Image Gradient-based Monocular Visual-Inertial Odometry

Tae Ihn Kim
Robotics Lab, Hyundai Motor Company,
Uiwang-si, Gyeonggi-do, 16082, Republic of Korea

Jae Hyung Jung
Dept. of Aerospace Engineering, Seoul National University / ASRI
Seoul, 08826, Republic of Korea

Chan Gook Park
Dept. of Aerospace Engineering, Seoul National University / ASRI
Seoul, 08826, Republic of Korea
E-mail: natigerya@snu.ac.kr, lastflowers@snu.ac.kr, chanpark@snu.ac.kr

Abstract

This paper presents image gradient-based monocular visual-inertial odometry (VIO) algorithm, using image gradient measurements, robust to illumination change. We expand the measurements from the reprojected feature locations on the image coordinates to the corresponding image gradients. The iterated EKF and low-pass pyramid are adapted to reduce the linearization error in the multi-state constraint Kalman filter (MSCKF) measurement update process. We verify that our proposed algorithm outperforms both conventional indirect and direct MSCKF-based VIO algorithms by evaluating the pose estimation performance.

Keywords: Illumination Change, Image Gradient, Iterated Extended Kalman Filter, Low-pass Pyramid, Visual-Inertial Odometry

1. Introduction

An accurate ego-motion estimation of a vehicle in a global navigation satellite system (GNSS) denied environment has been a challenging task in robotics and autonomous driving for several decades [1]. To tackle this issue, the ego-motion estimation algorithm using camera information, so-called Visual Odometry (VO), has been studied [2], [3], [4]. VO incrementally estimates a relative 6-DOF pose between consecutive images.

The most representative fusion of camera and alternative sensors is visual-inertial odometry (VIO), which additionally uses IMU; accelerometers, and gyroscopes. Despite various VIO algorithm studies, several challenging environments exist for ego-motion estimation, still being tackled. One representative environment is where the photometric consistency assumption is violated due to the illumination change.

The main contribution of this paper is threefold. First, inspired by the studies on image gradient-based VO algorithms [5]. we adapted the novel measurement, image gradient, to the VIO framework. Image gradient is known to be robust to illumination change. We utilize image gradients in the horizontal and vertical image directions as measurements of the multi-state constraint Kalman filter (MSCKF)-based VIO algorithm. Second, unlike the conventional MSCKF-based VIO algorithm, we apply iterated extended Kalman filter (EKF) and low-pass pyramid to minimize linearization error. Lastly, we evaluate our proposed algorithm at both simulation and real-world datasets and demonstrate that our proposed algorithm outperforms both conventional indirect and direct methods in the environment where illumination changes.

2. Preliminaries

In this paper, the 3D pose is described at the current IMU-affixed frame, $\{B_k\}$, at time k with respect to a global frame, $\{G\}$. The global frame is the initial body frame aligned with gravity. Unlike EKF, MSCKF considers multiple images and utilizes the geometric constraint to estimate the ego-motion of the vehicle. Therefore, MSCKF includes the previous camera poses in the state. $\{C_i\}$ is a camera frame of the ith element in the sliding window of MSCKF.

The mathematical expression is described as follows. For example, $p^{C_l}_{C_l f_i}$ is the position of camera frame described as the translation from the lth camera frame, C_l to the ith 3D feature position, f_i with respect to C_l.

3. State Representation

The MSCKF state, X_{MSCKF} is formulated as a combination vector of the IMU state, X_{IMU}, and the previous camera state, X_{cam}, in the sliding window.

