2016 IEEE International Conference on Robotics and Biomimetics
December 3-7, 2016 Qingdao, China

北京航空航天大学 BEIHANG UNIVERSITY

Stereo-Inertial Pose Estimation and Online Sensors Extrinsic Calibration

Fumin Pang and Tianmiao Wang

School of Mechanical Engineering and Automation, Beihang University (BUAA)

Abstract

The fusion of visual and inertial measurement has been popular in mobile robotics community for decades due to the complementary properties of two sensors. In this paper ,we propose a method which fuses stereo visual and inertial cues based on Multi-State Constraint Kalman Filter (MSCKF), to estimate 6DOF pose of mobile robot. In addition, the proposed method includes extrinsic parameters in state vector to do online calibration. Experimental results on real-world datasets demonstrate that proposed method is consistent and substantially improves the accuracy of pose estimation, as well as the calibration between sensors.

Introduction

In this paper, we propose a stereo visual inertial navigation system. Stereo visual measurement can give a real metric perception of environment, which results in a better initialization of the estimator. A key contribution of proposed method is we give the derivation of stereo visual measurement model, which is different from the existing monocular one. In addition, to improve the performance, the system includes online extrinsic calibration between IMU and two cameras.

Summarized as follows:

- A stereo-inertial VIO based on Multi-State Constraint Kalman Filter (MSCKF).
- Online extrinsic calibration of sensors is included to improve the performance.
- The proposed method attains substantially higher accuracy than monocular MSCKF.

The frame system is defined as Fig.1 illustrated:

- $\{G\}$: Global frame
- $\{B\}$: Body frame fixed to IMU
- $\{C_0\}$: Camera frame CAM0
- $\{C_1\}$: Camera frame CAM1

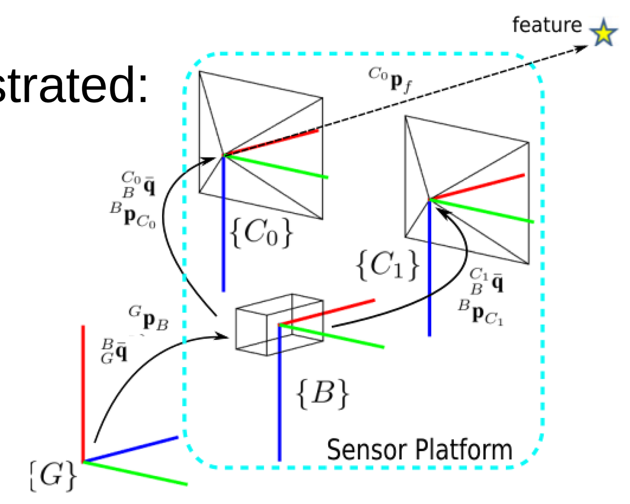

Fig.1 Coordinate frames involved in sensor platform and visual feature

A. MSCKF State Parametrization

The *error state*[1] of the estimator includes three part at time k:

$$\tilde{\mathbf{x}}_k := \Big[\underbrace{\tilde{\mathbf{x}}_{I,k}^T}_{IMU\ State} \quad \underbrace{\tilde{\mathbf{x}}_{Calib,k}^T}_{Calibrations} \quad \underbrace{\tilde{\mathbf{x}}_{B_0,k}^T \ \cdots \ \tilde{\mathbf{x}}_{B_{N-1},k}^T}_{Body\ States}\Big]^T$$

The IMU readings are used to propagate the IMU state , and the stereo visual measurements are used for update the filter , correcting the estimator means and covariance[2].

B. Measurement Model

In the update step the reprojection errors are calculated in a sliding window of images in which the feature is co-visible. Comparing to the monocular-MSCKF, the stereo reprojection errors are calculated:

$$\boldsymbol{r}_l^{(j)} = \begin{bmatrix} \boldsymbol{z}_{l,0}^{(j)} \\ \boldsymbol{z}_{l,1}^{(j)} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{z}}_{l,0}^{(j)} \\ \hat{\boldsymbol{z}}_{l,1}^{(j)} \end{bmatrix}$$

For the Kalman Filter update step , the measurement model is linearized with First-Order Approximation.:

$$\boldsymbol{r}_l^{(j)} = \boldsymbol{H}_{\boldsymbol{x}B_l}^{(j)} \tilde{\boldsymbol{x}} + \boldsymbol{H}_{\boldsymbol{f}_l}^{(j)} {}^G\tilde{\boldsymbol{p}}_{f_j} + \boldsymbol{n}_l^{(j)}$$

The related Jacobians are given by:

$$\boldsymbol{H}_{\boldsymbol{x}B_l}^{(j)} = \begin{bmatrix} \mathbf{0}_{4\times15} & \boldsymbol{\Pi}_l & \mathbf{0}_{4\times9(l-1)} & \boldsymbol{H}_{B_l} & \mathbf{0}_{4\times9(N-l)} \end{bmatrix}$$

