Optical Flow Monocular Visual-Inertial Odometry with Online Photometric Calibration

Quanpan Liu, Zhengjie Wang* and Huan Wang
School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China
*Email: wangzhengjie@bit.edu.cn

Abstract. Kanade-Lucas-Tomasi (KLT) optical flow algorithm based on the brightness constancy assumption is widely used in visual simultaneous localization and mapping (SLAM) and visual odometry (VO). However, the automatic adjustment of camera exposure time, the attenuation factor of sensor irradiance caused by vignetting, and the nonlinear camera response function will cause the same feature point to have different brightness values on different image frames, thus breaking this assumption. Hence, we propose a gain-adaptive KLT optical flow algorithm with online photometric calibration, and on this basis, design a monocular visual-inertial odometry which is insensitive to brightness changes. This method can calibrate the photometric parameters online in real time, meet the assumption of constant brightness in practical applications, and make the algorithm more robust and accurate in the case of dynamic changes in brightness. Experimental results on the TUM Mono and EuRoC datasets show that the proposed algorithm can reliably calibrate the photometric parameters of any video sequence and perform well in the environment with varying brightness.

1. Introduction
The visual SLAM and VO algorithms can provide the robot with accurate and robust state estimation in GPS-denied environments, which is the cornerstone of the robotic completely autonomous navigation [1,2]. According to the different ways of visual information processing, SLAM and VO can be divided into three categories: feature-based methods, direct methods, and semi-direct methods. The feature-based methods extract a set of sparse point or line features in an image, and match the features in multiple frames of images by feature descriptors or optical flow algorithms [3]. The direct methods consider pixels of the entire image or local pixels with large gradients, estimate the camera pose and scene structure by minimizing the photometric errors, so feature extraction and matching are not required [4]. The semi-direct methods have been considered to be a hybrid of the direct method and the feature-based method [5]. This method obtains the camera pose by directly aligning the feature point image blocks, and estimates the feature point depth through depth estimation. Bundle adjustment (BA) is used to further optimize the pose and depth to reduce the cumulative error caused by incremental estimation.

The direct methods, semi-direct methods and feature-based methods based on optical flow algorithm all rely on the brightness constancy assumption, that is, the same scene point maintains the same pixel brightness value on different images. However, in practical applications, the brightness constancy assumption is usually invalid [6]. Automatic exposure, vignetting factor and camera response function will break this assumption. Therefore, they need to perform photometric calibration to obtain more competitive results [7].
In this paper, we propose a gain-adaptive KLT optical flow algorithm with online photometric calibration, and on this basis, design a brightness insensitive monocular visual-inertial odometry based on nonlinear optimization. The main contributions of our work are as follows:

- Traditional photometric calibration methods need accurate feature point matching as input, but accurate feature point matching also requires accurate photometric calibration, which leads to chicken-and-egg problems. We combine the photometric calibration with the optical flow matching algorithm, so online calibration of photometric parameters and optical flow matching can be performed at the same time, thus solving the chicken and egg problem.
- We have designed a novel visual-inertial odometry, which can achieve better performance in an environment with dynamic changes in brightness.

2. Related Work

Many photometric calibration methods have been proposed, but they are mainly used in image stitching and mosaicking [8-11]. With the development of direct SLAM algorithm based on the brightness constancy assumption, photometric calibration has been applied to SLAM.

Daniel proposed an on-line photometric calibration method for visual slam based on direct method, which can improve the performance of the algorithm [12]. For cameras without photometric calibration, DSO can estimate photometric parameters online [13]. This process models the imaging process of the camera, so there will be better performance for the changes in image brightness caused by different camera exposures.

Since the direct method needs to compare image information and the result is easily disturbed by changes in illumination, the photometric calibration is mainly applied to SLAM based on the direct method. However, the SLAM algorithm based on the optical flow method also relies on the brightness constancy assumption, so it is necessary to apply photometric calibration algorithm to optical flow method. Therefore, we propose a gain-adaptive KLT optical flow algorithm with online photometric calibration, and integrate the photometric calibration algorithm into the optical flow method to make it perform better in complex and dynamic environments.

