ROVO：面向低空和中等高度无人机的鲁棒仅旋转视觉里程计

Xiaokai SONG[1,2], Li ZHANG[1,2*], Xichao TENG[1,2],Erting PAN[1,2],Yibin YE[1,2],Kun WANG[3], Leqi LIU[1,2] & Guangcheng DAI[1,2]

[1]国防科技大学航空航天科学与工程学院，中国长沙 *410000*[2]国防科技大学湖南省图像测量与视觉导航重点实验室，中国长沙 *410000*[3]湖南先进技术研究院，中国长沙 *410000*

arXiv:2611.32475v1 [cs.RO] 24 Nov 2026

摘要 对于在低空和中等高度飞行的无人飞行器（UAV），图像噪声会干扰单目视觉导航中的旋转和平移估计，即使微小扰动也可能导致平移方向出现较大误差。受远距离视觉感知中角度线索的启发，我们针对单目航空影像提出了仅旋转视觉里程计（ROVO）。ROVO根据匹配图像估计成对旋转，在重叠窗口上对其进行鲁棒平均，并通过共享帧对连续窗口进行对齐。通过在轨迹状态中仅保留姿态，它可以防止不确定的地标深度传播至后续估计。在轨迹仿真中，滑动窗口式ROVO的平均姿态方向误差仅为传统SLAM基线的10.7%。在真实飞行数据上，采用原始Bell412 设置时，ROVO的旋转精度与ORB-SLAM3和DPVO相当。当视场较窄或飞行高度较高时，与各项比较中的次优方法相比，ROVO将相对旋转均方根误差降低了19.7–69.4%。代码和仿真工具可在此获取https://github.com/sssxxxkkkk/ROVO.

Keywords 无人飞行器、视觉里程计、旋转平均、姿态估计、视觉导航 n

1 引言

无人机依靠位置和姿态估计来执行航空测量、建图、配送和应急响应等任务 [1–3]。一种常见的导航系统结合了全球导航卫星系统和惯性测量单元的测量值 [4]。在城市峡谷、山地地形和茂密森林中，卫星信号可能会受到遮挡。在卫星信号长时间中断期间，惯性导航解中的误差会不断累积，因此需要使用机载相机作为运动信息的额外来源 [5]。

视觉导航可以提供绝对参考或相对运动估计 [6, 7]。场景匹配导航利用地理配准地图对机载图像进行定位 [8, 9]。其精度取决于地图的可用性，以及匹配在不同光照条件、不同季节或不同视角下获取的图像的能力 [10]。

相对视觉导航（RVN）方法，包括视觉里程计和同时定位与建图（SLAM） [11],，无需预先建立地理配准地图 [12]即可估计图像间的相机运动。视觉导航已应用于室内点目标导航 [13] 以及低空航空视觉惯性导航 [14]。将相对视觉导航扩展到更高高度会带来两个几何问题。对于相机位置之间的固定距离，更远的地形会产生更小的角视差。更长的焦距可以保留地面细节，但在传感器尺寸一定时会缩小相机的视场（FoV）。当这两个条件同时出现时，平移估计可能会对特征坐标中的微小误差十分敏感 [15]。随后，平移和深度误差可能通过三角测量传播，并进一步影响相机定位 [16]。

在远距离场景中，旋转仍然可以测量，因为观察方向的变化仍会使图像特征发生位移。这一观察也与利用视觉流感知自运动的相关研究相联系 [17]。因此，我们针对低空和中等高度的无人机飞行，提出了仅旋转视觉里程计（ROVO）。图 1（a）展示了促成本项工作的 120–3000 m 高度范围。（b）和（c）对比了传统建图流程与 ROVO。对于每个新帧，重叠视图会对其姿态提供相互独立的预测。ROVO 利用这些预测之间的一致性，在经过局部验证的双视图估计中进行选择。只有选定的相对旋转会进入滑动窗口，从而避免传播不确定的地标深度或初始化全局平移尺度。

本文有三项贡献。

- 我们提出了 ROVO，这是一种用于单目姿态估计的滑动窗口框架。该框架在每个窗口内采用鲁棒旋转平均，通过共享帧上的共识对齐重叠窗口，并使用

* 通讯作者（电子邮箱：lizhang08@nudt.edu.cn）

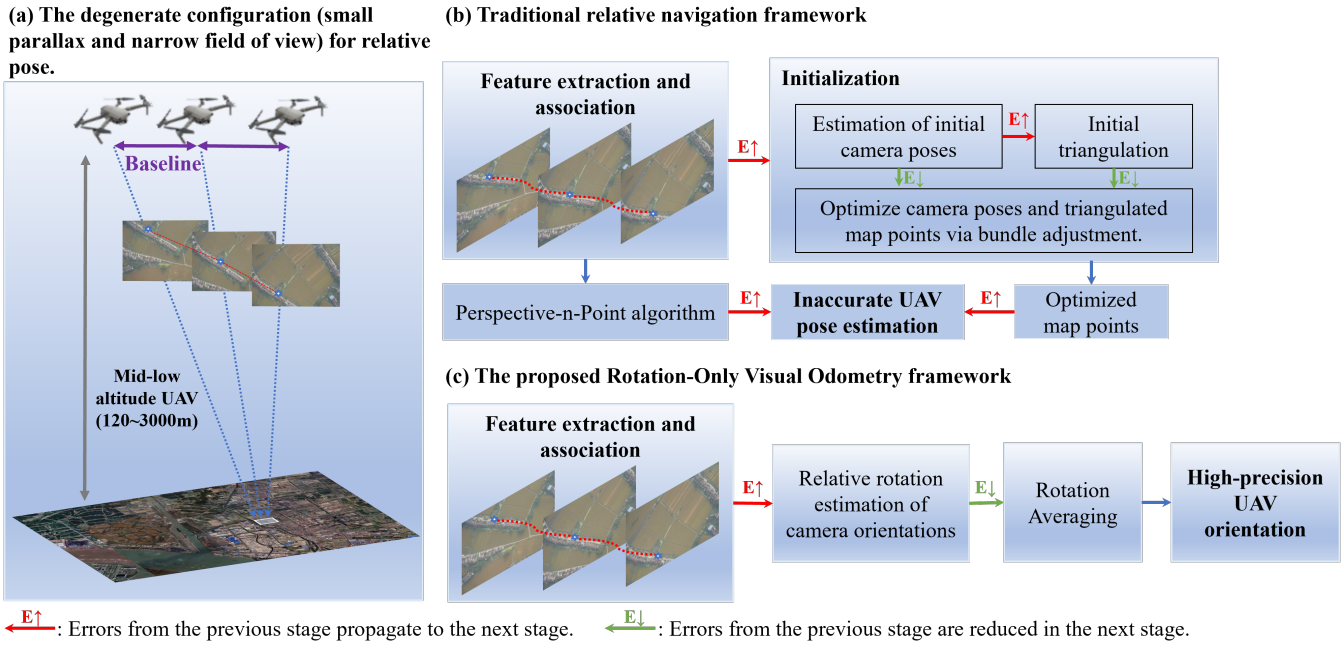

图1 航空影像中的相对位姿估计。(a) 相对于场景深度较短的基线几乎不产生视差，狭窄的视场限制了视线方向分布。(b) 在建图流程中，初始位姿的误差会影响三角测量地标以及后续的透视-*n*-点（PnP）位姿估计。(c) ROVO为每个被接受的图像对估计一个旋转，并利用旋转平均恢复一系列相机方向。

有界角度修正。

- 我们提出一种仅旋转估计器，用于选择amon g双视图旋转候选，衡量重叠视图对当前帧姿态预测的接近程度。该选择会适应可用的图像几何，仅使用图像测量值和先前估计的旋转。

- 我们评估观察几何和图像噪声如何在仿真中影响旋转与平移方向估计。在MUN-FRL和VPAIR上的实验分别在狭窄视场和更高飞行高度下验证了ROVO，表明其旋转精度得到提升，并具有较高的跟踪覆盖率。

第 2 节回顾相关工作，第 3节介绍 ROVO。第4 节通过仿真考察成像几何和噪声的影响。第 5节评估飞行数据集，第6 节讨论结果的适用范围。

2 相关工作

2.1 相对视觉导航

相对视觉导航通过关联重叠视图中的观测来估计相机运动，而无需环境的地理配准地图 [12]。直接法和半直接法利用图像强度 [18, 19]，而基于特征的系统则检测并匹配具有显著特征的图像点 [16, 20, 21]。视觉同步定位与建图系统联合估计相机位姿和地图 [22, 23]，，而视觉惯性系统则结合视觉观测与惯性测量，以支持在视觉信息较弱的时间段进行运动估计 [16,24,25]。学习式里程计方法则提供了另一种根据图像序列估计运动的方式 [26]。这些方法推动了多种平台上的视觉导航发展，但其估计结果仍取决于观测场景几何结构所包含的信息。

