A Performance Analysis of Pose Estimation Based on Two-View Tracking and Multi-State Constraint Kalman Filter Fusion

Tae Ihn Kim, Jae Hyung Jung, Chan Gook Park
Department of Mechanical and Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea
E-mail: natigerya@snu.ac.kr, lastflowers@snu.ac.kr, chanpark@snu.ac.kr

Abstract

This paper presents a performance analysis of two-view tracking and Multi-State Constraint Kalman Filter (MSCKF) fusion for a pose estimation. The system and measurement model of both two-view tracking and MSCKF are derived based on the fusion condition. The simulation result of the fused algorithm using the Drone Racing dataset, collected from an aggressive flight of micro aerial vehicle (MAV), shows the performance improvement of both attitude and position estimation compared to the performance of MSCKF.

Keywords: Visual-Inertial Odometry, Two-View Tracking, Multi-State Constraint Kalman Filter, Sensor Fusion

1. Introduction

Visual-Inertial Odometry (VIO) is an algorithm for a pose estimation using images from a camera sensor and linear accelerations and angular velocities from an Inertial Measurement Unit (IMU). The pose estimation algorithms using each sensor are known as Visual Odometry (VO) and Inertial Navigation System (INS). Both VO and INS suffer from drift over time because the pose is estimated incrementally in both algorithms.[1] While the pose estimation of VO results in high precision in a slow motion, that of INS results in high accuracy in a rapid motion. Therefore, VIO, which is the fusion of VO and INS, complements the strengths of each sensor and improves the performance of pose estimation.

Among various VIO algorithms, Multi-State Constraint Kalman Filter (MSKCF) is chosen as the main algorithm. Different from other general VIO algorithms such as Extended Kalman Filter (EKF) based algorithms, MSCKF does not include three-dimensional feature positions in the filter states, which result in a drop of computation complexity. MSCKF uses a geometric constraint obtained from the poses of previous camera frames included in a sliding window of filter state to estimate the pose.[2] However, the performance of MSCKF pose estimation is comparatively inaccurate in a rapid motion. To overcome this weakness, a fused algorithm of two-view tracking and MSCKF is proposed. Two-view tracking uses an optical flow measurement obtained from consecutive image frames in the measurement update. The strength of two-view tracking is that the pose estimation is comparatively accurate even in a rapid motion. Therefore, the fusion of two algorithms compensates for the weakness of MSCKF.

This paper presents the performance analysis of two-view tracking and MSCKF fusion for pose estimation throughout an application to the Drone Racing dataset, which images, linear accelerations and angular velocities are collected in a rapid motion.[3]

2. System Model

The error state of the fused algorithm is a combination of two-view tracking and MSCKF error state. The error state of two-view tracking is described as Eq.(1).[4]

$$\tilde{X}_{TV} = \begin{bmatrix} \tilde{p}_{GB}^{G}{}^{T} & \tilde{v}_{GB}^{B}{}^{T} & \tilde{\theta}_{GB}^{T} & \tilde{b}_{a}{}^{T} & \tilde{b}_{g}{}^{T} & \tilde{\alpha}^{T} \end{bmatrix}^{T} \quad (1)$$

where $\tilde{p}_{GB}^G$ is a position error expressed in a global-frame, $\tilde{v}_{GB}^B$ is a velocity error expressed in a body-frame, $\tilde{\theta}_{GB}$ is an attitude error expressed in a global-frame, $\tilde{b}_a$ is an accelerometer bias error, $\tilde{b}_g$ is a gyroscope bias error and $\tilde{\alpha}$ is an inverse-scene-depth error. The error state of MSCKF system model is described as Eq.(2).[5]

$$\tilde{X}_{MSCKF} = \begin{bmatrix} \tilde{X}_{IMU}{}^T & \tilde{X}_{slw}{}^T \end{bmatrix}^T \tag{2}$$

where

$$\tilde{X}_{IMU} = \begin{bmatrix} \tilde{p}_{GB}^G{}^T & \tilde{v}_{GB}^G{}^T & \tilde{\theta}_{GB}{}^T & \tilde{b}_a{}^T & \tilde{b}_g{}^T \end{bmatrix}^T \tag{3}$$

$$\tilde{X}_{slw} = \begin{bmatrix} \tilde{p}_{GC_1}^G{}^T & \tilde{\theta}_{GC_1}{}^T \vdots & \dots & \vdots & \tilde{p}_{GC_N}^G{}^T & \tilde{\theta}_{GC_N}{}^T \end{bmatrix}^T \tag{4}$$

