无偏 PnP 立体视觉测距法：可证明的一致性和大规模定位

Guangyang Zeng*, Yuan Shen*, Ziyang Hong, Yuze Hong, Viorela Ila, Guodong Shi, Junfeng Wu

***Abstract*—在本文中，我们首先提出了一种具有可证明一致性的消除偏差加权（Bias-Eli-W）透视点（PnP）估计器，用于立体视觉里程计（VO）。具体来说，利用统计理论，我们开发了一种渐近无偏且 $\sqrt{n}$ 一致的 PnP 估计器，该估计器考虑了不同的 3D 三角测量不确定性，确保随着特征数量的增加，相对姿态估计能收敛至真实值。接着，在立体视觉里程计工作流程方面，我们提出了一个框架，能够连续三角化当前特征以追踪新帧，有效地解耦姿态和 3D 点误差之间的时间依赖关系。我们将 Bias-Eli-W PnP 估计器集成到提出的立体视觉里程计工作流程中，形成协同效应以增强对姿态估计误差的抑制。我们在 KITTI 和 Oxford RobotCar 数据集上验证了该方法的性能。实验结果表明，我们的方法：1）在大规模环境中相对姿态误差和绝对轨迹误差方面实现了显著提升；2）在不稳定和不可预测的机器人运动下提供可靠定位。Bias-Eli-W PnP 在立体视觉里程计中的成功应用表明，在高不确定性测量的机器人估计任务中进行信息筛选的重要性，并揭示了 PnP 作为关键成分的多样化应用。**

***Index Terms*—Stereo visual odometry, PnP pose estimation, large-scale localization, consistent estimator.**

I. 介绍

V是 里程计（VO）指的是通过摄像机拍摄的连续图像来估计移动摄像机在三维空间中的姿态。VO 的重要性在于它具有无需基础设施、成本低、重量轻、能效高等优势 [? ? ?]。它使机器人能够自主感知和导航其环境。与单目 VO 相比，立体 VO 具有若干优势，例如尺度一致性、更高的精确度和增强的鲁棒性，因为它能够直接感知深度 [? ?]。

现有的 VO 方法通常同时优化相机姿态和 3D 地图点，通过透视 n 点（PnP）算法利用地图来跟踪新帧。这些方法往往缺乏对点对应关系的精确不确定性估计，并忽略结合估计器优化和分析，以提供具有理论保证的统计基础 PnP 估计。准确估计点不确定性的一个关键挑战源于常用的框架，在该框架中，初始帧被固定，所有后续的姿态和 3D 点都是相对于这一帧表示的。在这种设置下，不确定性估计本质上变得不可靠，因为通过转换链的线性误差传播变得越来越不准确。为了解决不确定性无限增长的问题，一些方法采用了一种相对表示，每个帧通过相对姿态与其前一帧连接，3D 点则固定在首次观察到它们的帧上。虽然这些方法减轻了不确定性无限增长的问题，但它们采用的同时定位与地图构建（SLAM）管道引入了姿态和 3D 点误差的时间耦合。这种耦合阻碍了精确的不确定性估计，并可能导致次优的相对姿态估计。

有趣的是，SOFT2 [?]，即在 KITTI 里程计基准测试中排名最高的方法，采用了一种相对姿态表示并专注于纯里程计，而非 SLAM，即它不涉及 3D 点的优化。虽然作者并未明确提及这一点，但这是其在 KITTI 数据集上表现出色的关键因素之一，因为里程计在最小化相对姿态误差（RPE）方面表现优异。然而，SOFT2 是一个高度专业化的解决方案，针对自动驾驶场景进行了量身定制，因为它结合了专门为地面车辆设计的技术。此外，该算法尚未开源，限制了其可获取性和重复性。

受 SOFT2 的启发，本文聚焦于一个纯视觉里程计框架，该框架强调相对位姿估计，同时不包括 3D 点优化。具体来说，我们提出了 CurrentFeature Odometry，它仅利用来自当前关键帧（KF）的三角化特征点进行基于 PnP 的相对位姿跟踪。该框架有效地打破了位姿和 3D 点误差之间的时间耦合。基于这种解耦，我们精确地建模 3D 点的不确定性，并从统计的角度优化估计器。这导致了一致的相对位姿估计，即随着点数的增加，估计值收敛到真实值。通过将一致的 PnP 估计器集成到所提出的立体视觉里程计框架中，我们展示了 CurrentFeature Odometry 不仅显著降低了 RPE，还在绝对轨迹误差（ATE）方面超越了最先进（SOTA）的 SLAM 算法。

本文的主要贡献有三点：

- 提出了一种偏差消除加权（Bias-Eli-W）PnP 估计器，该估计器考虑了不同的 3D 三角测量不确定性，并具有可证明的一致性。理论保证确保相对位姿估计随着特征数量的增加而收敛到真实值。
- 提出了一种新的立体视觉里程计框架，称为 Current-Feature Odometry。该框架仅使用来自当前关键帧的三角化点进行 PnP 位姿跟踪，有效地打破了 3D 三角化与未来位姿估计之间的时间耦合，这有助于不确定性传播分析和位姿估计中的误差抑制。
- 所提出的框架能够在特征匹配和三角测量中进行有效的不确定性分析。特征匹配误差协方差可以被一致地估计，提供适应各种特征检测和跟踪方案的灵活性。三角测量不确定性根据特征点位置进行量身定制估计。这些估计结果反过来又反馈到我们的 Bias-Eli-W PnP 中。

在 KITTI 和 Oxford RobotCar 数据集上的实验结果表明，我们的方法：1）在 KITTI 彩色序列中，RPE 达到显著改进（与第二好的相比提高了 24% ）和 ATE（平均提高了 28% ）；2）在不稳定和不可预测的机器人运动下提供可靠的定位。我们的代码将在 https://github.com/LIAS-CUHKSZ/CurrentFeature-Odometry 开源。

符号说明：对于向量 v，$[v]_i$ 表示它的第 i 个元素，而 $[v]_{i:j}$ 表示包含其第 i 到第 j 个元素的子向量。在本文中，我们遵循的惯例是，上标 $(\cdot)^o$ 表示变量 $(\cdot)$ 的真实（或无噪声）值，而 $\hat{(\cdot)}$ 表示对 $(\cdot)$ 的估计。

Guangyang Zeng, Yuan Shen, Ziyang Hong, Yuze Hong, and Junfeng Wu are with the School of Data Science, Chinese University of Hong Kong, Shenzhen 518172, China.

