Proceedings of Machine Learning Research vol 278:1–7, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

6D Pose Estimation of Camera Based on the Fusion of Checkerboard and ICP Algorithm

Ziqiang Li
School of Automation, China University of Geosciences, Wuhan 430074, China

Siyv Fu
School of Automation, China University of Geosciences, Wuhan 430074, China

Wenxing Liao
School of Automation, China University of Geosciences, Wuhan 430074, China

Tao Wu* WUTAO@CUG.EDU.CN
School of Automation, China University of Geosciences, Wuhan 430074, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Nowadays, camera pose estimation, as a core technology in the field of 3D vision, plays a crucial role in various applications such as autonomous driving. This paper presents a camera pose estimation method that combines checkerboard pattern and ICP (Iterative Closest Point) algorithm. Based on the chessboard calibration plate captured by the mobile camera, by extracting two-dimensional feature points through pixel-level corner detection and then performing sub-pixel optimization on the feature points, and then precisely matched with the three-dimensional point cloud data obtained by the Gemini2 depth camera. A registration model based on the ICP algorithm is constructed to simultaneously solve the rotation matrix and translation vector of the camera. The experimental results demonstrate that this method exhibits high accuracy and achieves a root mean square error (RMSE) of 0.0109 and the fitness is 1. By merely utilizing 40 key feature points (derived from an 8×5 checkerboard), this method reduces the computational load, enables ICP to converge within 40 iterations, and enhances the real-time efficiency.

Keywords: Checkerboard, ICP, Camera pose estimation.

1. Introduction

In recent years, technological developments in the field of 3D vision highlight the core value of camera pose estimation. In augmented reality, pose estimation enables seamless integration of virtual and real through precise spatial alignment; in autonomous driving, it provides basic support for 3D scene reconstruction and dynamic obstacle perception. Cross-domain requirements promote research breakthroughs in high-precision and robust pose estimation algorithms, gradually solve key challenges such as complex lighting and dynamic interference, and promote the deep application of 3D vision technology in industrial detection, intelligent navigation and other scenes.

In the field of camera pose estimation, according to the difference of sensor data modes, existing methods can be divided into three categories: Epipolar Geometry based on 2D-2D feature correspondence, PnP method based on 3D-2D projection relationship, and ICP method for 3D-3D point cloud registration. When only multiple pairs of 2D pixel corresponding points in two views can be obtained, the pose estimation algorithm based on Epipolar Geometry principle is usually adopted. However, Epipolar Geometry operation is often complicated and easily affected by initialization problems, resulting in low accuracy of pose estimation (White and Beard, 2020). When a group of 3D spatial coordinates and their projection positions on the camera are known, the PnP problem solution method is generally adopted to estimate the camera pose, and EPnP algorithm is commonly used. Chen et al. (2023) proposed an attitude estimation algorithm based on YOLOv5 and EPnP for the telescopic boom of boarding bridge in the process of docking with offshore wind turbine. Although EPnP algorithm performs well in certain scenarios, EPnP algorithm has unstable factor values in coplanar control point scenarios, resulting in solution degradation, nonlinear amplification of 2D detection errors into pose results, and inability to deal with dynamic targets or complex scene disturbances (Qiao et al., 2023). Compared with the first two methods, ICP algorithm (Qin et al., 2023) realizes pose estimation by directly aligning the geometric structure of 3D point cloud, without relying on 2D feature matching or camera internal parameters, can adapt to partially missing or dense point cloud scenes, and directly outputs complete pose parameters including absolute scale through iterative optimization. Clotet and Palacín (2023) monitored the detection of obstacles in the navigation of mobile robots through outlier detection after ICP positioning. However, the traditional ICP algorithm is also easy to fall into local optimization due to lack of feature guidance, low computational efficiency, and sensitive to initial pose, and its all-point cloud processing mechanism has poor noise immunity and is difficult to ensure accurate correspondence (Zhang et al., 2022; Li et al., 2020).