$$X_{MSCKF} = \begin{bmatrix} X_{IMU}{}^T & X_{camera}{}^T \end{bmatrix}^T \in \mathbb{R}^{(16+7N)\times 1} \tag{1}$$

$$X_{IMU} = \begin{bmatrix} q_{GB}{}^T & p_{GB}^{G}{}^T & v_{GB}^{G}{}^T & b_a{}^T & b_g{}^T \end{bmatrix}^T \in \mathbb{R}^{16\times 1} \tag{2}$$

$$X_{camera} = \begin{bmatrix} q_{GC_1}{}^T & p_{GC_1}^{G}{}^T & \cdots & q_{GC_N}{}^T & p_{GC_n}^{G}{}^T \end{bmatrix}^T \in \mathbb{R}^{7N\times 1} \tag{3}$$

q_{GB} is the unit quaternion describing the rotation from body frame $\{B\}$ to the global frame $\{G\}$ (or the body attitude with respect to $\{G\}$, p_{GB}^G and v_{GB}^G are the body position and velocity with respect to $\{G\}$, respectively, b_a and b_g are accelerometer and gyroscope bias, respectively, and q_{GC_l} and $p_{GC_l}^G$ are the l th camera attitude and position, respectively.
True position, velocity, accelerometer and gyroscope bias are expressed as $X = \hat{X} + \tilde{X}$, where $\hat{X}$ and $\tilde{X}$ are estimated and error state, respectively, and true attitude as $q = \delta q \otimes \hat{q}$, where $\otimes$ denotes quaternion multiplication. Then, the quaternion error is derived as follows.

$$\delta q = q \otimes \hat{q}^{-1} \approx \begin{bmatrix} 1 \\ \frac{1}{2}\delta\theta \end{bmatrix} \tag{4}$$

Therefore, the error state is defined as below.

$$\tilde{X}_{MSCKF} = \begin{bmatrix} \tilde{X}_{IMU}^T & \tilde{X}_{cam}^T \end{bmatrix}^T \in \mathbb{R}^{(15+6N)\times 1} \tag{5}$$

$$\tilde{X}_{IMU} = \begin{bmatrix} \delta\theta_B{}^{G^T} & \tilde{p}_{GB}^{G}{}^T & \tilde{v}_{GB}^{G}{}^T & \tilde{b}_a{}^T & \tilde{b}_g{}^T \end{bmatrix} \in \mathbb{R}^{15\times 1} \tag{6}$$

$$\tilde{X}_{cam} = \begin{bmatrix} \delta\theta_{C_1}{}^{G^T} & \tilde{p}_{GC_1}^{G}{}^T & \cdots & \delta\theta_{C_N}{}^{G^T} & \tilde{p}_{GC_N}^{G}{}^T \end{bmatrix} \in \mathbb{R}^{6N\times 1} \tag{7}$$

4. Measurement Model

In our proposed algorithm, the measurement of the conventional MSCKF, which is the reprojected feature location on the image coordinate, is extended to the image gradients along the horizontal and vertical direction of the image, $z_{gradient}$,

$$z_{gradient,i,l} = \begin{bmatrix} \nabla_x I_{i,l}(\pi(p_{C_l f_i}^{C_l})) \\ \nabla_y I_{i,l}(\pi(p_{C_l f_i}^{C_l})) \end{bmatrix} + n_{gradient,i,l} \tag{8}$$

where $p_{C_l f_i}^{C_l} = h(p_{Gf_i}^G, \theta_{C_l}^G, p_{GC_l}^G)$. $z_{gradient,i,l}$ is the measurement of the ith feature on the lth image. $\pi(\cdot)$ is a reprojected feature location, $I_{i,l}(\pi(\cdot))$ is an intensity at the reprojected feature location, $\nabla_x I_{i,l}(\pi(\cdot))$ and $\nabla_y I_{i,l}(\pi(\cdot))$ are gradients along horizontal and vertical directions of image. $n_{gradient,i,l}$ is the noise of the ith feature on the lth image.

