Computationally Efficient Visual-inertial Sensor Fusion for GPS-denied Navigation on a Small Quadrotor

Chang Liu & Stephen D. Prior
Faculty of Engineering and the Environment, University of Southampton, Southampton, UK

ABSTRACT: Because of the complementary nature of visual and inertial sensors, the combination of both is able to provide accurate and fast six degree-of-freedom (DOF) state estimation, which is the fundamental requirement for robotic (especially unmanned aerial vehicle) navigation tasks in GPS-denied environments. This paper presents a computationally efficient visual-inertial fusion algorithm, by separating orientation fusion from the position fusion process. It is designed to perform 6 DOF state estimation, based on a gyroscope, an accelerometer and a monocular visual-based simultaneous localization and mapping (mSLAM) algorithm measurement. It also recovers the visual scale for the mSLAM. In particular, the fusion algorithm treats the orientation fusion and position fusion as two separate processes, where the orientation fusion is based on a very efficient gradient descent algorithm, and position fusion is based on a 13-state linear Kalman filter. The elimination of a magnetometer avoids the problem of magnetic distortion, which makes it a power-on-and-go system once the gyroscope and accelerometer are factory calibrated. The resulting algorithm shows a significant computation reduction over the conventional extended Kalman filter with competitive accuracy. Moreover, the separation between the orientation and position fusion process enables the algorithm to be easily implemented into separate hardware, thus allowing the two fusion processes to be executed concurrently.

1 INTRODUCTION

The combination of visual and inertial sensors has been shown to be viable, and the significant performance improvement over single sensor has attracted many researchers to get into the field. After the success of (Weiss & Siegwart 2011), which enables world's first autonomous unmanned aerial vehicle (UAV) in GPS-denied environments. (Blösch et al. 2010)

In the past five years, world top research institutes paid high attention to developing advanced monocular visual- based simultaneous localization and mapping (mSLAM) algorithms based on structure from motion (SFM) theory (Klein & Murray 2007; Engel et al. 2014; Forster et al. 2014; Pizzoli et al. 2014; Vogiatzis & Hernández 2011; Roussillon et al. 2011),which are suitable to modern onboard embedded computer. Moreover, the visual scale problem, which was the main challenge of involving monocular vision into control loop, is addressed to various extents by fusing onboard inertial measurements (accelerometer and gyroscope), which is named as visual inertial navigation system (VINS) (Kelly & Sukhatme 2009; Lynen et al. 2013; Lobo & Dias 2003; Li & Mourikis 2013; Jones & Soatto 2011; Weiss et al. 2012).

Almost all the visual-inertial fusion algorithms, to our knowledge, rely on nonlinear Kalman filter (extended Kalman filter, unscented Kalman filter etc.) to process both orientation and position measurement in the same process, which results in a large state vector (mostly more than 20 states) and complex nonlinear system model. Nevertheless, recent advance in computationally efficient IMU orientation estimation (Madgwick et al. 2011), shows a competitive accuracy against Kalman-based algorithm. Thus, in this paper, a computationally efficient visual- inertial fusion algorithm is proposed by separating orientation and position fusion process, which maintains the same level of accuracy with nonlinear Kalman filter. The algorithm is designed to perform six degree of freedom state estimation, based on a gyroscope, an accelerometer and a mSLAM measurement. It also recovers the visual scale for the mSLAM.

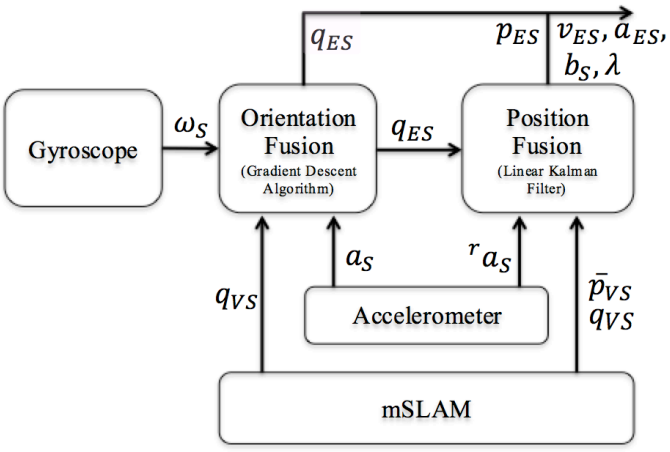


Figure 1 Algorithm Overview

2 ALGORITHM OVERVIEW

As shown in Figure.1, the visual-inertial fusion algorithm assumes rotation, as well as an mSLAM algorithm, which is treated as a black box, provides the un-scaled position. Moreover, it receives angular rates measurement from gyroscope, acceleration measurement from accelerometer. The output of the fusion process is to estimate the true rotation and position of sensor frame in the earth frame, Furthermore, the position filter also estimates the linear velocity, linear acceleration, and accelerometer bias, as well as the metric scale of the mSLAM position measurement.