3. Front End

3.1. Algorithm Framework

As shown in Figure 1, our algorithm processes the uncalibrated image frame information from the visual sensor and the inertial information from the IMU. To accurately estimate the vignetting factor and response function, a lot of image frames need to be collected for calibration, which will result in a decrease in optimization speed. However, for online photometric calibration, it is necessary to estimate photometric correction parameters in real time. Therefore, we put the optimization of the vignetting factor and the response function in the background independently to ensure the real-time performance of the algorithm. The system has a database to store vignetting factor, response function, feature points and calibrated image frames.
When the latest image frame is collected, it is first calibrated by the vignetting factor and response function stored in the database. Then through the gain adaptive KLT optical flow algorithm, the feature point optical flow tracking and exposure time estimation are completed simultaneously. Since the exposure time can be estimated from one frame to another, our algorithm can be implemented in real time. Finally, we store the calibrated image frames and feature points in the database. As more frames arrive, the vignetting factor and response function stored in the database will be continuously updated in the background.

The inertial information from the IMU is first processed by IMU pre-integration, and then combined with the visual information for visual-inertial alignment. We minimize the re-projection error, IMU residual, and prior information in the back-end to obtain a more accurate state estimation [14].

3.2. Gain-adaptive KLT Optical Flow Algorithm
The nonlinear optimization of the back end and the on-line photometric calibration need a certain number of feature points and accurate feature matching. Therefore, we extract new FAST feature points from the latest frame to keep at least 200 feature points per frame, and use gain-adaptive KLT optical flow algorithm to match them. We adopt the reverse optical flow tracking method to eliminate the outliers generated in the tracking process. In order to accurately estimate the vignetting, the feature points are required to be evenly distributed on the image, so we divide the image into several network units. In addition, we focus on selecting the points of long tracking trajectory, which is necessary for the estimation of vignetting factor.

The spatial feature points will reflect light to space after receiving the illumination from the light source. The reflected light \( L \) captured by the camera is affected by the exposure time \( e \), vignetting factor \( V \), response function \( R \), and then the light intensity \( O \) is output on the image.

\[
O = R(eV(x)L)
\]

Suppose that the same feature point can be observed by two adjacent frames \( I, J \), the following formula can be obtained from the brightness constancy assumption.

\[
f(J)\left( x + \frac{dx}{2} \right) - f(I)\left( x - \frac{dx}{2} \right) - K = 0
\]

where \( f = \log \left( \frac{R}{V} \right) \), \( K = \log \frac{e_2}{e_1} \). The \( K \) is the logarithm of the exposure ratio between the two images. Vignetting factor and response function can be known from the database. Therefore, we only need to track the feature points and estimate the exposure ratio of two images.

We apply the Taylor expansion to Eq. (2).

\[
f(J) + f'(I)\nabla f^T \frac{dx}{2} - f(I) - f'(I)\nabla^T f^T \frac{dx}{2} - K = 0
\]
In order to increase the number of feature points, we consider all pixels in the small image block around the feature \( P_i \). We can get the displacement \([dx_i, dy_i]^T\) and exposure ratio \( K \) by minimizing the following error function:

\[
E(dx_i, dy_i, K) = \sum_{x \in P_i} (c + a \frac{dx_i}{2} + b \frac{dy_i}{2} - K)^2
\]

where

\[
\begin{align*}
a &= f(J) J_x + f(I) I_x \\
b &= f(J) J_y + f(I) I_y \\
c &= f(J) - f(I)
\end{align*}
\]

If the partial derivative of unknown quantity in Eq. (4) is zero, the following error equation can be obtained.

\[
\begin{bmatrix}
U_i \\
w_i \\
\lambda_i
\end{bmatrix}
= \begin{bmatrix}
\sum_{P_i} a^2 & \sum_{P_i} ab \\
\sum_{P_i} ab & \sum_{P_i} b^2 \\
-\sum_{P_i} a & -\sum_{P_i} b
\end{bmatrix} \begin{bmatrix}
x_i \\
y_i
\end{bmatrix}
= \begin{bmatrix}
v_i \\
m_i
\end{bmatrix}
\]

where

\[
\begin{align*}
X_i &= [dx_i, dy_i, K]^T \\
U_i &= \frac{1}{2} \sum_{P_i} a^2 + \frac{1}{2} \sum_{P_i} ab \\
w_i &= \frac{1}{2} \sum_{P_i} ab + \frac{1}{2} \sum_{P_i} b^2 \\
v_i &= -\sum_{P_i} a + \sum_{P_i} b \\
m_i &= 2 \sum_{P_i} c
\end{align*}
\]

Compared with the standard KLT optical flow algorithm, our algorithm can jointly optimize the feature point tracking and the exposure rate between image frames under the condition of image response function and vignetting factor compensation, and has better performance when the brightness changes greatly between consecutive frames.

