Visual-inertial Odometry Using Iterated Cubature Kalman Filter

Jianhua Xu[1], Huan Yu[1], Rui Teng[1]

1. School of Automation, Beijing Institute of Technology, Beijing 100081
E-mail: jianhuaxu_bit@163.com

Abstract: In recent years, there are many excellent algorithms for visual-inertial odometry. However, in practical engineering application, we must consider the computational cost. High-precision optimization-based methods usually can not meet the requirement of real-time. In this paper, we proposed a visual-inertial odometry algorithm, which is based on iterated cubature Kalman filter. Compared with EKF-based method, it can reduce the influence of linearization and improve localization precision. The computational complexity of this method is similar with other Kalman filtering based methods. Experimental results are presented for a real-world dataset captured on a Beijing street with a land vehicle and the results show that the method proposed can attain a better accuracy than other methods.

Key Words: Visual-inertial Odometry, Cubature Kalman filter, Localization

1 INTRODUCTION

Visual SLAM (simultaneous localization and mapping) has been a hot subject for research in the last two decades in the Computer Vision and Robotics. Localization and mapping are two fundamental problems in the research area of robot navigation and control [1]. Visual inertial odometry (VIO) is a technique to estimate the change of a mobile platform in position and orientation over time using measurements form cameras and IMU sensor, which can track a vehicle's motion in GNSS-denied environment [2]. Visual and inertial measurements offer complementary properties, which make them particularly suitable for fusion. The rich representation of structure projected into an image, together with the accurate short-term estimates by gyro-scopes and accelerometers contained in an inertial measurement unit (IMU) have been acknowledged to complement each other. There are many promising use-cases in automotive and airborne localization.

There are two main approaches for the visual-inertial data fusion algorithms: nonlinear optimization methods and filtering based methods [3, 4, 5]. The former approach can repeatedly linearize the nonlinear cost in the visual-inertial state, therefore those methods can limit linearization errors. But the iteration also incurs high computational cost. In some practical application, the computational resources are limited, thus we often choose filtering based methods. Based on the Bayesian estimation framework, there are many suboptimal Kalman filters used for nonlinear system filtering, such as extended Kalman filter (EKF), unscented Kalman filter (UKF) and cubature Kalman filter (CKF) [6]. EKF may generate large approximation error when the system has significant nonlinearities. CKF and UKF are Sigma-point based Kalman filter, which usually used for high dimension state estimation. Recently, the most popular VIO algorithms are EKF-based algorithms, which can obtain high accuracy and the computational cost are lower than iterative-minimization algorithms.

To reduce the influence of linearization on parameter estimation, in this paper we proposed an iterated cubature Kalman filter (ICKF) for the data fusion. In section two, state of art works about visual-inertial odometry are introduced. In section three, we introduced the model of the ICKF. In section four, we introduced the system model of the visual-inertial system. The last section provides the simulation and real-world data experiment results, which show that the method proposed in this paper can really improve the localization accuracy compared with EKF-based method.

2 RELATED WORKS

Filtering approaches enable efficient estimation by restricting the inference process to the latest state of the system. In filtering based approaches, according to whether the image feature information is added to the state vector, the visual-inertial data fusion methods can be divided into two categories: loosely coupled and tightly coupled approaches [7, 8, 9]. For loosely coupled methods, to improve accuracy of image-based motion estimation, Roumeliotis S. et al. proposed a Kalman filter based method for the fusion of visual and inertial data [10]. Weiss et al. use IMU model for state propagation with vision only-pose estimation for measurement update [11, 12]. Comparing both approaches, recent researches show that tightly couple approaches can achieve higher motion estimation precision. A notable EKF-based method is proposed by Mourikis [13]. This approach is called MSCKF. Bloesch et al. proposed a direct EKF-based approach [14]. Kelly and Sukhatme also proposed an UKF-based method [16].

