

## 目标3D模型不准确条件下的单目位姿测量

郭良超, 陈霖, 张倬, 孙晓亮\*, 于起峰

(国防科技大学 空天科学学院, 湖南长沙 410073)

**摘要:**已有单目视觉引导中平台间位姿高精度测量方法需要准确的目标平台3D模型,无法消除3D模型误差给位姿测量带来的影响。针对此问题,对目标平台3D模型和位姿进行迭代优化,并提出一种新的单目视觉测量方法:采用稀疏3D关键点集合建模目标平台3D模型,利用序列图像中多视图几何约束信息,将目标稀疏3D关键点集合和6D位姿作为待求解参数,以最小化物方残差建立目标函数,通过求解该最优化问题,迭代优化稀疏3D关键点集合及位姿,通过采用滑动窗口结合关键帧筛选策略,实现实时、在线的高精度单目视觉测量。实验结果表明,通过迭代优化稀疏3D关键点集合及位姿,所提方法实现了目标平台3D模型不准确条件下实时、在线的高精度单目位姿测量,同时提升了目标3D模型精度。

**关键词:**单目位姿测量;多视图几何约束;迭代优化;物方残差

**中图分类号:**TP391.4 **文献标志码:**A **文章编号:**1001-2486(2025)06-178-11



论  
文  
拓  
展

## Monocular pose measurement method for inaccurate 3D model of a target

GUO Liangchao, CHEN Lin, ZHANG Zhuo, SUN Xiaoliang\*, YU Qifeng

(College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** In monocular vision-guided high-precision inter-platform pose measurement, existing methods require an accurate 3D model of the target platform and are unable to eliminate the impact of 3D model errors on pose measurement. To address this issue, iterative optimization was performed on the 3D model of the target platform and pose, and a new monocular vision measurement method was proposed. Specifically, the target platform's 3D model was modeled using a set of sparse 3D keypoints. By leveraging multi-view geometric constraint information in sequential images, the sparse 3D keypoint set of the target and 6D pose were treated as parameters to be solved. An objective function was established to minimize object-space residuals, and through solving this optimization problem, iterative optimization of the sparse 3D keypoint set and pose was achieved. Additionally, a sliding window combined with a keyframe selection strategy was adopted to realize real-time and online high-precision monocular vision measurement. Experimental results demonstrate that, through iterative optimization of the sparse 3D keypoint set and pose, the proposed method achieves real-time, online high-precision monocular pose measurement under the condition of an inaccurate 3D model of target platform, while simultaneously improving the accuracy of the target's 3D model.

**Keywords:** monocular pose measurement; multiple view geometry constraint; iterative optimization; object-space residual

单目视觉引导仅利用平台搭载的单个视觉测量相机完成平台间相对6D位姿参数(3D旋转和3D平移)精确测量<sup>[1-3]</sup>,具有配置简单、成本低、自主性高、抗干扰能力强等优势,在飞行器着舰、空间平台交会对接等双动平台交会场景中具有广阔应用前景,得到了研究人员的广泛关注。

单目位姿测量是单目视觉引导中的关键技术,当前基于单目视觉的位姿测量方法依据是否

使用合作标志可主要归结为合作和非合作两类方法:合作方法通过在目标平台上布设合作标志实现平台间位姿测量<sup>[4-7]</sup>,算法简单、处理效率高,但应用场景较为受限。非合作方法无须布设合作标志,利用目标自身特征与图像信息完成平台间位姿参数解算,硬件配置简单、应用限制少,但可靠特征提取、定位难度大,是当前单目位姿测量领域研究的热点。已有基于非合作模式的位姿视觉

收稿日期:2024-05-01

基金项目:国家自然科学基金资助项目(12272404);湖南省科技创新计划资助项目(2023RC3023)

第一作者:郭良超(1995—),男,河南三门峡人,博士研究生,E-mail:guolianghchao19@nudt.edu.cn

\*通信作者:孙晓亮(1989—),男,安徽亳州人,副研究员,博士,硕士生导师,E-mail:alexander\_sxl@nudt.edu.cn

引用格式:郭良超,陈霖,张倬,等.目标3D模型不准确条件下的单目位姿测量[J].国防科技大学学报,2025,47(6):178-188.

Citation: GUO L C, CHEN L, ZHANG Z, et al. Monocular pose measurement method for inaccurate 3D model of a target[J]. Journal of National University of Defense Technology, 2025, 47(6): 178-188.

测量方法可进一步分为传统方法和深度学习相关方法。传统方法通过检测图像中目标自身点、直线、轮廓等特征,并结合目标3D先验信息,实现位姿求解<sup>[8-10]</sup>,此类方法难以处理弱纹理目标,且易受复杂环境干扰等影响。得益于深度神经网络强大的特征提取与表征能力,深度学习相关方法已成为当前单目位姿测量的主流方法,并在多个公开测试数据集上取得了优于传统方法的性能。已有基于深度学习的位姿视觉测量方法依据网络输出形式可归结为直接法和间接法<sup>[11]</sup>。直接法<sup>[12-14]</sup>获取输入图像并通过回归网络直接得到6D位姿参数,实现了端到端的网络训练与推理,但位姿测量的精度不高<sup>[15]</sup>。间接法引入目标的中间模型,结合迭代最近点(iterative closest point, ICP)、透视 $n$ 点(perspective- $n$ -point, PnP)等算法求解位姿参数。常用的中间模型包括稠密模型、归一化坐标和稀疏关键点。基于稠密模型的位姿测量方法匹配图像或场景中的每个像素<sup>[16-18]</sup>,具有较高的精度和稳定性,但计算效率较低。基于归一化坐标的方法通过归一化对象坐标空间将同类别物体进行归一化,无须获取目标的精确3D模型<sup>[19-22]</sup>,但是依赖深度图像去恢复物体的真实尺度。稀疏关键点表示方法将目标建模为稀疏关键点集合,此类方法简单高效,适用于实时位姿测量任务<sup>[23-28]</sup>,当前,基于稀疏关键点集合的方法已被广泛使用。相较于直接法,间接法通过建立稀疏或稠密的2D-3D对应关系,并求解PnP问题,能够获得更加精确的位姿测量。已有单目位姿测量方法均依赖准确的目标3D先验信息,通过建立2D-3D对应或3D-3D对应解算位姿参数,无法实现3D模型不准确情况下的高精度位姿求解。然而,在实际应用中,准确的目标3D模型可能难以获取,如对于飞行器、卫星等敏感目标,此外,部分交会对象由于加工等因素也会出现理论设计3D模型和目标实物不符的情况,难以保证目标3D模型的准确性。针对目标3D模型不准确条件下的高精度位姿测量,当前尚无有效的解决方法。