Here we give more detail about the Jacobians with respect to *extrinsic calibration* parts and the *body states* parts:

$$\boldsymbol{\Pi}_l = \underbrace{\begin{bmatrix} \boldsymbol{\Pi}_{\boldsymbol{\theta}_{l,0}} & \boldsymbol{\Pi}_{\boldsymbol{p}_{l,0}} & \mathbf{0}_{2\times3} & \mathbf{0}_{2\times3} \\ \mathbf{0}_{2\times3} & \mathbf{0}_{2\times3} & \boldsymbol{\Pi}_{\boldsymbol{\theta}_{l,1}} & \boldsymbol{\Pi}_{\boldsymbol{p}_{l,1}} \end{bmatrix}}_{Jacobian\ w.r.t\ Calibrations} \qquad \boldsymbol{H}_{B_l} = \underbrace{\begin{bmatrix} \boldsymbol{H}_{\boldsymbol{\theta}_{l,0}} & \boldsymbol{H}_{\boldsymbol{p}_{l,0}} & \mathbf{0}_{2\times3} \\ \boldsymbol{H}_{\boldsymbol{\theta}_{l,1}} & \boldsymbol{H}_{\boldsymbol{p}_{l,1}} & \mathbf{0}_{2\times3} \end{bmatrix}}_{Jacobian\ w.r.t\ Body\ States}$$

$$\boldsymbol{\Pi}_{\boldsymbol{\theta}_{l,i}} = \boldsymbol{J}_l^{(j,i)}\, {}^{C_i}_{B}\hat{\mathbf{R}} \lfloor {}^{B}_{G}\hat{\mathbf{R}} ({}^G\boldsymbol{p}_{f_j} - {}^G\tilde{\boldsymbol{p}}_B)\times \rfloor \qquad \boldsymbol{H}_{\boldsymbol{\theta}_{l,i}} = \boldsymbol{J}_l^{(j,i)}\, {}^{C_i}_{B}\hat{\mathbf{R}}\, {}^{B}_{G}\hat{\mathbf{R}} \lfloor ({}^G\boldsymbol{p}_{f_j} - {}^G\tilde{\boldsymbol{p}}_B)\times \rfloor$$

$$\boldsymbol{\Pi}_{\boldsymbol{p}_{l,i}} = \boldsymbol{J}_l^{(j,i)} \qquad \boldsymbol{H}_{\boldsymbol{p}_{l,i}} = -\boldsymbol{J}_l^{(j,i)}\, {}^{C_i}_{B}\hat{\mathbf{R}}\, {}^{B}_{G}\hat{\mathbf{R}}$$

where J is the Jacobian matrix of the perspective model.

Following MSCKF pipeline, all the residuals of all measurements of the feature is stacked. In order to transform the linearized measurement equation into the standard form to update , we eliminate the Jacobian w.r.t the 3D point $H_{f_l}^{(j)}$.

Finally, we stack all the errors for all features selected for update and make a EKF update.

Experiments

We tested the proposed method and it is compared with the monocluar-MSCKF method on the *EuRoC dataset* recorded by ETH Autonomous Systems Lab .

In the front end, we extracted between 100 and 200 salient point features using Oriented FAST and Rotated BRIEF (ORB) detector of OpenCV from the stereo pairs and tracked them temporally using Kanade-Lucas-Tomasi (KLT) tracking.

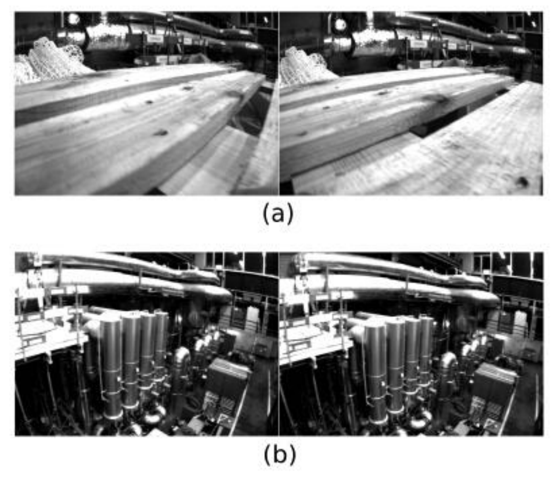

Fig.2 Sample images from dataset used in pose estimation experiment. (a) is the first pair of stereo images. (b) is the last pair.