$\tilde{X}_{IMU}$ is IMU error state, described as Eq.(3), $\tilde{X}_{slw}$ is temporarily included previous camera pose error state in the sliding window, described as Eq.(4), $\tilde{v}_{GB}^G$ is a velocity error expressed in a global-frame, $\tilde{p}_{GC_N}^G$ and $\tilde{\theta}_{GC_N}$ are N-th camera position error and attitude error in the sliding window. The error state of the fused algorithm is described as Eq.(5). For the convenience of Jacobian derivation, $\tilde{v}_{GB}^B$ is chosen instead of $\tilde{v}_{GB}^G$.

$$\tilde{X}_{fused} = \begin{bmatrix} \tilde{X}_{TV}{}^T & \tilde{X}_{slw}{}^T \end{bmatrix}^T \tag{5}$$

The system model is described as Eq.(6).

$$\dot{\tilde{p}}_{GB}^G = \hat{R}_B^G \tilde{v}_{GB}^B - \lfloor \hat{R}_B^G \hat{v}_{GB}^B \times \rfloor \tilde{\theta}_{GB}$$
$$\dot{\tilde{v}}_{GB}^B = \hat{R}_G^B \lfloor g^G \times \rfloor \tilde{\theta}_{GB} - \lfloor (\omega_m - \hat{b}_g) \times \rfloor \tilde{v}_{GB}^B$$
$$\quad -\tilde{b}_a - \lfloor \hat{v}_{GB}^B \times \rfloor \tilde{b}_g - n_a - \lfloor \hat{v}_{GB}^B \times \rfloor n_g$$
$$\dot{\tilde{\theta}}_{GB} = -\hat{R}_B^G \tilde{b}_g - \hat{R}_B^G \tilde{n}_g$$
$$\dot{\tilde{b}}_a = n_{wa}$$
$$\dot{\tilde{b}}_g = n_{wg}$$
$$\dot{\tilde{\alpha}} = n_\alpha \tag{6}$$

where n_a and n_g are zero-mean, white Gaussian noise of accelerometer and gyroscope, respectively, and n_{wa} and n_{wg} are random walk rate of accelerometer and gyroscope, respectively.

3. Measurement Model

The measurement model of the fused algorithm is divided into two-view tracking and MSCKF measurement model. The measurement update related term of two-view tracking is described as Eq.(7).

$$\tilde{y}_i = M_i \{ \hat{\alpha} R_B^C \tilde{v}_{GB}^B + (\lfloor \bar{P}_{Cf_i}^C \times \rfloor R_B^C + \hat{\alpha} R_B^C \lfloor P_{BC}^B \times \rfloor) \tilde{b}_g$$
$$R_B^C (\hat{v}_{GB}^B + \lfloor (\omega_m - \hat{b}_g) \times \rfloor P_{BC}^B) \tilde{\alpha} \} + \tilde{n}_{v_i} \tag{7}$$

where, $\tilde{y}_i$ is a two-dimensional innovation term, M_i is a nullspace reprojection matrix and ω_m is a gyroscope measurement. The measurement update related term of MSCKF is described as Eq.(8).[5]

$$r_i^{(j)} = z_i^{(j)} - \hat{z}_i^{(j)} \cong H_{X_i}^{(j)} \tilde{X} + H_{f_i}^{(j)} \tilde{p}_{Gf_i}^G + n_i^{(j)} \tag{8}$$