3 ITERATED CUBATURE KALMAN FILTER

For generalized iterated Gaussian filter, consider a nonlinear discrete-time system

$$\begin{cases} x_k = f(x_{k-1}) + \omega_k \\ z_h = h(x_k) + v_k \end{cases} \tag{1}$$

978-1-5386-1243-9/18/$31.00 ©2018 IEEE

Where x_k is the state vector and z_k is the measurement vector at time step k. ω_k and ν_k are Gaussian white noise with covariance Q_k and R_k. $f(\cdot)$ and $h(\cdot)$ denote the dynamic model and measurement model of the system. The approximation and update stage of the Gaussian filters can be described as the following equations

$$\hat{x}_{k|k-1} = \int f(x_{k-1}) N(x_{k-1} | \hat{x}_{k-1|k-1}, P_{k-1|k-1}) dx_{k-1} \quad (2)$$

$$P_{k|k-1} = \int (f(x_{k-1}) - \hat{x}_{k|k-1})(f(x_{k-1} - \hat{x}_{k|k-1})^T N(x_{k-1} | \hat{x}_{k-1|k-1}, P_{k-1|k-1}) \times dx_{k-1} + Q_k \quad (3)$$

$$\hat{z}_{k|k-1} = \int h(x_k) N(x_k | \hat{x}_{k|k-1}, P_{k|k-1}) dx_k \quad (4)$$

The iterated cubature Kalman filter used in this paper includes two stages: Time update and Measurement update. Let $x_{k-1} \sim N(\hat{x}_{k-1|k-1}, P_{k-1|k-1})$, $\omega_{k-1} \sim N(\hat{q}_{k-1}, \hat{Q}_{k-1})$ and $v_k \sim N(r, R)$. The cubature point number $m = 2n$. The integrals involved in (2), (3) and (4) can be calculated by the following equation

$$I(f) = \int_{\Re^n} f(x_k) p(x_k | z_{1:k-1}) dx1_k \approx \sum_{i=1}^{m} w_i f(\xi_i) \quad (5)$$

Where $w_i = \frac{1}{m}$, $i = 1, 2, ..., 2n$ and $\xi_i = \sqrt{\frac{m}{2}}[1]_i$. The cubature Kalman filter can be described as the following two steps:

Time update:

Firstly, factorize $P_{k-1|k-1} = S_{k-1|k-1} S^T{}_{k-1|k-1}$,generate cubature points and propagate them by system function

$$x_{i,k-1|k-1} = S_{k-1|k-1} \xi_i + \hat{x}_{k-1|k-1} \quad (6)$$

$$x_{i,k|k-1} = f(x_{i,k|k-1}) + \hat{q}_{k-1} \quad (7)$$

Then, estimate the predicted state and error covariance

$$\hat{x}_{k|k-1} = \sum_{i=1}^{m} w_i X_{i,k|k-1} \quad (8)$$

$$P_{k|k-1} = \sum_{i=1}^{m} w_i X_{i,k|k-1} X^T{}_{i,k|k-1} - \hat{x}_{k|k-1} \hat{x}^T_{k|k-1} + \hat{Q}_{k-1} \quad (9)$$

Measurement update:

Firstly, factorize $P_{k|k-1} = S_{k|k-1} S^T{}_{k|k-1}$, generate cubature points and propagate them through measurement function

$$\zeta_{i,k|k-1} = S_{k|k-1} \xi_i + \hat{x}_{k|k-1} \quad (10)$$

$$Z_{i,k|k-1} = h(\zeta_{i,k|k-1}) + r \quad (11)$$

Then, estimate the predicted measurement, innovation covariance and cross-covariance matrix

$$\hat{z}_{k|k-1} = \sum_{i=1}^{m} w_i Z_{i,k|k-1} \quad (12)$$

$$P^{zz}_{k|k-1} = \sum_{i=1}^{m} w_i Z_{i,k|k-1} Z^T{}_{i,k|k-1} - \hat{z}_{k|k-1} \hat{z}^T_{k|k-1} + R \quad (13)$$

$$P^{xz}_{k|k-1} = \sum_{i=1}^{m} w_i \zeta_{i,k|k-1} Z^T{}_{i,k|k-1} - \hat{x}_{k|k-1} \hat{z}^T_{k|k-1} \quad (14)$$

