

热点追踪

## 融合特征轨迹的相机全局平移估计方法

陶沛霖 崔海楠 荣梦琪 申抒含

中国科学院自动化研究所

本文是中国科学院自动化研究所三维视觉研究组团队的研究成果，发表在 CVPR 2024 的工作 HETA<sup>[1]</sup>。在三维视觉领域中，准确估计摄像机位姿并从图像集生成场景点云是基础性任务，在自动驾驶<sup>[20]</sup>、增强现实<sup>[18]</sup>、和神经辐射场<sup>[19]</sup>等领域，具有广泛的应用。一般而言，运动恢复结构算法作为实现这些目标的一种常见且有效的方法脱颖而出。论文研究的问题是如何通过融合特征轨迹约束，解决全局式运动恢复结构算法中在短基线、相机近似共线运动场景下，平移平均精度低鲁棒性差的问题。先前的方法要么仅依赖相机间的相对平移，在相机近似共线移动时无法准确估计相机间的相对尺度；要么仅依赖特征轨迹约束，对特征轨迹中的错误匹配敏感。针对上述问题，团队提出一种新颖的混合约束显式平移平均框架(HETA)，通过混合使用相对平移和特征轨迹约束，同时显式优化场景三维点。通过在顺序 KITTI 测距基准和互联网无序数据集 1DSfM 上测试，发现我们的方法表现超过了许多最先进的全局运动恢复结构算法。

## 一、引言

从运动恢复结构方法的基本步骤是首先通过特征点检测和匹配构建一个场景图<sup>[12,13]</sup>，图中节点表示摄像机，边连接具有充足特征匹配的摄像机。随后，估计摄像机位姿并三角化得到场景三维结构。

通常摄像机位姿估计的主要方法是采用增量式模式，例如 COLMAP<sup>[10]</sup>。该方法首先通过精心选择图像对来创建初始模型。然后，使用 PnP 算法<sup>[17]</sup>注册包含足够数量 2D-3D 对应关系的图像。最后，通过迭代的方法，包括三角测量、捆绑调整 (BA)<sup>[11]</sup>和 PnP 步骤，同时估计场景结构和摄像机位姿。尽管增量式方法在精

度和对抗异常值的稳健性方面表现出色，但它们对图像注册顺序的变化较为敏感，可能导致误差积累和漂移。此外，重复的非线性捆绑调整显著影响效率，使其不适用于大规模场景。为解决增量式方法中的这些问题，全局式方法<sup>[2-9]</sup>被提出，通过从摄像机的相对位姿估计其全局旋转和平移以实现对所有摄像机进行注册，随后，对场景结构进行三角测量和优化，其全部流程如图 1 所示。由于仅需一次的捆绑调整优化，从而显著提高了效率，并实现了所有相机间均匀的误差分布。具体而言，摄像机的全局姿态 $R_i, t_i$ 和相对姿态 $R_{ij}, t_{ij}$ 满足以下方程：

$$R_j R_i^T = R_{ij}, \quad \frac{t_i - t_j}{\|t_i - t_j\|_2} = R_j^T t_{ij} = v_{ij} \quad (1)$$

其中符号 $v_{ij}$ 表示全局坐标系中的相对平移。对于全局旋转估计，现有基于李代数的方法<sup>[15]</sup>已经得到充分研究。然而，由于相对平移估计对低视差特征匹配敏感<sup>[9]</sup>，并且存在尺度不确定性问题，这使得全局平移估计比全局旋转估计更为困难。仅依赖相对平移的方法局限于注册在平行刚性图中的摄像机，并在摄像机经历共线运动时遇到退化问题。即使摄像机运动轨迹几乎共线，相对平移中的轻微扰动也可能导致估计摄像机位置发生显著变化，使得仅使用相对平移无法实现准确的估计。

为解决这些挑战，一些方法将特征轨迹的约束纳入目标函数中。根据在优化过程中是否估计了特征轨迹对应的三维点，这些方法可以被分为隐式方法和显式方法。大多数隐式方法估计摄像机基线尺度，或者利用来自隐式三维点的相机到点的约束，但这些方法都对相对平移的异常值敏感。例如，LiGT<sup>[8]</sup>旨在仅使用特征轨迹构建约束。然而，由于特征轨迹通常表现出比相对平移更高