In order to address the limitations of the traditional ICP algorithm. Firstly, we generate 3D point cloud data from the depth images, and obtain the corner point data of the checkerboard pattern from the color images. Eliminate noise points by using depth mapping and symmetrical filtering. Then, we use the ICP algorithm to match the corner points with the three-dimensional point cloud, thereby completing the camera pose estimation. This method enhances the efficiency and robustness of feature matching, effectively suppresses noise, and ensures the stability of feature correspondence. It avoids the local optimization and mismatch problems caused by random sampling in the traditional ICP, thereby improving the accuracy and convergence speed of pose estimation in structured scenes and reducing the computational complexity.

2. Pose Estimation Based on Point Cloud Matching

2.1. Scheme Design for Pose Estimation

In this chapter, the strategy for camera pose estimation is as follows: Obtain color images and depth images synchronously through the Gemini2 depth camera. It begins by detecting the 2D pixel coordinates of checkerboard corners in the color image and enhances coordinate accuracy through sub-pixel optimization. Using the camera's internal parameters and depth values, we convert the corner pixel coordinates into 3D spatial coordinates, generating a corresponding 3D point cloud. Coordinate mapping is performed between the detected corners and the 3D points, filtering out noise to retain effective corners. The ICP algorithm registers the current corner point cloud with the target point cloud by iteratively finding the closest point correspondences and calculating the rotation matrix and translation vector, ultimately outputting the camera's 6-DOF pose transformation parameters. With color image feature guidance and depth information fusion, we establish high-precision feature constraints prior to registration. Strong geometric feature localization is achieved through checkerboard corner detection, achieving pixel coordinate accuracy of 0.1 pixels by leveraging sub-pixel optimization. This results in a low-noise feature point cloud via depth mapping and symmetric filtering, ensuring clear feature correspondences and high-quality initial alignments during optimization. This approach effectively mitigates traditional ICP's tendency to fall into local optimization and its sensitivity to initial poses in feature-free clouds, significantly enhancing registration accuracy and convergence speed in complex scenes.

2.2. Realization Process

The mobile camera takes pictures of the same scene from different angles. Each set of images included color images and depth images. Color map is used to extract checkerboard corners and provide 2D pixel coordinate information. Depth map stores depth value d corresponding to each pixel, which is converted into 3D coordinates by camera internal parameter matrix K.

Detects checkerboard corners in color maps. For images, the corner pixel coordinates are:

$$p_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}, i = 1, 2, \ldots, \tag{1}$$

Where $N = 8 * 5$ (checkerboard size 8*5). The sub-pixel-level corner detection can capture the subtle variations of corner points between pixels, thereby improving the positioning accuracy of corner points. To improve accuracy, corner positions are further corrected using a sub-pixel pole optimization algorithm:

$$\hat{p_i} = \text{argmin} \sum_{q \in N(p)} w_q \|I(q) - I(p)\|^2 \tag{2}$$

Where $N(p)$ is the corner domain and w is the Gaussian weight. Sub-pixel-level corner detection achieves this by enhancing the accuracy of corner location from integer pixels to fractional pixels.

According to the camera imaging model, the three-dimensional coordinates $P(u, i) = [X, Y, Z]^T$ of the pixel (u, v) in the depth map are calculated as:

$$\begin{cases} X = \frac{(u-u_0)\cdot d}{f_x}, \\ Y = \frac{(v-v_0)\cdot d}{f_y}, \\ Z = d, \end{cases} \tag{3}$$

Where f_x, f_y are the focal lengths, and u_0, v_0 are the principal point coordinates, which are calibrated in advance using the Zhang Zhengyou calibration method.