The Kalman gain and state can be calculated

$$K_k = P^{xz}_{k|k-1} \left(P^{zz}_{k|k-1} \right)^{-1} \quad (15)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - \hat{z}_{k|k-1}) \quad (16)$$

Finally, update system noise $\omega_k \sim N(\hat{q}_k, \hat{Q}_k)$,

$$\hat{Q}_k = (1 - d_k) \hat{Q}_{k-1} + d_k [K_k \varepsilon_k \varepsilon_k^T K_k^T + P_{k|k} - (\sum_{i=1}^{m} w_i X_{i,k|k-1} X^T_{i,k|k-1} - \hat{x}_{i,k|k-1} \hat{x}^T_{k|k-1})] \quad (17)$$

$$\hat{q}_k = (1 - d_k) \hat{q}_{k-1} + d_k [\hat{x}_{k|k} - \sum_{i=1}^{m} w_i f(x_{i,k-1|k-1})] \quad (18)$$

Where $\varepsilon_k = z_k - \hat{z}_{k|k-1}$, $d_k = (1-c)/(1-c^k)$, with $0.95 < c < 0.99$

4 SYSTEM MODEL OF VISUAL-INERTIAL ODOMETRY

4.1 IMU data driven dynamic model

To use the IMU measurement for state propagation, we define the following IMU state vector,

$$X_I = [{}^I_W \overline{q}^T \quad {}^W p_I^T \quad {}^W v_I^T \quad b_g^T \quad b_a^T]^T \quad (19)$$

Where ${}^I_W \overline{q}^T$ is the unit quaternion describing the rotation from world frame {W} to IMU frame {I}, ${}^W p_I$ and ${}^W v_I$ are the position and velocity with respect to {I}, b_g and b_a are 3×1 vectors that describe the biases affecting the gyroscope and accelerometer measurements, respectively. Figure 1 shows the spatial relations between frames.

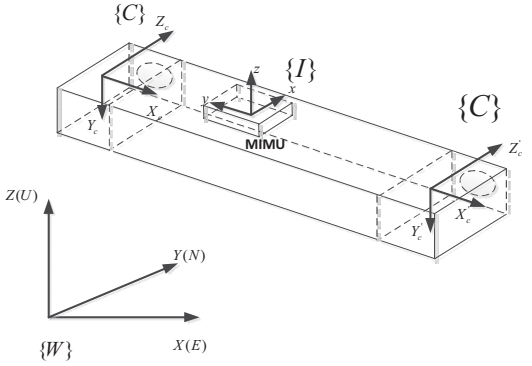

Fig 1. Coordinate frames.

The IMU (gyroscope and accelerometer) measurements are given by

$$\omega_m = \omega + b_g + n_g \tag{20}$$

$$a_m = a + b_a + n_a \tag{21}$$

Where n_g and n_a are zero-mean Gaussian noises. ω and a denote the IMU angular rate and linear acceleration respectively, b_g and b_a are measurement biases modeled as random walk processes.

The continuous time dynamic model of IMU is represented by the equations:

$$\begin{aligned} {}^{I}_{W}\dot{\bar{q}}(t) &= \frac{1}{2}\Omega(\omega)\,{}^{I}_{W}\bar{q} \qquad {}^{W}\dot{v}_I = C^T_{{}^{I}_{W}\bar{q}}a - g \\ {}^{W}\dot{p}_I &= {}^{W}v_I \qquad \dot{b}_g = n_{\omega g} \qquad \dot{b}_g = n_{\omega a} \end{aligned} \tag{22}$$

Where $C^T_{{}^{I}_{W}\bar{q}}$ is a rotational matrix described by ${}^{I}_{W}\bar{q}$. $\omega = [\omega_x \quad \omega_y \quad \omega_z]^T$ is the angular velocity expressed in IMU frame $\{I\}$. $\Omega(\omega) = \begin{bmatrix} -\lfloor \omega \times \rfloor & \omega \\ -\omega^T & 0 \end{bmatrix}$ is the quaternion kinematic matrix. The IMU biases are modelled as random walk process and driven by the Gaussian noise $n_{\omega g}$ and $n_{\omega a}$.