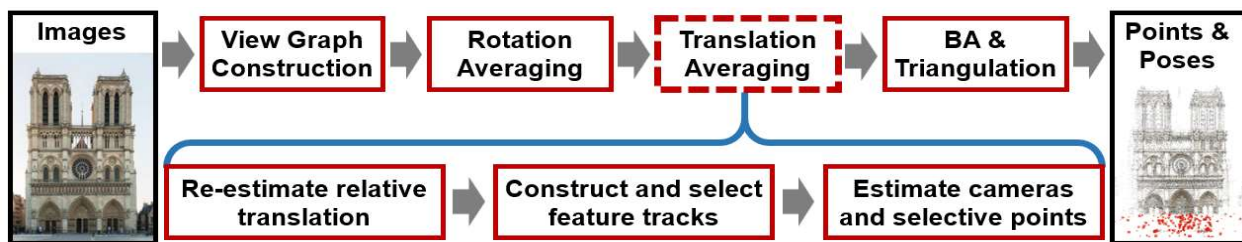


图1 使用 HETA 的全局式 SfM 方法流程框架。其中 HETA 算法的流程为：给定相机的全局姿态，首先重新估计相机相对平移，然后构造并挑选一些特征轨迹，最后使用一种鲁棒的目标函数同时估计相机和部分三维点的位置。

的外点率，因此其鲁棒性仍然不足。1DSfM<sup>[6]</sup>作为一种经典的显式方法，同时利用摄像机到摄像机和摄像机到点的约束，以估计三维点和摄像机位置，但在处理特征轨迹异常值时也会产生嘈杂的解。1DSfM<sup>[6]</sup>的失败主要归因于使用不恰当的目标函数和不准确的观测值作为输入。在本研究中，我们重新审视了这些问题，并引入了一种新颖的混合约束显式平移平均框架(HETA)。

我们的贡献涵盖了三个关键方面：(1) 将使用特征轨迹的全局平移估计方法分为显式和隐式方法，并重新审视它们的优势和劣势。(2) 对两种形式的线性目标函数进行了比较分析，并引入了一种新颖的混合约束显式方法，通过鲁棒的 $L_1$ 范数优化，随后进行无偏的 $L_2$ 范数优化，在两个步骤中同时估计摄像机和点。(3) 为了提高相对平移的准确性，我们使用极线几何中的共面约束对其进行了重新估计。为了增强这种重新估计鲁棒性，我们分析了视差角影响，并滤除了不稳定特征匹配。

## 二、融合特征轨迹的平移平均相关工作

我们重新审视了显式和隐式方法，并对它们的优势和劣势进行了彻底分析。假设从三维点 $P_k$ 发出的光线在 $n$ 个相机平面上生成 $n$ 个投影特征点。对于一个特征点，我们将其在局部归一化摄像机坐标系中的坐标表示为 $X_{ki}^T = (x_{ki}, y_{ki}, 1)^T$ ，其中 $k$ 是特征轨迹或三维点的索引， $i$ 是相机的索引。在全局相机坐标系中，三维点和相机之间的关系满足：

$$\frac{P_k - t_i}{\|P_k - t_i\|_2} = R_i^T \frac{X_{ki}}{\|X_{ki}\|_2} = f_{ki} \quad (2)$$

其中 $t_i$ 表示相机位置， $f_{ki}$ 表示从相机 $t_i$ 到三维点 $P_k$ 的归一化特征射线。从公式(1)和公式(2)可知，相机到相机约束和相机到点约束的数学表达式是等价的。显式方法的核心思想是使用与估计相机平移相同的目标函数

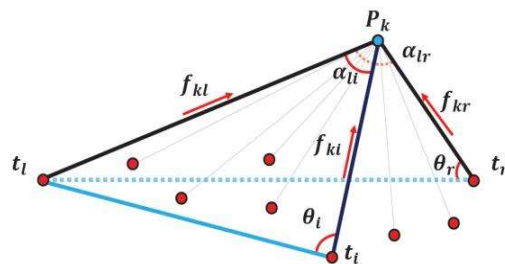


图2 摄像机与摄像机和摄像机与三维点间的约束示例。其中红色的点表示摄像机，蓝色点表示一个一个特征轨迹对应的三维点，红色的箭头代表特征射线的方向，基相机 $t_l, t_r$ 在特征轨迹中拥有最大的视差角。

来估计三维点。与低视差场景下误差较大的相对平移相比，特征射线作为从图像中导出的原始信息，自然表现出更高的精度。因此，在显式方法中使用特征射线在理论上可以提供比仅依赖相对平移的方法更高的精度。隐式方法主要通过两种方式约束相机。一类方法利用特征点的深度一致性来计算相机基线尺度。另一类方法用已有的观测量如特征射线和相对平移线性表示三维点，并基于这些三维点及其相应的特征射线来约束相机。对于第一类方法，我们以2016年Cui等人<sup>[2]</sup>的方法为例。如图2所示，通过三维点到其可见相机的连接构造了两个相邻三角形 $\{P_k - t_l - t_r\}$ 和 $\{P_k - t_l - t_i\}$ 。根据正弦定理，两个相机基线尺度的比例为：

$$\frac{\|t_l - t_r\|_2}{\|t_l - t_i\|_2} = \frac{\sin \theta_i \cdot \sin \alpha_{lr}}{\sin \theta_r \cdot \sin \alpha_{li}} \quad (3)$$