Viorela Ila and Guodong Shi are with the Australian Center for Robotics and School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Australia.

* Equally contributed.

II. 相关工作

大量的立体视觉里程计算法是基于几何的，利用经典计算机视觉模型中的几何刚性来估计相机运动。同时，越来越多的研究工作使用深度学习直接从原始图像序列中以端到端的方式推断运动。这些工作包括在有监督学习管道中的单目视觉里程计 [? ? ?]，无监督学习管道中的单目视觉里程计 [? ?]，以及立体视觉里程计 [?]。此外，混合方法不仅在特征检测和匹配中集成了深度学习，还在几何优化的上下文中将其用于特征跟踪和深度估计 [? ? ? ?]。在本文中，我们主要关注基于几何的立体视觉里程计（VO），该方法可以大致分为两类：基于特征的方法和直接方法。基于特征的立体视觉里程计方法通常检测并匹配连续帧之间的特征点，通过最小化重投影误差等度量来估计运动 [? ? ?]。如文献中所述 [? ?]，平移估计主要受靠近相机的 3D 点影响，而旋转估计对靠近和远离相机的 3D 点都很敏感，其中远离相机的点被定义为深度超过立体基线的 40 倍。因此，对于专门针对地面车辆的工作 [? ?]，检测接近地面的特征对于准确的平移估计至关重要。为了解决点稀疏或分布不佳的情况，有些研究结合了点和线特征 [? ?]。最近，Fontan 等人 [?] 引入了一种可以在六种不同特征类型之间无缝切换的自动化管道，提供了增强的适应性。直接方法跳过了特征检测，而是通过最小化光度误差直接优化位姿 [? ? ?]。利用图像中的像素强度可以提高运动估计的准确性。然而，对亮度恒定性假设的依赖可能使这些方法对噪声敏感。此外，通常用于每个像素的深度估计可能导致计算效率低下。为了应对这些限制，一些工作结合了基于特征和直接方法的优势，从而产生了半直接算法 [? ?]。

另一种不太常见的分类方法是基于用于运动估计的几何关系 [?]。3D-3D 方法通过立体当前帧对 3D 点进行三角测量，并使用迭代最近点算法将其与之前的 3D 点云对齐 [?]。2D-2D 方法则依赖于用于姿态跟踪的对极约束 [?]。然而，大多数研究采用 3D-2D 方法，即通过相机投影模型和待估计姿态将之前的 3D 点与当前 2D 点对齐 [? ? ?]。Cvi š ić 等 [?] 应用了纯 2D-2D 方法，并在 KITTI 立体测距基准上达到了 SOTA 性能。然而，2D-2D 算法实际上对平移距离或点分布的变化不够稳健。与 3D-2D 方法的结合会是更好的选择。

III. 系统概述

CurrentFeature 里程计系统概述如图 1 所示。在前端，当一个新的立体帧到达时，首先对单个图像进行特征检测（在我们的情况下，选择了左图像）。然后，依次执行特征跟踪和剔除离群点，在立体关键帧（KF）与当前帧（CF）的左图像之间生成点对应关系。如果成功跟踪的点数低于预定义的阈值或点的分布退化，则创建一个新的 KF。在后端，使用立体 KF 中的匹配特征进行三角化 3D 点。同时估计 2D 特征噪声的方差，并利用这些方差来计算三角化 3D 点的不确定性。借助这些元素，使用去偏加权 PnP 估计器（Bias-Eli-W）计算 CF 相对于 KF 的姿态。最后，执行滑动窗口极线束调整（BA）以优化里程计。接下来，我们将详细介绍涉及的关键模块。

IV. 前端

前端模块作为初始处理管道，处理原始图像并创建 KFs。与 OV ^{2} SLAM [?] 类似，该管道由几个关键组件组成：图像预处理、特征关联、异常值去除和 KF 生成。

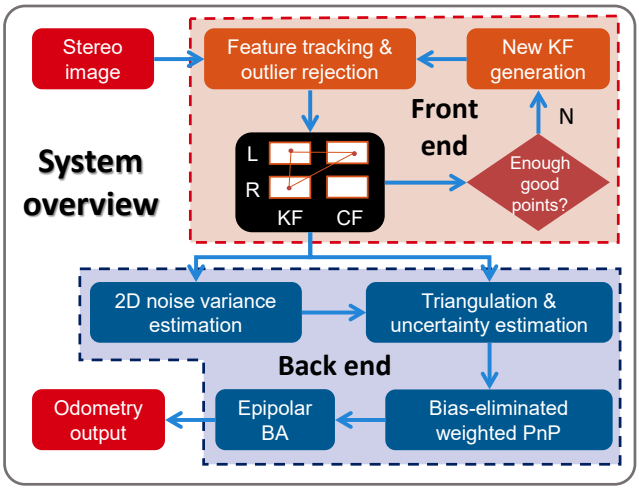


Figure 1: 系统概览。L 和 R 分别指左图像和右图像，KF 和 CF 分别表示关键帧和当前帧。

为了增强特征检测和跟踪的鲁棒性，我们首先应用 CLAHE [?] 进行图像质量增强。我们的跟踪采用滑动窗口方式，其中每个新帧相对于最新的 KF 进行定位，KF 作为局部坐标原点。具体来说，我们使用金字塔 Lucas-Kanade 光流对新检测到的特征进行跟踪，以在连续帧的左图像之间建立 2D-2D 匹配。当这些二维特征不断跟踪时，它们在 KF 中的关联三维坐标已经储存。