多视图几何约束在结构重建中有着广泛的应用,其利用图像间的内在关系恢复物体或场景的三维结构,并同时计算相机位姿。例如,同步定位和建图<sup>[29-32]</sup>(simultaneous localization and mapping, SLAM)、运动恢复结构<sup>[33-36]</sup>(structure from motion, SFM)、目标定位<sup>[37]</sup>和飞行器自身高精度定位<sup>[38]</sup>等。类似地,在单目位姿测量中,由于相对运动的存在,图像序列提供了多视图间的

几何约束。因此,本研究尝试使用多视图几何约束来处理缺少精确的目标3D模型的单目位姿测量问题。传统的SLAM和SFM等方法均利用特征点提取和匹配建立帧间匹配,通过优化求解位姿进行定位。然而,目标3D平台无法基于传统特征点建立对应关系,此外,即使成功建立特征点对应关系,也难以将其与具体的3D模型进行关联。因此,复杂的特征点提取和匹配方法被放弃,转而利用现有的基于神经网络的关键点检测算法直接回归与预定义3D关键点对应的2D投影。

综上所述,单目图像序列内部的多视图几何约束被运用来解决上述问题,在此基础上提出了一种新的单目位姿测量方法,对目标的3D模型和位姿进行迭代优化。与人体姿态估计中使用的关键点类似,使用一组稀疏关键点来表示目标的3D模型,并使用训练好的神经网络从输入图像中检测关键点,结合2D-3D对应关系,通过求解PnP问题得到初始位姿估计。将以线性参数方程形式表示的3D关键点和位姿作为待优化参数,最小化3D关键点和观测视线之间的物方残差来构建优化目标函数,通过求解该优化问题,对目标的3D模型和位姿进行迭代优化。所提方法在目标3D模型不准确的情况下,也能得到高精度的位姿测量结果。此外,滑动窗口结合关键帧筛选策略被采用以实现实时、在线的高精度单目位姿测量。

## 1 方法

为在目标3D模型不准确的情况下实现可靠的单目位姿测量,提出一种利用图像序列内多视图几何约束的迭代优化方法。需要说明的是,所描述的方法以飞行器着舰为应用背景,但同样适用于其他平台交会场景中的视觉引导任务。

### 1.1 问题公式和符号

$T_0^c$ 表示从物体坐标系到相机坐标系的刚体变换。

$$T_0^c = \begin{bmatrix} R_0^c & t_0^c \\ \mathbf{0} & 1 \end{bmatrix} \in SE(3) \quad R_0^c \in SO(3), t_0^c \in \mathbb{R}^3 \quad (1)$$

式中, $R_0^c$ 和 $t_0^c$ 分别是 $T_0^c$ 的旋转和平移分量(为书写方便,后文 $R_0^c$ 简写为 $R$ , $t_0^c$ 简写为 $t$ ), $SE(3)$ 为特殊欧氏群, $SO(3)$ 为特殊正交群, $\mathbb{R}^3$ 为三维向量空间。相机内参矩阵 $K$ 被预先标定且为一固定不变值,如式(2)所示。

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

目标 3D 模型的顶点  $V^0$  在二维图像上的透视投影被建模为:

$$x = \pi[K(T_0^c \tilde{V}^0)_{3 \times 1}] \quad (3)$$

式中,  $\tilde{V}^0$  为  $V^0$  的齐次坐标形式,  $x = [u \ v]^T$  为二维图像点, 并且  $\pi(X) = [x/z \ y/z]^T$ 。

传统上, 单目位姿测量提取图像中的 2D 特征, 结合目标的精确 3D 信息建立 2D-3D 对应关系并进行高精度 6D 位姿求解。然而, 在目标 3D 模型不准确时, 传统方法无法利用准确的 2D-

3D 对应获得高精度位姿。因此, 本研究旨在解决 3D 模型不准确情况下的单目位姿测量问题。

### 1.2 3D 模型与位姿迭代优化的单目位姿测量

提出使用一组稀疏 3D 关键点来表示目标的 3D 模型, 通过检测到的 2D 关键点结合求解 PnP 问题得到初始位姿, 并通过光束法平差 (bundle adjustment, BA) 对位姿和 3D 模型进行迭代优化。该迭代优化方法的整体架构如图 1 所示, 所提出的单目位姿测量方法主要包括基于稀疏关键点的 3D 模型表示、初始位姿估计、3D 模型与位姿的迭代优化、滑动窗口和关键帧提取。为简洁起见, 图 1 中仅显示部分用红点标注的关键点。