然而，在低视差场景中，这种方法对异常值非常敏感。一方面，在低视差场景中，相对平移估计不准确，导致诸如 $\theta_i$ 和 $\theta_r$ 等角度不正确。另一方面，分母中的低视差角，例如 $\alpha_{li}$ ，导致数值不稳定性。这意味着视差角的轻微变化会导致比率计算中的显著变化。

对于第二类隐式方法，为看到共同场景点的相机导出了线性约束。然而，在 2015 年 Cui 等人<sup>[3]</sup>的方法和 PGILP<sup>[9]</sup>中，对三维点的表示仍然依赖于相对平移，其误差累积到所表示的三维点中。为了解决这些问题，2021 年 Cai 等人提出了一种 LiGT<sup>[8]</sup>约束，通过在每个特征轨迹中选择具有最大视差角的两个基相机中的特征射线，线性表示其对应的三维点。如图 2 所示，对于一个特征轨迹，其中两个基相机 $t_l, t_r$ 的，特征点在相机 $l$ 中的深度计算为：

$$\begin{aligned} \|P_k - t_l\|_2 &= \frac{\|t_l - t_r\|_2 \cdot \sin \theta_r}{\sin \alpha_{lr}} \\ &= \frac{\|f_{kr} \times (t_l - t_r)\|_2}{\|f_{kl} \times f_{kr}\|_2} \\ &= \frac{((f_{kl} \times f_{kr}) \times f_{kr}) \cdot (t_l - t_r)}{\|f_{kl} \times f_{kr}\|_2^2} \end{aligned} \quad (4)$$

由公式 (4) 可知，三维点可由两个基相机及其特征射线表示：

$$P_k = t_l + \frac{f_{kl}((f_{kl} \times f_{kr}) \times f_{kr})^T}{\|f_{kl} \times f_{kr}\|_2^2} (t_l - t_r) \quad (5)$$

结合公式 (2)，可以通过隐式表达三维点和其对应特征轨迹里其他相机（如相机 $i$ ）建立相机和点之间的约束。

我们从两个方面比较显式与隐式方法。在鲁棒性方面，显式方法通常优于隐式方法，因为在显式方法中，每个特征轨迹的三维点是通过所有特征射线进行估计的，而在隐式方法中，三维点仅由两个基相机的特征射线表示。在效率方面，尽管隐式方法避免了显式地优化三维点，但新引入摄像机到点约束增加了相机之间的连接性，从而破坏了传统全局平移平均方法的优化矩阵的稀疏特性。通过实验发现显式方法与隐式方法效率相当。

### 三、混合约束显式平移平均框架(HETA)

尽管引入特征轨迹约束带来了许多好处，但当特征轨迹中存在很多异常值时，求解的相机平移也会包含较大的噪声。与特征射线相比，相对平移可以在摄像机之间提供更直接和严格的约束。为了增强鲁棒性和效率，与 PGILP<sup>[9]</sup>和 LiGT<sup>[8]</sup>等方法使用所有的特征轨迹不同，

我们不但利用相对平移来提供相机之间直接约束，而且选择更可靠特征轨迹提供相机与点之间约束。

具体思路为，首先构建一个场景及轨迹图 $G = \{V \cup P, E_v \cup E_p\}$ ，其中 $V$ 中的每个节点表示一个相机， $P$ 中的每个节点表示一个三维点， $E_v$ 中的每个边连接 $V$ 中的相机对， $E_p$ 中的每个边表示相机到三维点的特征射线。设 $C$ 为相机到相机约束， $P$ 为相机到点约束。目标函数可以表达为：

$$\min_{V, P} \sum_{E_v} \rho(\|C\|_p) + \sum_{E_p} \rho(\|P\|_p) \quad (6)$$

其中 $p$ 表示优化范数， $\rho(\cdot)$ 表示鲁棒估计函数。考虑到鲁棒性，有三个主要任务：(1) 增强低视差场景中相对平移的精度；(2) 选择可靠的特征轨迹；(3) 为两种类型的约束定义鲁棒的目标函数。我们将在接下来几小节中解决这些任务，并在最后给出我们方法的完整框架。

#### 3.1 相对平移的重估计

在对极几何中，共面约束可以表示为 $X_{kj} \cdot (t_{ij} \times R_{ij} X_{ki}) = 0$ 。在给定全局相机姿态时，可以改写为：

$$\begin{aligned} (R_i^T X_{ki} \times R_j^T X_{kj}) \cdot R_j^T t_{ij} &= 0 \\ \Leftrightarrow (f_{ki} \times f_{kj}) \cdot v_{ij} &= 0. \end{aligned} \quad (7)$$