为了确保对潜在不匹配的稳健性，我们实施了两阶段的几何验证策略。我们首先使用五点算法结合 RANSAC 来估计本质矩阵，并从跟踪的特征中过滤掉初始离群值。其次，利用 KF 中的三角化 3D 点，我们执行 ℓ_1 范数 PnP 以获得稳健的位姿估计，并去除重投影误差最大的 10 个% 点。新的 KF 会根据两个主要标准策略性插入：跟踪数量和几何质量。系统监控成功跟踪的特征数量和平均特征位移，以决定 KF 的插入。当创建一个新的 KF 时，会使用均匀网格采样策略检测附加特征，以保持图像中一致的特征分布，确保在后续帧中光流跟踪有足够的特征。

V. 后端

A. 三角测量与不确定性估计

如图 2 所示，在特征跟踪和异常值拒绝之后，我们在 KF 和 CF 的左图像之间获得了点对应关系，记为 $\{x_i, y_i, z_i\}_{i=1}^n$，其中，$x_i$ 在 KF 的右图像，y_i 在 KF 的左图像，z_i 在 CF 的左图像。点坐标表示为归一化图像坐标。我们假设特征匹配的误差服从高斯分布 $\mathcal{N}(0, \sigma^2 I_2)$。换句话说，我们有

$$x_i = x_i^o + \epsilon_{x_i}, \quad y_i = y_i^o + \epsilon_{y_i},$$

其中 $\epsilon_{x_i}, \epsilon_{y_i} \sim \mathcal{N}(0, \sigma^2 I_2)$ 是特征匹配误差。各向同性误差假设是对具有距离和方位观测的 SLAM 的良好近似，包括针孔相机模型 [?]。这种假设已在许多现有研究中被广泛采用 [? ? ?]。根据 [?] 中的定理 1，可以通过求解特征值问题构建对 σ^2 的一致估计 $\hat{\sigma}^2$。这意味着随着点数量的增加，2D 匹配不确定性估计会收敛到真实值。

对于三角测量，我们采用线性最小二乘闭式解，而不是常用的基于视差的三角测量 [? ?] 或基于奇异值分解的方法 [?]，这种方法更具普遍性并且更适合不确定性分析。具体来说，假设 3D 点在左侧关键帧中的坐标

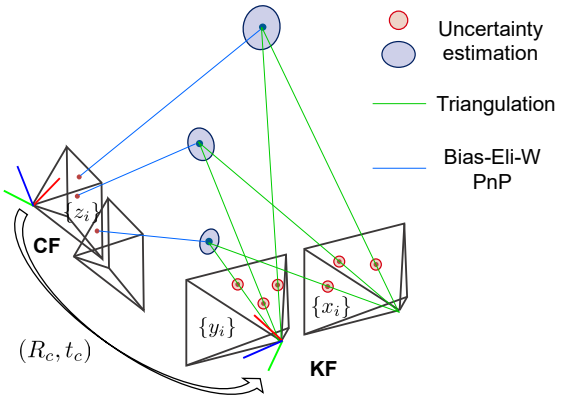


Figure 2: 帧跟踪的示意图。橙色圆圈表示特征匹配的不确定性，而蓝色椭圆表示三角测量的不确定性。

为 p_i 。我们用 x_i^h 和 y_i^h 表示齐次坐标。在无噪声的情况下，我们有 y_i^{ho} 平行于 p_i 且 x_i^{ho} 平行于 $R_0^\top p_i - R_0^\top t_0$ 。然后，通过引入误差，我们得到 $(y_i^h - \epsilon_{y_i}) \times p_i = 0$ 和 $(x_i^h - \epsilon_{x_i}) \times (R_0^\top p_i - R_0^\top t_0) = 0$ ，其中 (R_0, t_0) 是立体基线。令

$$A_i = \begin{bmatrix} y_i^{h\wedge} \\ x_i^{h\wedge} R_0^\top \end{bmatrix}, \quad b_i = \begin{bmatrix} 0_3 \\ x_i^{h\wedge} R_0^\top t_0 \end{bmatrix},$$

，其中 $(\cdot)^\wedge$ 表示将向量映射到斜对称矩阵的函数，上述平行几何约束具有紧凑的形式 $b_i = A_i p_i + \varepsilon_i$ ，其中 ε_i 是误差项。因此，最小二乘三角化公式为

$$p_i = (A_i^\top A_i)^{-1} A_i^\top b_i. \tag{1}$$

由于 A_i 和 b_i 包含噪声 ϵ_{x_i} 和 ϵ_{y_i} ，难以获得 p_i 的准确不确定性 Σ_i 。我们使用以下不确定性传播公式进行估计：

$$\Sigma_i = J_{p_i} \Sigma J_{p_i}^\top, \tag{2}$$

，其中 J_{p_i} 是 p_i 关于 $\epsilon_i \triangleq [\epsilon_{x_i}^\top, \epsilon_{y_i}^\top]^\top$ 的雅可比矩阵，$\Sigma = \hat{\sigma}^2 I_4$ 是 ϵ_i 协方差的一致性估计。 (2) 中的误差传播基于一阶近似。当噪声足够小时，高阶项变得可以忽略不计，使其能够收敛到真实协方差。我们观察到采用的前端产生了较小的特征匹配误差。例如，在 KITTI 数据集中估计的噪声标准差约为 0.2 像素（见章节 **??** ）。因此， (2) 中的 Σ_i 可以被认为是高度准确的。当噪声水平较大时，一阶近似不再给出准确的结果。这种情况下，可以应用统计学文献中开发的用于测量误差模型的基于仿真的模拟外推方法 [**? ?**] 。该方法的进一步扩展以提高计算效率及验证其有效性留待未来研究。

如图 2 所示，三角化的 3D 点的不确定性受其深度的强烈影响。一般来说，点越远，不确定性就越大。当一个点过远时，立体相机会在功能上退化为单目相机，失去对平移尺度的可观测性 [**?**] 。通过结合估计的不确定性，我们可以有效减少远处点的影响。

