Multi-sensor Fusion and Application In SlAM

Haotian Gu
Literature Review Paper
ME 656: Autonomous Navigation for Mobile Robots, Spring 2020
Stevens Institute of Technology

ABSTRACT

There has been an ever-increasing interest in multi-disciplinary research on multi sensor data fusion technology, driven by its versatility and diverse areas of application. The Simultaneous Localization and Mapping is one of the major applications in the field of sensor data fusion. Therefore, there seems to be a real need for an analytical review of recent developments in the target application for combined sensor is Simultaneous Localization and Mapping (SLAM). This paper proposes a short review about analysis of different ways of integration about the visual inertial odometry.

INTRODUCTION

Multi sensor data fusion is the process of combining observations from a number of different sensors to provide a robust and complete description of an environment or process of interest. Due to a single sensor does not work well for all scenarios. So fusing the information from two or more measurement channels allows us to mutually compensate for some of the unwanted effects so that achieve the accurate positioning through the fusion of multiple sensors. Obviously sensor fusion is a necessary trend. The target application for our combined sensor is Simultaneous Localization and Mapping (SLAM).

There is a popular application that is fusing camera and IMU. The camera produces pictures, The six-axis gyroscope produces acceleration and angular velocity. Second, Visual Inertial Odometry VIO implements a more complex and effective Kalman filter, such as MSCKF (Multi-State-Constraint-Kalman-Filter), Which focus on fast pose tracking instead of maintaining global maps. Google's Project Tango Flyby Media, which has been purchased by Apple, represents two famous commercial implementations. Visual sensors work well in environment with most textured scenes. But it basically can not work under the environment with less feature such as white wall， glasses and so on. IMU has a large cumulative error if it keeps on working for a long time. However the estimation about its relative displacement is accurate in a short period of time. Therefore, when the vision sensor fails, the fusion of IMU data can improve the accuracy of positioning. Through using differential GPS, IMU and lidar together is another scheme for positioning. Differential GPS can achieve good position accuracy under environment in good weather and less blocking. Otherwise the positioning accuracy reduce if in urban with high-rise building or encountering tough whether conditions. Therefore fusing IMU and camera is able to complement shortages of application of only GPS in use.

The measurement of the angular velocity and acceleration of the sensor body is considered to be significantly complementary to the camera sensor. And there is great potential to get a better SLAM system after fusion.

1. Although IMU can measure angular velocity and acceleration, but these amounts will drift as time pass by.The position data that makes the integrals in two times very unreliable. However, IMU can provide some good estimation for fast movement in a short period of time. This character can complement the weakness of the camera. When it moving too fast, the camera screen will be blurred. Even too few overlap areas between the two continuous frames makes matching features impossible. So pure visual SLAM can not work in fast motion. Under support of IMU can also maintain a better position estimate even during the time when the camera data is invalid.

2. Compared with IMU, the camera data will not drift. If the camera is stationary, then the position estimation of the visual SLAM is also stationary (in static scene).So Camera can effectively estimate and correct drift error in the IMU reading, making the position estimation still valid under slow motion.

3. There is still another application limitation of pure visual SLAM in dynamic obstacles. Image change can not identify whether the camera move or camera is under the changed external conditions. The IMU can measure its own motion to some extent to offset the shortage of camera in dynamic objects.

ALGORITHM ANALYSIS AND DEVELOPMENT ABOUT RESOLVING POSE THROUGH USING SINGLE IMU

Google Cardboard's nine-axis fusion algorithm - extended Kalman filter based on Li Qun The nine-axis fusion algorithm refers to the method of obtaining the attitude of the object by fusing the accelerometer (three-axis), gyroscope (three-axis), and magnetic field meter (three-axis) in the IMU. It is a vital part of developing VR head-on. Virtual reality head show must get the user's head pose accurately in real time and then render the picture on the screen with respect to equal pose so that let the user to immerse in the VR world.

[1]Madgwick algorithm, complementary filtering algorithm is a filter which fuses angular velocities, accelerations, and

(optionally) magnetic readings from a generic IMU device into an orientation quaternion. Google Cardbord's nine-axis fusion algorithm, Madgwick algorithm, [2]complementary filtering algorithm are about sensor fusion on SO3, which means the output is only a purely rotating pose.

The current VIO framework has been categorized into loosely couple and tightly couple. Fig 1 and Fig 2 show the loosely couple and tightly couple respectively. Loosely couple means to estimate IMU and camera motivation respectively and then fuse the estimation result about pose. Tightly couple means combine the state of IMU and camera together so that building motion and observation equation to estimate state. Otherwise tight coupling theory can be divided into two directions namely filter based and optimization based. Optimization based method also has its development branch. Optimization based method has become mainstream in pure visual SLAM. However due to high frequency of IMU data, optimizing state needs too much calculation. So the prevalent method is cooperating filter and optimization together. VIO is potential direction for miniaturization and low cost.