The IMU error state can be defined as:

$$\tilde{X}_I = [{}^{I}\tilde{\theta}^T \quad {}^{G}\tilde{p}^T \quad {}^{G}\tilde{v}^T \quad \tilde{b}_g^T \quad \tilde{b}_a^T]^T \tag{23}$$

Where, the standard error definition is defined (e.g. ${}^{W}\tilde{v} = {}^{W}v - {}^{W}\hat{v}$). The attitude error ${}^{I}\tilde{\theta}$ satisfies the following equation:

$${}^{I}_{G}R \cdot (I_3 - \lfloor {}^{I}\tilde{\theta} \times \rfloor)\,{}^{I}_{G}\hat{R} \tag{24}$$

Visual-inertial measurement model

In iterated cubature Kalman filter based alogorithm, the filter state includes the current IMU state X_{Ik} and a set of feature positions. The filter state is defined as

$$X_k = \begin{bmatrix} X_{Ik}^T & f_1^T & \mathrm{L} & f_{n_k}^T \end{bmatrix}^T \tag{25}$$

Where f_i, $i = 1,\mathrm{K}\,,n_k$ is the position of the *i-th* feature.

The measurement of the *i-th* feature at time step k is defined as

$$z_{i,k} = h({}^{C_k}p_{f_i}) + n_{i,k} \tag{26}$$

$${}^{C_k}p_{f_i} = {}^{C}_{I}RR_k({}^{W}p_{f_i} - {}^{W}p_{I_k}) + {}^{C}p_I \tag{27}$$

Where $\{{}^{C}_{I}\mathrm{R}, {}^{C}\mathrm{p}_I\}$ is the rotation and translation between the camera and the IMU.

5 SIMULATION AND EXPERIMENT RESULTS

In the simulation part, we compared the localization result using EKF and ICKF. In the real-world experiment part, we first describe the hardware used for the data acquisition and then introduce the implementation. Finally, we show the localization results.

5.1 Two-Dimensional Localization simulation

To verify the effectiveness of the proposed algorithm, the comparison of EKF and ICKF is introduced under the same conditions.

Assume a point robot is moving in two-dimensional plane. The state of the robot is $X = [x, y, \theta]$, including two-dimensional plane position coordinate $[x, y]$ and heading angle θ . Suppose the visual sensor can sense landmarks within a limited range and bearing.

We firstly generate the moving trajectory of the point robot, and then according to the trajectory, generating IMU data [17]. Finally, we get the localization results based on the method described in section 3 and 4.

As shown in Figure 2. , the red plus signs are visual landmarks. The red curve denotes the true path; the blue curve is the trajectory estimated by EKF; the green one is the path estimated by ICKF. As in Figure 3, the pose error of the point robot both are very small, but the pose error generated by ICKF is smaller than EKF. The root mean square error (RMSE) of pose are 0.395m and 0.2492m for EKF and ICKF respectively.

The simulation results proved the validity and effectiveness of ICKF.

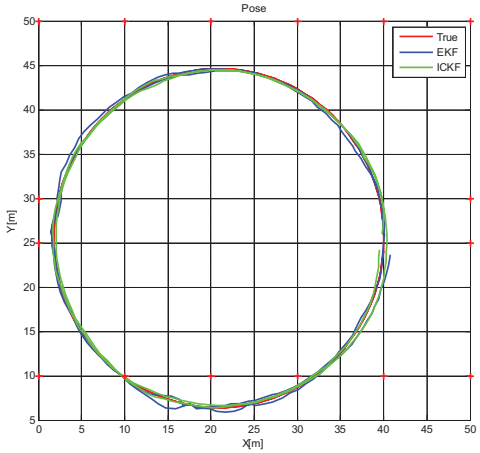

Fig 2. Two-dimensional localization result.