在三角测量的基础上，我们使用在 CF 中匹配点的二维坐标来估计 CF 相对于 KF 的姿态，记为 (R_c, t_c) 。令 $R_c = [r_1, r_2, r_3]^\top$ 、$t_c = [t_1, t_2, t_3]^\top$ 和 $\theta \triangleq \alpha[r_3^\top, r_1^\top, t_1, r_2^\top, t_2]^\top$ ，其中 α 是一个正缩放因子。参考 [**?**] 中的公式 (12)，我们可以得到关于 θ 的最小二乘估计如下：

$$\hat{\theta}^{\mathrm{B}} = (H^\top H)^{-1} H^\top d, \tag{3}$$

其中 $d = [z_1^\top, \ldots, z_n^\top]^\top$ ，

$$H = \begin{bmatrix} -z_1 \otimes (p_1 - \bar{p})^\top & I_2 \otimes [p_1^\top, 1] \\ \vdots & \vdots \\ -z_n \otimes (p_n - \bar{p})^\top & I_2 \otimes [p_n^\top, 1] \end{bmatrix},$$

和 $\bar{p} = \sum_{i=1}^n p_i / n$ 。

文献 [**?**] 中的工作没有估计和考虑三维点的不确定性。在我们的场景中，三角化的点 p_i 受到了噪声的影响，并与协方差矩阵 Σ_i 相关联。结果是，回归矩阵 H 与噪声项相关。根据估计理论，这种相关性通常导致估计量 $\hat{\theta}^{\mathrm{B}}$ 既不渐进无偏也不一致 [**?**] 。根本问题在于项 $p_i p_i^\top$ 的均值偏离了其无噪声对应物 $p_i^o p_i^{o\top}$ ，而 $p_i p_i^\top$ 是 $H^\top H$ 的一个组成部分。该偏差由 Σ_i 量化，它为估计过程引入了偏差。为消除估计量的偏差，我们需要从 $p_i p_i^\top$ 中减去 Σ_i 。这导致了偏差消除的解决方案

$$\hat{\theta}^{\mathrm{BE}} = \left(\frac{H^\top H}{n} - G \right)^{-1} \frac{H^\top d}{n}, \tag{4}$$

，其中 $G = \frac{1}{n} \sum_{i=1}^n G_i^\top G_i$ 达到 $G_i = [-z_i \otimes \Sigma_i^{\frac{1}{2}}, I_2 \otimes [\Sigma_i^{\frac{1}{2}}, 0_{3\times 1}]]$ 。从 $\hat{\theta}^{\mathrm{BE}}$ 恢复 $\hat{R}_c^{\mathrm{BE}}$ 和 $\hat{t}_c^{\mathrm{BE}}$ 可以参考 [**?**] 中的 (14)-(17)。我们的偏差消除估计量 $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 具有一致性的性质。在正式推导定理之前，我们引入 $\sqrt{n}$ 一致性的定义。

Definition 1 ($\sqrt{n}$ -Consistency in Probability)**.** *一个估计量 $\hat{\boldsymbol{\gamma}}$ 被称为 $\sqrt{n}$ -一致的估计量 $\boldsymbol{\gamma}^o$ ，如果 $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^o = O_p(1/\sqrt{n})$ ，即对于任何 $\varepsilon > 0$ ，存在有限的 M 和有限的 N ，使得对于任何 $n > N$ ，$\mathbb{P}(\|\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^o)\| > M) < \varepsilon$ 。*

“$\sqrt{n}$ 一致性”的概念包括两个含义：估计量是一致的，即随着 n 增加，它收敛到真实值；收敛速度与 $1/\sqrt{n}$ 一样快。

Theorem 1. *消除偏差的估计量 $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 是 $\sqrt{n}$ 一致的，即 $\hat{R}_c^{\mathrm{BE}} - R_c^o = O_p(1/\sqrt{n})$ ，$\hat{t}_c^{\mathrm{BE}} - t_c^o = O_p(1/\sqrt{n})$ 。*

Proof: 证明主要基于以下引理：在无噪声情况下，$(H^{o\top} H^o)^{-1} H^{o\top} d$ ，或等价地，$(\frac{H^{o\top} H^o}{n})^{-1} \frac{H^{o\top} d}{n}$ 结果为真实值 θ^o 。证明的思路是展示 $\frac{H^\top H}{n} - G$ 和 $\frac{H^\top d}{n}$ 分别收敛到 $\frac{H^{o\top} H^o}{n}$ 和 $\frac{H^{o\top} d}{n}$ 。设 $\Delta H = H - H^o$ 。首先，由于 ΔH 只包含 3D 点噪声 ϵ_{p_i} 的一阶项，其均值为零且协方差是有限的，基于引理 **??** ，我们有其次，$\Delta H^\top \Delta H$ 包含 ϵ_{p_i} 的一阶和二阶项。对于二阶项 $\epsilon_{p_i} \epsilon_{p_i}^\top$ ，我们可以通过减去 Σ_i 来实现零均值。可以验证 $-G$ 实际上完成了这一过程。因此，根据引理 **??** ，我们得出结论最终，通过结合 (**??**) 和 (**??**) ，我们得到也就是说，$\hat{\theta}^{\mathrm{BE}}$ 是 $\sqrt{n}$ 的一致估计量。由于从 $\hat{\theta}^{\mathrm{BE}}$ 恢复 $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 仅涉及连续函数，参见 [**?**] 中的方程 (14)-(17)，$(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 也是 $\sqrt{n}$ 的一致估计量，从而证明完成。 ■

请注意，在去偏估计量 (4) 中，没有对旋转矩阵施加约束。因此，尽管该估计量是一致的，它可能不具备最小方差。因此，我们进一步将 $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 作为初始值，并执行加权的 PnP 迭代细化。设 $h(\cdot)$ 表示针孔相机投影模型，即，对于 $p \in \mathbb{R}^3$ ，$h(p) = [p]_{1:2}/[p]_3$ 。分配给第 i 个点的权重是 $\bar{\Sigma}_i^{-\frac{1}{2}}$ ，其中 $\bar{\Sigma}_i = J_{h_i} \Sigma_i J_{h_i}^\top$ ，并且 J_{h_i} 是 $h(\hat{R}_c^{\mathrm{BE}} p_i + \hat{t}_c^{\mathrm{BE}})$ 关于 p_i 的雅可比矩阵。因此，第 i 个点的加权残差是 $Re_i = \bar{\Sigma}_i^{-\frac{1}{2}} (h(R_c p_i + t_c) - z_i)$ ，我们解决