5. Measurement Update

The measurement update of our proposed algorithm proceeds when either a feature is lost from tracking or the length of the feature track exceeds the size of the sliding window, which are the same measurement update conditions of the conventional MSCKF-based VIO algorithm [6], [7]. Since the novel measurement, image gradients along the horizontal and vertical direction of the image are utilized, the residual is expanded to the difference between the anchor and the lth image gradient along two directions as,

$$\begin{aligned} r_{i,l} &= z_{i,l} - \hat{z}_{i,l} \\ &= \begin{bmatrix} \nabla_x I_{i,A} - \nabla_x I_{i,l}(\pi(\hat{p}_{C_l f_i}^{C_l})) \\ \nabla_y I_{i,A} - \nabla_y I_{i,l}(\pi(\hat{p}_{C_l f_i}^{C_l})) \end{bmatrix} \\ &\approx H_{X_{i,l}} \tilde{X}_k + H_{f_{i,l}} \tilde{p}_{Gf_i}^G + n_{i,l} \end{aligned} \tag{9}$$

where $\hat{p}_{C_l f_i}^{C_l} = h(\hat{p}_{Gf_i}^G, \hat{\theta}_{C_l}^G, \hat{p}_{GC_l}^G)$. $\nabla_x I_{i,A}$ and $\nabla_y I_{i,A}$ are the image gradients along the horizontal and vertical direction of the anchor image. The anchor frame is defined as the camera image where the ith feature is captured in the sliding window. $r_{i,l}$, $z_{i,l}$ and $\hat{z}_{i,l}$ are the residuals, measured gradients and estimated gradients of the ith feature on the lth image, respectively. $\mathrm{H}_{\mathrm{X}_{i,l}}$ and $\mathrm{H}_{\mathrm{f}_i}$ are state and feature position Jacobian matrix, respectively. $\tilde{X}_k$ is the error state at step k and $\tilde{p}_{Gf_i}^G$ is the error feature position.
The residuals of the ith feature on all image frames are accumulated to form a block vector of residuals of the ith feature. Since the measurement is extended from the image feature projection location on the image coordinates to image gradients, Jacobian matrices of state, $H_{X_{C_l}}$, and feature position, $H_{p_{Gf_i}^G}$, are newly derived as below.

$$\begin{aligned} H_{X_{C_l}} &= \frac{\nabla^2 I_l}{\partial(u_d, v_d)} \frac{\partial(u_d, v_d)}{\partial p_{C_l f_i}^{C_l}} \frac{\partial p_{C_l f_i}^{C_l}}{\partial X_{C_l}}, \\ H_{p_{Gf_i}^G} &= \frac{\nabla^2 I_l}{\partial(u_d, v_d)} \frac{\partial(u_d, v_d)}{\partial p_{Gf_i}^G}, \\ \frac{\nabla^2 I_l}{\partial(u_d, v_d)} &= \begin{bmatrix} \frac{\nabla_x(\nabla_x I_l)}{\partial(u_d)} & \frac{\nabla_y(\nabla_x I_l)}{\partial(v_d)} \\ \frac{\nabla_x(\nabla_y I_l)}{\partial(u_d)} & \frac{\nabla_y(\nabla_y I_l)}{\partial(v_d)} \end{bmatrix} \end{aligned} \tag{10}$$

∇_x and ∇_y are gradients along horizontal and vertical direction of the image, respectively. u_d and v_d are undistorted image pixels along horizontal and vertical direction of image, respectively. X_{C_l} is the lth camera pose. $\frac{\nabla^2 I_l}{\partial(u_d, v_d)}$ implies the gradients of the gradient image.
Then, the residual is projected to a left null-space to eliminate the terms related to the 3D feature position. The block vectors of the residuals of the ith feature are accumulated to form a block vector of all residuals and proceed QR decomposition to reduce computational complexity.
Unlike EKF, which utilizes prior estimate as a linearization point, iterated EKF iteratively computes posterior estimate, which is more accurate since the measurement is reflected. The MSCKF measurement update equation is summarized as follows.

$$\begin{aligned} \delta X_k^+ &= K_k(y_k - h(\hat{X}_k^-)) \\ K_k &= P_k^- T_{H_k}{}^T (T_{H_k} P_k^- T_{H_k}{}^T + Q_{1_k}{}^T R_{0_k} Q_{1_k})^{-1} \\ r_{n_k} &= y_k - h(\hat{X}_k^-) = Q_{1_k}{}^T r_{0_k} \end{aligned} \tag{11}$$

δX_k^+ is a posterior estimated state, which is the correction term with respect to the error state from $\hat{X}_k^+ = \hat{X}_k^- \boxplus \delta X_k^+$. $\boxplus$ is the operator defining the error state as the estimated error state and the correction term. $h(\hat{X}_k^-)$ is the measurement equation of the system, K_k is the

Kalman gain, and P_k^- is the prior covariance of the estimated error. T_{H_k} is the upper triangular matrix computed from the QR decomposition. Q_{1_k} is the unitary matrix computed from the QR decomposition, whose columns form bases for the range of H_X . y_k is the measurement, and R_{0_k} is the covariance matrix of the measurement.

