $(a_g(k) + w_g(k))$ which is the acceleration in local NED frame. w and v are input and measurement noises. With the designed filter, the system measurement updates at 50 HZ.

3 VISION BASED NAVIGATION FRAMEWORK

As discussed in previous sections, for the purpose of navigating MAVs in GPS-denied environment, additional perception sensors are mandatory for sensing the environment. Visual SLAM or visual odometry is the process of estimate the ego motion from continuous image sequence. The supplied abundant information and light weight of camera makes it a perfect choice for MAV navigation. However, neither monocular camera nor stereo camera can provide the depth information of the environment directly, which is critical to the perception and navigation of MAVs. Different from traditional cameras, our onboard Realsense camera is a infrared depth camera that able to measure the depth of pixel directly even in low illumination environment. Therefore, we can save the computational cost of onboard computer to the largest degree. Our proposed vision based navigation framework aims to solve the MAV perception issue in harsh environment such as dark tunnel.

Our depth estimation module contains two parts:(1) direct depth sensing (2) triangular depth sensing. For the direct depth sensing, the depth measurement is based on the output depth image. However, the depth information is not continuous and smooth enough for pixel based depth sensing. Therefore, a series of following processing modules are used to smooth and register available depth to image pixels.

1. **Median Filtering**
   The depth image captured by the RealSense is quite noisy. A fast, patch based median filter is utilized to

IMAV2017-38
http://www.imavs.org/pdf/imav.2017.38

reduce the noise. The path has a size of 4 *by* 4. After this process, the depth measurement tends to be smooth with less noise.

2. **Depth Registration**
The great advantage of the depth camera is that the depth can be directly captured and stored into depth image. Yet, the color image should be aligned with the depth image since there is a spatial difference between the color image and depth image. Through coordinate transformation, the color image can be registered with depth image.

3. **Pass Through and Noise Filter**
In consideration of the outlier noise caused by the complicated environment, a statistical outlier remove algorithm is applied to remove the small isolated point blobs generated by the camera[10].

Figure 3: color image of forest

Figure 4: depth image of forest

Since not all the pixel depth can be referred from the direct depth, classical triangular depth sensing approach is required. The depth of the sparse visual features are based on the matched feature pair from consecutive valid image sequence. Concurrently, the odometry information can be obtained from multiple observations. During the whole navigation process, a sensing and control technique is needed to extract valuable information from outside environment for the purpose of controlling the MAV pose and localizing itself in an unknown environment. The onboard camera is not only light weighted but also give a comprehensive information in various environments. With the rapid development of monocular camera based visual SLAM, the MAVs equipped with onboard camera are able to provide spatial localization information from the visual odometry system [11]. However, due to the natural disadvantage of the monocular camera which converts a 3D world into a 2D image, the real scale of world MAV facing is lost. This drawback can be alleviated to some extent during landing since the MAV is trying to track a fixed target: either a feature based landmark or a suitable landing area.

Until now, several visual odometry techniques are available for MAVs' navigation and close the control loop. In [12], a realtime optical flow method implemented on monocular camera is presented. However, only the velocity information is estimated which can not give a full state information for control and the variation of light condition in outdoor environment can lead to bad performance. Andrew et al. first realize a realtime visual SLAM on a monocular camera in [13]. Yet this methodology also suffer the issue of depth lost from the beginning.

Different from other approaches, our proposed approach is a depth aware visual odometry without explicit loop closure. During the initialization, the scale and initial motion is obtained from the known depth based on the sparse feature optical flow. Followed by continuously depth and motion estimation from sparse feature matching and association. A local sparse feature with known depth is kept for feature matching.

The classical sparse feature based odometry using monocular camera has established a sophisticated motion estimation framework. Starting from feature detection, the depth (up to a scale) of the detected feature is continuously estimated from feature association. In addition, the camera pose is optimized according to the tracking error. To improve the accuracy and robustness of the tracking, the detected feature is inserted into a 3D map. The camera pose is further refined from matching with built map. The designed process is quite computational intensive. Therefore, we propose a multi-sensor fusion based visual odometry framework to improve the feature perdition accuracy and utilize the direct depth sensing to reduce computational load.

3.1 IMU Preintegration

The tightly-coupled approach is widely utilized in current vision based estimation. The motion between camera frames is precomputed through sensing information from IMU. After this, the IMU error term is computed and optimized with reprojection error to achieve a fast and highly accurate motion estimation. However, the optimization process requests the IMU information to be inserted into a precise timestamp.