以下问题：

$$(\hat{R}_c, \hat{t}_c) = \arg\min_{(R_c,t_c)} \sum_{i=1}^{n} \rho_\delta \left(Re_i \right), \tag{5}$$

，其中 ρ_δ 是截断最小二乘（TLS）稳健核函数。问题 (5) 使用 LM 算法求解。由于初始值 $(\hat{R}_c^{\mathrm{BE}}, \hat{t}_c^{\mathrm{BE}})$ 的一致性和 LM 算法在全局最小值附近的二次收敛速度，当点数 n 较大时，单次 LM 迭代就足以实现最小方差，这在计算上是高效的 [**?**] 。

生成新的关键帧（KF）时，我们执行一个局部 BA，其中包括最新的两个 KF 和中间的普通帧（OFs）。如图 3 所示，需要优化的参数是六自由度的相对姿态 $\xi_k \in \mathbb{R}^6, k = 1, \ldots, K+1$ 。每个 ξ_k 由三个欧拉角和一个平移向量组成。由于在 OF 中跟踪特征仅涉及左图像，因此它们的右图像在局部 BA 中没有被使用。值得注意的是，大多数方法通过施加重投影约束，来在局部 BA 中同时优化相机姿态和可见 3D 点的坐标。然而，在里程计背景下，3D 点是潜在变量，并不是主要关注点。将它们包括在优化过程中不一定能提高姿态估计的准确性，但确实引入了额外的计算开销。作为替代方案，极线约束不涉及 3D 点，直接通过相对姿态连接两个匹配的特征点。假设相对姿态为 (R, t) ，对于两个匹配的特征点 x 和 y ，极线约束是

$$y^{h\top} E x^h = 0, \tag{6}$$

，其中 $E = t^\wedge R$ 是本质矩阵。令 $l = Ex^h$ 被称为极线。按照 SOFT2 [**?**] ，我们使用点到极线的距离作为残差：

$$Re = \frac{y^{h\top} l}{\| [l]_{1:2} \|}. \tag{7}$$

在 [**?**] 中，滑动窗口内的帧仅涉及左图像参与 BA 优化。这种仅依赖于时间刚性而没有基线约束的方法，无法细化尺度。在本文中，通过额外结合基线引起的刚性（使用关键帧的右图像），可以同时细化所有六个自由度。

令 $\xi = [\xi_1^\top, \ldots, \xi_{K+1}^\top]^\top$ 。我们搜索涉及的任何图像对的极线约束并解决以下问题：

$$\hat{\xi} = \arg\min_{\xi} \sum_{j \in \mathcal{E}} \sum_{i=1}^{n_j} \rho_\delta \left(Re_{i,j} \right), \tag{8}$$

，其中 $\mathcal{E}$ 是所有图像对的集合，n_j 是第 j 对中的特征匹配数量。问题 (8) 使用 LM 算法解决。

VI. 仿真与实验

我们进行模拟有三个目的：1）验证我们的去偏差 PnP 估计器是一致的（定理 **??** ）；2）证明仅使用从当前关键帧三角化的 3D 点可以提高里程计精度；3）展示极线 BA 的效果。在整个模拟过程中，我们将立体相机的基线设置为 $R = I$ 和 $t = [0.5, 0, 0]^\top$ 米。我们假设相机可以看到深度在 $[1, 40]$ 米内的 3D 点。焦距设置为 $f = 800$ 像素，主点偏移为 $u_0 = 320$ 像素，$v_0 = 240$ 像素。除非另有说明，特征匹配噪声的标准差为 1 像素。

1) 消除偏差的 PnP 的一致性： 回忆一下，PnP 跟踪估计的是 CF 和最新的 KF 之间的相对姿态。我们让特征对应的数量分别从 30、60、120、240、480 和 960 变化。对于每种情况，我们进行 1000 次蒙特卡洛测试来计算均方根误差 (RMSE)，每次测试中姿态和特征位置是随机生成的。我们让 $\sigma = 0.5, 1$ 像素，分别。结果绘制在图 4 中。噪声标准差、旋转、和平移的 RMSE 在双对数图中都随着特征点数量线性减小，验证了我们估计器的一致性。此一致性特性确保只要特征丰富，我们的 PnP 跟踪就能作为后续迭代优化的良好初始值。

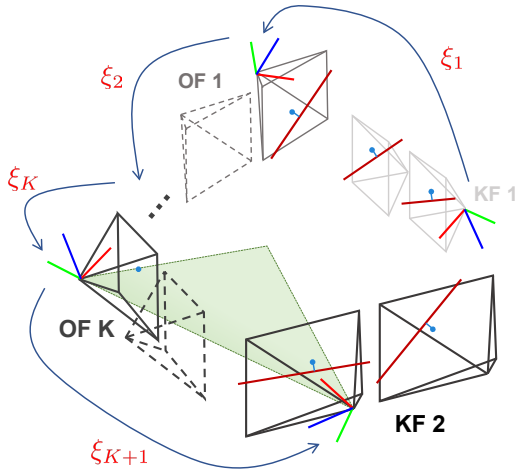
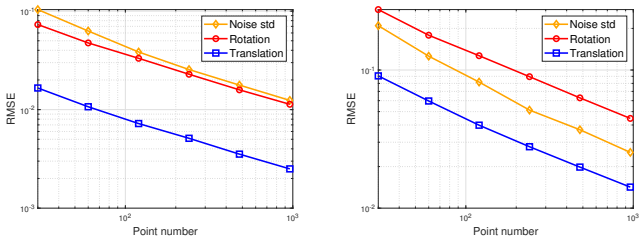


Figure 3: 对极束调整的示意图。对于两个关键帧，我们使用立体图像，而对于光流帧，我们仅使用左图像。在束调整中，包含匹配特征的每对图像之间的对极约束，并优化点到对极线的距离。