3.3. Online Photometric Calibration

We adopt the empiric model of response (EMoR) proposed by Grossberg and Nayar [8] to model the response function.

\[
R(x) = R_0(x) + \sum_{k=1}^n c_k h_k(x)
\]

The model has been successfully applied to the estimation of response function in panorama stitching system. We select the parameter \( c_k \), and then estimate the average response function \( R_0(x) \) and the basis function \( h_k(x) \), and combine them linearly to form the global response function \( R(x) \).

We assume that the vignetting factor is symmetric around the image center, and the vignetting center is located in the image center. Model the vignetting factor as a sixth-order polynomial.

\[
V(x) = 1 + v_1 r(x)^2 + v_2 r(x)^4 + v_3 r(x)^6
\]

where \( r(x) \) is the normalized radius of the image point \( x \) with respect to the image center.

In order to accurately calibrate the photometric parameters, a lot of features matching is needed as input, which will lead to the slow optimization speed. However, the exposure time can be estimated frame by frame in real time. Therefore, we decouple the estimation of exposure time from the other two parameters. We use the gain-adaptive KLT optical flow for fast feature point tracking and exposure estimation in a sliding window. The database stores 200 tracking frames and uses one fifth of them to estimate the response function and vignetting.

Given the feature points \( p \in P \) on a series of frames \( F_p \), we can get the following formula through formula Eq. (1)
\[ E = \sum_{p \in P} \sum_{i \in F_p} w_i^p \| O_i^p - r(V(x_i^p) L_p) \|_2 \] (8)

where \( O_i^p \) represents the output intensity of \( p \) in image \( i \), \( L_p \) represents the radiance of \( p \) and \( x_i^p \) represents the spatial location of the projection of \( p \) onto image \( i \). \( w_i^p \) define a weighting factor for each residual \( r \). Solving the above formula, the vignetting coefficient and response function can be estimated.

4. Back End

The convergence speed and results of the nonlinear visual-inertial odometry depend heavily on reliable initial values. Therefore, we need to initialize the system by aligning the IMU pre-integration with the image to estimate the metric scale, gyroscope bias, gravity vector, and initial velocity for the system [15,16].

The back-end based on tightly-coupled optimization is performed in the sliding window. After tracking the current frame, it will be jointly optimized with other frames in the sliding window. To ensure the real-time performance of the algorithm, a fixed number of keyframes are optimized in the sliding window, and marginalization is used to process the keyframes removed from the sliding window [17]. The prior information generated by the marginalization will also be combined with the IMU residual error, monocular or stereo vision re-projection errors to jointly optimize the camera pose and feature point position to obtain high-precision positioning and mapping.

4.1. State Variables

In the back end, the estimated state variables are as follows:

\[
X = [X_0, X_1, \ldots, X_n, \rho_0, \rho_1, \rho_2, \ldots, \rho_m]
\]
\[
X_k = [P_{B_k}^W, V_{B_k}^W, q_{B_k}^W, b_{a_k}^B, b_{o_k}^B], \quad k \in [0,n]
\]

where the state variables \( X \) is composed of the optimization variables \( X_n \) related to the IMU and feature point depths \( \rho_m \), \( n \) represents the size of the sliding window, and \( m \) is the number of feature points observed in the keyframe in the sliding window. The \( X_k \) includes the translation \( P_{B_k}^W \), velocity \( V_{B_k}^W \), rotation quaternions \( q_{B_k}^W \), the bias of gyroscope \( b_{a_k}^B \) and accelerometer \( b_{o_k}^B \).

We can get all state variables by minimizing the following formula:

\[
\min_X \left\{ \sum \| r_B(Z_{B_{t+1}^i}, X) \|_2^2 + \sum \| r_C(Z_{F_j}^i, X) \|_2^2 + \| r_p H_p X \|_2^2 \right\}
\]

where \( r_B(Z_{B_{t+1}^i}, X), r_C(Z_{F_j}^i, X) \) and \( \{r_p, H_p\} \) are IMU residuals, re-projection errors and prior information respectively.