The EKF-based MSCKF uses the prior estimate state as the linearization point. Therefore, the measurement is not reflected on the linearization point. However, the iterated EKF-based MSCKF first computes the posterior estimate state using the prior estimate state to reflect the measurement on the linearization point and iteratively computes the posterior estimate state to reduce the linearization point. Therefore, the adaptation of the iterated EKF to MSCKF is made through modification of equation 11 to

$$\begin{aligned}
\delta X_{k,i+1}^{+} &= K_{k,i}(y_k - h(\hat{X}_{k,i}^{+}) + T_{H_{k,i}}(\delta X_{k,i}^{+})) \\
K_{k,i} &= P_k^- {T_{H_{k,i}}}^T (T_{H_{k,i}} P_k^- {T_{H_{k,i}}}^T + {Q_{1_{k,i}}}^T R_{0_{k,i}} Q_{1_{k,i}})^{-1} \\
r_{n_{k,i}} &= y_k - h(\hat{X}_{k,i}^{+}) + T_{H_{k,i}}(\delta X_{k,i}^{+}) = {Q_{1_k}}^T r_{0_k} + T_{H_{k,i}}(\delta X_{k,i}^{+})
\end{aligned} \tag{12}$$

$\delta X_{k,i+1}^{+}$ is the correction term at the $i+1$th iteration and $h(\hat{X}_{k,i}^{+})$ is the measurement equation of system, where posterior is reflected. $K_{k,i}$, $T_{H_{k,i}}$, and $Q_{1_{k,i}}$ are the Kalman gain, the upper triangular matrix, and the unitary matrix, respectively, at the ith iteration. The combination of the iterated EKF and the low-pass pyramid results in a reduction of the residual, which implies the improvement of linearization error due to the high non-linearity of the image gradients.

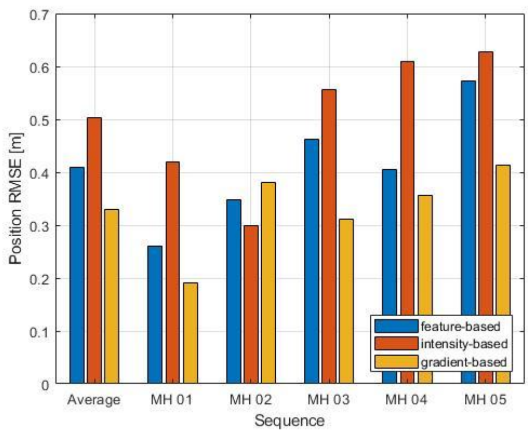

Fig. 1. Attitude and position RMSE

6. Experimental Results

The proposed algorithm is evaluated at a real-world dataset to verify that it is more robust to the illumination change than the feature-based and intensity-based methods. EuRoC real-world dataset is collected from an on-board hex-rotor helicopter. This dataset provides stereo images, IMU data, and 6-DOF ground-truth. Illumination change due to the camera's automatic exposure time is included in the image. The performance of our proposed algorithm is analyzed on five different sequences collected in the same indoor environment, and MAV travels 94.86 m in 134.8 s on average.