(a) $\sigma = 0.5$ 像素　　(b) $\sigma = 1$ 像素

Figure 4: 消除偏差的 PnP 估计器的一致性。噪声标准差、旋转和位移的 RMSE 单位分别为像素、° 和米。

2) 利用最新的 KF 进行姿态跟踪的合理性： 我们部署了两种轨迹：直线和圆。每种轨迹包含 500 帧。一般情况下，每帧是关键帧。3D 点均匀且随机地分布在空间中，以使每张图像中大约有 100-200 个点可见。离群点比例设置为 2 % 。对比的方法中，我们假设它们分别可以使用最新的 $m = 2, 3$ 个关键帧。在这个模拟中不包括 BA。我们运行 10 次蒙特卡洛测试来计算平均相对位姿误差（RPE）和绝对轨迹误差（ATE） [**?**] 。结果如表 I 所示。我们还绘制了线性轨迹中每帧的平均位姿误差，如图 5 所示。我们发现，使用最新的关键帧具有最佳精度，而整合来自前三个关键帧的更多 3D 点反而显示出最差精度，尤其在 RPE 方面。使用更多关键帧的位姿误差也波动得更加严重。这些结果表明，从先前位姿误差传播来的 3D 点误差会抵消因特征数量增加而获得的优势，甚至可能导致里程计性能的恶化。

除了 PnP 位姿跟踪，该仿真还结合了具有四帧滑动窗口的极线束优化。实验设置与之前的模拟保持一致，并测试了两种类型的轨迹。ATE 和 RPE 的比较在表 I 中展示。可以观察到，结合束优化显著改善了 ATE，虽然在 RPE 方面略有代价。这是因为没有束优化时，偶尔大的相对位姿

Table I: 使用不同数量的 KFs 进行 PnP 跟踪的平均错误。t 和 R 的误差单位是 m，° 。

Trajectory	KF used	ATE		RPE	
		t	R	t	R
Line	latest	1.161	2.452	0.046	0.048
	two	2.193	0.947	0.176	0.175
	three	2.858	1.859	0.170	0.163
	latest+BA	1.068	1.307	0.044	0.057
Circle	latest	20.313	6.379	0.084	0.084
	two	55.840	18.784	0.289	0.316
	three	63.209	20.355	0.387	0.395
	latest+BA	9.415	1.398	0.170	0.080

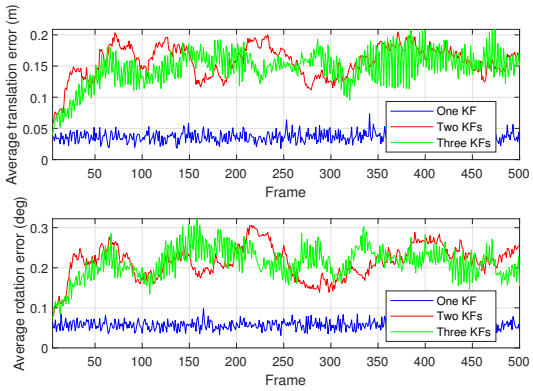


Figure 5: 当使用不同数量的关键帧进行线性轨迹时，每帧的平均姿态误差。

误差会沿着轨迹传播，降低其准确性并增加 ATE。通过校正这些偶尔的大偏差，极线束优化增强了轨迹的一致性和整体准确性。

我们在两个公开可用的数据集上评估了 CurrentFeature Odometry 的性能：KITTI 和 Oxford RobotCar 。为了比较，我们包括了 ORB-SLAM3 和 OV SLAM ，这两个是 KITTI 数据集上排名前两位的开源立体视觉里程计方法。为了确保公正的比较并专注于里程计性能，两个方法中的闭环检测模块都被禁用了。对于评估指标，我们采用了在 ORB-SLAM3 和 OV SLAM 中也使用的 ATE RMSE 和 RPE RMSE。

3) KITTI 数据集： KITTI 数据集是在各种户外驾驶场景中捕获的，包括城市、郊区和农村环境。它提供了同步和校正的立体图像对，既有彩色格式，也有灰度格式，录制帧率为 10 Hz。该数据集还包括从 OXTS3003 GPS/IMU 单元中获取的 11 个训练序列（seq00-10）的真实位姿，支持轨迹评估和基准测试。 [**?**] 中已经证明立体设备不是严格刚性的，随着车辆移动右摄像头相对于左摄像头的相对旋转发生变化。因此，在极线 BA 中，我们还优化设备的旋转。在所有序列中参数设定保持不变。PnP 和极线 BA 的 TLS 阈值分别设置为 5×10^{-5} 和 3×10^{-4} 。每次极线 BA 中使用的光流数量最多为 5 。

ATE 和 RPE 的比较结果汇总在表 II 中。我们观察到平均来看，所有算法在灰度图像上的表现优于彩色图像。这主要是因为彩色图像更容易受到严重曝光变化的影响，从而导致特征关联或跟踪不准确。如预期般，我们提出的 CurrentFeature Odometry 在 RPE 方面始终优于 ORB-SLAM3 和 OV ^{2} SLAM，灰度 08 序列除外。与第二好的方法相比，我们的算法在彩色序列上平均 RPE 降低了 24 %。这一提升归因于位姿和 3D 点误差的解耦，以及我们精心设计的 Bias-Eli-W PnP 估计器。有趣的是，我们的方法在 ATE 性能上也优于 SOTA SLAM 通道。具体而言，在 10 个彩色序列中，它提供了 9 个序列中最高的准确性，与第二好的方法 ORB-SLAM3 相比，平均 ATE 提高了 28 %。在灰度情况下，CurrentFeature Odometry 在一半的序列中取得了最佳性能，并在另一半中排名第二。我们的算法对彩色图像挑战的敏感性降低，归因于 ℓ_1 范数和 TLS 技术提供的增强鲁棒性。我们方法优异的 ATE 性能在很大程度上归功于滑动窗极线 BA 的集成，这改善了整个轨迹的一致性。轨迹可视化展示在图 6 中。由于极低的漂移速度，CurrentFeature Oodmetry 的估计轨迹在这些大规模场景中与真实轨迹匹配良好。