Matching the extracted corner points in the color map with the corresponding three-dimensional coordinate values, filtering the three-dimensional coordinate data of the corner points considering that there are invalid points without depth values in the obtained depth map and the depth information corresponding to a single pixel point is unstable, taking the two-dimensional sub-pixel coordinates of the corner points as the center, taking the mean value of the corresponding three-dimensional coordinate points in a certain radius area, wherein the invalid points and their central symmetric points are eliminated, so as to obtain three-dimensional point cloud data corresponding to two groups of corner points:

$$\overline{p_i} = \frac{1}{M} \sum_{k=1}^{M} P_k \tag{4}$$

ICP algorithm uses point-to-point mode to register point clouds and iteratively solve the transformation matrix of camera. Let source point cloud $P = \{P1, P2, \ldots, PN\}$ (corresponding to the first set of images) and target point cloud $Q = \{Q1, Q2, \ldots, QN\}$(corresponding to the second set of images), ICP algorithm solves rigid body transformation (R, t) by minimizing the Euclidean distance of point pairs:

$$\sum_{i=1}^{N} \|RP_i + t - Q_i\|^2, R \in SO(3), t \in R^3 \tag{5}$$

Where $SO(3)$ is a three-dimensional rotation group.

$$\bar{p} = \frac{1}{N}\sum_{i=1} P_i, \bar{Q} = \frac{1}{N}\sum_{i=1} Q_i \tag{6}$$

Calculate point cloud centroid:

$$p_i = P_i - \bar{P}, q_i = Q_i - \bar{Q} \tag{7}$$

Calculate covariance matrix:

$$W = \sum_{i=1}^{N} p_i q_i^T \tag{8}$$

Perform singular value decomposition (SVD) on W:

$$W = U \sum V^T \tag{9}$$

Solve rotation matrix R and translation vector t:

$$R = VU^T, t = \bar{Q} - R\bar{P} \tag{10}$$

3. Analysis of Experimental Results

This experiment involves four input images. The scene is still designed with the black and white checkerboard as the landmark. Two sets of color images and depth images were obtained by using the Gemini2 depth camera to shoot with small-angle rotation. As shown in Figures 1 to 4.

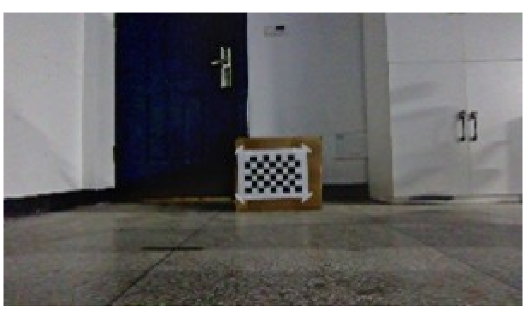

Figure 1: Color Figure 1

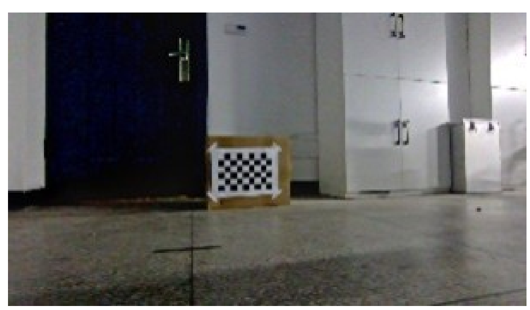

Figure 2: Color Figure 2

For the color images, obtain the corner points of the checkerboard pattern., as shown in Figure 5 and Figure 6.

For the depth map, first convert it into point cloud data that stores the three-dimensional position information and read it. Then, optimize the extracted corner points to sub-pixel level and