Therefore, a dedicated IMU triggered camera capture system is highly demanding which most of commercial camera system does not achieve. In our proposed framework, we only implemented an IMU preintegration between each selected frame to speed up the feature association process. In this framework, we define the world reference frame as W and IMU frame as I. Thus, the rotation, velocity and position of IMU in world reference frame between time interval k is calculated as:

$$R_{WB}^{k+1} = R_{WB}^{k} Exp((\omega_B - b_g)\Delta t) \tag{2}$$

The velocity and position can also be calculated as following equations respectively:

$$v_{W,I}^{k+1} = v_{W,I}^{k} + g_W \Delta t + R_{WB}^{k}(a_I^k - b_a^k)\Delta t \tag{3}$$

$$p_{W,I}^{k+1} = p_{W,I}^{k} + v_{W,I}^{k}\Delta t + \frac{1}{2} g_W \Delta t^2 + \frac{1}{2} R_{WB}^{k}(a_I^k - b_a^k)\Delta t^2 \tag{4}$$

Where b_g and b_a are the varying biases of gyroscope and accelerometer respectively. ω and a are the measurement from IMU.

3.2 Visual Odometry

We define the camera frame as C. Thus, the 3D coordinate of a detected feature i in current frame k is defined as $X_C^{k,i}$. Therefore, if feature i can be detected in selected frame k and $k+1$, the rigid motion relationship can be expressed as

$$X_C^{k+1,i} = R X_C^{k,i} + T \tag{5}$$

The detected features and corresponding depth map for indoor and outdoor condition are shown in Fig. 5 and Fig. 6, where the green points stand for direct depth measurement and blue points stand for indirect depth measurement.

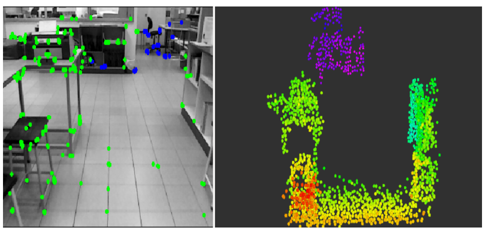

Figure 5: Visual features and depth map in indoor environment

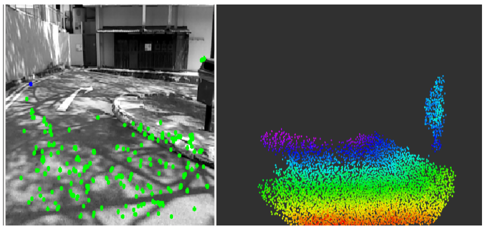

Figure 6: Visual features and depth map in outdoor environment

For the feature with known depth the above equation can be directly transformed into

$$(R_1 - \overline{x_i} R_3) X_C^{k,i} + T_1 - \overline{x_i} T_3 = 0 \tag{6}$$

$$(R_1 - \overline{y_i} R_3) X_C^{k,i} + T_1 - \overline{y_i} T_3 = 0 \tag{7}$$

Where $\overline{x_i}$ and $\overline{y_i}$ are the normalized coordinate by depth measurement z_i.

Using multiple detected features, a nonlinear function can be built using above equations

$$f(R,T) = d \to 0 \tag{8}$$

Moreover, the square error function can be established as

$$S = (f(R,T))^2 \tag{9}$$

Based on the derived error function, the Trust-Region-Reflective method (TRR) [14] is utilized to minimize the error and obtain the transformation matrix.

4 EXPERIMENTAL RESULTS

The proposed framework is verified in challenging wind gust, open outdoor and long corridor indoor environment to demonstrate the accuracy and robustness. Although we are not able to create a real harsh environment to demonstrate the performance of the designed framework, we claim that the designed experiments can test the system to a certain degree.

4.1 Hovering in wind gust

Wind gust is occasionally happened in harsh environment. Therefore, MAV with stable flight capability even in wind gust is highly demanded. The designed wind gust includes random wind with speed of 2 m/s, 4 m/s and 5 m/s. In this autonomous hovering experiment, our designed platform shows outstanding stability using proposed vision based navigation framework. The result is shown in Fig. 7. From the result we can see that the MAV is continuously resist wind to maintain the position and maintain the position error in a reasonable range.