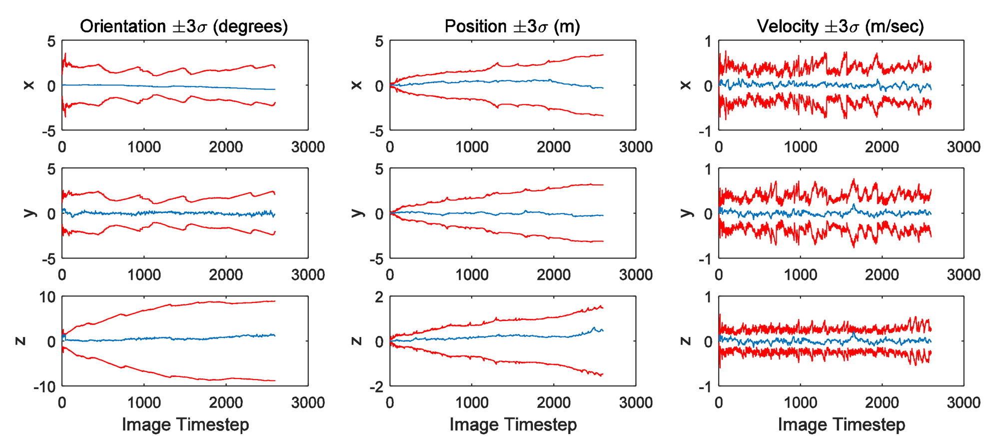

Fig.3 The orientation errors and $\pm3\boldsymbol{\sigma}$ bounds about axes x (roll), y (pitch), and z (yaw). Position and velocity errors and $\pm3\boldsymbol{\sigma}$ bounds about axes x, y, and z.

A. Pose Estimation

The results shown in Fig.3 show that the error remains bounded by $\pm3\boldsymbol{\sigma}$, meaning that the filter is consistent. And the proposed estimator has correct observability properties for camera/IMU system. The roll , pitch and velocities are observable, the error bounds do not increase indefinitely. In contrast, those for the position and yaw do continuously increase because they are not observable[3].

In Fig.4, the RMSE of the proposed method is generally smaller than that of m-MSCKF. The results show that the proposed method owns better precision than m-MSCKF method.

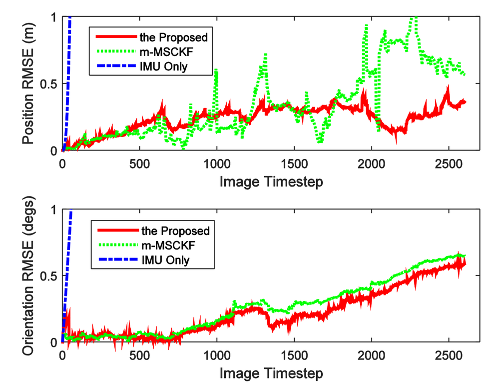

Fig.4 Comparison of Root Mean Squared Error (RMSE) of the proposed method, m-MSCKF and IMU-only integration.

B. Sensors Calibration

We added initial alignment error for translation and rotation to the known extrinsic calibration between sensors to validate the effectiveness of extrinsic calibration. Fig.5 shows the convergence of IMU-CAM0 extrinsic parameters. Table.1 demonstrate that the proposed calibration is effective in real-world environment.

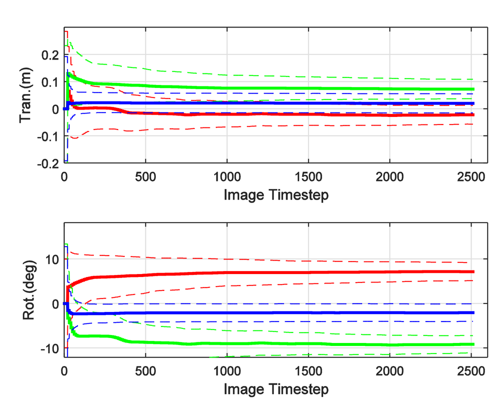

Fig.5 IMU-camera translation error and $\pm3\boldsymbol{\sigma}$bounds (Top). Translation along axes x(red), y(green), and z(blue). Rotation error and$\pm3\boldsymbol{\sigma}$bounds (bottom). Rotation about axes x (roll, red), y (pitch,green), and z (yaw,blue).

error		x(cm)	y(cm)	z(cm)	r(°)	p(°)	y(°)
CAM0	Initial	-3.10	7.20	-2.10	7.10	-9.50	-2.10
	Final	-0.77	0.26	0.31	0.12	-0.00	0.02
CAM1	Initial	-3.20	5.90	1.91	5.00	-4.00	-1.00
	Final	-0.53	0.22	0.29	0.12	-0.013	-0.01

Table.1 Final errors of the IMU-Cameras parameters after 2500 image timestamp

Contact

Fumin Pang
Beihang University (BUAA)
Email: fuminpang@buaa.edu.cn

Tianmiao Wang
Beihang University (BUAA)
Email: itm@buaa.edu.cn

References

[1] Roumeliotis S I, Sukhatme G S, Bekey G A. Circumventing dynamic modeling: Evaluation of the error-state kalman filter applied to mobile robot localization[C]//Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on. IEEE, 1999, 2: 1656-1663.
[2] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE, 2007: 3565-3572.
[3] Kelly J, Sukhatme G S. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration[J]. The International Journal of Robotics Research, 2011, 30(1): 56-79.