4.2. IMU Residuals

We use IMU pre-integration to integrate these IMU measurements into constraints for keyframe selection, visual inertial alignment, and back-end optimization [18-20]. IMU residuals are obtained by IMU pre-integration.

\[
r_B \left( Z_{B_{t+1}^i}, X \right) = \begin{bmatrix}
\delta \theta_{B_{t+1}^i}^B \\
\delta \beta_{B_{t+1}^i}^B \\
\delta \gamma_{B_{t+1}^i}^B \\
\delta \omega_{B_{t+1}^i}^B \\
\delta \alpha_{B_{t+1}^i}^B
\end{bmatrix}
\begin{bmatrix}
R_{W}^B(\vec{P}_{B_{t+1}^i}^W \vec{P}_{B_{t}^W}^W \Delta t + \frac{1}{2} \vec{g}^W \Delta t^2) \vec{\alpha}_{B_{t+1}^i}^B \\
R_{W}^B(\vec{V}_{B_{t}^W}^W \vec{V}_{B_{t+1}^W}^W + g^W \Delta t) \vec{\beta}_{B_{t+1}^i}^B \\
2 \left[ q_W^{-1} \otimes \vec{g}_B \otimes (\vec{b}_{B_{t+1}^i}^B)^{-1} \right]_{xyz} \\
\vec{b}_{a_{t+1}^B} - \vec{b}_{a_{t}^B} \\
\vec{b}_{o_{t+1}^B} - \vec{b}_{o_{t}^B}
\end{bmatrix}_{15 \times 1}
\]

(11)

where \( [\vec{\alpha}_{B_{t+1}^i}^B, \vec{\beta}_{B_{t+1}^i}^B, \vec{\gamma}_{B_{t+1}^i}^B]^T \) are the IMU pre-integration between two adjacent frames at time \( t \) and time \( t + 1 \).
4.3. Visual Re-Projection Errors

In the monocular system, when the feature point $F_1$ is first observed at the time $t$, the monocular visual re-projection error at the time $t + 1$ can be defined as:

$$r_C(Z_{F_1}^{t+1}, X) = \pi^{-1}(\begin{bmatrix} u_{F_1}^{Ct} \\ v_{F_1}^{Ct} \end{bmatrix}) - R_B^W(R_W^B)^{-1} \pi^{-1}(\begin{bmatrix} u_{F_1}^{Ct} \\ v_{F_1}^{Ct} \end{bmatrix}) + P_B^W - P_B^{Wt+1}$$

where $[u_{F_1}^{Ct}, v_{F_1}^{Ct}]$ and $[\tilde{u}_{F_1}^{Ct+1}, \tilde{v}_{F_1}^{Ct+1}]$ represents the coordinates of the pixel at time $t$ and time $t + 1$ respectively.

5. Experiment

In this section, we evaluate the performance of the algorithm and the accuracy of the online photometric calibration part on the TUM Mono and EuRoC datasets. Table 1 shows the absolute pose error (APE) root mean square error (RMSE) of PC-VIO and VINS-Fusion [21,22].

| Dataset       | VINS-Mono | PC-VIO |
|---------------|-----------|--------|
| MH_01_easy    | 0.254651  | 0.263676 |
| MH_04_difficult | 0.550573 | 0.537219 |
| MH_05_difficult | 0.678140 | 0.493640 |
| V1_01_easy    | 0.229267  | 0.209508 |
| V1_02_medium  | 0.436770  | 0.415312 |
| V1_03_difficult | 0.375504 | 0.343641 |
| V2_01_easy    | ×         | 0.170920 |
| V2_02_medium  | 0.395870  | 0.411921 |
| V2_03_difficult | 0.435668 | 0.396309 |

Optical flow algorithm is based on the brightness constancy assumption, but we find that there are strong exposure changes in some datasets (such as V1_01_easy), so the standard KLT tracker performs badly in these cases. The exposure time change of dataset V1_01_easy is shown in Figure 2.

![Figure 2. The exposure time change of dataset V1_01_easy.](image)

As shown in Table 1, in these datasets, VINS-Mono with standard KLT tracker does not perform well. In contrast, our algorithm uses gain adaptive KLT optical flow algorithm, which combines photometric calibration with optical flow matching algorithm, and realizes online calibration of photometric parameters and optical flow matching at the same time. Therefore, it can achieve good results in the environment of brightness changes. The vignetting factor and response function of the dataset V1_01_easy are shown in Figure 3. The trajectory heat map estimated by PC-VIO and VINS-Mono in MH_05_difficult are shown in Figure 4.
Figure 3. The vignetting factor and response function of the dataset V1_01_easy.

Since the EuRoC dataset does not provide the true values of photometric parameters, we evaluate the online photometric calibration of the algorithm on the TUM Mono dataset. Figures 5 and 6 show the photometric parameter calibration of sequence 50 of the TUM Mono dataset. It can be seen that our photometric calibration algorithm can converge the response function, vignetting factor and exposure time to a satisfactory solution.