The fusion is separated into two fusion processes: orientation fusion process and position fusion process. The orientation fusion is based on very efficient gradient descent algorithm (Madgwick et al. 2011), and position fusion is based on a 13-state linear Kalman filter. The following two sections will present the mathematical expression of the two algorithms respectively.

3 ORIENTATION FUSION PROCESS

The orientation fusion algorithm is based on the gradient descent algorithm in quaternion representation. The origin of the algorithm comes from (Madgwick et al. 2011), where the detailed mathematical derivation and proof is presented. However, different from the original algorithm, the following fusion derivation eliminates the magnetometer sensor, while, instead, the rotation correction about gravity vector is compensated by fusing the vision measurement. Therefore, it avoids the problem of magnetic field distortion, thus only factory calibration is required once for gyroscope and accelerometer before the system become fully self-contained. The essential mathematical expression of one iteration at time t is shown as follows. Note that the orientation estimation from last iteration is assumed to be known, and the sampling period is denoted as Δt.

As stated in (Madgwick et al. 2011), given that the convergence rate of the estimated orientation is equal or greater than the angular rate of the physical orientation, only one iteration is required to be computed per sample time, Δt. Therefore an unconventional gradient descent algorithm is derived to fuse all the three sensor measurements. The process to compute the orientation in next time stamp is summarized as

$$q_{ES,k+1} = q_{ES,k} + \dot{q}_{ES,k+1}\Delta t,$$

$$\dot{q}_{ES,k+1} = \dot{q}_{\omega,k+1} - \beta\Delta f,$$

where β is the only adjustable parameter of this filter. It represents the magnitude of the gyroscope measurement error, which is removed in the direction according to the accelerometer and vision sensor. Moreover, since IMU and the vision sensor operate in different speed, Δf is the error direction when comparing with accelerometer or mSLAM vision orientation estimation, depending on which sensor measurement is available.

4 POSITION FUSION PROCESS

This position fusion algorithm assumes the orientation of the sensors is known, thus it takes three inputs: (1) the orientation estimation in earth frame from the result of the orientation fusion process; (2) the raw sensor acceleration measurement from accelerometer; (3) the un-scaled position and orientation in vision frame from the mSLAM. It outputs its state vector, which contains: position estimation, velocity estimation and acceleration estimation, in earth frame, and accelerometer bias, as well as the metric scale $\lambda > 0$ of the mSLAM position estimation. The position fusion algorithm is formed of a coordinate frame management process and a 13-state linear Kalman filter. The Kalman filter conducts in the earth frame, thus, all the sensor measurement values have to be converted to earth frame in the coordinate frame management process.

The conventional Kalman Filter (KF) framework consists of a prediction step, which performs the state vector time update in constant time interval; and a measurement update step, which performs the correction of the state vector based on the new sensor measurement. Here in order to encounter the asynchronies measurements from both accelerometer and mSLAM algorithm, two different measurement update models are constructed, and will be executed depending on which sensor measurement is available.

The state of the Kalman filter is represented as a state vector x:

$$x = [p_{ES}^T, v_{ES}^T, a_{ES}^T, b_S^T, \lambda]^T, \tag{3}$$

where position estimation p_{ES}, velocity estimation v_{ES} and acceleration estimation a_{ES} are in earth frame, and accelerometer bias b_S^T is in sensor frame, as well as the metric scale $\lambda>0$ of the mSLAM position estimation.

The state vector is updated once every time interval, following the rule defined by the prediction model, which defines the physics of the inertial system. It is summarized as:

$$\dot{p}_{ES} = v_{ES}, \tag{4}$$

$$\dot{v}_{ES} = a_{ES}, \tag{5}$$

$$\dot{a}_{ES} = n_a, \dot{b}_S = n_b, \dot{\lambda} = n_\lambda. \tag{6}$$

where n_a, n_b and n_λ are independent zero-mean normal distribution Gaussian process noise.

The measurement model is derived in the form of:

$$z_* = H_* x + e_*, \tag{31}$$

where z_* is the measurement from the mSLAM vision sensor or the IMU, H_* is the measurement model matrix, and e_* denotes the measurement error from the sensor, where $\star$ can be '*as*' or '*vs*' depending on which sensor measurement is available between acceleration sensor measurement and vision sensor measurement. Here, e_* is also modeled as independent zero-mean normal distribution Gaussian process noise.