由公式 (7) 可知，每个相对平移可以使用极线平面向量重新估计，该法向量由 $f_{ki} \times f_{kj}$ 计算得到。由于相机内参数和全局旋转的不准确性，从特征射线估计的法向量不可避免地具有一些角度误差。LUD<sup>[5]</sup>中提出一种方法通过最小化相对平移和法向量之间余弦角来使用所有归一化法向量重新估计相对平移。然而，由于法向量的准确性也受到视差角的影响，因此在估计过程中对每个法向量采用相同的权重是不合理的。与 LUD<sup>[5]</sup>中归一化法向量 $f_{ki} \times f_{kj}$ 的方法相比，通过理论推导和实验发现，在估计过程中使用非归一化的法向量估计相对平移，可以为每个特征匹配保留合理的权重 $\|f_{ki} \times f_{kj}\|_2 = \sin \alpha$ ，改进的目标函数如下所示：

$$\min_{v_{ij}} \sum_k \rho(\|(f_{ki} \times f_{kj}) \cdot v_{ij}\|_2) \quad \text{s.t.} \quad \|v_{ij}\|_2 = 1 \quad (8)$$

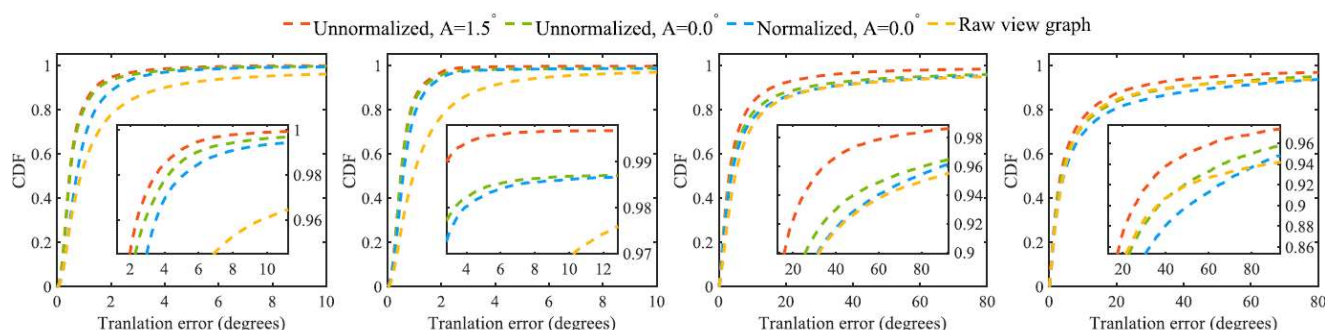


图3 在部分 KITTI<sup>[16]</sup>和 1DSfM<sup>[6]</sup>数据集上, 比较原始场景图、是否归一化法向量, 以及过滤小视差角相对平移后相对平移角度误差的累积分布函数结果, 从左至右, 数据集分别为 KITTI-06, KITTI-09, 1DSfM-PIC, 1DSfM-ROF。

此外, 当低视差角的法向量误差变得非常大时, 基于共面性一致性来估计相对平移或验证特征匹配变得无效。因此, 在估计相对平移之前, 我们预先定义一个阈值  $A$  过滤视差角低于该阈值的特征匹配。

### 3.2 特征轨迹的挑选策略

通过重新估计相对平移, 我们筛选出违反共面性或可见性约束的特征匹配, 并使用并查集算法构建特征轨迹。由于特征轨迹可能包含高比例的特征射线异常值, 因此只选择部分特征轨迹以提高效率和鲁棒性。根据公式 (8), 具有较大视差角的特征匹配的共面性一致性更可靠。因此, 根据它们最大视差角, 对所有特征轨迹进行降序排序。然后, 依次检查每个特征轨迹, 确定它是否能与覆盖次数不足的图像之间建立连接。此过程一直持续, 直到所选特征轨迹子集至少覆盖所有相机  $N$  次。