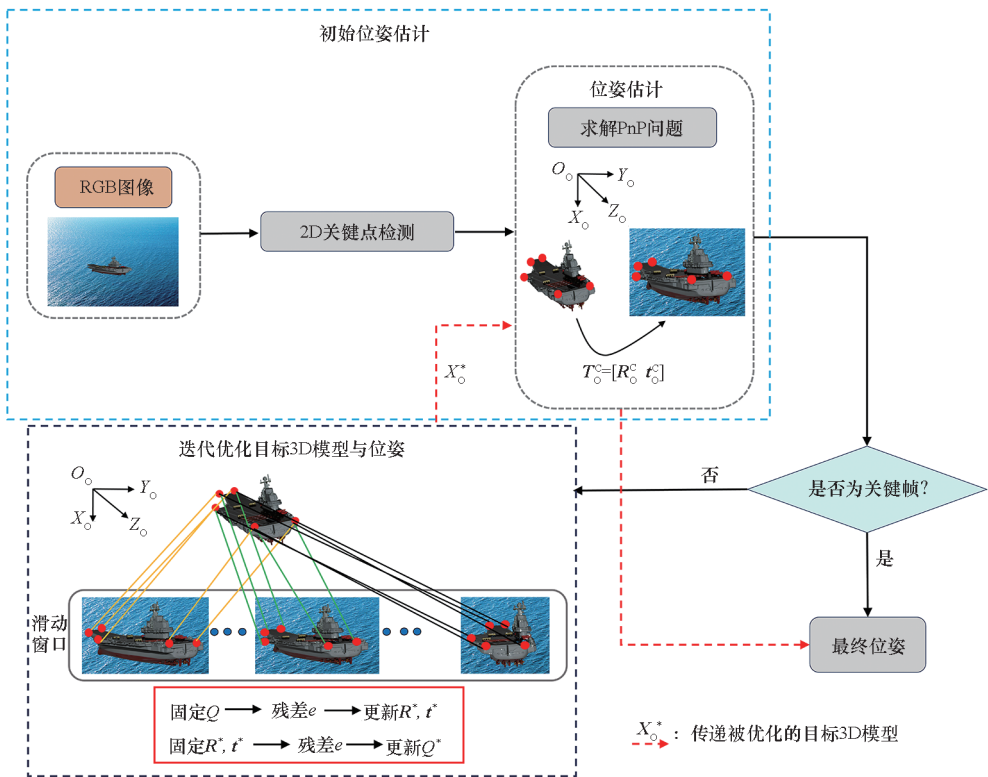


图 1 本文方法的整体架构

Fig. 1 Overall architecture of the proposed method

#### 1.2.1 基于稀疏关键点的 3D 模型表示

以飞行器着舰视觉引导中的位姿测量为例说明所提方法, 并以图 2 所示的舰船 3D 模型为目标平台。图 2 为对舰船简化物理模型进行精确三维扫描得到的 3D 模型。为了使所提出的迭代优化方法高效可靠, 提出使用一组如图 2 中红点所示的稀疏 3D 关键点  $\{\tilde{X}_j^0\}_{j=1,2,\dots,N} = \{[x_j^0 \ y_j^0 \ z_j^0 \ 1]^T\}_{j=1,2,\dots,N}$  表示目标的 3D 模型, 其中  $N$  为 13。与传统 SLAM、SFM 等领域使用的 2D 特征点不同, 采用语义级关键点来表示目标的 3D 模型。为减少所提方法中要优化的参数数目, 采用线性参数方程形式表示 3D 稀疏关键点集合。在

后续描述中, 所提到的目标 3D 模型即指建立的目标 3D 关键点集合, 所提方法不再对两者进行区分。

#### 1.2.2 初始位姿测量

在初始帧中利用一瞥即识目标检测模型的扩展版 (you only look once extended, YOLOX) 算法<sup>[39]</sup>检测目标区域, 为了提高效率, 在获得初始帧的目标区域后, 利用模板与逐像素学习器的集合 (sum of template and pixel-wise learners, STAPLE) 算法<sup>[40]</sup>从后续中输入图像中跟踪目标。基于 1.2.1 节提出的稀疏关键点集合的表示形式, 借鉴在人体姿态估计任务中表现优异的实时多人姿态估计 (real-time multi-person pose estimation,

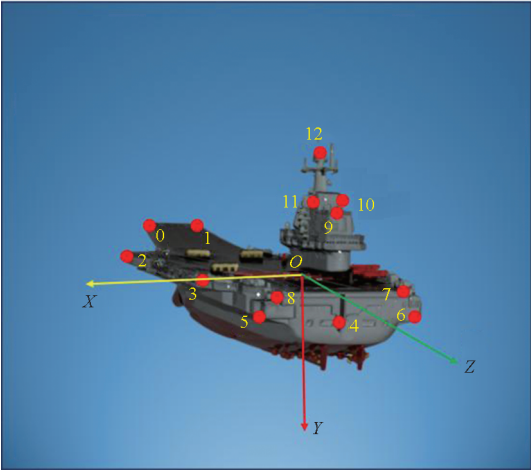


图 2 目标 3D 模型和关键点标注

Fig. 2 3D model and keypoint annotation of target

RTMPose)算法<sup>[41]</sup>,实现图像中目标关键点的高效准确检测。RTMPose 算法通过简单的坐标分类输出 2D 关键点的位置,简洁、高效且易于部署。最后,采用高效透视  $n$  点 (Efficient Perspective- $n$ -Point, EPnP) + 随机抽样一致性 (random sample consensus, RANSAC) 方法求解初始位姿。

1.2.3 3D 模型和位姿的迭代优化

基于上述得到的初始位姿测量结果,建立优化目标函数,通过最小化物方残差,迭代优化目标 3D 模型和位姿。

所提出的优化方法需要处理大量的数据和计算过程,参数过多可能导致计算时间过长,同时会使解空间过于复杂,使得优化算法难以收敛到全局最优解。因此,如图 3 所示,使用线性参数方程形式表示目标 3D 关键点的坐标,以减少优化目标函数中待优化参数的数目。图 3 中,  $e_k, e_M$  分别

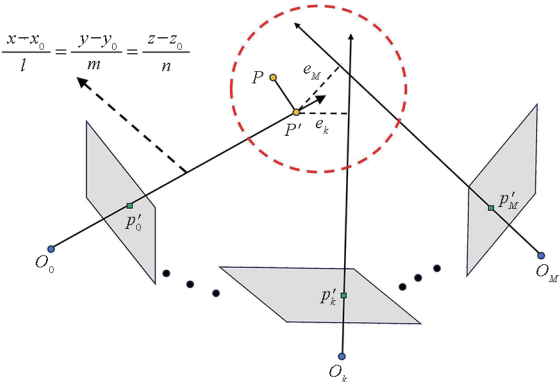


图 3 基于线性参数方程形式的 3D 关键点表示和空间同名视线交会

Fig. 3 3D keypoint representation based on linear parametric equation and the intersection of corresponding sight line in space

为点  $P'$  与  $\vec{O_k P'_k}$  和  $\vec{O_M P'_M}$  之间的空间距离。

如图 3 所示,  $P$  为一个表示目标 3D 模型的待优化的初始关键点,  $\vec{O_0 P'_0}, \vec{O_k P'_k}$  和  $\vec{O_M P'_M}$  分别为初始帧,第  $k$  帧和第  $M$  帧中关于点  $P$  的观测视线。在 BA 中,初始帧的位姿通常作为固定参考的真实位姿。因此,在不考虑首帧 2D 关键点观测误差的情况下,认为关键点  $P$  真实位于  $\vec{O_0 P'_0}$  上。进一步,如图 3 所示,将关键点  $P$  投影到  $\vec{O_0 P'_0}$  上,记为  $P'$ ,并以参数方程形式表示  $P'$ 。

$$\begin{cases} x = x_0 + lq \\ y = y_0 + mq \\ z = z_0 + nq \end{cases} \quad (4)$$

其中,  $[l \ m \ n]$  为  $\vec{O_0 P'_0}$  的方向矢量,  $[x_0 \ y_0 \ z_0]$  为  $\vec{O_0 P'_0}$  上任意一点,  $q$  为确定关键点  $P'$  位置的参数。通过式(4),单个 3D 关键点的表示由三维参数  $[x \ y \ z]$  简化为一维参数  $q$ ,从而减少了需要优化的参数,同时增加了对目标 3D 关键点优化方向的约束。