航空飞行会随高度、基线和相机视场改变这种几何关系。Liu 和 Shen 研究了高空单目视觉惯性估计的初始化和传感器融合 [27]。Warren 和 Upcroft 研究了长距离高度飞行中的双目视觉里程计，包括相机基线和场景深度的影响 [28,29]。后续研究探讨了航空视觉惯性里程计中的深度不确定性 [30]，，而低空飞行数据集则支持在近距离条件下评估视觉惯性导航 [14]。综合来看，这些研究表明，性能应结合飞行几何关系进行考量。在本文中，仿真独立改变高度和焦距，飞行实验则在原始视场和较窄视场的图像条件下考察旋转精度。

2.2 小视差几何与旋转估计

传统的单目建图流程通过图像对应关系估计相对位姿，对 3D 路标点进行三角测量，并通过束调整优化位姿和路标点 [16]。当基线相对于场景深度较小时，平移引起的图像运动很小。较窄的

视场也会限制观测方向的分布。在这些条件下，微小的图像坐标误差可能导致估计的平移方向和地标深度发生较大变化，进而影响后续的位姿估计 [15, 29, 31]。

双视图几何已通过多种相对位姿求解器进行研究，包括八点法、七点法和五点法 [15, 32, 33]。直接进行帧间旋转优化，是根据匹配视线方向估计旋转的另一种方法 [34]。由于求解器的表现取决于图像噪声和观察几何，单个图像对上的误差较低，并不能保证长轨迹上的方向保持一致。因此，有必要同时评估成对旋转估计及其在各帧间的累积效果。

仅旋转方法在估计姿态时，减少了对三角测量场景结构的依赖 [35]。旋转平均通过约束成对旋转之间的一致性，在统一坐标系中恢复相机方向；鲁棒方法则能减轻不一致测量值的影响 [36]。一种层次化方法利用环一致性来初始化鲁棒旋转平均 [37]。仅旋转光束法平差是一种相关方法：消除成对平移方向后，根据图像视线残差优化方向 [38]。ROVO 在这类研究的基础上采用在线滑动窗口。该方法在局部验证备选旋转估计，根据重叠视图之间的一致程度筛选边测量，并通过鲁棒平均合并筛选出的旋转。平移可用于局部候选项检查，但轨迹状态中只传播姿态。

3 方法论

3.1 系统概述

ROVO 将标定单目图像流映射为带时间戳的相机方向（图 2）。每个当前帧都会与缓冲的关键帧进行匹配，以生成相对旋转候选项。局部几何验证会检验每个候选项，而联合选择则利用历史视图间的一致性来选定旋转测量值。在检查选定的边是否一致后，ROVO 会决定是否将当前帧保留为关键帧。随后，旋转平均会估计窗口内的姿态。共享帧将窗口与现有轨迹对齐，鲁棒融合则更新存储方向，以供后续帧使用。

该缓冲区包含活动关键帧及其图像特征。当达到配置的长度后，ROVO 会求解相关旋转图，对解进行对齐和融合，并通过移除指定数量的最旧关键帧及其关联边来推进窗口。最终可能较短的窗口会在序列末尾进行求解。

具有已接受旋转边的帧可以属于该图并获得姿态，即使它未保留在关键帧缓冲区中。

我们用 $\mathbf{R}_{cw,\,i}$ 表示从世界坐标系到相机 i 的旋转，用 $\mathbf{R}_{wc,\,i} = \mathbf{R}_{cw,\,i}^{\top}$ 表示其逆变换。每个跟踪段都有自己的世界参考系。相对旋转 $\mathbf{R}_{ij}$ 将相机 j 映射到相机 i，带帽符号表示估计相对旋转。上标 loc 和 prev 分别表示当前窗口参考系中的估计值，以及通过融合先前处理的窗口得到的存储估计值。在为当前帧选择候选项期间，这些存储的历史姿态保持不变。

所有旋转矩阵均属于 SO(3)。映射 Log 返回以弧度表示的主旋转向量，而 Exp 将旋转向量映射为其旋转矩阵。我们使用角距离

$$d(\mathbf{A},\mathbf{B}) = \frac{180}{\pi}\left\|\mathrm{Log}(\mathbf{A}^{\top}\mathbf{B})\right\|_2 \tag{1}$$

来以度为单位表示候选项之间的差异。涉及旋转向量的方程使用弧度，并相应地转换以度为单位的设定值。

3.2 特征提取与匹配

对于每个当前帧，在其灰度图像中检测尺度不变特征变换（SIFT）关键点 [39] 。如果检测到的关键点超过设定的特征预算，则保留一个空间均匀子集，以保持图像覆盖范围。

描述子匹配将这些特征与每个缓冲关键帧关联起来。我们保留数量最多的一对一子集，其中每个图像位置仅使用一次。至少包含 15 个独立匹配的图像对将进入相对旋转估计阶段。由此得到的对应关系既提供拟合数据，也提供用于验证旋转候选解的留出测量值。

3.3 相对旋转候选解

对于在帧i和j中观测到的场景点，设其齐次像素坐标为 $\tilde{\mathbf{u}}_i$ 和 $\tilde{\mathbf{u}}_j$，相机内参矩阵为$\mathbf{K}$。相应的单位视线向量可通过对$\mathbf{K}^{-1}\tilde{\mathbf{u}}_i$和$\mathbf{K}^{-1}\tilde{\mathbf{u}}_j$进行归一化得到。将它们分别记为$\mathbf{b}_i$和$\mathbf{b}_j$，则极线约束为 [31]

$$\mathbf{b}_i^{\top}[\mathbf{t}_{ij}]_{\times}\mathbf{R}_{ij}\mathbf{b}_j = 0, \tag{2}$$

其中，$\mathbf{t}_{ij}$ 表示从相机 i 到相机j的平移方向，并以相机 i 为表达坐标系；$[\cdot]_{\times}$ 表示叉积矩阵。因此，本质矩阵为 $\mathbf{E}_{ij} = [\mathbf{t}_{ij}]_{\times}\mathbf{R}_{ij}$。平移方向支持局部验证，而轨迹状态仅携带姿态。

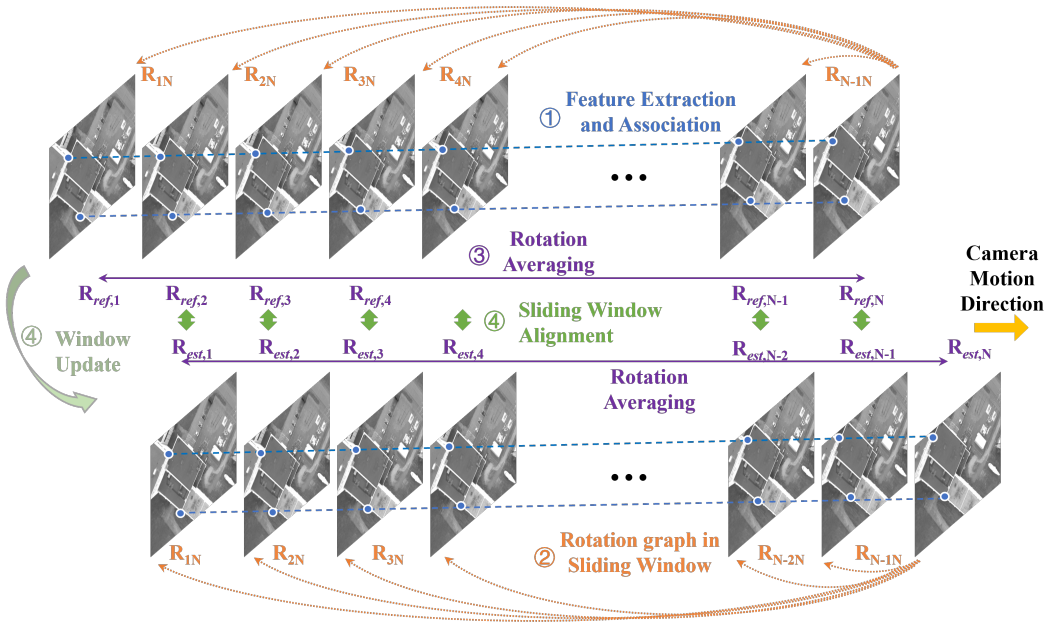

图 2 ROVO 框架。每条 橙色边表示一个选定的相对旋转。蓝色特征匹配生成候选项，而 pur ple 方向通过旋转 平均进行估计，并在各窗口间对齐。

对于具有 n 个独立匹配的匹配对，当$n \geqslant 23$时，我们预留一个空间分布的验证集，其中包含 $m = \max(8, \lfloor n/5 \rfloor)$ 个匹配。否则$m = 0$。RANSAC 使用其余匹配拟合本质矩阵，并要求至少有八个拟合内点。对于当前帧 k 和历史关键帧 i，Nist´er 在 OpenGV [40]中的五点法求解器 [15]提供一个基准旋转 $\hat{\mathbf{R}}^{(0)}_{ik}$。RANSAC 初始化得到的本质矩阵分解结果和归一化线性估计结果提供了额外的候选项。我们根据支持度对这些候选项进行排序，并以中位重投影误差打破平局。选定的基准旋转初始化 Kneip 和 Lynen [34], 的迭代特征值求解器，该求解器根据相同的拟合视线方向生成精化后的候选项 $\hat{\mathbf{R}}^{(1)}_{ik}$。