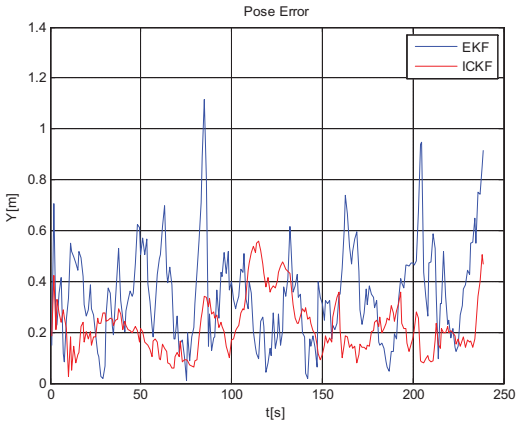

Fig 3. Two-dimensional localization pose error .

5.2 Real-world Experiment and Result

We collected a dataset on a street of Beijing. The movement trajectory is the red curve (see Figure 4).

Fig 4. The movement trajectory.

The experimental set-up is mainly composed of a ZED Stereo Camera, a Xsens MTi unit (MTi-G-710 GNSS/INS) and a data collecting computer. The Resolution of the ZED camera is 1920x1080; Sample Rates is 30FPS; Pixel value is 200MP; Data transmission interface is USB3.0; Base line is 12cm. For the gyro of MTi-G-710 GNSS/INS, the Bias repeatability (1 year) is 0.2[deg/s]; in-run bias stability is 10[deg/h]; Non-orthogonality is 0.05[deg]. For the accelerometer, the Bias repeatability (1 year) is 0.03[m/s2]; Non-orthogonality is 0.05[deg]. Figure 5 shows sensors used in the test and the installation mode.

Fig 5. M The hardware system used for data acquisition.

Our implementation consists of a front-end based on ORB-SLAM2 [17] and a back-end based on ICKF. ORB-SLAM2 is a complete SLAM system for monocular, stereo and RGB-D cameras. In our implementation, the Local Mapping and Loop Closing threads are deactivated.

Figure 6 shows the picture of the street, there are many moving cars in the scene, which will increase the number of outlier. The following localization result approved that ICKF has very good adaptability for Dynamic Scene.

Fig 6. Images of the street.

Figure 7 shows the trajectory estimates computed by three algorithms (ICKF-VIO, MSCKF, and OKVIS) on a map of the area where the vehicle drove.

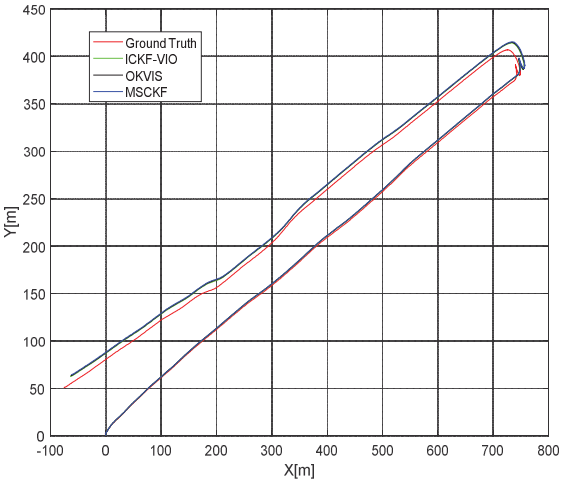

Fig 7. Comparison of the localization result reported by ICKF-VIO, MSCKF, and EKF-IMU.

MSCKF (Multi-State Constraint Kalman Filter) is a notable contribution in the area of filtering-based visual-inertial odometry, which is an EKF-based real-time fusion method [19]. OKVIS (Open Keyframe-based Visual-Inertial SLAM) is based on nonlinear optimization, which offer the advantage of repeated linearization of the inherently non-linear cost terms involved in the visual-inertial state estimation problem, but the optimization based method takes more computational resources.

The total distance traveled is approximately 1.8km. The average translational error of ICKF-VIO, OKVIS and MSCKF are 1.45%, 1.68% and 1.70% respectively. We see that the method proposed in this paper can provide more accurate motion estimation.