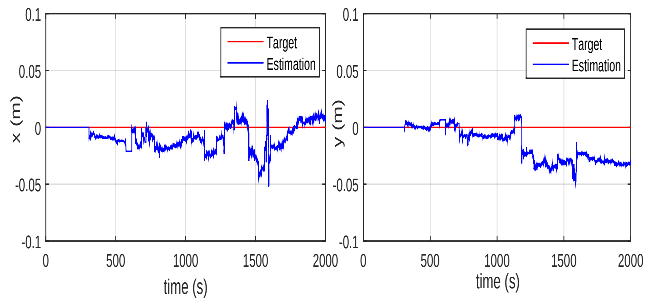

Figure 7: Horizontal motion estimation for MAV hovering in wind gust.

4.2 Flying in outdoor environment

A large outdoor close loop experiment is conducted to demonstrate that our design framework can help MAV to localize and reconstruct the surrounding environment. Since the ground truth is not available, the close loop path is designed to verify the accuracy of the framework. The detected feature and reconstructed dense environment is shown in Fig. 8 and Fig. 9.

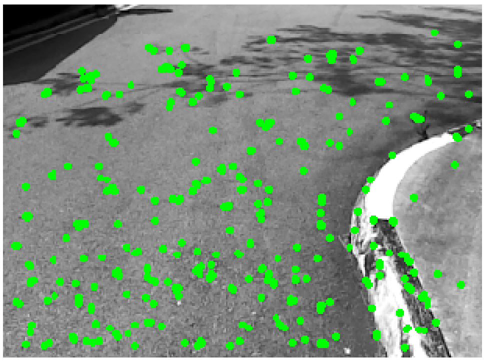

Figure 8: Detected feature in outdoor experiment.

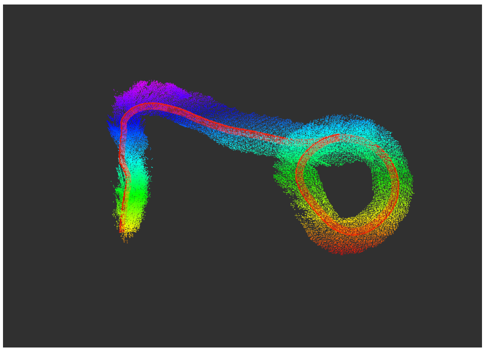

Figure 9: Reconstructed environment of outdoor environment(Estimated trajectory in red).

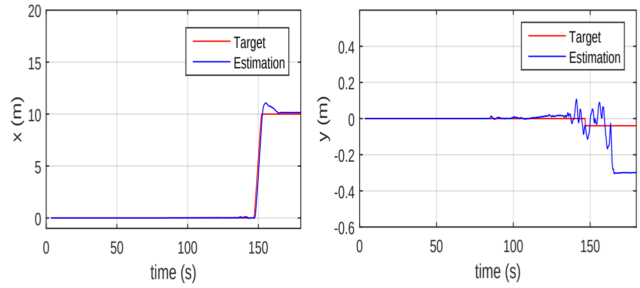

Figure 10: Horizontal motion estimation of high speed flight in long corridor.

Figure 11: Detected features in long corridor.

Due to limited sensing range of RGBD camera we can see only see limited field of view as shown in Fig. 8. However, uniform distributed visual feature can still be detected to achieve motion estimation. In addition, the large scale circular motion also demonstrate the loop closure accuracy of our proposed framework as shown in Fig. 9.

4.3 Flying long corridor environment

Finally, a challenging autonomous flight in long corridor environment is conducted. Due to the limited sensing range of visual sensors, distinct visual feature is preferred to build robust feature alignment. However, this requirement is hard to achieve in homogeneous environment such as long corridor with textureless wall. We demonstrate that our proposed framework can still achieve stable performance in this challenging environment with a average flight speed of 2.3 m/s.

The motion estimation result is shown in Fig. 10. As we can see from results, MAV closely tracks the target reference even in high speed motion with reasonable overshoot. In addition, few feature can be detected in this environment as shown in Fig. 11. Yet, the low number infeature will not greatly affect the performance of our proposed framework.

In addition, a back and forth autonomous flight experiment is conducted to demonstrate the stability of the proposed framework in high speed motion. the motion estimation result and reconstructed dense environment is shown in Fig. 12 and Fig. 13.