4) 牛津机器人汽车数据集： 牛津数据集是在牛津市中心使用一个 Point Grey Bumblebee XB3 三目立体摄像头采集的，图像分辨率为 1280 × 960 像素，帧率为 16 Hz。车辆的真实轨迹是通过一个以 50 Hz 运行的 GPS/INS 系统记录的。然而，GPS 接收和融合后的 INS 解决方案的精度根据环境显著变化 [**?**] 。类似于 SOFT2 [**?**] ，我们选择了该数据集中的四个片段，这些片段提供了有效的传感器数据、可靠的 GPS/INS 轨迹和高质量图像用于评估。此外，为了去除无关的视觉信息例如天空和车辆底盘，我们从图像的顶部和底部各裁剪了 140 像素。所有四个序列中的参数设置与 KITTI 试验中的参数设置相同。ATE 和 RPE 的比较结果在表中展示。我们的算法在所有序列的这两个指标中都取得了最佳表现。值得注意的是，与 KITTI 数据集相比，牛津数据集由于诸如强曝光等问题带来了更大的挑战。因此，ORB-SLAM3 在两个序列中无法估计完整的轨迹。相比之下，我们 ℓ_1 -TLS 增强方法显示出卓越的鲁棒性，在第四个序列中实现了显著更小的 ATE。

与依赖恒速（CV）运动模型进行位姿预测的 OV

A. 抗不稳定运动的鲁棒性

SLAM 不同，我们的方法通过使用 3D-2D 对应关系的 PnP 直接估计相机位姿，而不对运动模型进行任何假设。当实际运动偏离匀速运动时，OV ^{2} SLAM 中的 CV 假设可能会引入额外误差。为了验证这一点，我们分析了来自 KITTI 彩色序列 00 的一个具有挑战性的片段（帧 226-230）。如图 ^{2} 所示，在短短的 0.5 秒时间内（5 帧），这个片段的线速度从 5.7 m/s 增至 6.2 m/s，并且角速度波动至 5 deg/s。在这样的动态条件下，OV 7 SLAM 的 CV 预测显示出一些不稳定性，位置误差达到 0.05 m，旋转误差累积达 0.4 度。相比之下，我们的方法通过直接求解 PnP 而不进行运动假设，在这个具有挑战性的片段中保持了相对稳定的误差。重投影结果将三角化的 3D 点使用真实位姿、我们的 PnP 位姿和 CV 预测位姿投影到 CF 中，也显示了我们 PnP 估计器的优越性。在 BA 优化后，我们的方法实现了位置误差低于 0.01 m 和旋转误差低于 0.05 度，显示出相较于 OV ^{2} SLAM 的初始和最终结果的精度提升。

在本文中，我们重新审视了立体视觉里程计，并提出了一个一致的支持 PnP 的框架，CurrentFeature Odometry。我们的方法利用当前关键帧的三角化点进行 PnP 跟踪，有效地打破了姿态和 3D 点误差之间的耦合。由于这种解耦，我们根据统计理论和误差传播公式准确地建模了 3D 点的不确定性。这些不确定性随后在优化过程中用于偏差消除和加权。我们证明了去偏差的 PnP 姿态估计以 $1/\sqrt{n}$ 的速度收敛到真实值，这为后续迭代精细化提供了良好的初始值。此外，加入极线束调整进一步增强了轨迹的一致性。

Table II: 在 KITTI 数据集中不同序列间 ATE 和 RPE 的比较。ORB3 表示 ORB-SLAM3，OV2 表示 OV ^{2} SLAM。用 蓝色加粗 标出的值表示最小值， 蓝色 中的值表示次小值。