两个候选项均在相同的拟合集和验证集上进行测试。在视差较弱时，若某个匹配的去旋转视线方向在对应于两个像素的角度容差内一致，则该匹配支持该候选项。当从两个空间子集估计出的平移方向在 15°（至多相差一个符号）范围内一致时，我们还会测试正深度，并以两个像素为阈值检查双视图重投影误差。拟合数据用于确定平移符号。一个局部有效的候选项必须至少获得八个拟合内点的支持，且支持数至少占拟合内点的一半。当 $m > 0$ 时，它还必须至少获得 $\min(8, \lceil m/2 \rceil)$ 个验证匹配的支持。其支持数为拟合匹配与验证匹配中支持该候选项的匹配总数。

3.4 旋转测量的联合选择

局部图像支持度可能使两个候选项都看似合理。我们通过要求针对不同历史关键帧估计的旋转对同一当前帧预测出一致的姿态来消除这种歧义。我们将一个候选项与其他传入边上的候选项一同评估，从而可以联合选择一组一致的细化结果。

对于具有存储姿态的历史帧i，候选项$a \in \{0, 1\}$ 预测当前的相机到世界姿态

$$\mathbf{Q}^{(a)}_{k|i} = \mathbf{R}^{\text{prev}}_{wc,i} \hat{\mathbf{R}}^{(a)}_{ik}. \tag{3}$$

对于同一跟踪片段中的历史帧i和j，定义$\mathbf{H}_{ij} = (\mathbf{R}^{\text{prev}}_{wc,i})^{\top} \mathbf{R}^{\text{prev}}_{wc,j}$。该旋转将相机$j$映射到相机$i$。候选项$a$ 和b之间的分歧为

$$\epsilon^{(k)}_{ij}(a, b) = d(\hat{\mathbf{R}}^{(a)}_{ik}, \mathbf{H}_{ij} \hat{\mathbf{R}}^{(b)}_{jk}). \tag{4}$$

当两个存储方向均可用时，它等于$d(\mathbf{Q}^{(a)}_{k|i}, \mathbf{Q}^{(b)}_{k|j})$。在系统启动期间，或在历史帧融合之前，可改由已接受的历史边提供$\mathbf{H}_{ij}$。这会闭合一个三帧环，使得在获得首个窗口解之前即可进行选择。来自不同已知跟踪片段的帧对，或缺少任一历史关系的帧对，均不予考虑。

令 $\mathcal{P}_k$ 包含入边均具有局部有效基线的可比较历史关键帧对。二值标签s_i 在边(i,k)上选择候选项0或1，$\mathbf{s}$ 则汇集这些标签。我们最小化

$$J_k(\mathbf{s}) = \sum_{(i,j) \in \mathcal{P}_k} \rho_{\delta_{\text{sel}}}(\epsilon^{(k)}_{ij}(s_i, s_j)), \tag{5}$$

其中， ρ_δ 是 Huber损失，对于非负残差$x^2/2$，其他情况下则为 $\delta(x - \delta/2)$。选择尺度为 $\delta_{\text{sel}} = 3°$，残差和尺度均以度为单位进行评估。每条边的所有项共用一个标签，

这确保了其各项比较之间的一致性。

联合选择器仅在某条边至少有八个留出匹配时才考虑对其进行细化。搜索会枚举符合条件的二元标签。若一条边有两个局部有效的候选项，且至少有另外两条具有有效基线的可比较边，则可切换到细化结果。对于每条符合条件的边，我们先将其标签固定为 0，再固定为 1，并在每种情况下优化其他标签，从而最小化 J_k。只有当该成本降低量除以具有基线支持的可比较邻边数量后超过 0.1 deg^2 时，细化结果才仍符合条件。对于具有有效基线但不符合条件的边，我们将其标签固定为 0，并对其余符合条件的标签进行最小化。孤立边保留局部有效的基线。当只有细化结果通过局部验证时，若其与基线支持的相邻边上至少两个已选候选项的差异在 8° 以内，则接受该结果。这些邻边还必须通过下文所述的一致性检查。

3.5 滑动窗口旋转平均

3.5.1 图更新与关键帧选择

联合选择后，我们将非连续边与其间连续旋转的乘积进行比较。如果一条完整链中的每条边的支持计数都至少与该边相同，且它们的角度不一致超过 8°，则拒绝该直接边。保留下来的边及其支持计数决定图更新和关键帧决策。

第一帧用于初始化关键帧缓冲区。如果新到达的帧没有保留下来的旋转边，或者其保留边上的最大支持计数低于关键帧阈值，则该帧成为关键帧。未匹配的关键帧会予以保留，以便后续帧能够将其连接到图中。支持计数足够的帧会将其接受的约束作为关键帧缓冲区外的图顶点。窗口按照第 节所述的调度进行求解和推进 3.1。

3.5.2 旋转平均目标

令 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 表示更新后图中包含最高度顶点的连通分量。其顶点为帧索引，边则携带所选测量值 $\mathbf{R}_{ij}$。在当前窗口参考系中，这些测量值满足

$$\hat{\mathbf{R}}_{ij} \approx \mathbf{R}^{\mathrm{loc}}_{cw,i}(\mathbf{R}^{\mathrm{loc}}_{cw,j})^{\top}. \tag{6}$$

对于每条边，定义旋转向量残差

$$\mathbf{r}_{ij} = \mathrm{Log}\Big((\mathbf{R}^{\mathrm{loc}}_{cw,i})^{\top}\hat{\mathbf{R}}_{ij}\mathbf{R}^{\mathrm{loc}}_{cw,j}\Big). \tag{7}$$

令 $\mathcal{E}_{\mathrm{in}} \subseteq \mathcal{E}$ 表示经过下文的初始化和筛选后保留的边。我们将第一个保留帧 r 固定在其初始旋转 $\bar{\mathbf{R}}^{\mathrm{loc}}_{cw,\,r}$，并求解

$$\begin{aligned} &\underset{\{\mathbf{R}^{\mathrm{loc}}_{cw,i}\}_{i\in\mathcal{V}\setminus\{r\}}}{\operatorname{argmin}} \quad \sum_{(i,j)\in\mathcal{E}_{\mathrm{in}}} \|\mathbf{r}_{ij}\|_2^{1/2}, \\ &\text{subject to} \quad \mathbf{R}^{\mathrm{loc}}_{cw,r} = \bar{\mathbf{R}}^{\mathrm{loc}}_{cw,r}. \end{aligned} \tag{8}$$

固定锚点可消除公共旋转歧义。$l_{1/2}$ 损失会限制不一致边的影响，而迭代重加权最小二乘（IRLS）则细化初始方向。

3.5.3 初始化与鲁棒细化

我们采用鲁棒旋转平均的分层方法（HARA） [37] 来初始化每个连通窗口。该方法对每条边最多采样十个公共邻居。对于一条边（*i*,*j*）及一个采样得到的公共邻居 l，无量纲的三视图一致性误差定义为

$$\zeta_{ij\ell} = \left\|\hat{\mathbf{R}}_{ij} - \hat{\mathbf{R}}_{i\ell}\hat{\mathbf{R}}_{j\ell}^{\top}\right\|_F^2. \tag{9}$$

HARA 根据低于 1 的采样误差的第 10、20 和 30 百分位数确定其初始化阈值。从最高度顶点及其单位旋转开始，该方法沿着得到一致回路支持的边传播方向。若传播停滞，则放宽初始的十个回路支持要求。对于任何剩余顶点，已初始化邻居通过投票决定下一个加入的顶点。得票最多的顶点采用其预测旋转的弦式L_1 均值。如果通过初始测试的采样回路少于五个，则由边传播提供初始方向，随后所有边均进入细化阶段。

对于以 HARA 初始化的窗口，仅当所有采样回路误差的中位数不超过 1 时，我们才会筛选边。随后，若某条边的测量值与初始化端点预测之间的弗罗贝尼乌斯范数平方差小于 1，则保留该边；否则保留所有边。这些无量纲检查确定 $\mathcal{E}_{\mathrm{in}}$。迭代重加权最小二乘最多使用四元数迭代 250 次，并在平均角度更新小于 10^{-3} 弧度时停止。

细化后，每个帧i 都会获得一个无量纲质量评分q_i ，该评分基于其总关联匹配支持度、图度和角度残差的中位数计算。这些量在 $\mathcal{E}$上进行评估，其中包括经 HARA 过滤移除的边。评分在窗口内按其正值中位数归一化，并裁剪至 $[0.05, 4]$。所得q_i 更偏向于由多个支持充分且一致的测量值约束的方向，并为窗口间对齐与融合提供权重。

3.6 窗口间对齐与姿态融合

3.6.1 基于共享帧的对齐

旋转平均返回的是当前窗-口参考系中的方向。共享帧将该参考系与先前估计的轨迹段关联起来。我们首先在共享帧中选择总对齐权重最大的轨迹段。设 O 包含其共享帧索引。对于每个$i \in O$，$\mathbf{R}_{cw,i}^{\mathrm{loc}}$ 是局部估计值，而 $\mathbf{R}^{\mathrm{prev}}{}_{wc,i}$ 是其存储的相机到世界姿态。它们的候选对齐旋转为

$$\mathbf{G}_i = (\mathbf{R}_{cw,i}^{\mathrm{loc}})^{\top}(\mathbf{R}_{wc,i}^{\mathrm{prev}})^{\top}, \tag{10}$$