Measurement update process handles different sampling rate between mSLAM and IMU estimation, by only updating state with the corresponding measurement, which becomes available. Thus by assuming the orientation fusion reaches steady state, the state vector x can be effectively estimated over time.

5 IMPLEMENTATION

In this section, we will demonstrate the fusion performance based on real data, on embedded platform. We used FreeIMU v0.4.3 hardware, which includes an MPU6050 gyroscope- accelerometer combo chip, an HMC5883 magnetometer and MS5611 high-resolution pressure sensor. However, in this experiment, only the MPU6050 is used. We performed orientation estimation in Teensy 3.12, which features an ARM Cortex-M4 processor running at 96 MHz. Besides, we run the SVO mSLAM framework as black box on Odroid-U3 single board embedded Linux computer, which features 1.7GHz Quad-Core processor and 2GByte RAM. In the same Odroid-U3 computer, we conduct position fusion in parallel with the SVO. The video is captured by an uEye global shutter monocular camera. The communication between software packages is realized by Robot Operating System (ROS).

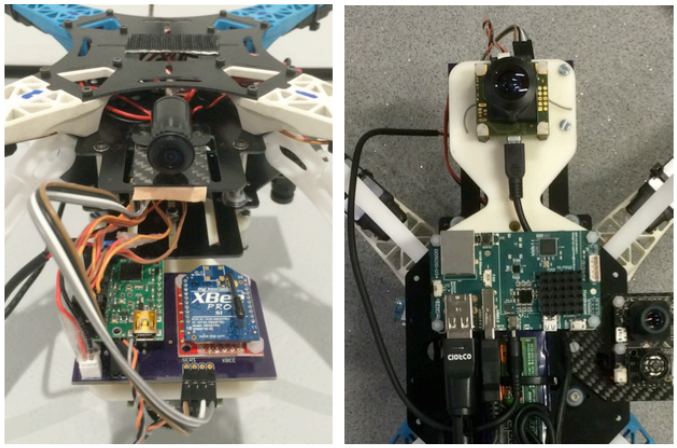
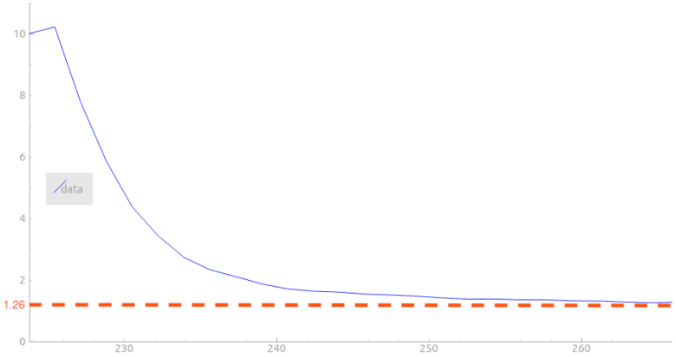
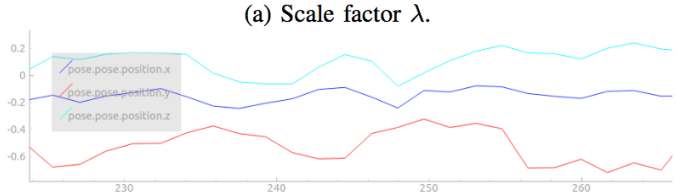


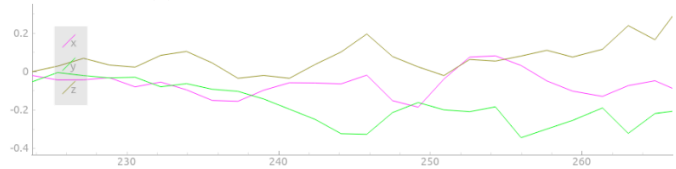
Figure 2 Hardware Implementation.



(a) Scale factor λ.



(b) Position measurement from SVO.



(c) Position measurement from KF position estimator.

Figure 3 Fusion Results.

The entire system is installed onto a quadrotor platform, as shown in Figure 2. The Autopilot board including the FreeIMU, Teensy processor, servo controller and XBee Radio are shown in left side of Figure 4. The uEye camera and Odroid-U3 computer is installed underneath the quadrotor as shown in right side of Figure 4.

6 TEST RESULTS

The Teensy processor is capable of executing the orientation fusion alongside with autopilot control algorithm at 300 Hz, while communicating with Odroid-U3 computer with ROS protocol, including publishing orientation estimation and acceleration measurement at 100 Hz and subscribing the pose estimation from SVO mSLAM framework in Odroid-U3. Moreover, in the Odroid-U3 computer, the SVO mSLAM is executed at 40 FPS with the KF position fusion algorithm running at 100 Hz in parallel.