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**

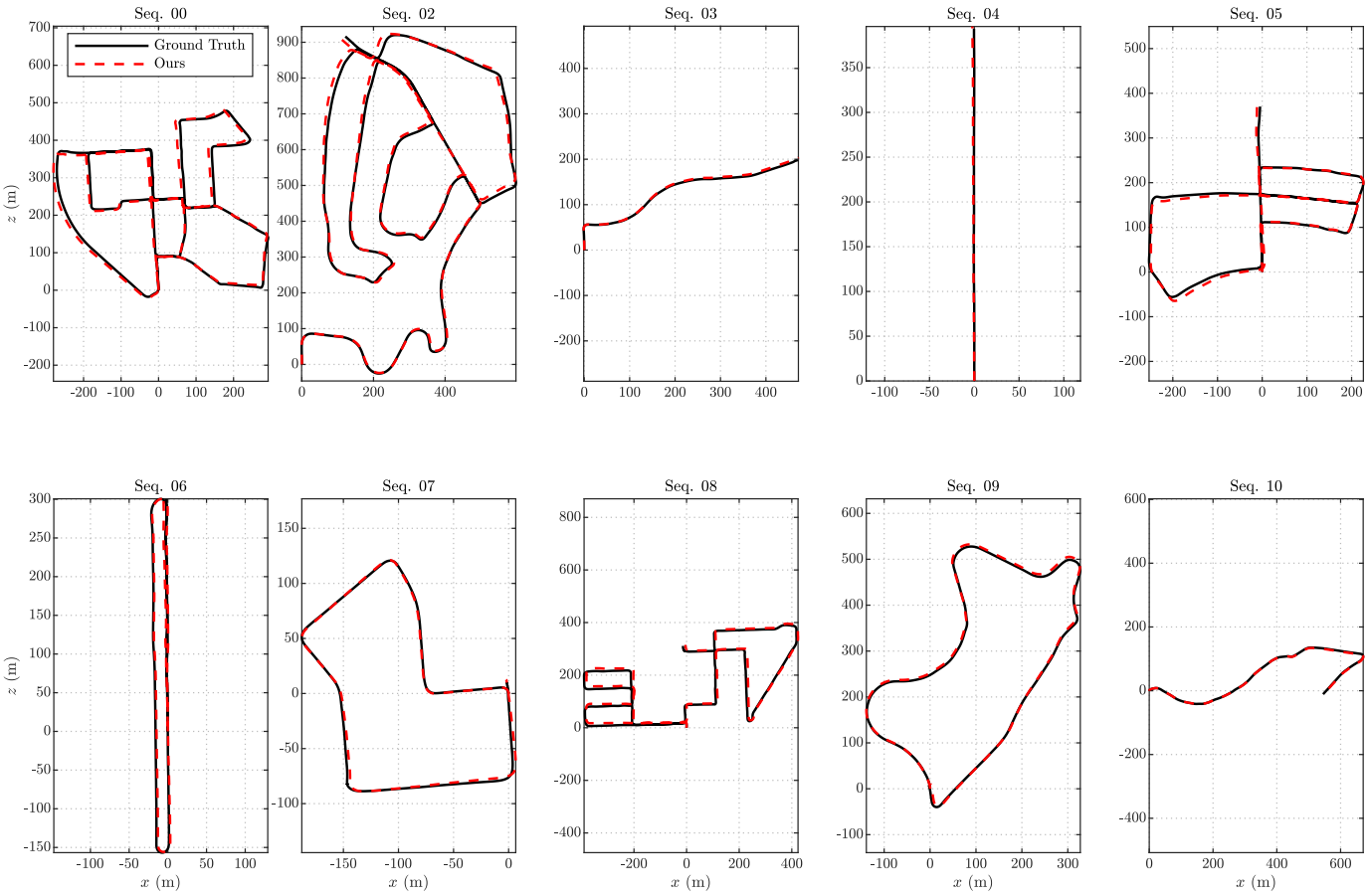


Figure 6: KITTI 序列在灰度图像上的顶视图轨迹。

Table III: 对比 Oxford 数据集中不同序列的 ATE 和 RPE。用 蓝色粗体 突出显示的数值代表最小值，而 蓝色 中的数值表示第二小值。符号 - 表示无法估计整个轨迹。

Sequence	Start frame	Stop frame	Length (m)	ATE (m)			RPE (m)		
				ORB3	OV2	Ours	ORB3	OV2	Ours
2014-05-14-13-59-05	1400075963932033	1400076150344330	945.49	34.5311	34.6471	**33.7454**	0.5126	0.5146	**0.5061**
2014-05-19-12-51-39	1400503987511809	1400504194609323	931.58	-	28.9658	**28.3328**	-	0.4761	**0.4682**
2014-05-19-13-20-57	1400505700597042	1400506098919265	1989.92	-	72.4767	**64.9658**	-	0.5768	**0.5680**
2014-11-14-16-34-33	1415985043842007	1415985331240621	1531.52	24.1157	15.5798	**6.2296**	0.5773	0.5872	**0.5674**

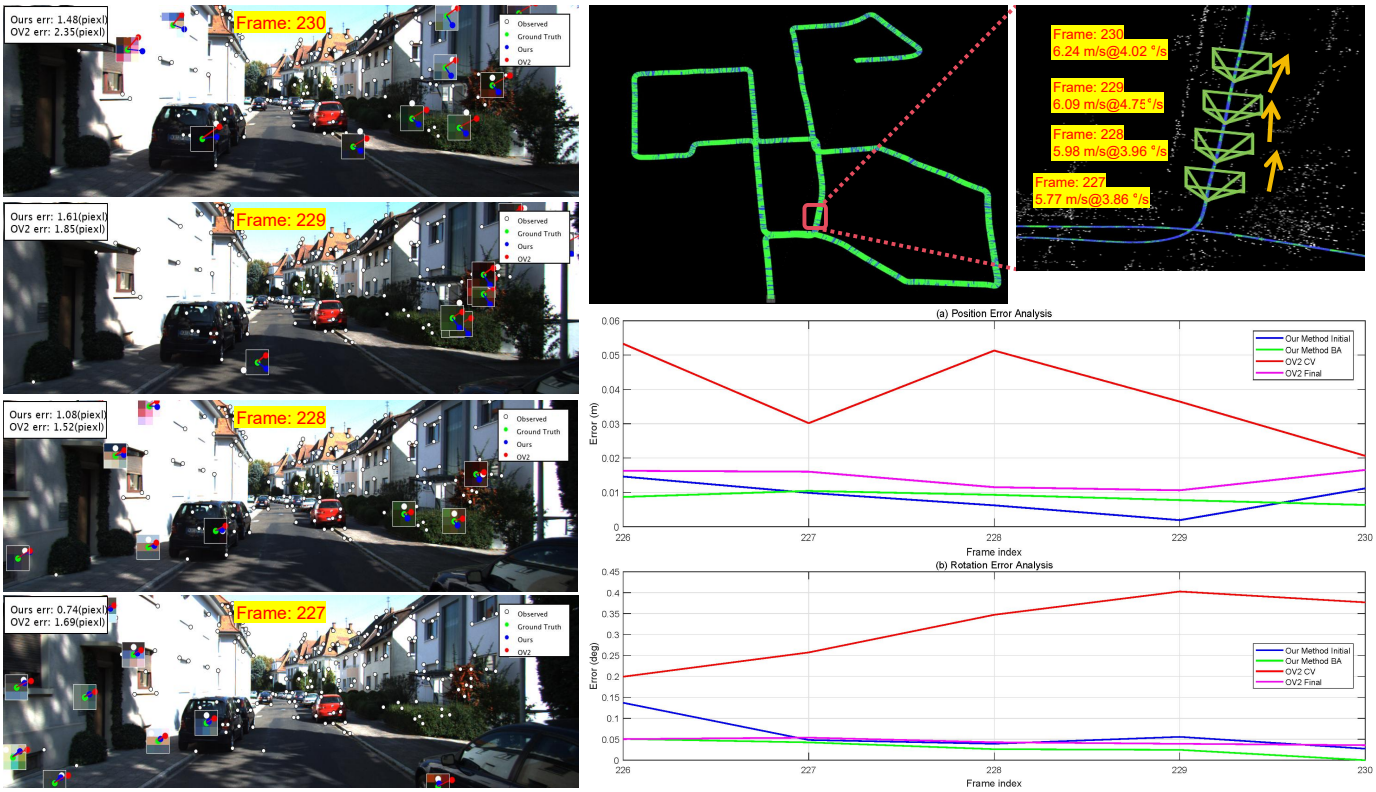


Figure 7: 对具有变化速度的挑战性段落进行误差分析。该图由四个关键部分组成：左侧是带有重投影结果的顺序图像，中间顶部显示了带有放大部分的轨迹鸟瞰图，右上角展示了带有速度标注的相机运动的详细可视化，底部绘制了定量误差。

实验结果表明，CurrentFeature Odometry 在大规模环境中表现优于现有的最好算法，在 ATE 和 RPE 方面表现出色，同时在不确定运动中展现出更大的鲁棒性。