Bias-Eliminated PnP for Stereo Visual Odometry: Provably Consistent and Large-Scale Localization

Guangyang Zeng*[1], Yuan Shen*[1], Ziyang Hong[1], Yuze Hong[1], Viorela Ila[2], Guodong Shi[2], Junfeng Wu[1]
[1]School of Data Science, Chinese University of Hong Kong, Shenzhen, China
[2]Australian Center for Robotics and School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Australia

***Abstract*—We propose a bias-eliminated weighted (Bias-Eli-W) perspective-n-point (PnP) estimator for stereo visual odometry (VO). This estimator leverages statistical theory to handle varying 3D triangulation uncertainties, ensuring consistent relative pose estimates. Our stereo VO framework uses only triangulated points from the current keyframe, decoupling temporal dependencies between pose and 3D point errors. We integrate the Bias-Eli-W PnP estimator into the proposed stereo VO pipeline, creating a synergistic effect that enhances the suppression of pose estimation errors. Experiments show significant odometry performance improvement in large-scale environments.**

I. Introduction

Most VO methods are based on SLAM frameworks, jointly optimizing camera poses and 3D map points [1, 2, 3]. These methods often lack accurate uncertainty estimation for point correspondences and fail to incorporate estimator optimization with theoretical guarantees. The primary challenge of precise uncertainty estimation in a SLAM framework lies in the temporal coupling between pose and 3D point errors.

This paper introduces a pure odometry framework, CurrentFeature Odometry, which focuses on relative pose estimation without 3D point optimization. It leverages only triangulated feature points from the current keyframe for PnP-based tracking, breaking the temporal coupling between pose and 3D point errors. Building on this decoupling, we accurately model point uncertainties and optimize the estimator from a statistical perspective, resulting in a consistent PnP pose estimator that converges to the true value as the point number increases. CurrentFeature Odometry not only achieves significantly lower relative pose error (RPE) but also surpasses SOTA SLAM algorithms in terms of absolute trajectory error (ATE).

II. CurrentFeature Oodmetry

The system overview is illustrated in Figure 1.

1) Front end: We use pyramidal Lucas-Kanade optical flow to track features and establish 2D-2D matches between consecutive frames. To enhance robustness, a two-stage geometric verification is applied: (1) the five-point algorithm with RANSAC filters initial outliers via essential matrix estimation, and (2) an ℓ_1-norm PnP refines the pose estimate and discards the 10% of points with the largest reprojection errors. Keyframe (KF) insertion is determined based on the number of successfully tracked features and average feature displacement.

* Equally contributed

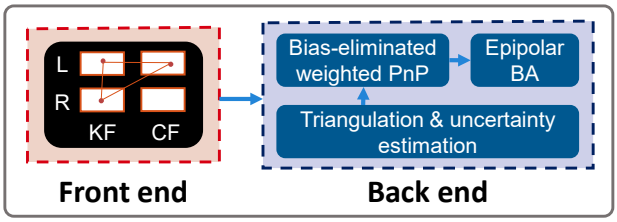

Figure 1: System overview. L and R refer to the left and right images, respectively, and KF and CF denote the keyframe and current frame, respectively.

2) Back end: As illustrated in Figure 2, after feature tracking and outlier rejection, we obtain point correspondences among the KF and the left image of the current frame (CF), denoted as $\{x_i, y_i, z_i\}_{i=1}^n$. According to Theorem 1 in [4], a consistent estimator $\hat{\sigma}^2$ for 2D feature noise variance is derived by solving an eigenvalue problem, converging to the true variance as the feature number increases.

For triangulation, we use a linear least-squares closed-form solution instead of parallax-based methods [1, 2] or singular-value-decomposition approaches [5], as it is more general and better suited for uncertainty analysis. Specifically, the solution is $p_i = (A_i^\top A_i)^{-1} A_i^\top b_i$, where A_i and b_i depend on x_i, y_i, and the stereo baseline. The uncertainty of the 3D point p_i is estimated as $\Sigma_i = J_{p_i} \Sigma J_{p_i}^\top$, where $\Sigma = \hat{\sigma}^2 I_4$ and J_{p_i} is the Jacobian matrix of p_i with respect to 2D feature noise.