$$r^{(j)} \cong H_X^{(j)} \tilde{X} + H_f^{(j)} \tilde{p}_{Gf}^G + n^{(j)} \tag{9}$$

$$r_0^{(j)} = A^T (z^{(j)} - \hat{z}^{(j)}) \cong A^T H_X^{(j)} \tilde{X}^{(j)} + A^T n^{(j)} \tag{10}$$

$$= H_0^{(j)} \tilde{X}^{(j)} + n_0^{(j)} \tag{11}$$

$$r_0 = H_X \tilde{X} + n_0 \tag{12}$$

where $r_i^{(j)}$ is a measurement residual, $z_i^{(j)}$ is a measurement, $\hat{z}_i^{(j)}$ is a reference, $H_{X_i}^{(j)}$ and $H_{f_i}^{(j)}$ are state and feature position Jacobians of $z_i^{(j)}$, respectively, $\tilde{X}$ is the error state, $\tilde{p}_{Gf_i}^G$ is a feature position error and $n_i^{(j)}$ is a noise vector. Residuals for each previous camera frames in Eq.(8) are stacked up to form Eq.(9) and reprojected on the left nullspace to form Eq.(10) and Eq.(11). Residuals for each feature are finally stacked up and H_X is used in the EKF update.

The keypoint of the fused algorithm is that both two-view tracking and MSCKF measurement updates proceed during the pose estimation. MSCKF proceeds the measurement update when one of the two conditions satisfies. Those two conditions are a failure of feature tracking and an excess of the size of the sliding window. However, the minimum number of tracks is also assigned in order to avoid an error in residuals. The fused algorithm follows the same measurement trigger as MSCKF, but two-view tracking measurement update proceeds when there are exactly two tracks in the sliding window. For example, as shown in Fig. 1., when N equals two, the optical flow is measured from C_N and C_{N-1} camera frames included in the sliding window.
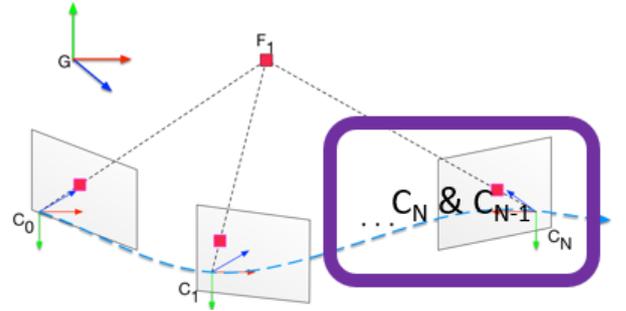
Fig. 1. Two-View Tracking Frames in the Sliding Window

4. Performance Analysis

The performance analysis of the fused algorithm is proceeded using the Drone Racing dataset, which provides images, accelerometer and gyroscope measurements and ground truth collected using a laser tracking system.[3] The total distance traveled by MAV in the dataset is 270.7448 m and the maximum

instantaneous velocity is 25.0499 m/s. Other drone flight datasets such as EuRoC and Zurich Urban MAV dataset exist, but the Drone Racing is chosen because only this dataset provides measurements collected by an aggressive drone flight. The performance difference between MSCKF and the fused algorithm is expected to be clearly observable in bad condition in terms of the visual environment.

For the performance analysis, the altitude error and the position error of the fused algorithm are compared to those of MSCKF. The result is shown in Table 1, Fig.2 and Fig.3. The attitude error has improved from 0.2670 rad to 0.2306 rad, and the position error has improved from 6.9418 m to 6.4576 m, which are 13.6 % and 7.0 % improvement, respectively.