### 3.3 优化目标函数的定义

相机到相机和相机到点的约束都可以表示为以下公式:  $\mathbf{s}_i - \mathbf{s}_j = \|\mathbf{s}_i - \mathbf{s}_j\|_2 \cdot \mathbf{s}_{ij}$ , 其中  $\mathbf{s}_i, \mathbf{s}_j$  表示相机或点,  $\mathbf{s}_{ij}$  表示从  $\mathbf{s}_j$  到  $\mathbf{s}_i$  的已知归一化向量, 例如特征射线或相对平移。我们比较了两种线性目标函数, 包括叉积形式  $\|\mathbf{s}_{ij} \times (\mathbf{s}_i - \mathbf{s}_j)\|_2$  和尺度形式  $\|\mathbf{s}_i - \mathbf{s}_j - \lambda_{ij} \mathbf{s}_{ij}\|_2$ , 其中  $\lambda_{ij}$  是一个尺度变量,  $\times$  表示为叉乘。为了消除尺度和方向的不确定性, 分别对叉积形式和尺度形式的目标函数使用不等式约束  $\mathbf{s}_{ij} \cdot (\mathbf{s}_i - \mathbf{s}_j) \geq 1$  和  $\lambda_{ij} \geq 1$ 。

假设  $\mathbf{s}_{ij}^G$  是  $\mathbf{s}_{ij}$  的真值。在大多数情况下, 当  $\mathbf{s}_{ij} \cdot \mathbf{s}_{ij}^G \geq 0$  时, 两个不等式约束都定义了正确的可行区域。在最优解的情况下, 两个目标函数中残差的大小是相同的。与此同时, 尺度形式中的  $\lambda_{ij}$  等于  $\mathbf{s}_{ij} \cdot (\mathbf{s}_i - \mathbf{s}_j)$ , 表示  $\mathbf{s}_i - \mathbf{s}_j$

在  $\mathbf{s}_{ij}$  上的投影的大小。因此,  $\lambda_{ij}$  是一个冗余变量, 因为它完全由当前的  $\mathbf{s}_i, \mathbf{s}_j$  和已知的  $\mathbf{s}_{ij}$  确定。此外, 当尺度变量的变化范围很大, 例如在基线长度或特征射线深度差异较大的情况下, 它们通常难以收敛到最优解。这显著影响了实际优化的整体精度和效率。

当  $\mathbf{s}_{ij}$  存在显著误差导致  $\mathbf{s}_{ij} \cdot \mathbf{s}_{ij}^G < 0$  时, 尺度形式中的  $\lambda_{ij}$  等于最低界限 1, 以最小化惩罚。叉积形式的不等式约束提供了一个错误的可行区域, 导致一个偏置的解决方案。然而, 通过我们的相对平移重新估计, 整体相对平移的精度得到提升。此外, 1DSfM<sup>[6]</sup>过滤方法通过多个随机的一维投影很好地过滤方向误差较大的相对平移, 可以过滤掉大多数存在显著误差的相对平移, 因此, 使用叉积形式的目标函数以获得更好的收敛性。

### 3.4 优化框架

为了避免来自特征射线的冗余和不正确的约束, 仅使用相机到相机的不等式约束来消除尺度和方向的模糊性。为了提高鲁棒性, 目标函数在  $L_1$  范数下使用 ADMM 方法<sup>[14]</sup>进行优化, 如下所示:

$$\begin{aligned} \min_{\substack{\mathbf{t}_i, i \in V; \\ \mathbf{P}_k, k \in P}} \quad & \sum_{ij \in E_v} \|\mathbf{v}_{ij} \times (\mathbf{t}_i - \mathbf{t}_j)\|_1 + \sum_{ki \in E_p} \|\mathbf{f}_{ki} \times (\mathbf{P}_k - \mathbf{t}_i)\|_1 \\ \text{s.t.} \quad & \sum_{i \in V} \mathbf{t}_i = \mathbf{0}, \quad \mathbf{v}_{ij} \cdot (\mathbf{t}_i - \mathbf{t}_j) \geq 1, \quad \forall ij \in E_v \end{aligned} \quad (9)$$

其中这两个约束被用来消除内在的位置和尺度模糊。然而, 正如 BATA<sup>[7]</sup>中所提到的, 对于相机基线和特征点深度尺度不同的目标函数, 这个优化结果是有偏的。因此, 可以使用一个无偏的基于角度的目标函数来进一步优化当前的结果。在每次迭代中, 使得:  $\hat{\mathbf{v}}_{ij} = \frac{\mathbf{t}_i - \mathbf{t}_j}{\|\mathbf{t}_i - \mathbf{t}_j\|_2}$ ,