在理想情况下,不同视角的同一关键点的观测视线应相交于空间中同一点,该点即为 3D 关键点的真实位置。因此,如图 3 所示,将传统位姿与模型优化中基于最小化重投影残差的优化目标修改为基于最小化多视图内空间同名视线交会残差的优化目标。然而,直接计算多个视线的交会残差较为复杂,且后续优化过程中难以确定目标 3D 关键点的位置。因此,如图 3 和图 4 所示,计算  $P'$  与剩余同名观测视线间的物方残差,并利用该物方残差间接表示同名视线间的交会残差。

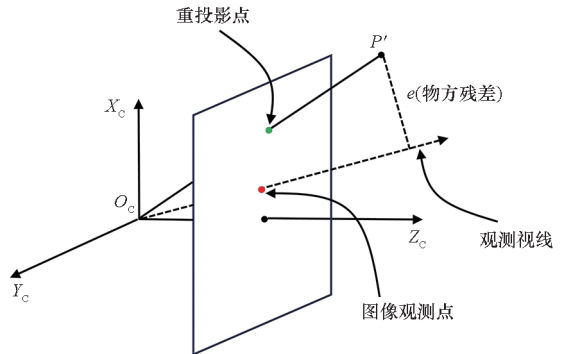


图 4 空间物方残差

Fig. 4 Object-space residual

图 4 中的物方残差  $e$  的计算如式(5)所示。

$$\begin{cases} e = (\mathbf{I} - \hat{\mathbf{V}})(\mathbf{R}\mathbf{P}' + \mathbf{t}) \\ \hat{\mathbf{V}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^T / (\hat{\mathbf{v}}^T \hat{\mathbf{v}}) \end{cases} \quad (5)$$

其中,  $\hat{\mathbf{v}}$  为 2D 关键点观测视线与归一化平面的交点。进一步,提出如下优化目标函数:



$$\begin{cases} Q^* = \arg \min_Q \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i r(q_j) + t_i]\|_2^2 \\ R^* = \arg \min_R \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i r(q_j) + t_i]\|_2^2 \\ t^* = \arg \min_t \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i r(q_j) + t_i]\|_2^2 \end{cases} \quad (6)$$

其中,  $Q = \{q_j\}_{j=1,2,\dots,N}$  为线性参数表示形式的 3D 关键点集合,  $r(q_j)$  为基于参数  $q_j$  求解第  $j$  个待优化 3D 关键点的三维坐标。通过式(4)和式(6), 优化目标中待优化参数的个数由传统优化方法中的  $6(M-1) + 3N$  减少为  $6(M-1) + N$ 。同时, 对于单个 3D 关键点, 仅需要一个观测方程, 因此观测方程的个数由  $MN$  减少为  $N$ 。

由于位姿与模型在优化过程中约束不足, 且关键点的误差范围通常较小, 采用阈值  $\sigma_1$  限制 3D 关键点的优化范围, 防止关键点优化增量过大: 若初始关键点与优化后的关键点间的欧氏距离大于  $\sigma_1$ , 则放弃关键点优化结果, 如式(7)所示。

$$q = \begin{cases} q^* & \|r(q^*) - r(q_{\text{init}})\|_2 \leq \sigma_1 \\ q_{\text{prev}}^* & \text{其他} \end{cases} \quad (7)$$

其中,  $q_{\text{prev}}^*$  为上一次优化中关键点的优化结果。在实际应用中,  $\sigma_1$  通常根据预估的 3D 模型误差大小来设置。

位姿与模型参数同步进行优化通常会致优化参数的约束较少和优化参数数量过多, 难以得到最优解。因此, 采用位姿与模型的迭代优化, 如式(8)所示。

$$\begin{cases} R^* = \arg \min_R \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i r(q_j) + t_i]\|_2^2 \\ t^* = \arg \min_t \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i r(q_j) + t_i]\|_2^2 \\ Q^* = \arg \min_Q \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^M \|(I - \hat{V}_i^j)[R_i^* r(q_j) + t_i^*]\|_2^2 \end{cases} \quad (8)$$

式(8)中, 在每一次优化时, 首先固定目标的 3D 关键点参数  $Q$  作为参考, 并对 6D 位姿  $(R, t)$  进行优化, 随后将被优化的位姿  $(R^*, t^*)$  再次代入优化目标函数中并保持固定不变, 对目标 3D 关键点进行优化得到  $Q^*$ 。

#### 1.2.4 滑动窗口结合关键帧筛选策略

单目视觉引导需要实时、在线和连续的位姿测量。然而, 连续图像的数量会随着时间的推移而增加, 此外, 在建立的优化目标函数中, 残差和位姿优化参数也在不断增加。因此, 采用滑动窗

口策略约束优化规模, 使所提方法只在固定长度  $L$  的时间窗口内优化图像的位姿。此外, 时间窗口外的位姿保持固定不变。

同时, 为了更有效利用多视图间几何约束信息, 提高处理效率, 采用关键帧的处理策略: 在序列图像中选取关键帧, 利用关键帧序列迭代优化目标平台 3D 模型和位姿, 针对非关键帧, 依据优化得到的目标平台 3D 模型, 通过求解 PnP 问题, 实现高精度位姿求解。参考文献[27]中的关键帧选取策略, 根据相机转动的角度大小进行关键帧筛选。

通过采用滑动窗口结合关键帧筛选策略, 有效降低了优化过程中的计算规模和冗余, 将式(8)修改为:

$$\begin{cases} R_{\text{key}}^* = \arg \min_R \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^L \|(I - \hat{V}_{\text{key}_i}^j)[R_{\text{key}_i} r(q_j) + t_{\text{key}_i}]\|_2^2 \\ t_{\text{key}}^* = \arg \min_t \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^L \|(I - \hat{V}_{\text{key}_i}^j)[R_{\text{key}_i} r(q_j) + t_{\text{key}_i}]\|_2^2 \\ Q^* = \arg \min_Q \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^L \|(I - \hat{V}_{\text{key}_i}^j)[R_i^* r(q_j) + t_{\text{key}_i}^*]\|_2^2 \end{cases} \quad (9)$$

其中, 下标  $\text{key}_i$  表示第  $i$  个关键帧。

为了确保位姿测量的准确性, 在迭代优化后, 使用被优化的目标 3D 模型再次求解 PnP 问题从而计算当前帧位姿。除首个关键帧外, 滑动窗口内其余经过优化的关键帧和新加入的关键帧被用于下一次优化。

## 2 实验结果与分析

为了验证所提出方法的性能, 首先对所提出方法的鲁棒性进行测试分析, 随后在仿真数据中对所提出方法进行验证, 并与当前部分单目位姿估计方法进行对比。

### 2.1 实验设置

#### 2.1.1 参数设置

在鲁棒性测试及仿真对比实验中, 将模型优化更新阈值设置为模型预估误差, 滑动窗口长度设置为 20 帧。本文方法主要采用 C++ 语言编程实现, 所有实验的计算机配置为 3.2 GHz AMD Core R7-5800H CPU、RTX 3060 GPU、16 GB RAM。