The loosely coupled method use a inertial positioning module and a positioning navigation module . The frequency of updates between the two modules is inconsistent meanwhile these two models exchange information with each other.In loose coupling mode, inertial data is used as the main part and the visual measurement data to correct the cumulative error of inertial measurement data.

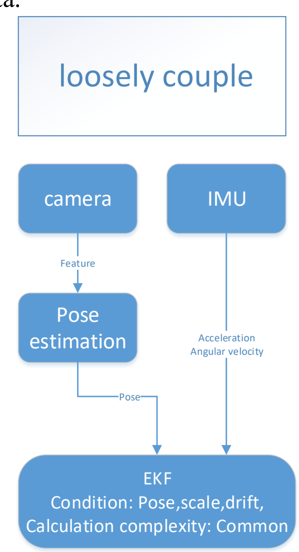

Fig 1

The visual positioning method in the loose coupling method acts as a black box module without the assistance of IMU information. Therefore it is not robust enough if it is hard to use visual position and this method cannot correct the drift introduced by visual measurement.

Tightly coupled uses IMU to complete motion estimation in the visual VO. Integrals of IMU between image frames have a small margin of error, and data from imu can be used to predict motion between two continuous frames so that speed up completion about point matching and position estimation in VO. Compared with loose coupling, another advantage of tight coupling is that IMU's scale measurement information can be used to aid the estimation of scales in vision .

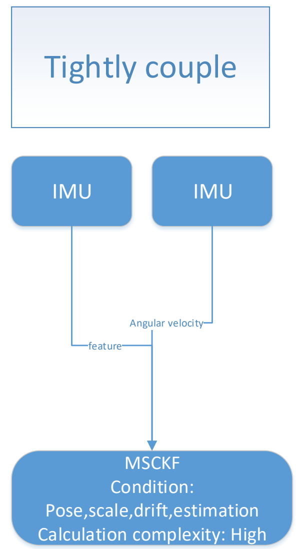

Fig 2

INTRODUCTION ABOUT DIFFERENT COUPLE METHODS

1.Filter-based Tightly Coupled method

The image feature should be contained into the feature vector in tight coupling Therefore, the dimension of the state vector of the whole system is very high. Correspondingly a very high calculation is required. MSCKF, ROVIO are all classic algorithm. One of example about tight couple is MSCKF. Google Tango. also adopt MSCKF algorithm framework. In the traditional EKF-SLAM framework, the information of the feature point is added to the feature vector and covariance matrix. The disadvantage of this approach is that the information of the feature point gives an initial depth and an initial covariance. It is extremely easy to lead to the subsequent non-convergence and inconsistent. MSCKF maintains a pose FIFO, in chronological order, and can be called a sliding window.If a feature point is observed in several poses (frames) in the sliding window, Constraints are established between these poses to update KF.

EKF-SLAM: Multiple feature points constrain a camera position at the same time for updating KF.
MSCKF: One feature point constrains multiple camera positions simultaneously (optimization of multi-camera observation, multi-frame optimization in window) for KF updating.
When using traditional EKF-based SLAM to achieve IMU fusion, the system state vector for each moment contains the current pose, velocity, and 3D map points coordinates (IMU fusion is generally joined with IMU's bias (floating: zero and

drifting) and then using IMU information to predict step. Then through using the observation error of the 3D map points observed in the image frame to update step.

The motivation about MSCKF's improvement is that each update (similar optimization) step of the EKF is based on 3D map points observed in a single frame. It would be good if its observations is under support of multiple frames (like the idea of local bundle adjustment).So MSCKF improvements are as follows:
prediction step at the prediction stage is the same as EKF update step in the update phase defers to a 3D map point after it has been observed in multiple frames. Each frame is received before update simply by expanding the state vector and adding the estimation of current frame's pose. This idea is basically similar to the local bundle (or sliding window smoothing). This is equivalent to optimizing both pose and 3D map points based on multiple observations in updating step

2.Filter-based Loosely Coupled method

The method of loose coupling is much simpler, which avoiding the image's feature into the state vector. Instead, the image is treated as a black box, and then the IMU data is fused after the vo processing is calculated. Ethz's Stephen Weiss has done a lot of research on this. His ssf and msf[3] are excellent open source algorithms in this area