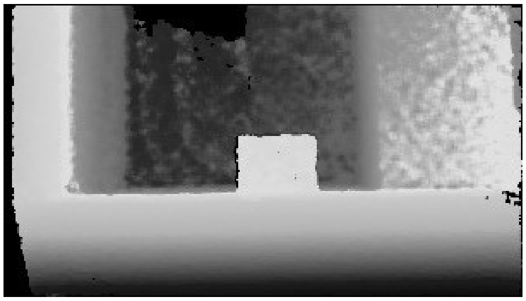

Figure 3: Depth Map 1

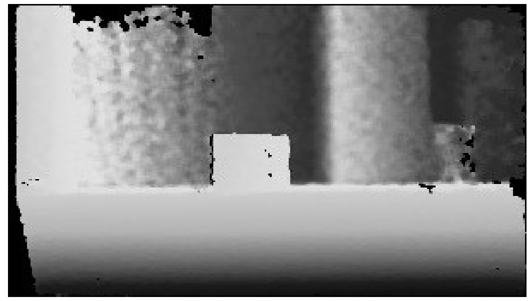

Figure 4: Depth Map 2

Figure 5: corner points 1

Figure 6: corner points 2

match them with the corresponding three-dimensional coordinate values. After many tests, set the maximum allowable distance threshold as 0.01, the maximum iteration times as 50, and after a total of 40 iterations, it tends to stabilize. The RMSE of all interior points matching is 0.01092932, and the fitness is 1, indicating that the fitting degree is good. R and t are extracted from the obtained transformation matrix, i.e. the pose of the camera results as follows:

$$
\begin{aligned}
R = [&0.95357842, -0.02578756, -0.30003865;\\
&0.00907712, 0.99833545, -0.05695561;\\
&0.30100797, 0.05158815, 0.95222522]\\
t = [&0.70333006; 0.13303514; 0.24402243]
\end{aligned}
$$

Apply the above results to the source point cloud, align it with the target point cloud, and visualize the original point cloud and the registered point cloud as shown in the figures 7 and 8.

The red points represent the source point cloud, corresponding to Figure 1 and Figure 3, while the blue points indicate the target point cloud following camera pose transformation, corresponding to Figure 2 and Figure 4. Figure 8 shows that the corner points in the middle three columns match well, whereas the left and right sides exhibit large matching errors. These discrepancies may stem from inaccuracies in the captured depth map, where external noise can lower the accuracy of the generated point cloud data, leading to poor matching results. To improve this, techniques such as averaging multiple images and further filtering the point cloud data can be employed. Additionally, substituting R and t will more clearly illustrate the posture transformation in the camera's corresponding coordinate system, as shown in the figure 9:

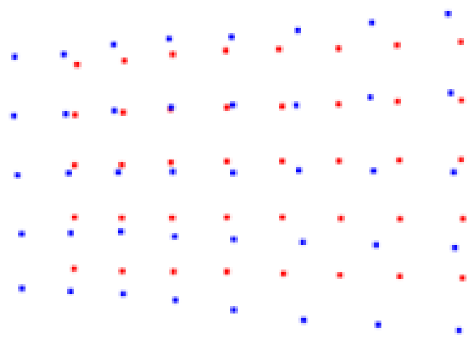

Figure 7: Original point cloud distribution

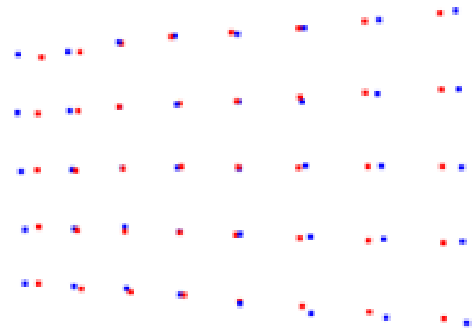

Figure 8: Point cloud matching visualization

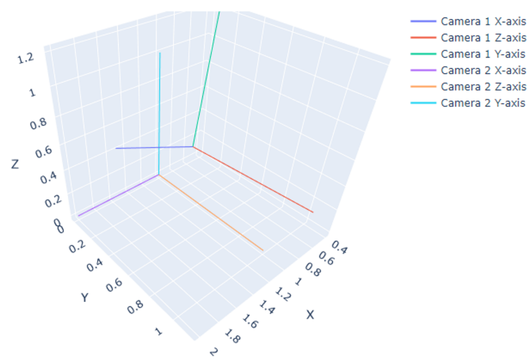

Figure 9: camera attitude transformation estimation result

The chessboard grid corner points are selected as feature points. Through sub-pixel coordinate extraction and local depth mean filtering, the noise and invalid points in the depth map are significantly reduced. Furthermore, the strategy of retaining only a few dozen key feature points while maintaining the registration accuracy limits the number of iterations to 40, which promotes the rapid convergence of the ICP algorithm in the point cloud registration process. It has greatly enhanced the real-time performance.