在典型参数设置下, 得益于关键帧筛选和滑动窗口处理策略, 本文方法在单帧上的平均处理耗时约为 32 ms, 其中目标跟踪耗时约为 4 ms、关键点检测耗时约为 23 ms、优化求解耗时约为 4.5 ms、PnP 问题求解耗时约为 0.15 ms, 因此满足实时应用需求。特别说明, 对于首帧输入图像,

采用目标检测算法定位图像中的目标位置,目标检测耗时长于目标跟踪耗时,约为 30 ms,首帧图像处理耗时约为 53 ms。

### 2.1.2 评价指标

采用旋转误差  $E_R$  和归一化平移误差  $E_T$  来评价 6D 位姿中旋转量和平移量的测量精度,并采用 3D 关键点误差均值  $E_L$  来评价目标 3D 模型精度,如式(10)所示。

$$\begin{cases} E_R = 2\arccos\{\min[\|\mathbf{R}_g^T \mathbf{R}_p\|, 1]\} \frac{180}{\pi} \\ E_T = \frac{\|\mathbf{t}_g - \mathbf{t}_p\|}{\|\mathbf{t}_g\|} \times 100\% \\ E_L = \sum_{j=1}^N \|\mathbf{L}_g - \mathbf{L}_p\|_2 / N = \sum_{j=1}^N L_2 / N \end{cases} \quad (10)$$

其中,  $\mathbf{R}_g$  和  $\mathbf{t}_g$  分别为旋转和平移分量的真值,  $\mathbf{R}_p$  和  $\mathbf{t}_p$  分别为旋转和平移分量的估计值,  $\mathbf{L}_g$  和  $\mathbf{L}_p$  分别为稀疏 3D 关键点的真值和估计值,  $L_2$  为单个关键点误差。

## 2.2 鲁棒性测试

采用基于多视图几何约束的迭代优化方法来处理缺乏精确 3D 模型的位姿估计问题。迭代优化方法的效果主要取决于 2D 关键点检测的准确性和目标 3D 模型的误差。因此,对所提方法在不同的 2D 关键点噪声和目标 3D 模型误差下的表现进行测试。

本节模拟飞行器搜索舰船目标、绕飞调整位姿等过程,并同步生成测试数据和对应的位姿真值。为便于控制 3D 模型误差程度,依据均匀分布  $U(a, b)$  为每个 3D 关键点在空间的三个维度

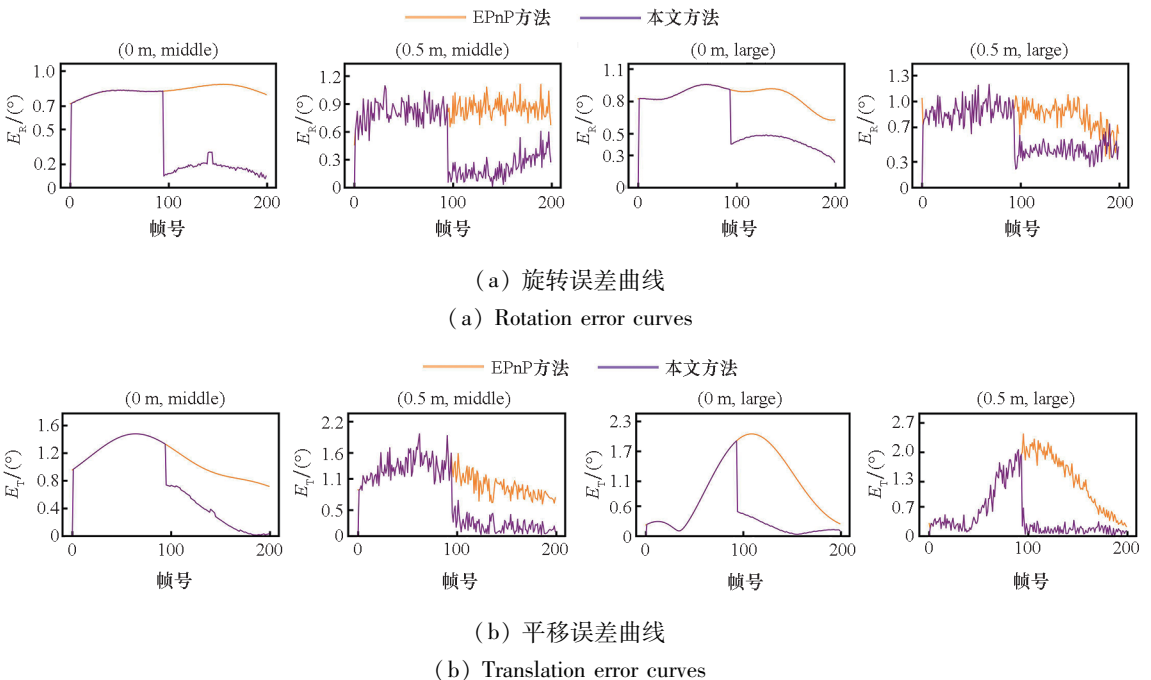
上分别添加  $a$  和  $b$  之间的误差。为了模拟 2D 关键点检测误差,向测试数据中的 2D 关键点加入图像处理领域常用的高斯误差  $N(0, \sigma_2)$ 。  $N(0, \sigma_2)$  表示加入 2D 关键点上的高斯观测误差的均值为 0, 标准差为  $\sigma_2$ 。随后,将不同水平误差的 2D 关键点和 3D 关键点样本成对组合,形成不同的仿真测试数据,如表 1 所示。

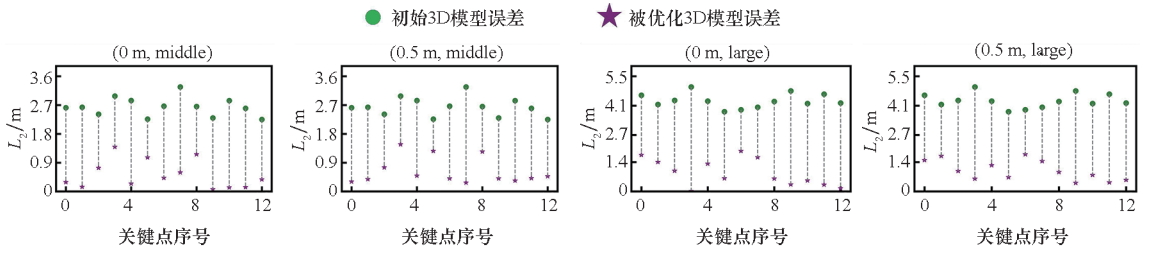
表 1 2D 关键点和 3D 关键点添加误差的参数设置  
Tab. 1 Parameter settings for errors added on 2D keypoints and 3D keypoints