---Filter-based loose coupling example-ssf

The filter's state vector x is 24 dimensions, which is much simpler than the tight coupling method. Ssf_core mainly deal with state data, which contain both prediction and update processes. Ssf_update processes data from another sensor, mainly completing the measurement process

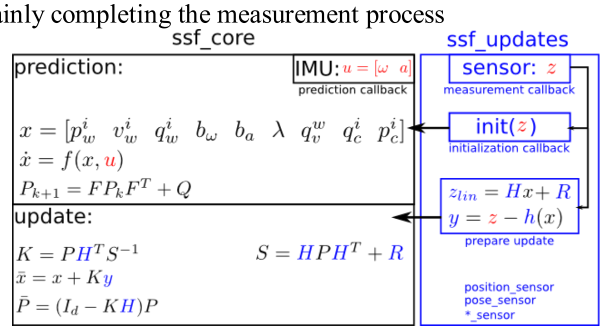

Fig 3

The red font part is data obtained from the sensor and send into the prediction and update phases. The blue font changes in the updating phase. The black part is constrained part and meanwhile is constant.

Variables introduction：

p for pwi: IMU position in the world frame

v for vwi: IMU velocity in the world frame

q for qwi: IMU pose in the world frame

b_w for bw: the gyro biases

b_a for ba: the accelerometer biases

L for λ : the visual scale factor with pmetric* λ = psensor

---Filter-based loose coupling example-msf

The basic model is shown in the following image Fig4.MSF is easy to generalize to the application of new sensors such as GPS, lidar, encoder, etc.

First, there is no gravity in its core state under the world coordinate system. The world coordinate system is based on horizontal coordinate systems. The initial position is calculated based on the measurements and theoretical gravity of the current accelerometer, with a covariance at the initial position. Moreover, IMU's world coordinate system and camera's world coordinate system are not tied together to build. The camera can acquire its world coordinate system after the IMU has been turned on for a short time. But the time between the interval should be as short as possible, because the displacement depends on the IMU accelerometer integral. As time pass by the accuracy will reduce. if there is a wheel encoder, then it can be better.

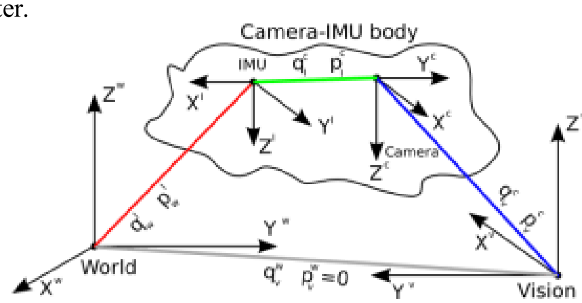

Fig 4

3.Loosely coupling based on optimization

With the development and improvement of the computing platform, The optimized-based method is applied into use for SLAM. There are not too much research work done for loose couple. Gabe Sibley[4] mention this method in his paper.

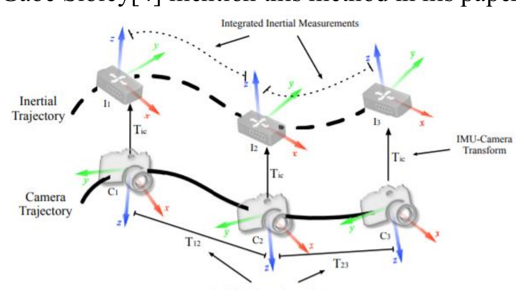

Fig 5

Fig. 5. A windowed bundle adjuster takes frame-to-frame relative estimates, and its corresponding covariance, as binary constraints and jointly optimizes poses with integrated inertial measurements. The sensor rig calibration file, which contains camera intrinsics as well as the camera-to-IMU transform, is also provided.

Tracking is performed in a loosely coupled sliding windowed dense/semi-dense visual-inertial bundle adjuster[5]. The bundle adjuster receives as input relative constraints between poses from the visual odometry engine as well as IMU measurements to form visual and inertial residuals.

Visual Residuals (e_v): Visual only frame-to-frame tracking is performed in a 2.5D Lucas-Kanade [7] style photometric minimization using the Efficient Second Order minimization [8] technique.

$$e_v = \rho(I_{cur}\left(\varphi(k, u_p, d, T_{cp})\right) - I_{pre}(u_p), \sigma)$$

Finally, the visual frame-to-frame relative transform T_{cp}estimated by the optimization is transferred into the IMU's reference frame – which is the privileged frame – and added into the bundle adjuster as a binary constraint.