The experiment result is summarized in Fig. 1. It reports that the average 3D position RMSE of the feature-based method is less than that of the intensity-based method. This result implies that the illumination change strongly influences the image intensity, while the reprojected feature position is less influenced. The feature-based method is less influenced because the extracted feature position's accuracy depends on motion-blur, not the illumination change. Average 3D position RMSE of the intensity-based method is reduced by utilizing the image gradient

7. Conclusion

In this paper, we have proposed the utilization of image gradients as measurements in the monocular MSCKF-based VIO algorithm with an adaptation of iterated EKF with the low-pass pyramid. The novel measurement is used to tackle the illumination change environment, where the photometric-consistency assumption is violated. Unlike the conventional method, which utilizes the image feature coordinates, our proposed algorithm utilizes the image gradients along the horizontal and vertical direction of the image with the adaptation of iterated EKF and low-pass pyramid. The performance of the proposed algorithm is analyzed by comparing it to that of feature-based and intensity-based methods at the real-world dataset. Our experimental result reports that our proposed algorithm's accuracy of both 3D attitude and position estimation outperforms both feature-based and intensity-based algorithms.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT, the Republic of Korea. (NRF-2022R1A2C2012166)

References

1. M. Tang, Z. Chen, "Robot Tracking in SLAM with Masreliez-Martin Unscented Kalman Filter", International Journal of Control, Automation and Systems, vol. 18, no. 9, pp. 2315-2325, 2020.
2. F. Zhang, D. Clarke and A. Knoll, "Visual Odometry based on a Bernoulli Filter", International Journal of Control, Automation and Systems, vol. 13, no. 3, pp. 530-538, 2015.
3. W. Jeong, J. Moon, and B. Lee, "Error improvement in visual odometry using super-resolution", International Journal of Control, Automation and Systems, vol. 18, no. 2, pp. 322-329, 2020.
4. L. Wang, S. Xia, H. Xi, S. Li, and L. Wang, "Robust Visual Inertial Odometry Estimation Based on Adaptive Interactive Multiple Model Algorithm", International Journal of Control, Automation and Systems, vol. 20, no. 10, pp. 3335-3346, 2022.
5. J. Quenzel, R. A. Rosu, T. Labe, C. Stachniss and S. Behnke, "Beyond Photometric Consistency: Gradient-based Dissimilarity for Improving Visual Odometry and Stereo Matching", 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 272-278, 2020.

6. M. Li and A. I. Mourikis, “High-precision, Consistent EKF-based Visual-Inertial Odometry”, The International Journal of Robotics Research, vol. 32, no. 6, pp. 690-711, May 2013.
7. A. I. Mourikis and S. I. Roumeliotis, “A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation”, 2007 IEEE International Conference on Robotics and Automation (ICRA), pp. 3565-3572, 2007.

Authors Introduction

Mr. Tae Ihn Kim

He received the M.S. degree in the Department of Aerospace Engineering from Seoul National University, Republic of Korea, in 2021. He received the B.S. degree in Mechanical and Aerospace Engineering from Seoul National University, Republic of Korea, in 2019. His research interests include Visual Odometry and Visual-Inertial Odometry.

Mr. Jae Hyung Jung

He received the M.S., and Ph.D. degrees in the Department of Aerospace Engineering from Seoul National University, Republic of Korea. He received the B.S. degree in the Department of Aerospace Engineering from Pusan National University, Republic of Korea, in 2017. His research interests include visual odometry and vision-aided inertial navigation for mobile robots.

Dr. Chan Gook Park

He received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from the Seoul National University, Seoul, South Korea, in 1985, 1987, and 1993, respectively. He was as a Post-Doctoral Fellow with Prof. J. L. Speyer about peak seeking control for formation flight at the University of California at Los Angeles, Los Angeles, CA, USA, in 1998. From 1994 to 2003, he was an Associate Professor with Kwangwoon University, Seoul. In 2003, he joined the Faculty with the School of Mechanical and Aerospace Engineering, National University, Seoul, where he is currently a professor. In 2009, He was a Visiting Scholar with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. He served as a chair of IEEE AES Korea Chapter until 2009. His current research topics include advanced filtering techniques, high precision inertial navigation system (INS), visual-inertial odometry (VIO), INS/GNSS/IBN integration, and smartphone-based/foot-mounted pedestrian dead reckoning (PDR) systems.