Building on triangulation, we propose a bias-eliminated weighted (Bias-Eli-W) PnP algorithm to estimate the pose of the CF relative to the KF, denoted as (R_c, t_c). By referring to (12) in [6], a least-squares estimate for the pose can be obtained as $\hat{\theta}^{\mathrm{B}} = (H^\top H)^{-1} H^\top d$, where H and d are derived from 2D points z_i and 3D points p_i. However, since the regressor matrix H is correlated with noise, $\hat{\theta}^{\mathrm{B}}$ is inconsistent [7]. To address this, we design the bias-eliminated solution

$$\hat{\theta}^{\mathrm{BE}} = \left(\frac{H^\top H}{n} - G \right)^{-1} \frac{H^\top d}{n},$$

where G depends on z_i and 3D uncertainties Σ_i. The estimator $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ for rotation and translation can be recovered from $\hat{\theta}^{\mathrm{BE}}$; see (14)-(17) in [6].

Theorem 1. *The bias-eliminated estimator $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ is consistent, i.e., it converges to the ground truth as the feature number increases.*

Table I: Comparison of ATE and RPE across different sequences in KITTI dataset. ORB3 denotes ORB-SLAM3 and OV2 represents OV^{2}SLAM. Values highlighted in **blue bold** represent the smallest, and values in blue denote the second smallest.

Sequence	ATE (m)						RPE (m)					
	Color			Grayscale			Color			Grayscale		
	ORB3	OV2	Ours	ORB3	OV2	Ours	ORB3	OV2	Ours	ORB3	OV2	Ours
seq00	**4.042**	4.676	4.174	**4.263**	4.767	4.514	0.0287	0.0278	**0.0262**	0.0283	0.0262	**0.0260**
seq02	9.549	11.406	**5.756**	7.900	7.363	**3.900**	0.0286	0.0278	**0.0257**	0.0277	0.0263	**0.0257**
seq03	3.846	4.183	**0.551**	1.200	1.177	**1.030**	0.0250	0.0264	**0.0148**	0.0182	0.0166	**0.0158**
seq04	3.160	3.453	**2.328**	**0.213**	1.306	0.726	0.0445	0.0487	**0.0353**	0.0198	0.0239	**0.0197**
seq05	3.904	4.254	**3.332**	**2.115**	2.448	2.403	0.0264	0.0267	**0.0178**	0.0166	0.0163	**0.0124**
seq06	4.279	5.052	**2.400**	**1.791**	3.533	1.859	0.0360	0.0363	**0.0187**	0.0174	0.0183	**0.0138**
seq07	1.991	2.226	**1.593**	**1.222**	1.621	1.281	0.0235	0.0213	**0.0175**	0.0166	0.0124	**0.0123**
seq08	6.201	6.315	**5.866**	3.698	3.590	**3.430**	0.0439	0.0431	**0.0397**	0.0389	**0.0380**	0.0392
seq09	6.598	6.529	**5.245**	3.193	3.760	**2.169**	0.0324	0.0327	**0.0234**	0.0232	0.0249	**0.0181**
seq10	4.477	4.421	**3.088**	1.393	0.655	**0.638**	0.0261	0.0237	**0.0196**	0.0211	0.0181	**0.0172**
Ave	4.805	5.252	**3.433**	2.699	3.022	**2.195**	0.0315	0.0314	**0.0239**	0.0228	0.0221	**0.0200**

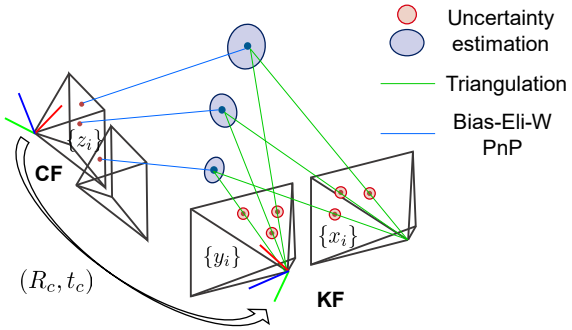

Figure 2: Illustration of frame tracking. The orange circles represent feature-matching uncertainties and the blue ellipses denote the triangulation uncertainties.

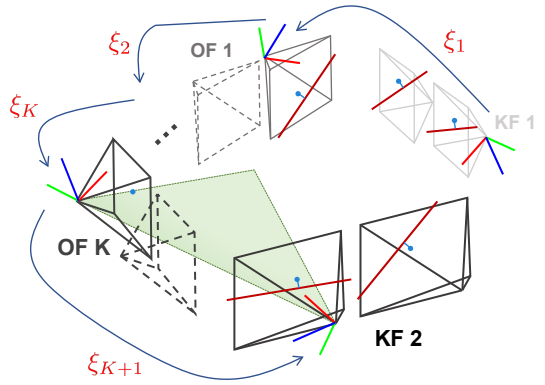

Figure 3: Illustration of epipolar BA. For two KFs, we utilize stereo images, while for OFs, we only use the left image.