Figure 4. The trajectory heat map estimated by PC-VIO and VINS-Mono in MH_05_difficult.

Figure 5. The vignetting factor and response function of the TUM Mono dataset.
6. Conclusions
In this paper, we propose a gain-adaptive KLT optical flow algorithm with online photometric calibration, and on this basis, design a tightly coupled monocular visual-inertial odometer based on nonlinear optimization. We evaluated the performance of the system in a public dataset. Compared with the traditional optical flow SLAM method, this algorithm can complete online calibration of photometric parameters and optical flow matching at the same time, so it can achieve better performance in complex and dynamic environment.

References
[1] Bailey T, Durrant-Whyte H 2006 Simultaneous localization and mapping (SLAM): Part II. IEEE Robot. Autom. Mag., 13 108–17
[2] Shiping W, Michael Z.Q. C, Zhigang Z, Tingwen H, Chaojie L, 2017 Adaptive Neural-Fuzzy Sliding-Mode Fault-Tolerant Control for Uncertain Nonlinear System. IEEE Transactions on Systems, Man and Cybernetics: Systems 47 2268-78
[3] Mur-Artal R, Montiel J M M and Tardos J D 2015 J ORB-SLAM: A Versatile and Accurate Monocular SLAM System IEEE Trans Robot 31 1147–63
[4] Engel J, Schöps T, and Cremers D 2014 LSD-SLAM: Large-scale direct monocular SLAM In Lecture Notes in Computer Science Proceedings of the European Conference on Computer Vision 834–49
[5] Forster C, Pizzoli M, and Scaramuzza D 2014 SVO: Fast semi-direct monocular visual odometry ICRA pp 15–22
[6] Yang N, Wang R, Gao X and Cremers D 2017 Challenges in monocular visual odometry: Photometric calibration, motion bias and rolling shutter effect arXiv:170504300
[7] Engel J, Usenko V, Cremers D 2016 A Photometrically Calibrated Benchmark For Monocular Visual Odometry arXiv preprint arXiv:160702555
[8] Goldmannan D B 2010 Vignette and exposure calibration and compensation IEEE Transactions on Pattern Analysis and Machine Intelligence 32 2276-88
[9] Grossberg M D and Nayar S K 2003 What Is the Space of Camera Response Functions CVPR 2
[10] Kim S J and Pollefeys M 2008 Robust Radiometric Calibration and Vignetting Correction IEEE Transactions on Pattern Analysis and Machine Intelligence 30 562-76
[11] Kim S J, Gallup D, Frahm J, Pollefeys M 2010 Joint Radiometric Calibration and Feature Tracking for an Adaptive Stereo System Computer Vision and Image Understanding 114 574-82
[12] Bergmann P, Cremers D, Wang, R 2018 Online photometric calibration of auto exposure video for realtime visual odometry and SLAM IEEE Robot 3 627-34
[13] Engel J, Koltun V, Cremers D 2018 Direct sparse odometry IEEE Trans Pattern Anal 40 611–25
[14] Qin T, Li P, Shen S, 2018 VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator IEEE Trans Robot 34 1004–20
[15] Liu Q, Wang Z, Wang H 2020 SD-VIS: A Fast and Accurate Semi-Direct Monocular Visual-Inertial Simultaneous Localization and Mapping (SLAM) Sensors 20 p1511
[16] Hartley R and Zisserman A 2003 Multiple View Geometry in Computer Vision (Cambridge University Press)
[17] Sibley G, Matthis L, Sukhatme G 2010 Sliding window filter with application to planetary landing J Field Robot 27 587–608
[18] Leutenegger S, Lynen S, Bosse M, Siegwart R, Furgale P 2015 Keyframe-based visual-inertial odometry using nonlinear optimization Int J Robot 34 314–34
[19] Lupton T and Sukkarieh S 2012 Visual-inertial-aided navigation for highdynamic motion in built environments without initial conditions IEEE Trans 28 61–76
[20] Forster C, Carlone L, Dellaert F, Scaramuzza D 2017 On-manifold preintegration for real-timevisual–inertial odometry IEEE Trans 33 1–21
[21] Sturm J, Engelhard N, Endres F, Burgard W, Cremers D 2012 *J A benchmark for the evaluation of RGB-D SLAM systems* IROS pp 573–580

[22] Michael G EVO 2019 Available online: https://githubcom/MichaelGrupp/evo (accessed on 8 August 2019)