它将片段世界参考系映射到窗口参考系。对齐权重为 $w_i = q_i \max(I^{\mathrm{prev}}{}_i, I_{\min})$，其中 $I^{\mathrm{prev}}{}_i$ 是存储的信息权重。正数 $I_{\min}$ 和$I_{\max}$ 分别在鲁棒加权前提供信息下限，并为存储的信息设定上限。

对于候选对齐 $\mathbf{G}_\ell$，其中$\mathrm{l} \in O$，定义其共识集

$$C_\ell = \left\{ i \in O : \left\| \mathrm{Log}(\mathbf{G}_i \mathbf{G}_\ell^{\top}) \right\|_2 \leqslant \tau_{\mathrm{align}} \right\}, \tag{11}$$

其中 τ_{align} 是以弧度表示的角度共识阈值。我们选择 $\sum_{i \in C_\ell} w_i$ 最大的假设；若出现并列，则根据截断至 τ_{align} 的角度误差加权和打破平局。令 C 表示获胜的共识集。将其加权矩阵均值通过奇异值分解投影到 SO(3)，以初始化对齐 $\Delta\mathbf{R}$。随后通过最多 30 次鲁棒迭代对其进行优化：

$$\begin{aligned} \mathbf{v}_i &= \mathrm{Log}(\mathbf{G}_i \Delta\mathbf{R}^{\top}), \\ \bar{w}_i &= w_i h(\|\mathbf{v}_i\|_2; \sigma_{\mathrm{align}}), \\ \Delta\mathbf{R} &\leftarrow \mathrm{Exp}\left(\frac{\sum_{i \in C} \bar{w}_i \mathbf{v}_i}{\sum_{i \in C} \bar{w}_i} \right) \Delta\mathbf{R}. \end{aligned} \tag{12}$$

这里，$\mathbf{v}_i$ 是以弧度表示的对齐残差，σ_{align} 是鲁棒角度尺度，Huber权重满足 $h(x;\sigma) = 1$ （当$0 \leqslant x \leqslant \sigma$ 时），并且 $h(x;\sigma) = \sigma/x$ （当 $x > \sigma$时）。将对齐应用于任意已求解的帧，可得

$$\mathbf{R}_{wc,i}^{\mathrm{cand}} = (\mathbf{R}_{cw,i}^{\mathrm{loc}} \Delta\mathbf{R})^{\top}, \tag{13}$$

其中，cand 表示选定片段世界参考系中的一个对齐候选项。

3.6.2 有界姿态融合

对于选定片段中存有估计值的帧，候选项会提供一个新观测。其残差为$\mathbf{e}_i = \mathrm{Log}((\mathbf{R}^{\mathrm{prev}}{}_{wc,i})^{\top}\mathbf{R}_{wc\,i}^{\mathrm{cand}})$。我们为该观测分配信息权重$I_i^{\mathrm{obs}}$，并将其融合比例 β_i 计算为

$$\begin{aligned} I_i^{\mathrm{obs}} &= \min\!\big(\max(q_i, I_{\min}), I_{\max}\big) \\ &\quad \cdot h(\|\mathbf{e}_i\|_2; \sigma_{\mathrm{fuse}}), \\ \beta_i &= \frac{I_i^{\mathrm{obs}}}{\max(I_i^{\mathrm{prev}}, I_{\min}) + I_i^{\mathrm{obs}}}, \end{aligned} \tag{14}$$

其中 σ_{fuse} 是以弧度表示的融合尺度。最大校正角度 $\theta_{\max}$ 通过以下方式限制更新

$$\alpha_i = \begin{cases} \beta_i, & \|\mathbf{e}_i\|_2 = 0, \\ \min\!\left(\beta_i, \dfrac{\theta_{\max}}{\|\mathbf{e}_i\|_2}\right), & \|\mathbf{e}_i\|_2 > 0, \end{cases} \tag{15}$$

其中 $\theta_{\max}$ 也以弧度表示。更新后的姿态和累积信息为

$$\begin{aligned} \mathbf{R}_{wc,i}^{\mathrm{new}} &= \mathbf{R}_{wc,i}^{\mathrm{prev}} \,\mathrm{Exp}(\alpha_i \mathbf{e}_i), \\ I_i^{\mathrm{new}} &= \min\!\big(I_{\max}, \max(I_i^{\mathrm{prev}}, I_{\min}) + I_i^{\mathrm{obs}}\big). \end{aligned} \tag{16}$$

信息上限限制了来自重叠窗口的相关估计的影响，同时保留接受后续校正的能力。上标 new 表示此次更新后的存储状态。处理后续帧时，它将成为 prev 状态。

选定片段中没有存储估计值的帧，将根据其对齐候选项进行初始化，信息量为 q_i，$I_{\min}$)，$I_{\max}$。没有有效重叠的求解窗口会使用单位对齐，开启一个新的轨迹段。这些存储方向构成第 3.4 节中的历史预测，并形成带时间戳的输出轨迹。

4 仿真实验

4.1 仿真设置

模拟地形是一个点云，包含最高达 300 m 的山丘，以及标准差为 0.5 m 的高斯表面高度扰动。每个点都有唯一标识符，因此即使在图像坐标中加入噪声，对应关系仍然已知。

名义飞行高度为 1300 m，并具有标准差为 50 m 的高斯垂直扰动。无人机沿正Y轴前进，平均步长为 10 m，步长扰动服从标准差为 3 m 的高斯分布。其横向位移增量服从标准差为 10 m 的高斯分布。每一步都会累积绕三个机体轴最多旋转 $\pm 3^{\circ}$。模拟相机的分辨率为 640×512 像素，名义焦距为 3600 像素。

设$\mathbf{X}_w = [X_w、\ Y_w、\ Z_w、\ 1]^{\top}$ 为齐次世界坐标中的一个点。在相机方向$\mathbf{R}_{wc}$ 和相机

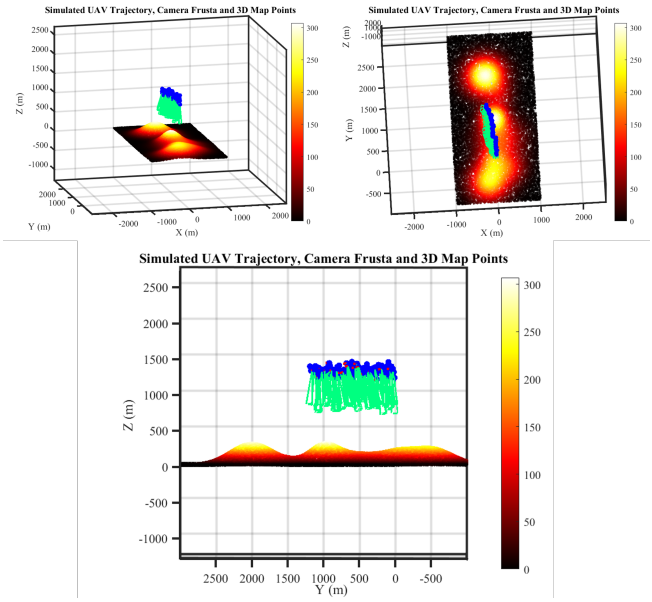

图 3 地形上方的模拟飞行轨迹和相机视锥。

中心$\mathbf{t}_{wc}$ 以世界坐标系表示时，该点的相机坐标为

$$\mathbf{X}_c = \begin{bmatrix} \mathbf{R}_{wc}^{\top} & -\mathbf{R}_{wc}^{\top}\mathbf{t}_{wc} \end{bmatrix} \mathbf{X}_w, \tag{17}$$

其中 $\mathbf{X}_c = [X_c、\ Y_c、\ Z_c]^{\top}$。对于相机前方的点，其透视投影为

$$\mathbf{u} = \mathbf{K}\frac{\mathbf{X}_c}{Z_c}, \tag{18}$$

其中 $\mathbf{u} = [u、\ v、\ 1]^{\top}$ 表示齐次像素坐标。图 3 展示了视锥如何沿模拟飞行轨迹与地形相交。

4.2 误差度量

我们以度为单位测量估计旋转$\mathbf{R}_{\text{est}}$ 与其真值$\mathbf{R}_{\text{gt}}$ 之间的角度差，

$$\theta_{\text{rot}} = \text{rad2deg}\left(\arccos\left(\frac{\text{trace}\left(\mathbf{R}_{\text{est}}^{\top}\mathbf{R}_{\text{gt}}\right) - 1}{2}\right)\right). \tag{19}$$

对于估计的平移方向和参考平移方向$\mathbf{T}_{\text{est}}$和$\mathbf{T}_{\text{gt}}$，我们取估计值取任一符号时所得的较小角度，

$$\phi_{\text{trans}} = \text{rad2deg}\left(\arccos\left(\frac{\left|\mathbf{T}_{\text{est}}^{\top}\mathbf{T}_{\text{gt}}\right|}{\|\mathbf{T}_{\text{est}}\|\|\mathbf{T}_{\text{gt}}\|}\right)\right). \tag{20}$$