6 CONCLUSION

In this paper, a tightly coupled ICKF-based method is proposed for the integration of stereo visual odometry and inertial navigation system, which exploits the complementary characteristics of the visual and inertial sensor. The performance of this method is proved by two-dimensional simulation and real-world dataset experiment. Although there are many moving objects in the scene, our method can still provide accurate motion estimation. Furthermore, the results show that ICKF-VIO attains better accuracy than other two state-of-the-art methods.

REFERENCES

[1] Cheng Y, Maimone M, Matthies L. Visual odometry on the Mars Exploration Rovers[C]// IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2006:903-910 Vol. 1.

[2] Scaramuzza D, Fraundorfer F. Visual Odometry [Tutorial][J]. Robotics & Automation Magazine IEEE, 2011, 18(4):80-92.

[3] Huang G, Kaess M, Leonard J J. Towards consistent visual-inertial navigation[C]// IEEE International Conference on Robotics and Automation. IEEE, 2014:4926-4933.

[4] Jia C, Evans B L. Probabilistic 3-D motion estimation for rolling shutter video rectification from visual and inertial measurements[C]// IEEE, International Workshop on Multimedia Signal Processing. IEEE, 2012:203-208.

[5] Gui J, Gu D, Wang S, et al. A review of visual inertial odometry from filtering and optimisation perspectives[J]. Advanced Robotics, 2015, 29(20):1-13.

[6] Arasaratnam I, Haykin S. Cubature Kalman Filters[J]. IEEE Transactions on Automatic Control, 2009, 54(6):1254-1269.

[7] Sirtkaya S, Seymen B, Alatan A A. Loosely coupled Kalman filtering for fusion of Visual Odometry and inertial navigation[C]// International Conference on Information Fusion. IEEE, 2013:219 - 226.

[8] Hildebrandt M, Kirchner F. IMU-aided stereo visual odometry for ground-tracking AUV applications[C]// OCEANS 2010 IEEE - Sydney. IEEE, 2010:1-8.

[9] Kong X, Wu W, Zhang L, et al. Tightly-Coupled Stereo Visual-Inertial Navigation Using Point and Line Features[J]. Sensors, 2015, 15(6):12816-12833.

[10] Roumeliotis S I, Johnson A E, Montgomery J F. Augmenting inertial navigation with image-based motion estimation[C]// IEEE International Conference on Robotics and Automation, 2002. Proceedings. ICRA. IEEE, 2002:4326-4333 vol.4.

[11] Weiss S, Siegwart R. Real-time metric state estimation for modular vision-inertial systems[C]// IEEE International Conference on Robotics and Automation. IEEE, 2011:4531-4537.

[12] Weiss S, Achtelik M W, Lynen S, et al. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments[C]// IEEE International Conference on Robotics and Automation. IEEE, 2012:957-964.

[13] Li, Mingyang, Mourikis, Anastasios I. High-precision, consistent EKF-based visual–inertial odometry[J]. International Journal of Robotics Research, 2013, 32(6):690-711.

[14] Bloesch M, Omari S, Hutter M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2015:298-304.

[15] Omari, Sammy, et al. "Dense visual-inertial navigation system for mobile robots." IEEE International Conference on Robotics and Automation IEEE, 2015:2634-2640.

[16] Kelly J, Sukhatme G S. Visual-Inertial Sensor Fusion: Localization, Mapping and Sensor-to-Sensor Self-calibration[J]. International Journal of Robotics Research, 2011, 30(1):56-79.

[17] https://github.com/valentinp/PROBE/blob/03d4911f7357cfc2ab3a2d2ebe05fa29a8b6cef2/testing/simulation/genTrajectoryCircle.m

[18] Mur-Artal, Raul, and J. D. Tardos. "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras." IEEE Transactions on Robotics PP.99(2016):1-8.

[19] Li, Mingyang, Mourikis, Anastasios I. High-precision, consistent EKF-based visual–inertial odometry[J]. International Journal of Robotics Research, 2013, 32(6):690-711.

[20] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, & Paul Furgale. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. International Journal of Robotics Research, 34(3), 314-334.