像点误差/像素	模型误差		
	mini: $U(0, 1)$	middle: $U(1, 2)$	large: $U(2, 3)$
0	(0 m, mini)	(0 m, middle)	(0 m, large)
$N(0, 0.5)$	(0.5 m, mini)	(0.5 m, middle)	(0.5 m, large)
$N(0, 1)$	(1 m, mini)	(1 m, middle)	(1 m, large)
$N(0, 1.5)$	(1.5 m, mini)	(1.5 m, middle)	(1.5 m, large)
$N(0, 2)$	(2 m, mini)	(2 m, middle)	(2 m, large)

依据式(10)中的误差度量,图 5 分别给出了多组不同水平误差组合下完整图像序列的位姿旋转误差曲线、平移误差曲线以及目标 3D 模型误差。测试序列长度为 200 帧。

在图 5 中,初始帧的位姿被设置为固定的真实参考值,且关键帧的数量达到滑动窗口长度时才输出位姿优化结果。因此,图 5(a)和图 5(b)中黄色曲线的前半部分与紫色曲线的前半部分重合。由于目标平台的 3D 模型和 2D 检测关键点





(c) 目标 3D 模型误差  
(c) Target's 3D model errors

图 5 测试序列中的位姿及 3D 模型误差

Fig.5 Pose and 3D model errors in test sequences

存在误差,传统的基于直接求解 PnP 问题的方法无法在整个图像序列上得到准确的位姿旋转和平移测量结果。所提方法利用多视图的几何约束信息迭代优化目标 3D 模型和位姿参数,能够有效优化具有不同 2D - 3D 误差水平的全图像序列位姿,实现目标 3D 模型不准确情况下高精度的位姿测量。此外,图 5(c)表明,本研究不

仅在目标 3D 模型不准确的情况下实现了高精度的单目位姿测量,而且能够有效减小 3D 模型的误差。

接下来,表 2 给出不同误差组合中所有测试数据的实验结果的均值,其中最小误差以粗体突出显示。表 2 中的 EPnP 方法表示直接通过 2D - 3D 对应求解 PnP 问题得到位姿测量结果。

表 2 实验数据不同误差组合下的相对位姿与 3D 稀疏关键点误差均值

Tab.2 The average error of relative poses and 3D sparse keypoints under different error combinations on experimental data

2D - 3D 误差组合	$E_R / (^\circ)$		$E_T / \%$		$E_L / m$	
	EPnP 方法	本文方法	EPnP 方法	本文方法	初始误差	本文方法
(0 m, mini)	0.28	<b>0.24</b>	0.23	<b>0.21</b>	0.95	<b>0.38</b>
(0 m, middle)	0.77	<b>0.52</b>	0.79	<b>0.65</b>	2.66	<b>0.60</b>
(0.5 m, middle)	0.78	<b>0.59</b>	0.80	<b>0.70</b>	2.66	<b>0.90</b>
(1 m, middle)	0.81	<b>0.67</b>	0.82	<b>0.79</b>	2.66	<b>1.33</b>
(1.5 m, middle)	0.84	<b>0.75</b>	0.86	<b>0.84</b>	2.66	<b>1.60</b>
(0 m, large)	0.89	<b>0.58</b>	0.81	<b>0.55</b>	4.34	<b>0.67</b>
(0.5 m, large)	0.90	<b>0.62</b>	0.82	<b>0.62</b>	4.34	<b>1.03</b>
(1 m, large)	0.93	<b>0.68</b>	0.84	<b>0.65</b>	4.34	<b>1.28</b>
(1.5 m, large)	0.96	<b>0.77</b>	0.88	<b>0.80</b>	4.34	<b>1.64</b>
(2 m, large)	1.02	<b>0.87</b>	0.93	<b>0.87</b>	4.34	<b>2.06</b>

如表 2 所示,目标 3D 模型误差固定时,2D 检测关键点误差越小,所提方法的优化效果越显著。这主要是因为本文方法的实质是通过最小化空间同名视线交会残差来对位姿和 3D 模型进行迭代优化,而 2D 检测关键点误差会导致观测视线出现偏差,进而影响后续视线交会的计算。因此,本研究需要 2D 检测关键点误差小于 3D 关键点重投影误差,否则无法对 3D 模型进行有效优化并提升位姿测量的精度。

### 2.3 仿真图像序列中的实验结果与分析

本节选择仿真图像序列进行实验,仿真图像通过目标 3D 模型与计算机图形引擎结合生成。具体地,采用视景仿真工具 BlenderProc<sup>[42]</sup> 加载舰船 3D CAD 模型。如图 6 所示,在仿真环境中,通过设定相机轨迹并进一步渲染生成相应的图像数据。

不同视角下的仿真测试样本如图 7 所示。此外,在本节实验中,通过均匀分布  $U(2,3)$  为 3D 模型添加误差,并随机生成 1 000 组样本进行测试。

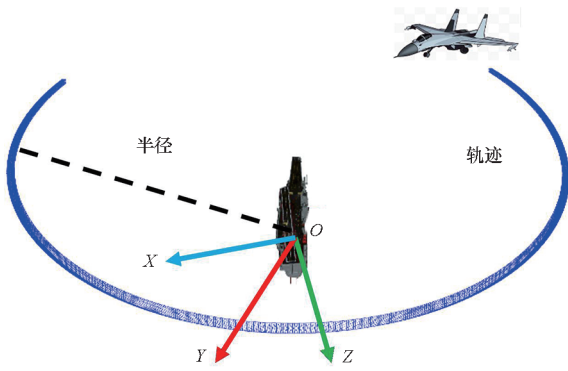


图 6 仿真相机轨迹可视化

Fig. 6 Synthetic camera trajectories visualization



(a) 左侧方视角 (b) 正前方视角 (c) 右侧方视角  
(a) Left side view (b) Front view (c) Right side view