在双视图实验中，对图像对上的 θ_{rot} 和 ϕ_{trans} 分别取平均，可得到平均旋转误差（MRE）和平均平移方向误差（MTE）。

4.3 相对旋转求解器的比较

我们使用上述名义飞行条件和相机配置，评估 OpenGV [40] 中的双视图求解器。在投影的特征坐标中加入零均值高斯噪声，其标准差范围为 0 至 3 像素。仿真提供已知的特征对应关系。求解器调用使用经过扰动的参考位姿，并利用参考旋转在候选解中进行选择。该流程在对应关系和参考辅助候选选择均受控的情况下，评估求解器对图像噪声的敏感性。

图 4表明，对于同一组对应点，较小的旋转误差可能与大得多的平移方向误差并存。当噪声标准差为 3 像素时，Nist´er 的五点法求解器所得双视图旋转误差均值为 1。65°，平移方向误差为 25。61°。在相同噪声水平下，特征值求解器的双视图旋转误差为 3。21°。

旋转会改变每条视线的方向，而平移造成的图像位移取决于基线与场景深度的相对关系。当这一比值较小时，图像扰动可能会掩盖平移视差，但旋转仍可测量。结果表明，当平移方向约束较弱时，应独立估计姿态；不同求解器的误差则说明有必要保留多个旋转候选解。

4.4 飞行高度和相机焦距的影响

我们在图像噪声标准差固定为 1 像素的条件下，分别改变标称高度和焦距。高度扫描范围为 580 至 1300 m，步长为 80 m；焦距扫描范围为 900 至 3600 像素，步长为 300 像素。其余配置以及双视图评估流程均与前述实验一致。

在高度扫描（图5）中，Nist´er 的旋转误差在两个端点之间有所下降，尽管其平移方向误差有所上升。其他求解器在所采样的高度范围内呈现出不同趋势。ff

图6 显示，随着视场变窄，平移方向误差通常会增大。Nist´er 的求解器在视场较宽时，平均旋转误差和平均平移误差均有所改善。

在基线固定的情况下，提高高度会降低基线与场景深度的比值。由此造成的平移视差损失，使平移方向对图像噪声更加敏感，同时视线也逐渐接近纯旋转时的视线。这解释了为何在高度扫描中，旋转精度可能提高，而平移方向精度却会下降。

改变焦距会同时影响角度噪声和视线方向分布。焦距较短时，固定像素误差对应的角度扰动会增大

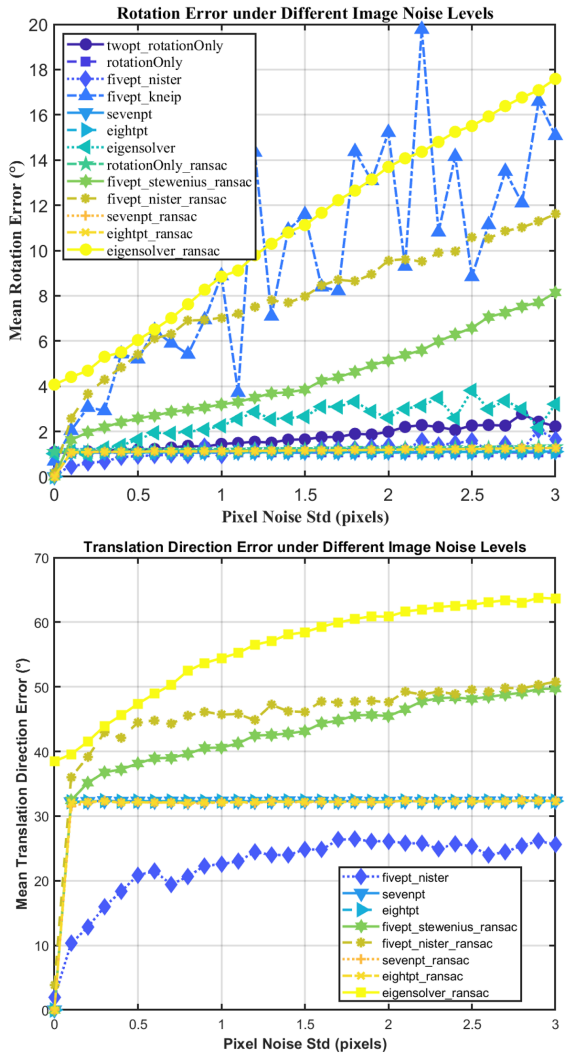

图4 随着高斯像素噪声标准差增大，各求解器的双视图旋转和平移方向误差。

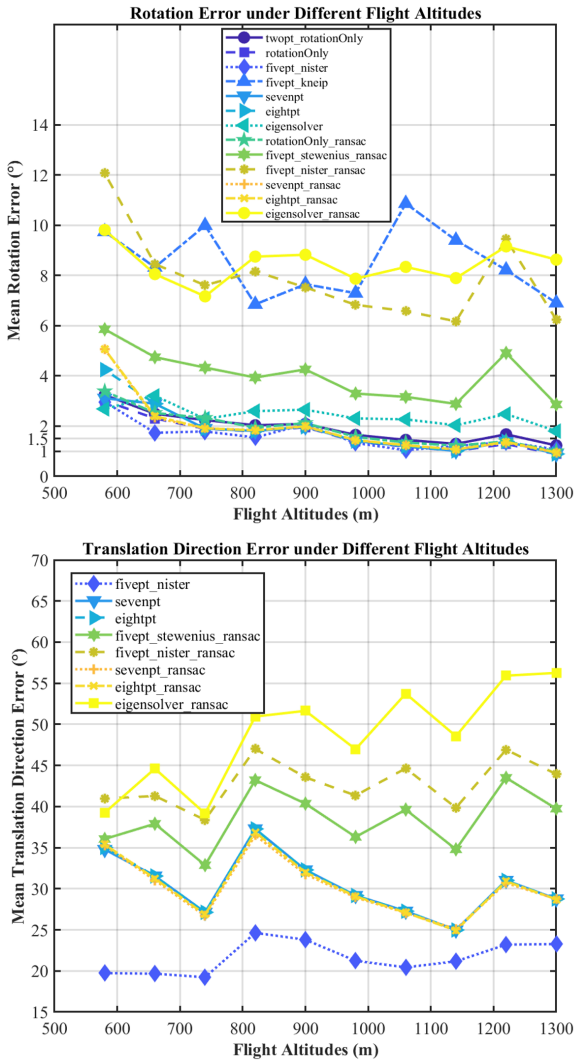

图5 在相机焦距固定、图像噪声标准差为 1 像素时，双视图旋转和平移方向误差随模拟飞行高度的变化。

，并能接纳更大方向范围内的视线方向。在仿真场景中，这种更广的分布有助于区分由旋转和平移引起的图像运动，这与视场更宽时观察到的性能提升一致。不同求解器的响应促使我们通过视图间的一致性来选择旋转。

4.5 与传统 SLAM 系统的仿真比较

模拟SLAM基线使用运动恢复结构进行初始化，将已知特征与三维点关联，并采用PnP和滑动窗口束调整更新位姿与地标。旋转图使用来自合成点标识符的对应关系。求解器调用遵循第4.3节所述的参考辅助协议。像素噪声标准差在相同范围内进行扫描，每次运行中固定双视图求解器。联合候选项选择的评估在第5节中，针对飞行图像单独进行。

对于每种旋转求解器，我们比较了三种聚合方案。全局旋转平均（RA）在一个图中使用所有可用帧。窗口化 RA 使用最多包含十个关键帧的重叠窗口，并在每次完整窗口更新后前进一个关键帧。链式累积按时间顺序组合连续的成对旋转。这些配置用于衡量成对求解器误差在初始帧对齐后对轨迹方向的影响。

图中 SLAM 基线7(a) 的方向误差呈现明显波动，偏航角也随总误差出现了几次幅度最大的偏离。图7(b) 显示，旋转平均显著降低了特征值求解器的累积误差。当噪声标准差为 3 像素时，窗口化全局 RA 的方向误差分别为 3.39°和 1.68°，而链式累积的误差为 25.12° 。五点法求解器的双视图旋转误差更低