4. Conclusion

This paper presents a camera pose estimation method that combines checkerboard pattern and ICP algorithm. aiming to address the limitations of traditional algorithms. The method first extracts two-dimensional feature points using a checkerboard calibration plate and matches them with the three-dimensional point cloud data in the depth image through sub -pixel corner detection. We adopt the ICP algorithm for point cloud registration, determining the precise inverse pose of the camera by iteratively solving the rotation matrix and translation vector to determine the camera pose. Traditional ICP needs to process tens of thousands or even millions of point clouds. However, the experimental results show that the method proposed in this paper only retains several tens of key feature points (for example, 8x5 checkerboard corresponds to 40 points). The reduction of data volume significantly improves the ICP iteration speed and greatly enhances the accuracy of pose estimation. It is particularly suitable for scenarios with high real-time requirements, and it provides new solutions for three-dimensional vision applications. For instance, in the robot inspection scenario, this method is of vital importance for the precise positioning and navigation of the inspection robot. By reducing the data volume, it enhances the real-time performance. When inspecting the status of equipment, high-precision pose estimation can ensure that the inspection robot accurately reaches the target location.

Although this method has its advantages, it still has some limitations: Firstly, checkerboard-based detection may fail under strong illumination or occlusion, causing pose estimation errors. Secondly, the number of feature points is limited by the size of the checkerboard, and in complex unstructured environments, the sparsity of point clouds may affect the robustness of registration. Future work will explore a dynamic mechanism for adaptive addition and deletion of feature points, and attempt to construct a multimodal fusion framework using data from multiple sensors to expand the applicability of the method in special scenarios.

References

Weizhuo Chen, Lijie Zhang, and Fangrui Luo. An attitude estimation algorithm for the telescopic arm of the boarding bridge based on yolov5 and epnp. *Journal of Physics: Conference Series*, 2492(1):012022, may 2023. doi: 10.1088/1742-6596/2492/1/012022.

Eduard Clotet and Jordi Palacín. Slamicp library: Accelerating obstacle detection in mobile robot navigation via outlier monitoring following icp localization. *Sensors*, 23(15), 2023. doi: 10.3390/s23156841.

Peng Li, Ruisheng Wang, Yanxia Wang, and Wuyong Tao. Evaluation of the icp algorithm in 3d point cloud registration. *IEEE Access*, 8:68030–68048, 2020. doi: 10.1109/ACCESS.2020.2986470.

Rui Qiao, Guili Xu, Ping Wang, Yuehua Cheng, and Wende Dong. An accurate, efficient, and stable perspective-n-point algorithm in 3d space. *Applied Sciences*, 13(2), 2023. doi: 10.3390/app13021111.

Lang Qin, Xin Chen, and Xuan Gong. An improved 3d reconstruction method for weak texture objects combined with calibration and icp registration. In *2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS)*, pages 1–5, 2023. doi: 10.1109/ICPS58381.2023.10128042.

Jacob H. White and Randal W. Beard. An iterative pose estimation algorithm based on epipolar geometry with application to multi-target tracking. *IEEE/CAA Journal of Automatica Sinica*, 7(4):942–953, 2020. doi: 10.1109/JAS.2020.1003222.

Juyong Zhang, Yuxin Yao, and Bailin Deng. Fast and robust iterative closest point. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3450–3466, 2022. doi: 10.1109/TPAMI.2021.3054619.