We use $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ as the initial value and apply a weighted PnP iterative refinement. Let $h(\cdot)$ represent the pinhole camera projection model. The weight for the i-th point is $\bar{\Sigma}_i^{-\frac{1}{2}}$, where $\bar{\Sigma}_i = J_{h_i} \Sigma_i J_{h_i}^\top$, and J_{h_i} is the Jacobian of $h(\hat{R}_c^{\mathrm{BE}} p_i + \hat{t}_c^{\mathrm{BE}})$ with respect to p_i. Due to the consistency of the initial estimator $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ and the quadratic convergence of the Levenberg-Marquardt (LM) algorithm near the global minimum, a single LM iteration is sufficient to achieve the minimum estimation variance when the number of points n is large [4]. This makes our method computationally efficient.

When a new KF is generated, we perform a local bundle adjustment (BA) that includes the latest two KFs and intermediate ordinary frames (OFs), as shown in Figure 3. The parameters to be refined are six-degree relative poses $\xi_k \in \mathbb{R}^6, k = 1, \ldots, K+1$, where each ξ_k consists of three Euler angles and a translation vector. Since feature tracking in OFs only involves left images, their right images are excluded from the local BA. Following SOFT2 [8], we use the point-to-epipolar-line distance as the residual. In [8], only the left images from frames within the sliding window are involved in BA optimization. However, this approach, relying solely on temporal rigidity without a baseline constraint, cannot refine the scale. In this paper, by incorporating baseline-induced rigidity (using the right images of keyframes), we achieve simultaneous refinement of all six degrees of freedom.

III. Experiments

We evaluate the performance of CurrentFeature Odometry on the KITTI [9] dataset. For comparison, we include ORB-SLAM3 [1] and OV^{2}SLAM [2], the top two open-source stereo VO approaches on the KITTI dataset. To ensure a fair comparison and focus solely on odometry performance, the loop closure modules in both methods are disabled. The ATE and RPE comparison results are summarized in Table I. We see that CurrentFeature Odometry significantly outperforms ORB-SLAM3 and OV^{2}SLAM. Specifically, compared to the second-best method, our algorithm achieves a 24% reduction in average RPE and a 28% reduction in average ATE on color sequences.

IV. Conclusion

We revisited stereo VO and proposed a consistent PnP-enabled framework, CurrentFeature Odometry. It breaks the coupling between the pose and 3D point errors. Based on the decoupling, we accurately modeled the uncertainties of 3D points and proposed a Bias-Eli-W PnP estimator to achieve consistent relative pose estimation. Finally, an epipolar BA is used to refine pose estimation. CurrentFeature Odometry achieves SOTA performance on the KITTI dataset.

REFERENCES

[1] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6): 1874–1890, 2021.

[2] Maxime Ferrera, Alexandre Eudes, Julien Moras, Martial Sanfourche, and Guy Le Besnerais. Ov^{2}slam: A fully online and versatile visual slam for real-time applications. *IEEE Robotics and Automation Letters*, 6(2):1399–1406, 2021.

[3] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015.

[4] Guangyang Zeng, Qingcheng Zeng, Xinghan Li, Biqiang Mu, Jiming Chen, Ling Shi, and Junfeng Wu. Consistent and asymptotically statistically-efficient solution to camera motion estimation. *arXiv preprint arXiv:2403.01174*, 2024.

[5] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[6] Guangyang Zeng, Shiyu Chen, Biqiang Mu, Guodong Shi, and Junfeng Wu. Cpnp: Consistent pose estimator for perspective-n-point problem with bias elimination. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1940–1946. IEEE, 2023.

[7] Biqiang Mu, Er-Wei Bai, Wei Xing Zheng, and Quanmin Zhu. A globally consistent nonlinear least squares estimator for identification of nonlinear rational systems. *Automatica*, 77:322–335, 2017.

[8] Igor Cvišić, Ivan Marković, and Ivan Petrović. Soft2: Stereo visual odometry for road vehicles based on a point-to-epipolar-line metric. *IEEE Transactions on Robotics*, 39 (1):273–288, 2022.

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.