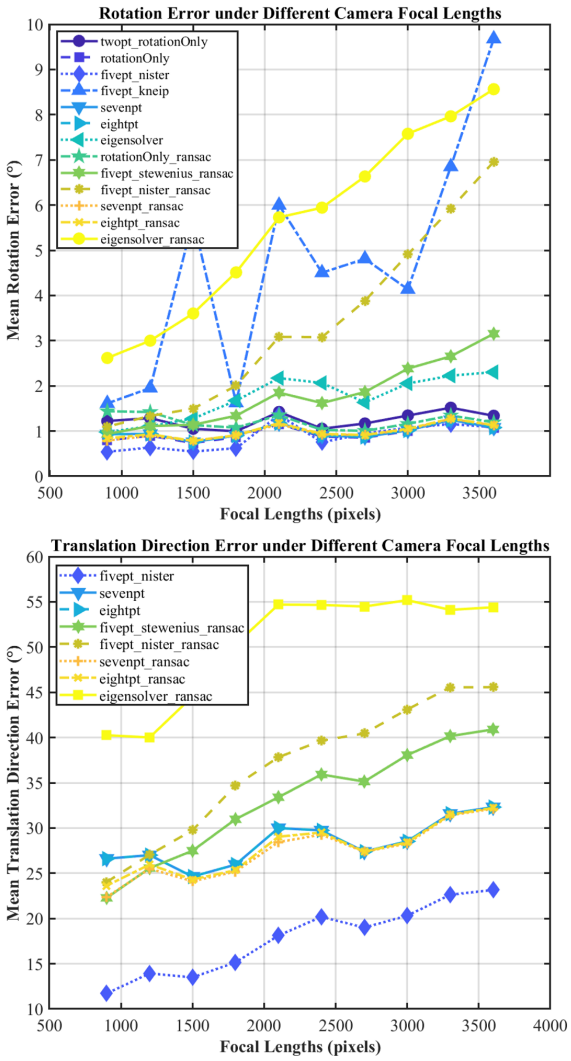

图 6 固定仿真飞行高度和 1 像素图像噪声标准差下，双视图旋转误差和平移方向误差随相机焦距的变化。

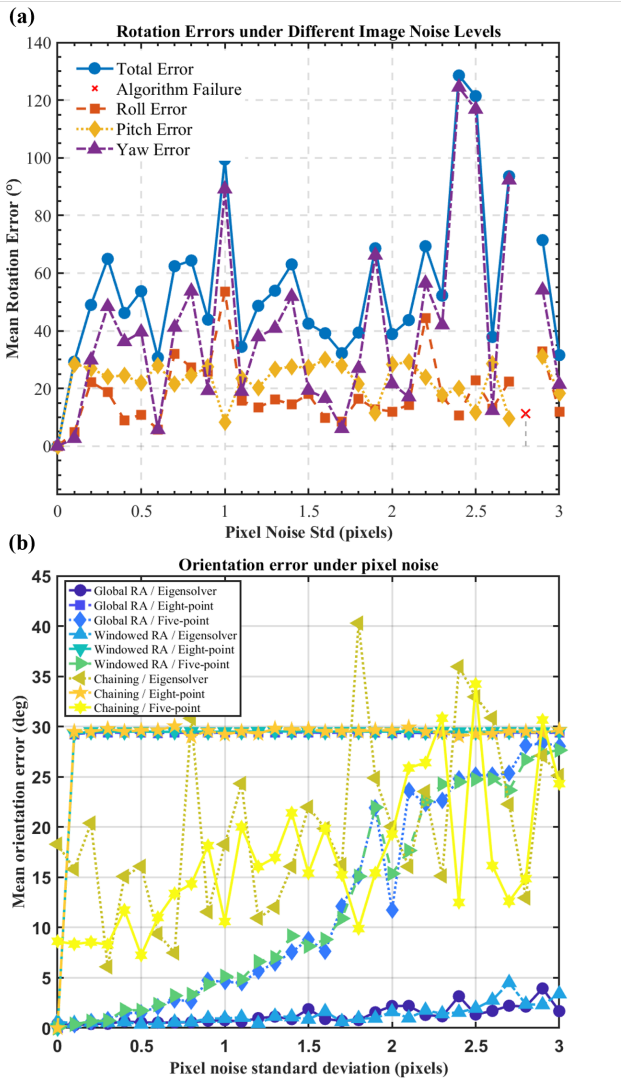

图7 轨迹姿态误差随像素噪声标准差的变化。(a) 模拟 SLAM 基线的总误差和欧拉角分量误差。叉号表示失败。(b) 三种固定的双视图求解器分别采用全序列旋转平均（全局旋转平均）、滑动窗口旋转平均（窗口化旋转平均）以及连续旋转的直接复合（链式累积）进行评估。五点法表示 Nist´er 的求解器。

图 4中，其窗口化轨迹误差却达到 27.65°。加入噪声后，八点法轨迹误差聚集在 29° 附近。

在三角测量角较小时，输入视线方向的扰动可能导致重建深度发生较大变化。PnP 根据这一不确定地图估计下一时刻的姿态，而得到的位姿又会参与后续的地图更新。这种耦合为图像噪声影响后续姿态估计提供了途径。

链式累积会将每个成对误差传递到后续方向中。旋转平均通过重叠路径约束每个方向，从而能够识别不一致边，并降低孤立偏差的权重。较小的窗口图仍保留了这一优势的大部分。两视图实验与轨迹实验中的求解器排名不同，这表明轨迹精度也取决于视图之间的一致性。这种依赖促使我们采用多视图旋转第 3.4 节中的选择。

5 实验评估

5.1 飞行数据集与评估协议

我们将 MATLAB R2023b 中实现的 ROVO 与单目 ORB-SLAM3 [16] 和深度图像块视觉里程计（DPVO） [26] 进行比较。MUN-FRL数据集 [41]提供了贝尔412直升机采集的视觉、惯性和激光雷达测量值，以及经过后处理的导航参考数据。其下视相机以 20 Hz 记录 1440 × 1080图像。我们使用了在不同地面场景上进行的三次飞行。对于第二种成像条件，我们将相同的视图重新投影，并将两个像素焦距（f_x 和f_y）均乘以三，同时相应地更新每个估计器的内参——

表 1 ROVO 飞行实验的运行特定设置。特征数量是每幅图像预算，描述符阈值是 MATLAB 匹配阈值，关键帧支持度是应用于最大存活匹配数的阈值。

Run	图像条件	每幅图像的特征数	描述符阈值	关键帧支持度	窗口/步长
Bell412-3	原始	1,000	1	800	10/1
Bell412-3	3× 焦距	1,000	1	1,000	10/1
Bell412-4	原始	1,000	1	800	10/1
Bell412-4	3× 焦距	800	1	800	10/1
Bell412-5	原始	1,000	2	800	10/1
Bell412-5	3× 焦距	1,000	5	1,000	10/1
VPAIR	离线校正	1,000	1	800	10/1

所有运行均使用 SIFT，分别将对比度阈值和边缘阈值设为 0.005 和 20，并采用联合旋转选择器。窗口间共享的设置为 $\tau_{\text{align}} = 4°$, $\sigma_{\text{align}} = 3°$, $\sigma_{\text{fuse}} = 4°$, $\theta_{\max} = 2°$，且 $(I_{\min}, I_{\max}) = (0.1, 4)$。

以此测试记录相机位置下更窄的视场。在两种条件下，相机到地面的距离以及曝光之间的基线均相同。

VPAIR 是一个用于空中视觉地点识别和定位的数据集 [42],，提供了一架轻型飞机飞越城市区域、农田和森林时拍摄的图像。已发布的 GNSS/INS 位姿中报告的飞行高度范围为 356 至 825 m，中位数为 489 m。其向下拍摄的相机以 25 Hz 记录 1600×1200 图像。我们使用以 800×600 像素分辨率和 1 Hz 发布的 2,706 张查询图像。该数据集还提供渲染参考视图、深度以及六自由度参考位姿。我们仅对查询图像运行里程计，使用离线校正的灰度查询图像及其对应的针孔相机标定参数进行特征提取和几何验证。我们使用已发布的滚转角、俯仰角和偏航角作为方向评估的真值。

表 1 列出了飞行实验中的 ROVO 设置。我们将关键帧支持阈值设置为接近或等于每图像特征预算的数值，使几乎每个输入帧都成为关键帧。经过几何验证后通常保留下来的匹配数量低于该阈值。在七次飞行中，99.99% 的输入帧成为了关键帧。该设置会更频繁地执行旋转平均以及窗口间对齐与融合，以优化每个当前帧的姿态。在 MATLAB R2023b 中使用英特尔酷睿 i5-12600KF 时，这些操作合计每次窗口更新平均仅耗时 8.6 ms，因此这种做法是切实可行的。

5.2 相对旋转误差

我们采用 evo [43] 的旋转角相对位姿误差约定来评估相对旋转误差（RRE）。估计位姿与参考位姿通过 0.05 s 内的最近时间戳进行关联。我们报告相对旋转角（单位为度）的均方根。

在 Bell412 上，每种成像条件都使用相邻的参考帧，且三个系统中这些帧关联的位姿各不相同。在 VPAIR 上，各方法的主要评分使用相邻的已发布查询，且两者均有可用位姿。ROVO 配对必须同时位于同一跟踪片段内，因为重启会引入独立的方向参考。我们单独报告覆盖率，并在三组有效对的交集上重复进行 VPAIR 比较。

表 2 Bell412 共享配对上的 RRE均方根误差（°）。高度为每次飞行记录的最大值。粗体表示误差最低。

飞行	高度（米）	ORB-SLAM3	DPVO	ROVO
原始焦距				
Bell412-3	154	0.511	**0.510**	0.511
Bell412-4	112	0.432	**0.428**	0.444
Bell412-5	171	0.600	**0.594**	0.687
三倍焦距				
Bell412-3	154	1.337	2.703	**0.918**
Bell412-4	112	2.129	2.160	**0.652**
Bell412-5	171	2.403	0.984	**0.789**