图 7 不同视角下的仿真测试样本

Fig. 7 Synthetic test samples under different viewing angles

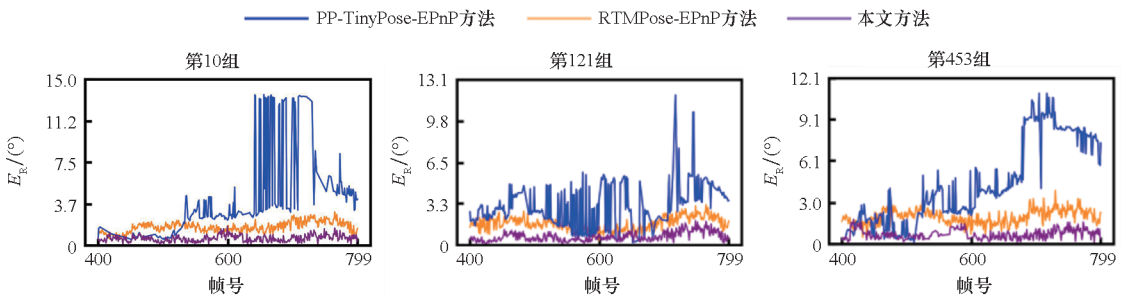
现有方法通常采用稀疏或密集的 2D-3D 对应关系,并直接求解 PnP 问题计算位姿。因此,分别采用当前先进的 RTMPose 和飞桨轻量级姿态估

计(paddle paddle tinypose, PP-TinyPose)方法<sup>[43]</sup>来检测 2D 关键点。将所提方法与 RTMPose-EPnP 和 PP-TinyPose-EPnP 两种方法进行对比。RTMPose-EPnP 和 PP-TinyPose-EPnP 方法均在关键点检测的基础上进一步求解 PnP 问题以获得位姿估计结果。

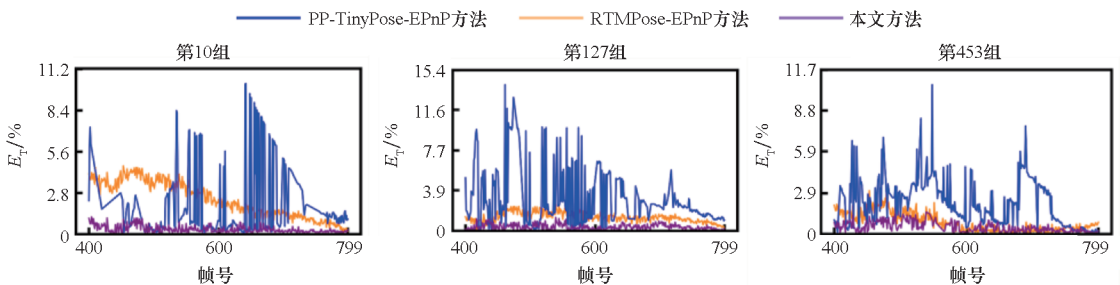
从仿真测试数据中选取三组(第 10 组、第 127 组、第 453 组)完整的位姿与 3D 模型估计结果进行展示,如图 8 所示,此外,图中仅展示了图像序列中的有效优化部分。

如图 8(a)和图 8(b)所示,由于 PP-TinyPose 方法的 2D 关键点误差较大,PP-TinyPose-EPnP 方法求解的位姿误差明显大于 RTMPose-EPnP 方法和所提方法。RTMPose 方法能够对 2D 关键点进行准确的检测,但由于目标 3D 模型误差,RTMPose-EPnP 方法求解的位姿误差仍然较大。本研究利用多视图几何约束来迭代优化目标的 3D 模型和位姿。如图 8(c)所示,本文方法通过迭代优化有效减小了目标 3D 模型误差,进而提高位姿测量精度。相较于两个对比方法,本文方法在整个图像序列中具有最高的位姿测量精度。

接下来,表 3 展示了所有样本的位姿与 3D 模型误差的均值统计结果,其中最小误差以粗体突出显示。



(a) 旋转误差曲线  
(a) Rotation error curves



(b) 平移误差曲线  
(b) Translation error curves



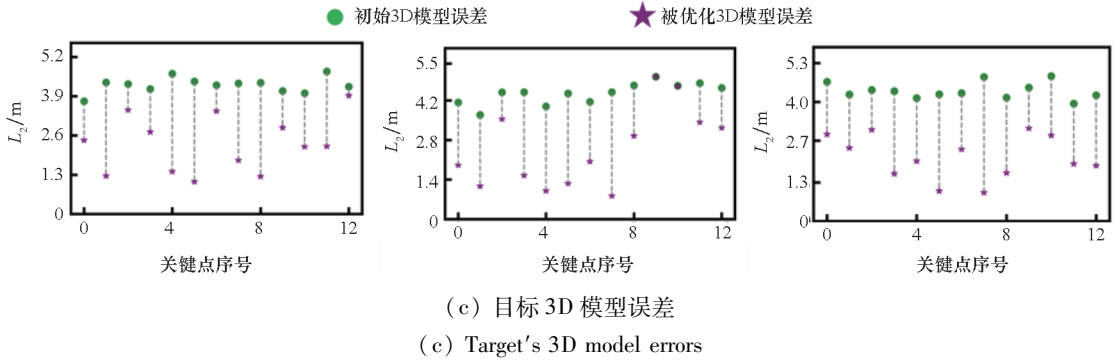


图 8 仿真图像序列中的位姿及 3D 模型误差

Fig. 8 Pose and 3D model errors in synthetic image sequences

如表 3 所示,本文方法通过迭代优化有效减小了目标 3D 模型的误差,相比于两个对比方法,在整个仿真实验数据中取得了最高的位姿测量精度。

表 3 仿真图像序列位姿与 3D 模型误差的均值统计结果  
Tab. 3 The average statistic results of pose and 3D model errors for synthetic image sequences

方法	$E_R/(^\circ)$	$E_T/\%$	$E_L/m$
PP-TinyPose-EPnP 方法	3.34	2.71	4.35
RTMPose-EPnP 方法	1.41	1.24	4.35
本文方法	<b>1.28</b>	<b>1.14</b>	<b>3.40</b>

### 3 结论

本文针对单目视觉引导中目标 3D 模型不准确条件下的目标平台间位姿测量问题,迭代优化目标平台 3D 模型及位姿,并提出一种新的单目视觉测量方法:采用稀疏 3D 关键点集合建模目标平台 3D 模型,并采用先进的关键点检测方法实现序列图像中关键点的鲁棒、高效检测。进一步,将以线性参数方程形式表达的目标 3D 关键点参数与位姿参数作为待优化参数,最小化物方残差来建立优化目标函数,通过求解该优化问题迭代优化目标 3D 模型与位姿,实现目标 3D 模型不准确条件下位姿的高精度求解。为了满足视觉引导中位姿测量的实时、在线应用需求,采用滑动窗口约束计算规模,并通过选取关键帧减少计算量和信息冗余,实现实时、在线的目标平台 3D 模型及位姿迭代优化求解。实验结果表明,本研究可以实现目标 3D 模型不准确条件下实时、在线的高精度位姿测量,并有效优化目标 3D 模型。

本文方法尽管取得了明显的位姿与模型迭代优化效果,但依然存在不适应较大 2D 关键点检测误差的问题,下一步工作中,将探索利用序列图

像多视图几何约束信息进一步减小 2D 检测关键点误差的方法。

### 致谢

国防科技大学空天科学学院应用力学系图像测量实验室博士生王秋富在仿真图像生成方面提供了指导和帮助,谨致谢意!