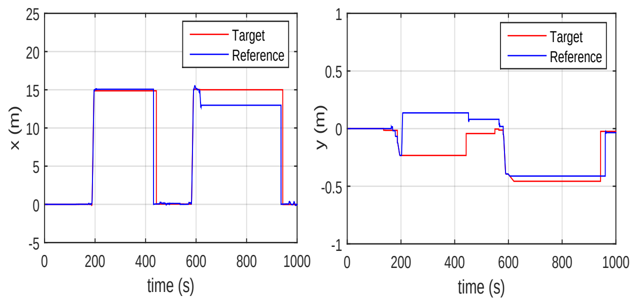

Figure 12: Horizontal motion estimation of high speed back and forth flight in long corridor.

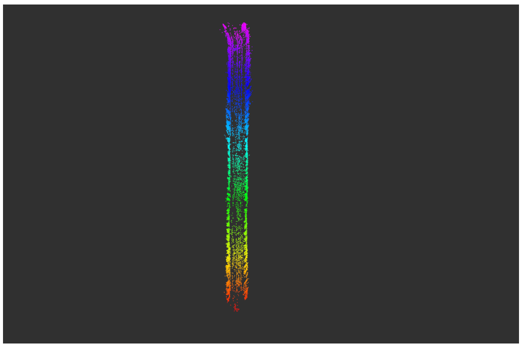

Figure 13: Reconstructed dense corridor environment.

Both the motion estimation with target tracking and clear reconstructed dense corridor verify the accuracy and robustness of our framework as wrong motion estimation will lead to corrupted reconstruction.

5 CONCLUSION

In this article, a complete RGBD camera based navigation framework for MAVs in harsh environments is presented. The vision estimated odometry is further optimized through an IMU preintegration approach. Finally, the realtime estimation is utilized as MAV localization information through a Kalman filter. We demonstrated the robustness of proposed approach in wind gust and efficiency in both indoor and outdoor GPS-denied environments. The experimental results clearly show the outstanding performance of our designed framework regardless of the harsh environments.

REFERENCES

[1] Lionel Heng, Dominik Honegger, Gim Hee Lee, Lorenz Meier, Petri Tanskanen, Friedrich Fraundorfer, and Marc Pollefeys. Autonomous visual mapping and exploration with a micro aerial vehicle. *Journal of Field Robotics*, 31(4):654–675, 2014.

[2] Stephan Weiss, Davide Scaramuzza, and Roland Siegwart. Monocular-slam–based navigation for autonomous micro helicopters in gps-denied environments. *Journal of Field Robotics*, 28(6):854–874, 2011.

[3] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *arXiv preprint arXiv:1607.02565*, 2016.

[4] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European Conference on Computer Vision*, pages 834–849. Springer, 2014.

[5] Stefan Kohlbrecher, Oskar Von Stryk, Johannes Meyer, and Uwe Klingauf. A flexible and scalable slam system with full 3d motion estimation. In *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*, pages 155–160. IEEE, 2011.

[6] Ji Zhang and Sanjiv Singh. Visual-lidar odometry and mapping: Low-drift, robust, and fast.

[7] Yihuan Zhang, Jun Wang, Xiaonian Wang, Chaocheng Li, and Liang Wang. A real-time curb detection and tracking method for ugvs by using a 3d-lidar sensor. In *Control Applications (CCA), 2015 IEEE Conference on*, pages 1020–1025. IEEE, 2015.

[8] Lorenz Meier, Petri Tanskanen, Friedrich Fraundorfer, and Marc Pollefeys. Pixhawk: A system for autonomous flight using onboard computer vision. In *Robotics and automation (ICRA), 2011 IEEE international conference on*, pages 2992–2997. IEEE, 2011.

[9] Guowei Cai, Ben M Chen, and Tong Heng Lee. *Unmanned rotorcraft systems*. Springer Science & Business Media, 2011.

[10] Radu Bogdan Rusu, Zoltan Csaba Marton, Nico Blodow, and Michael Beetz. Interpretation of urban scenes based on geometric features.

[11] Michael Blosch, Stephan Weiss, Davide Scaramuzza, and Roland Siegwart. Vision based mav navigation in unknown and unstructured environments. In *Robotics and automation (ICRA), 2010 IEEE international conference on*, pages 21–28. IEEE, 2010.

[12] Volker Grabe, HH Bulthoff, and Paolo Robuffo Giordano. On-board velocity estimation and closed-loop control of a quadrotor uav based on optical flow. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 491–497. IEEE, 2012.

IMAV2017-38
http://www.imavs.org/pdf/imav.2017.38

[13] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):1052–1067, 2007.

[14] Ya-xiang Yuan. A review of trust region algorithms for optimization.