5.3 在MUN-FRL上的结果

表2和图8 报告了原始和重投影的 Bell412 图像的结果。

在原始图像上，三条 Bell412 误差轨迹往往同步升降，其中 DPVO 在每种情况下的序列均方根误差都最低。更广阔的视野提供了更多机会，使空间分布式轨迹能够跨帧保持。利用这些轨迹，地图或图块约束可以作为联合位姿估计的一部分，帮助稳定旋转估计。仅凭误差轨迹无法将这一影响与求解器和轨迹选择的差异区分开来。

在窄视场配置下的共享帧对上，与各序列中 ORB-SLAM3 和 DPVO 中表现较好的方法相比，ROVO 在 Bell412-3、Bell412-4 和 Bell412-5 上的相对旋转均方根误差分别降低了 31.3%、69.4% 和 19.7%。这一优势与焦距仿真结果（图 6）一致：视场越窄，平移方向误差通常越大。在联合估计位姿和场景结构的方法中，平移方向误差可能通过估计的路标深度或图块深度传播到后续位姿估计中，包括旋转。ROVO 对重叠视图中的相对旋转取平均，并仅向前传递方向。通过避免这条误差传播路径，当平移方向受到较大

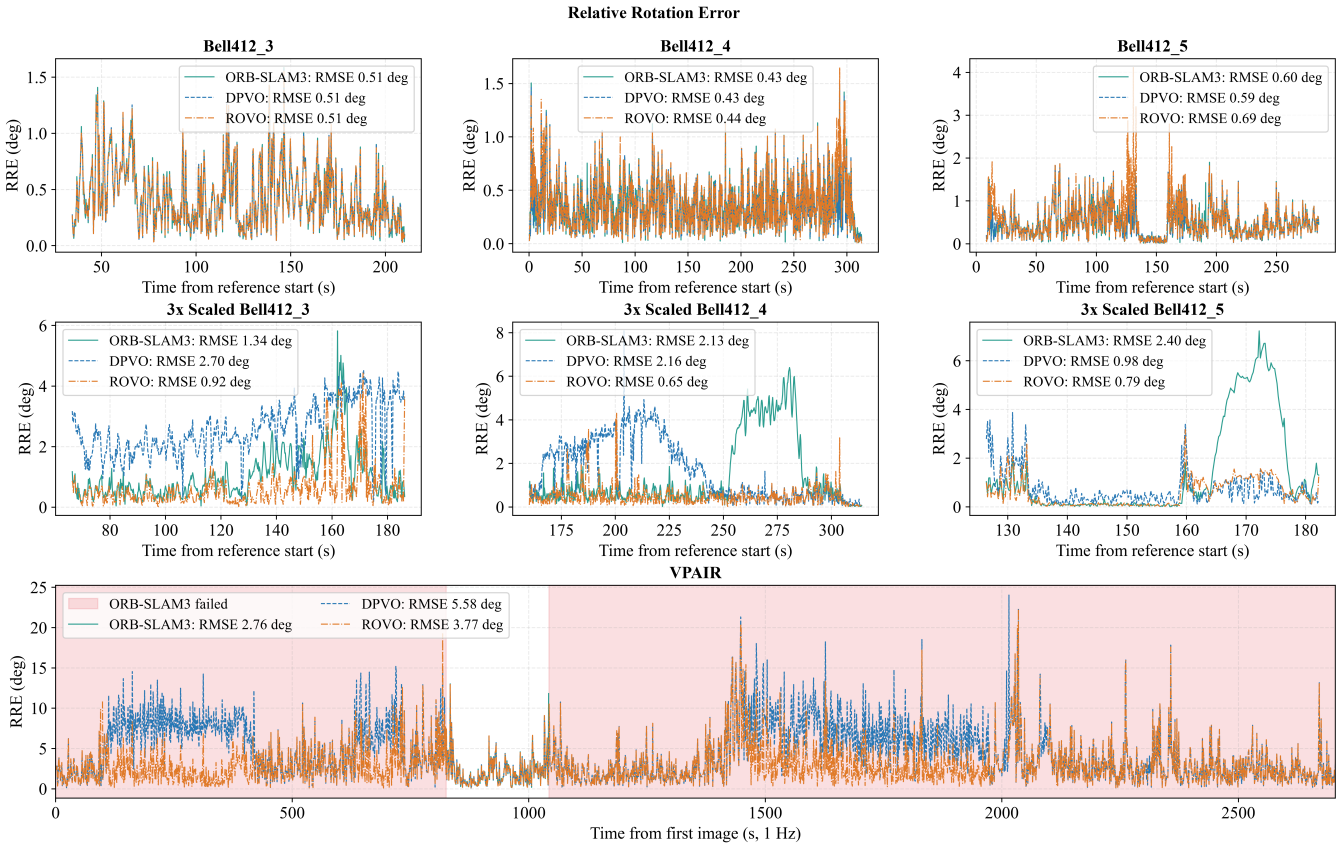

图 8 采用联合旋转选择的 ORB-SLAM3、DPVO 和 ROVO 的相对旋转误差。顶部和中间两行展示原始焦距和三倍焦距下共享的 Bell412 图像对。底部一行使用各方法有效的相邻 VPAIR 图像对。阴影表示 ORB-SLAM3 缺失的位姿。图例中的数值为各序列以度为单位的均方根误差。

受限。

5.4 VPAIR上的结果

在高度较高的 VPAIR序列上，ROVO 在 8 个跟踪区段中为 2,706 个查询中的 2,681 个返回了姿态（99.1%）（表 3）。如图 5 所示，高度扫描表明，在基线固定时，场景深度越大，平移视差越弱，平移方向也越容易受到图像噪声的影响。ROVO 根据几何验证的特征匹配估计并平均相对旋转。每个跟踪区段开始时，它都无需初始化三维地图或估计平移尺度；其纯姿态状态也避免了通过估计深度传播平移误差。ORB-SLAM3 在运行期间重启了 21 次，并在单个时间段内返回 217 个位姿（8.0%）。其单目初始化需要足够的视差来三角测量初始地图 [16]。DPVO 使用学习式图块跟踪以及相机位姿和图块深度的联合优化，为全部 2,706 个查询返回了位姿 [26]。

在整个 VPAIR 序列中，与 DPVO 在共享帧对上的表现相比，ROVO 将相对旋转均方根误差降低了 32.5%。在 ORB-SLAM3 成功跟踪的短区段上，ROVO 的均方根误差也低于另外两种方法。

5.5 消融研究

我们考察自适应旋转选择和鲁棒窗间融合的作用。表 4比较了四种 ROVO 变体在每种飞行条件下相同相邻参考对上的表现。排除了跨越跟踪区段边界的帧对。五点法、特征值求解器和混合连接变体采用所提出的融合方法；覆盖变体则采用混合连接，并直接替换对齐后的窗口估计值。

5.5.1 自适应旋转选择

混合连接的误差始终低于固定特征值求解器，并且与五点法基线相近。其相较于五点法的优势在不同序列之间有所变化，在部分 Bell412 飞行中，五点法仍保持最低误差。这一规律支持保留五点法候选项，同时在观测视图支持的情况下引入特征值求解器细化。该细化对观测方向的分布以及匹配方位的扰动较为敏感，因此其贡献取决于图像对和周围约束。

候选项记录显示，在较窄的视场下，混合连接更频繁地选择特征值求解器的细化结果。

表3 VPAIR 跟踪覆盖率及相对于已发布姿态参考的 RRE 均方根误差（°）。最后一列对所有方法均使用相同的 216 个相邻查询对。粗体表示最高覆盖率和最低的共同配对误差。

方法	位姿	覆盖率（%）	有效对	自身配对均方根误差	公共配对均方根误差
ORB-SLAM3†	217	8.0	216	2.761	2.761
DPVO	**2,706**	**100.0**	2,705	5.583	2.639
ROVO	2,681	99.1	2,673	3.767	**2.631**

†ORB-SLAM3 覆盖了 8.0% 的飞行距离。其两个均方根误差采用相同的区间。

表 4 相对旋转误差均方根误差（°）的消融结果。每行中的四个变体均使用相同的相邻参考对。粗体表示误差最低。

序列	五点法	特征值求解器	混合连接	覆盖
		原始焦距		
Bell412-3	0.5114	0.6107	**0.5109**	0.5137
Bell412-4	0.4445	1.0842	**0.4444**	0.4536
Bell412-5	**0.6831**	0.8924	0.6870	0.7044
		三倍焦距		
Bell412-3	**0.8735**	1.0592	0.9183	1.0681
Bell412-4	**0.6321**	0.9018	0.6525	0.8033
Bell412-5	0.8040	0.8276	**0.7796**	0.8230
		高海拔飞行		
VPAIR†	3.7925	3.9232	**3.7665**	3.7884

窄视场 Bell412-5 混合连接结果采用了完成后的重新运行。 †固定求解器 VPAIR 运行会在校正前提取特征，而两种混合连接变体均使用离线校正图像。

当两个候选项均通过验证时，选定的细化结果通常具有较低的参考旋转误差。这些观察表明，选择器通过几何支持和多视图一致性来响应观察几何的变化。第4节中的高度与噪声扫描说明了这种自适应的动机，因为这两个因素都会影响底层求解器的敏感性。对轨迹精度的影响还取决于所选测量与其余旋转图的相互作用，这有助于解释相较于五点法时增益的差异。