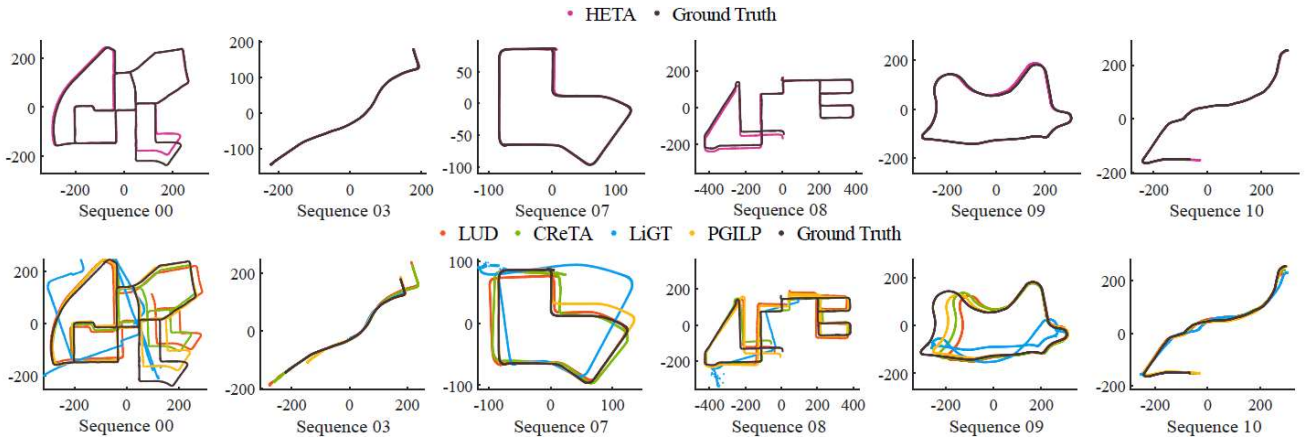


图 4 KITTI 数据集<sup>[16]</sup>上估计的相机运动轨迹。对比方法包括 LUD<sup>[5]</sup>、CReTA<sup>[4]</sup>、PGILP<sup>[9]</sup>和 LiGT<sup>[10]</sup>。

$\hat{f}_{ki} = \frac{P_k - t_i}{\|P_k - t_i\|_2}$ , 然后, 使用 IRLS 方法来优化目标函数。

$$\min_{\substack{t_i, i \in V; \\ P_k, k \in P}} \sum_{ij \in E_v} \rho(H(\hat{v}_{ij})) + \sum_{ki \in E_p} \rho(H(\hat{f}_{ki})), \quad s.t. \sum_{i \in V} t_i = 0,$$

$$\text{where } H(\hat{s}_{ij}) = \begin{cases} \|\hat{s}_{ij} \times \hat{s}_{ij}\|_2, & \hat{s}_{ij} \cdot \hat{s}_{ij} \geq 0; \\ 1, & \hat{s}_{ij} \cdot \hat{s}_{ij} < 0. \end{cases} \quad (10)$$

## 四、实验结果

### 4.1 相对平移的重估计

我们进行实验证明了使用未归一化法线向量和过滤低视差角特征匹配对重新估计的相对平移的影响。在公式 (9) 中, 误差损失宽度  $\beta$  设置为  $\sin 1^\circ \cdot \sin 5^\circ$ , 表明期望当视差角  $\alpha$  等于  $1^\circ$  时, 错误角度  $\gamma$  应小于  $5^\circ$ 。如图 3 所示, 相对平移误差的累积分布函数 (CDF) 表明使用未归一化法线向量显著提高了相对平移的准确性。通过过滤视差角低于  $1.5$  度的特征匹配, 进一步提高了准确性。如图 3 所示, 1DSfM-ROF 数据中, 由于相机全局旋转

精度不足, 使用归一化法线向量估计的相对平移的准确性不佳。然而, 在这种情况下, 我们的方法仍然产生了更好的结果。

### 4.2 有序数据集上的评估结果

KITTI 数据集<sup>[16]</sup>是使用安装在驾驶汽车上的两台摄像机收集的。因此, 大多数特征匹配的视差角受到限制, 相机运动轨迹倾向于近似共线, 这对全局平移估计系统提出了重大挑战。估计的相机运动轨迹如图 4 所示。HETA 实现了最高的准确性。尽管使用了非常精确的相对平移, LUD<sup>[5]</sup>和 CReTA-BATA<sup>[4]</sup>都难以产生准确的结果。LiGT<sup>[8]</sup>方法基于矩阵分解估计相机平移, 提高了效率但缺少鲁棒性。相比之下, PGILP<sup>[9]</sup>通过在  $L_1$  范数下优化每个相机到点的约束产生更好的结果。通过这些比较, 我们可以得出结论, 我们的方法 HETA 在准确性和鲁棒性方面均超过了所有比较方法。