Table 1. Root Mean Square Error (RMSE)

	MSCKF	Fused
Attitude [rad]	0.2670	0.2306
position [m]	6.9418	6.4576

The improvement of both attitude and position error is a result of the difference in the EKF update. In MSCKF, the measurement update does not proceed when the number of tracks in the sliding window is less than three. However, in the fused algorithm, the measurement update proceeds even when there are two tracks in the sliding window and two-view measurement update proceeds instead of MSCKF measurement update. Compared to other drone flight datasets, the Drone Racing dataset provides lower sized images, which fewer features are detected. MSCKF measurement update using a few features results in low accuracy. However, in the same condition, two-view tracking measurement update results in more accurate pose estimation compared to that of MSCKF. Therefore, the additional two-view tracking measurement update in MSCKF measurement update improves the performance of pose estimation of MAV in an aggressive flight condition.
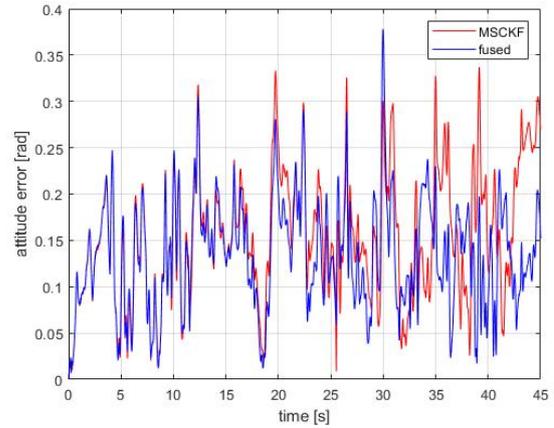
Fig. 2. Three-dimensional Attitude Error of Drone Racing
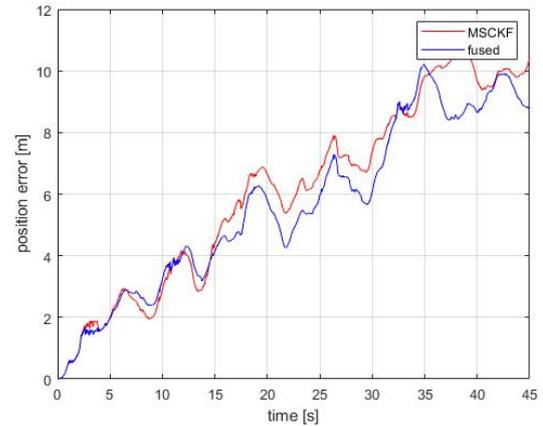
Fig. 3. Three-dimensional Position Error of Drone Racing

5. Conclusion

This paper presents the performance analysis of the fusion of two-view tracking and MSCKF for the pose estimation. The error state of the fused algorithm is selected to be the combination of two-view tracking and MSCKF error state. The system model and the measurement update related terms are newly derived since the velocity error is expressed in a body-frame. MSCKF measurement update proceeds when one of two conditions is triggered and two-view tracking measurement update proceeds when there are two tracks in the sliding window. The fused algorithm results in an improvement of the pose estimation.

Tae Ihn Kim, Jae Hyung Jung, Chan Gook Park

Acknowledgements

References

1. J. H. Jung, S. Heo and C. G. Park, Patch-based Stereo Direct Visual Odometery Robust to Illumintation Changes, *International Journal of Control, Automation and System*, 17(3), (2019), 743-751.
2. P. Kim, H. Lim and H. J. Kim, Visual Inertial Odometry with Pentafocal Geometric Constraints, *International Journal of Control, Automation and System*, 16(4), (2018), 1962-1970.
3. M. Bloesch, S. Omari, P. Fankhauser, H. Sommer, C. Gehring, J. Hwangbo, M. A. Hoepflinger, M. Hutter and R. Siegwart, Fusion of Optical Flow and Inertial Measurements for Robust Egomotion Estimation, *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2014), 3102-3107.
4. J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler and D. Scaramuzza, Are We Ready for Autonomous Drone Racing? The UZH-FPV Drone Racing Dataset, *IEEE International Conference on Robotics and Automation*, (2019).
5. A. I. Mourikis and S. I. Roumeliotis, A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation, *Proceedings 2007 IEEE International Conference on Robotics and Automation* (2007), 3565-3572.