We assume the timing error is negligible in the system. β is left as default value, 0.5, by assuming $\tilde{\omega}$ max is approximately 0.58 rad/s.

The KF position fusion algorithm is initialized with the state vector $x_0 = [0_{1\times 12}, 10]^T$, note that we initialize the scale factor λ to 10 as a arbitrary positive value to show how it converges to the true value. Also, the position fusion assumes the orientation fusion reaches the steady state before initialization.

The real time test result is shown in Figure 6. The initialization occurs at 227 second and the record shows a 39-second trail, indicating how the true scale factor is recovered, and how the output position estimation relates to the raw input position measurement over the same period of time.

It is clear that the scale factor λ is effectively discovered as 1.26, despite that its initial value is set to 10, as shown in Fig.6a. And during the converging period, the position estimation output from KF position estimator is scaled accordingly with λ over time. Since KF position estimator fuses the acceleration with SVO position measurement, the output position estimation performs better in dynamic operation.

7 CONCLUSION AND FUTURE WORK

This paper has shown the design, implementation work of a sensor fusion framework, which is capable of performing the six degree of freedom sensor state estimation, by fusion a 3 axis gyroscope, a 3 axis accelerometer and a vision based monocular simultaneous localization and mapping algorithm.

The future work includes further test evaluating of estimation error comparing with ground truth. And computational performance evaluation by benchmarking with other existing fusion algorithms.

8 REFERENCES

Blösch, M. et al., 2010. Vision based MAV navigation in unknown and unstructured environments. *Proceedings - IEEE International Conference on Robotics and Automation*, pp.21–28. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5509920 [Accessed December 11, 2013].

Engel, J., Sch, T. & Cremers, D., 2014. LSD-SLAM: Large-Scale Direct Monocular SLAM. *Eccv*, pp.1–16. Available at: http://link.springer.com/chapter/10.1007/978-3-319-10605-2_54 [Accessed November 25, 2014].

Forster, C., Pizzoli, M. & Scaramuzza, D., 2014. SVO : Fast Semi-Direct Monocular Visual Odometry. *IEEE International Conference on Robotics and Automation (ICRA)*. Available at: http://rpg.ifi.uzh.ch/docs/ICRA14_Forster.pdf [Accessed May 7, 2014].

Jones, E.S. & Soatto, S., 2011. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, pp.1–38. Available at: http://ijr.sagepub.com/content/30/4/407.short [Accessed November 6, 2013].

Kelly, J. & Sukhatme, G.S., 2009. Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration. *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA*, pp.360–368. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5423178.

Klein, G. & Murray, D., 2007. Parallel tracking and mapping for small AR workspaces. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*, pp.1–10. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4538852.

Li, M. & Mourikis, a. I., 2013. High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research*, 32(6), pp.690–711. Available at: http://ijr.sagepub.com/cgi/doi/10.1177/0278364913481251 [Accessed November 18, 2014].

Lobo, J. & Dias, J., 2003. Vision and Inertial Sensor Cooperation Using Gravity as a Vertical Reference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12), pp.1597–1608.

Lynen, S. et al., 2013. A robust and modular multi-sensor fusion approach applied to MAV navigation. *IEEE International Conference on Intelligent Robots and Systems*, pp.3923–3929. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6696917.

Madgwick, S.O.H., Harrison, A.J.L. & Vaidyanathan, R., 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. *IEEE International Conference on Rehabilitation Robotics*, (1945-7898), pp.1–7.

Pizzoli, M., Forster, C. & Scaramuzza, D., 2014. REMODE : Probabilistic , Monocular Dense Reconstruction in Real Time. *Proc. IEEE International Conference on Robotics and Automation (ICRA)*.

Roussillon, C. et al., 2011. RT-SLAM: A generic and real-time visual SLAM implementation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), 6962 LNCS, pp.31–40. Available at: http://link.springer.com/chapter/10.1007/978-3-642-23968-7_4 [Accessed October 28, 2013].

Vogiatzis, G. & Hernández, C., 2011. Video-based, real-time multi-view stereo. *Image and Vision Computing*, 29(7), pp.434–441. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0262885611000138 [Accessed March 26, 2014].

Weiss, S. et al., 2012. Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV. *Proceedings - IEEE International Conference on Robotics and Automation*, pp.31–38. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6225002.

Weiss, S. & Siegwart, R., 2011. Real-time metric state estimation for modular vision-inertial systems. *Proceedings - IEEE International Conference on Robotics and Automation*, 231855, pp.4531–4537. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5979982 [Accessed October 14, 2013].