Data	LUD			CReTA-BATA			LiGT			PGILP			1DSfM			HETA							
	BA			BA			BA			BA			BA			$L_1$		$L_2$		BA			
	$\bar{e}$	$\bar{e}$	$N_c$	$\bar{e}$	$\bar{e}$	$N_c$	$\bar{e}$	$\bar{e}$	$N_c$	$\bar{e}$	$\bar{e}$	$N_c$	$\bar{e}$	$\bar{e}$	$N_c$	$\bar{e}$	$\bar{e}$	$\bar{e}$	$\bar{e}$	$\bar{e}$	$N_c$		
ALM	497	0.1	0.5	483	0.1	0.4	487	0.3	1.8	422	0.1	0.5	486	0.3	4.6	389	0.5	1.2	0.5	1.2	0.1	0.4	<b>488</b>
ELS	217	0.2	0.4	212	0.2	0.4	215	0.2	0.4	204	0.2	0.4	<b>216</b>	0.2	0.4	196	2.4	3.9	2.1	3.8	0.2	0.4	<b>216</b>
GDM	590	0.1	3.4	560	0.1	3.6	561	5.1	1e3	504	0.2	4.1	556	0.4	66.5	475	2.8	10.4	2.2	10.1	0.2	2.7	<b>564</b>
MDR	178	0.2	6.3	168	0.2	5.6	170	8.6	16.4	137	0.2	9.7	170	0.8	9.8	122	1.4	9.7	1.4	9.6	0.2	7.0	<b>174</b>
MND	403	0.1	0.1	399	0.1	0.1	399	0.1	0.1	383	0.1	0.1	398	0.1	0.3	363	0.5	1.0	0.5	1.0	0.1	0.1	<b>400</b>
ND	479	0.1	0.6	457	0.1	0.3	468	6.2	7.3	397	0.1	0.3	462	0.1	47.5	374	0.3	1.4	0.3	0.9	0.1	0.3	<b>476</b>
NYC	296	0.1	0.2	290	0.1	0.3	<b>294</b>	0.1	1.5	222	0.1	0.2	285	0.1	5.7	261	0.7	1.8	0.5	1.5	0.1	0.1	290
PDP	295	0.1	0.4	287	0.1	0.1	286	7.4	2e2	108	0.1	0.3	290	0.1	1.9	249	1.1	2.9	1.1	2.9	0.1	0.2	<b>291</b>
PIC	1838	0.1	0.5	1797	0.1	0.5	<b>1811</b>	12.3	81.8	649	0.1	0.5	1774	0.5	3.1	1621	0.9	1.9	0.7	1.7	0.1	0.4	1807
ROF	918	0.1	0.2	875	0.1	0.2	899	0.6	3.1	732	0.1	0.5	892	0.8	23.5	725	1.9	3.9	1.2	3.3	0.1	0.1	<b>907</b>
TFG	3989	0.9	2.5	3864	0.7	1.8	3913	37.5	45.8	789	1.2	4.3	3860	12.1	18.9	3348	3.3	6.4	2.6	5.8	0.7	2.4	<b>3951</b>
TOL	396	0.5	3.6	<b>391</b>	0.2	4.8	387	70.3	76.0	152	0.3	4.3	380	3.2	7e2	276	2.5	4.9	2.1	4.5	0.4	1.5	387
USQ	637	0.3	2.8	582	0.3	4.4	603	6.7	1e2	336	0.5	4.5	602	0.4	1e2	505	4.2	7.5	3.6	7.2	0.2	2.1	<b>619</b>
VNC	713	0.2	5.7	672	0.2	9.7	<b>702</b>	20.5	28.2	474	0.2	9.1	664	0.2	4.5	556	1.8	4.2	1.7	4.0	0.1	0.8	686
YKM	337	0.1	0.1	327	0.1	0.1	329	0.1	0.3	318	0.1	0.2	323	2.9	23.9	262	1.2	2.4	1.1	2.1	0.1	0.2	<b>333</b>

表 1 1DSfM 数据集上相机位置误差。  $N_t$  表示场景图中摄像机的数目,  $N_c$  为捆绑调整 (BA) 后图像的注册数目, 其最优结果被加粗表示。  $\bar{e}$  和  $\bar{e}$  分别表示摄像机位置误差的中值和均值, 单位为米。

### 4.3 无序数据集上的评估结果

1DSfM 数据集<sup>[6]</sup>是通过许多不同类型的摄像机收集的。由于提供的相机内参数的有限精度和大量错误的特征匹配，估计的相对姿态存在较大误差。因此，由 Chatterjee 方法估计<sup>[14]</sup>的全局旋转的精度低于 KITTI 数据集<sup>[16]</sup>，使得全局平移估计更具挑战性。BA 后的标定结果如表 1 所示。从这个比较中，LUD<sup>[5]</sup>、CReTA-BATA<sup>[4]</sup>和 HETA 估计的相机位置的准确性相当。然而，对于大多数数据，HETA 注册的图像数量最多，表明与 LUD<sup>[5]</sup>和 CReTA-BATA<sup>[4]</sup>相比，它具有更高的鲁棒性。对于依赖于所有特征轨迹作为输入的隐式方法 LiGT<sup>[8]</sup>和 PGILP<sup>[9]</sup>，它们的性能不及 HETA。这种差异是由于