5.5.2 鲁棒窗间融合

以混合连接为前端时，在所有飞行条件下，鲁棒融合的误差都低于直接覆盖，且在视场较窄时收益最大。两种变体使用相同的窗口调度和基于重叠的对齐。随着窗口推进，其图像约束的变化可能会为轨迹中已有的姿态产生不同的估计。直接替换会将这种变化传递到存储位姿中。视场较窄时改进幅度更大，这与观测方向分布不够广泛时旋转估计的敏感性增加相符。

所提出的更新将对齐后的窗口估计与存储方向结合，并依据各自的信息权重进行加权。观测质量和旋转残差决定每次更新的贡献，而校正界限则限制与现有信息不一致的窗口估计所产生的影响。因此，随着新约束不断到来，较早的窗口信息仍会持续影响轨迹。对累积信息进行限制还可以保留后续观测的影响。由于存储方向会引导后续对齐和候选项选择，保留鲁棒历史信息不仅有益于当前更新，也有助于后续的窗口估计。

6 结论

ROVO 将自适应旋转选择、滑动窗口旋转平均和鲁棒窗间融合相结合，以根据单目航拍图像估计相机方向。仿真展示了在平移方向约束较弱时估计旋转的优势。飞行实验表明，在原始 Bell412 设置下，其精度与 ORB-SLAM3 和 DPVO 相当；在视场较窄以及高度更高的 VPAIR 数据上，精度有所提升。消融实验支持通过多视图一致性调整旋转测量，并在窗口之间保留信息。未来的工作将把 ROVO 与惯性测量和平移测量相结合，以实现全位姿导航。

致谢 本工作部分得到了国家自然科学基金委员会（批准号：12472189）、湖南省科技创新计划（批准号：2022RC1196）以及先进推进技术国家重点实验室基金（批准号：25-GC-ZZKY-021）的资助。

参考文献

1 Yan M、Dong R、Han Q L 等。保障低空经济：综述。《中国科学：信息科学》，2026，69：141202 2 Keshk H M、Abdallah A M、Almutairi S 等。基于无人机摄影测量与激光雷达集成的 KFUPM 高保真三维校园测绘。

《科学报告》，2026，16 8328 3：Deng Z、Wu F、Xu Y 等。基于无线电地图的无人机取送物流系统能量最小化。《IEEE车辆技术汇刊》，2024，73：17893–17898 4 Jiang C、Zhou X、Chen H 等。利用全球导航卫星系统进行无人机定位：现状综述。《无人机》，2026，10：91 5 Schmidt G T。GNSS受限与拒止环境中的导航传感器与系统。《中国航空学报》，2015，28：1–10 6 Zhang T、Hu X、Xiao J 等。视觉导航综述：从几何学到具身人工智能。《工程应用与人工智能》，2022，114 105036 7：Liu Y、Liu L、Zheng Y 等。具身导航。《中国科学：信息科学》，2025，68：141101 8 Mei C、Fan Z、Zhu Q 等。基于视觉惯性融合的无人机新型场景匹配导航系统。《IEEE传感器期刊》，2023，23：6192–6203 9 Ye Y、Teng X、Chen S 等。探索低空多视角观测条件下无人机视觉定位的最佳方法：一项基准测试。载于：IEEE/CVF计算机视觉与模式识别会议论文集，2026 1741 10 1731。– Yan H、Zhao R、Wu C 等。基于惯性导航信息与邻域结构一致性的适用于大视角SAR图像的快速匹配方法。《遥感》，2023，15：4084 11 Leonard J、Durrant-Whyte H。自主移动机器人的同步建图与定位。载于：IEEE/RSJ智能机器人与系统国际研讨会论文集，第3卷，1991 1442 1447。– 12 Abaspur Kazerouni I、Fitzgerald L、Dooly G 等。视觉SLAM最新研究进展综述。《专家系统及其应用》，2022，205 117734 13：Shen W、Meng Z、Gu P 等。VF-Nav：基于视觉平面图的点目标导航。《中国科学：信息科学》，2025，68 150207 14：Lyu P、Yong C、Lai J 等。基于视觉惯性导航的低空无人机数据集。《IEEE传感器期刊》，2024，24 41911 15 41904：– Nister D。五点相对位姿问题的高效求解。《IEEE模式分析与机器智能汇刊》，2004，26：756–770 16 Campos C、Elvira R、G´omez Rodr´ıguez J J 等。ORB-SLAM3：用于视觉、视觉惯性和多地图SLAM的精确开源库。《IEEE机器人学汇刊》，2021，37：1874–1890 17 Lappe M、Bremmer F、van den Berg A。通过视觉流感知自运动。《认知科学趋势》，1999，3 336 18 329：– Forster C、Pizzoli M、Scaramuzza D。SVO：快速半直接单目视觉里程计。载于：IEEE机器人与自动化国际会议论文集，2014。15–22 19 Mur-Artal R、Montiel J M M、Tard´os J D。ORB-SLAM：一种通用且精确的单目SLAM系统。《IEEE机器人学汇刊》，2015，31：1147–1163 20 DeTone D、Malisiewicz T、Rabinovich A。SuperPoint：自监督兴趣点检测与描述。载于：IEEE计算机视觉与模式识别会议研讨会论文集，2018。224–236 21 Lindenberger P、Sarlin P E、Pollefeys M。LightGlue：以光速进行局部特征匹配。载于：IEEE/CVF国际计算机视觉会议论文集，2023 17592 22 17581。– Liu C、Qin J、Wang S 等。基于动态视觉特征移除的动态环境下精确RGB-D SLAM。《中国科学：信息科学》，2022，65：202206 23 Yan C、Qu D、Xu D 等。GS-SLAM：基于三维高斯泼溅的稠密视觉SLAM。载于：IEEE/CVF计算机视觉与模式识别会议论文集

《计算机视觉与模式识别》，2024，19595–19604。24 Qin T、Li P、Shen S。VINS-Mono：一种鲁棒且通用的单目视觉惯性状态估计器。《IEEE机器人学汇刊》，2018，34：1004–1020。25 Mourikis A I、Roumeliotis S I。一种用于视觉辅助惯性导航的多状态约束卡尔曼滤波器。载于：IEEE国际机器人与自动化会议论文集，2007，3565–3572。26 Teed Z、Lipson L、Deng J。深度图像块视觉里程计。载于：Oh A、Naumann T、Globerson A 等编。《神经信息处理系统进展》第36卷。Curran Associates, Inc.，2023，39033–39051。27 Liu T、Shen S。高空单目视觉惯性状态估计：初始化与传感器融合。载于：IEEE国际机器人与自动化会议论文集，2017，4544–4551。28 Warren M、Upcroft B。高空双目视觉里程计。载于：机器人学：科学与系统会议论文集，德国柏林，2013。29 Warren M、Corke P、Upcroft B。用于无人飞行器扩展高度飞行的远距离双目视觉里程计。《国际机器人研究杂志》，2016，35：381–403。30 Dong J、Lei H、Yue Y 等。用于高空无人机的基于激光雷达的视觉惯性里程计。载于：Yan L、Duan H、Deng Y 编。《制导、导航与控制进展》，新加坡：Springer Nature Singapore，2025，419–431。31 Hartley R、Zisserman A。《计算机视觉中的多视图几何》。剑桥：剑桥大学出版社，2004，第2版。32 Hartley R。为八点算法辩护。《IEEE模式分析与机器智能汇刊》，1997，19：580–593。33 Faugeras O、Maybank S。由点匹配恢复运动：解的多重性。载于：IEEE视觉运动研讨会论文集，1989，248–255。34 Kneip L、Lynen S。帧间旋转的直接优化。载于：IEEE国际计算机视觉会议论文集，2013，2352–2359。35 Bustos A P、Chin T J、Eriksson A 等。视觉SLAM：为何需要光束法平差？载于：IEEE国际机器人与自动化会议论文集，2019，2385–2391。36 Chatterjee A、Govindu V M。鲁棒相对旋转平均。《IEEE模式分析与机器智能汇刊》，2018，40：958–972。37 Lee S H、Civera J。HARA：鲁棒旋转平均分层方法。载于：IEEE/CVF计算机视觉与模式识别会议论文集，2022，15777–15786。38 Lee S H、Civera J。仅旋转光束法平差。载于：IEEE/CVF计算机视觉与模式识别会议论文集，2021，424–433。39 Burger W、Burge M J。尺度不变特征变换（SIFT）。载于：《数字图像处理：使用Java的算法导论》，第609–664页，2016年，伦敦：Springer，第2版。40 Kneip L、Furgale P。OpenGV：一种统一且通用的实时标定几何视觉方法。载于：IEEE国际机器人与自动化会议论文集，2014，1–8。41 Thalagala R G、De Silva O、Jayasiri A 等。MUN-FRL：用于空中自主导航与建图的视觉惯性激光雷达数据集。《国际机器人研究杂志》，2024，43：1853–1866。42 Schleiss M、Rouatbi F、Cremers D。VPAIR：大规模户外环境中的空中视觉地点识别与定位。arXiv:2205.11567，2022。43 Grupp M。evo：用于评估里程计和SLAM的Python软件包。计算机软件，2017。