Inertial Residuals (eI): Inertial measurements between frames are integrated forming residuals against the estimated poses as seen in Fig. 5[7]. Each pose in the bundle adjuster contains world poses comprised of a 3-DOF translation and 4-DOF rotation vector (quaternion parametrization). It also stores a 3-dimensional velocity vector and two 3-dimensional vectors for the accelerometer and gyroscope biases.

In total, through adding the position change generated by VO calculation to the optimization framework of imu to minimize the visual and inertial residual so that reduce the error between the actual path and estimated path.

4.Tight coupling based on optimization

Examples of tight coupling based on optimization - Okvis Multi-Purpose[8] is shown bellows.

$$\mathrm{J}(x) \coloneqq \sum_{i=1}^{I} \sum_{K=1}^{K} \sum_{j \in J(i,k)} e_r^{i,j,k} W_r^{i,j,k} e_r^{i,j,k} + \sum_{k=1}^{K-1} e_s^{k^T} W_s^k e_s^k$$

The left side of the formula is the pure visual odemorty, and the right of the formula is the visual IMU fusion odemorty structure. This formula is about organization of frame and IMU. But the measurement of IMU takes a random bias. So every measurement is combined with this bias so that form the structure to the right of the formula. For this new part in formulation, we need to establish a unified loss function for joint optimization.

Corresponding to [9]MSCKF's filter-based SLAM faction, OKVIS is the keyframe-based SLAM faction to do visual-multiply sensor fusion. From MSCKF's algorithmic framework, it can be concluded that OKVIS is an optimization problem that combines image observation and imu observation into an optimization problem, and goes together to optimize the solution pose and 3D map point.

Indeed, [10]OKVIS's optimization target function includes a reprojection error term (reprojection error) and an imu integration error term (imu integral error). The known observations are the feature matching between each two frames and the integrals of all imu sampling data between the two frames.Note that imu sampling frequency is generally higher than the video frame rate, and the camera pose and 3D map point are to be asked.The optimization is for frames within a bounded window (including several recent frames and several keyframes).

It is important to note that in this optimization problem, the modeling of the uncertainty (uncertainty, similar variance) is still quite complex.The first is that constructing a model for the gyro of imu and the bias (drift) of accelerometer and the integral for uncertainty (uncertainty, similar variance) is still in need. Therefore when deduce the IMU integration error between the two frames, Covariance (covariance matrix) is required through using a method similar to the uncertainty propagation in Kalman filter step prediction (prediction phase). Motivation for keyframes (innovation point) in OKVIS is to filter out some frames with a small amount of information and left a few keyframes between the constraints due to the speed limit of the optimization algorithm and Optimization cannot be targeted at too many frames together.

CONCLUSION

This paper presented a critical review of focus on the introduction about fours schemes of sensor fusion from tightly coupling to loosely coupling method. Through compensating shortage in sensor, the accuracy of position in slam will increase. However the sensor fusion scheme is not limited in the cooperation mentioned in this short review. Data fusion is a multi-disciplinary research field with a wide range of potential applications in SLAM areas This has been and will continue to act as the driving force behind the ever-increasing interest in research community in developing more advanced data fusion methodologies and architectures.

REFERENCES

[1] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," pp. 1–7, July 2011.

[2] Valenti, R. G., Dryanovski, I., & Xiao, J. (2015). Keeping a good attitude: A quaternion-based orientation filter for IMUs and MARGs. Sensors (Switzerland), 15(8), 19302-19330

[3] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. International journal of computer vision, 56(3):221–255, 2004.

[4] N. Keivan, A. Patron-Perez, and G. Sibley. Asynchronous adaptive conditioning for visual-inertial slam. In International Symposium on Experimental Robotics, ISER '14, 2014

[5] S. Klose, P. Heise, and A. Knoll. Efficient Compositional Approaches for Real-Time Robust Direct Visual Odometry from

[6] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart and Paul Timothy Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. The International Journal of Robotics Research, 2015.

[7] RGB-D Data. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 2013

[8] http://wiki.ros.org/ethzasl_sensor_fusion/Tutorials/Introductory%20Tutorial%20for%20MultiSensor%20Fusion%20Framework

[9] lquez, J.M.; Kasper, M.; Sibley, G. Inertial aided dense & semi-dense methods for robust direct visual odometry. In Proceedings

[10] K. Sun, K. Mohtam B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, V. Kumar, IEEE Robotics and Automation Letters, Robust stereo visual inertial odometry for fast autonomous flight, 3, 2, 965-972 (2018)

[11] Liu, S.-C., Delbruck, T., Indiveri, G., Whatley, A., Douglas, R., Event-Based Neuromorphic Systems,Wiley. ISBN: 978-1-118-92762-5, 2014.