在优化过程中包含了大量来自特征轨迹外点的错误约束。

## 五、总结

我们重新审视了利用特征轨迹进行全局平移估计的问题，并提出了一种新颖混合约束显式平移平均框架。我们的方法在顺序和无序数据集上表现优越，超过了许多现有的最先进方法。然而，特征匹配异常值的普遍存在仍然限制了全局运动恢复结构算法广泛应用。在未来，我们期望利用神经网络，从数据中学习更多先验信息，以提高特征轨迹的匹配精度，提升算法的鲁棒性。

责任编辑 王金甲

## 参考文献

- [1] Tao P, Cui H, Rong M, et al. Revisiting Global Translation Estimation with Feature Tracks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 20686-20696.
- [2] Hainan Cui, Shuhan Shen, and Zhanyi Hu. Robust global translation averaging with feature tracks. In IEEE International Conference on Pattern Recognition, pages 3727–3732, 2016.
- [3] Zhaopeng Cui, Nianjuan Jiang, Chengzhou Tang, and Ping Tan. Linear global translation estimation with feature tracks. In British Machine Vision Conference, pages 46.1–46.13, 2015.
- [4] Lalit Manam and Venu Madhav Govindu. Correspondence reweighted translation averaging. In European Conference on Computer Vision, pages 56–72. Springer, 2022.
- [5] Onur Ozyesil and Amit Singer. Robust camera location estimation by convex programming. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2674–2683, 2015.
- [6] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In European Conference on Computer Vision, pages 61–75. Springer, 2014.
- [7] Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Baseline desensitizing in translation averaging. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4539–4547, 2018.
- [8] Qi Cai, Lilian Zhang, Yuanxin Wu, Wenxian Yu, and Dewen Hu. A pose-only solution to visual reconstruction and navigation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1):73–86, 2021.
- [9] Liyang Liu, Teng Zhang, Brenton Leighton, Liang Zhao, Shoudong Huang, and Gamini Dissanayake. Robust global structure from motion pipeline with parallax on manifold bundle adjustment and initialization. IEEE Robotics and Automation Letters, 4(2):2164–2171, 2019.
- [10] Johannes Lutz Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In IEEE Conference on Computer Vision and Pattern Recognition, pages: 4104-4113, 2016.

- [11] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms, pages 298–372. Springer, 2000.
- [12] Federica Arrigoni, Andrea Fusiello, Elisa Ricci, and Tomas Pajdla. Viewing graph solvability via cycle consistency. In IEEE International Conference on Computer Vision, pages 5540–5549, 2021.
- [13] Federica Arrigoni, Tomas Pajdla, and Andrea Fusiello. Viewing graph solvability in practice. In IEEE/CVF International Conference on Computer Vision, pages 8147–8155, 2023.
- [14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning, 3 (1):1–122, 2011.
- [15] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In IEEE International Conference on Computer Vision, pages 521–528, 2013.
- [16] 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.
- [17] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. Complete solution classification for the perspective-three-point problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8):930–943, 2003.
- [18] Rodrigo Chacón Quesada and Yiannis Demiris. Design and evaluation of an augmented reality head-mounted display user interface for controlling legged manipulators. In IEEE International Conference on Robotics and Automation, pages 11950–11956, 2023.
- [19] Linning Xu, Yuanbo Xiangli, Sida Peng, Xingang Pan, Nanxuan Zhao, Christian Theobalt, Bo Dai, and Dahua Lin. Grid-guided neural radiance fields for large urban scenes. In IEEE Conference on Computer Vision and Pattern Recognition, pages 8296–8306, 2023.
- [20] Eric Brachmann, Martin Humenberger, Carsten Rother, and Torsten Sattler. On the limits of pseudo ground truth in visual camera re-localisation. In IEEE International Conference on Computer Vision, pages 6218–6228, 2021.



## 陶沛霖

中国科学院自动化研究所硕士生，研究方向：全局式运动恢复结构，三维重建。  
Email: taopeilin2023@ia.ac.cn



## 崔海楠

中国科学院自动化研究所副研究员，研究方向：大规模场景三维建模，从运动恢复结构和视觉定位。  
Email: hncui@nlpr.ia.ac.cn



### 荣梦琪

中国科学院自动化研究所助理研究员，研究方向为三维场景理解，包括细粒度三维语义分割、多模态三维分割基础模型等。

Email: mengqi.rong@ia.ac.cn



### 申抒含

中国科学院自动化研究所研究员，研究领域为三维计算机视觉理论与应用，包括大规模场景三维重建、智能机器人三维环境感知、场景三维语义理解等。

Email: shshen@nlpr.ia.ac.cn