### 参考文献 (References)

- [1] ZHANG Z, WANG Q F, BI D M, et al. MC-LRF based pose measurement system for shipborne aircraft automatic landing[J]. Chinese Journal of Aeronautics, 2023, 36(8): 298–312.
- [2] 刘峰, 郭英华, 王霖, 等. 二维主动位姿引导的单目空间姿态测量方法[J]. 红外与激光工程, 2024, 53(2): 20211026.
- [3] LIU F, GUO Y H, WANG L, et al. Monocular spatial attitude measurement method guided by two dimensional active pose[J]. Infrared and Laser Engineering, 2024, 53(2): 20211026. (in Chinese)
- [4] 秦熠琳. 基于单目视觉的卫星编队实时相对位姿估计研究[D]. 上海: 中国科学院微小卫星创新研究院, 2023.
- [5] QIN Y L. Research on real-time relative pose estimation of satellite formation based on monocular vision[D]. Shanghai: Innovation Academy for Microsatellite, Chinese Academy of Sciences, 2023. (in Chinese)
- [6] SHARP C S, SHAKERNIA O, SASTRY S S. A vision system for landing an unmanned aerial vehicle[C]//Proceedings of 2001 ICRA. IEEE International Conference on Robotics and Automation, 2001: 1720–1727.
- [7] NIU G C, YANG Q K, GAO Y F, et al. Vision-based autonomous landing for unmanned aerial and ground vehicles cooperative systems[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6234–6241.
- [8] 罗威, 曾昱琳, 樊卫华, 等. 基于二维标志物的线夹检测及位姿测量[J]. 南京理工大学学报, 2023, 47(1): 41–47.
- [9] LUO W, ZENG Y L, FAN W H, et al. Cable clamp detection and pose measurement based on two-dimensional markers[J]. Journal of Nanjing University of Science and Technology, 2023, 47(1): 41–47. (in Chinese)
- [10] 陈汉瑜, 武俊峰, 康国华, 等. 基于激光灯塔的合作航天器位姿测量方法及其精度分析[J]. 仪器仪表学报, 2023, 44(1): 1–10.

- 2023, 44(12): 101–110.
- CHEN H Y, WU J F, KANG G H, et al. Pose measurement for cooperative spacecraft based on Lighthouse tracking system and the precision analysis[J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 101–110. (in Chinese)
- [8] SHARMA S, VENTURA J, D'AMICO S. Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous[J]. Journal of Spacecraft and Rockets, 2018, 55(6): 1414–1429.
- [9] ZHANG X, JIANG Z G, ZHANG H P, et al. Vision-based pose estimation for textureless space objects by contour points matching[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 54(5): 2342–2355.
- [10] 王世昌, 华宝成, 周依尔, 等. 基于 TEASER 算法的空间非合作目标位姿估计[J]. 空间控制技术与应用, 2024, 50(1): 25–34.
- WANG S C, HUA B C, ZHOU Y E, et al. Pose estimation of space non-cooperative target based on TEASER algorithm[J]. Aerospace Control and Application, 2024, 50(1): 25–34. (in Chinese)
- [11] HAN H, KIM H, BANG H. Monocular pose estimation of an uncooperative spacecraft using convexity defect features[J]. Sensors, 2022, 22(21): 8541.
- [12] ZHUANG B B, CHANDRAKER M. Fusing the old with the new: learning relative camera pose with geometry-guided uncertainty [C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 32–42.
- [13] LI Z G, WANG G, JI X Y. CDPN: coordinates-based disentangled pose network for real-time RGB-based 6-DoF object pose estimation [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 7677–7686.
- [14] KENDALL A, GRIMES M, CIPOLLA R. PoseNet: a convolutional network for real-time 6-DOF camera relocalization [C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015: 2938–2946.
- [15] WANG G, MANHARDT F, TOMBARI F, et al. GDR-net: geometry-guided direct regression network for monocular 6D object pose estimation[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 16606–16616.
- [16] PENG S D, ZHOU X W, LIU Y, et al. PVNet: pixel-wise voting network for 6DoF object pose estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3212–3223.
- [17] ZAKHAROV S, SHUGUROV I, ILIC S. DPOD: 6D pose object detector and refiner [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2020: 1941–1950.
- [18] SHUGUROV I, PAVLOV I, ZAKHAROV S, et al. Multi-view object pose refinement with differentiable renderer [J]. IEEE Robotics and Automation Letters, 2021, 6(2): 2579–2586.
- [19] WANG H, SRIDHAR S, HUANG J W, et al. Normalized object coordinate space for category-level 6D object pose and size estimation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 2637–2646.
- [20] LIU J H, CHEN Y K, YE X Q, et al. Ist-net: prior-free category-level pose estimation with implicit space transformation [C]//Proceedings of the International Conference on Computer Vision (ICCV), 2023: 13978–13988.
- [21] WAN B Y, SHI Y F, XU K. SOCS: semantically-aware object coordinate space for category-level 6D object pose estimation under large shape variations [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 14019–14028.
- [22] CHEN K, JAMES S, SUI C Y, et al. StereoPose: category-level 6D transparent object pose estimation from stereo images via back-view NOCS [C]//Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023: 2855–2861.
- [23] PÉREZ-VILLAR J I B, GARCÍA-MARTÍN Á, BESCÓS J. Spacecraft pose estimation based on unsupervised domain adaptation and on a 3D-guided loss combination [C]//Proceedings of Computer Vision-ECCV 2022 Workshops, 2023: 37–52.
- [24] WANG Z, ZHANG Z, SUN X L, et al. Revisiting monocular satellite pose estimation with transformer [J]. IEEE Transactions on Aerospace and Electronic Systems, 2022, 58(5): 4279–4294.
- [25] WANG Z, CHEN M L, GUO Y L, et al. Bridging the domain gap in satellite pose estimation: a self-training approach based on geometrical constraints [J]. IEEE Transactions on Aerospace and Electronic Systems, 2024, 60(3): 2500–2514.
- [26] WANG Q F, ZHOU J X, LI Z, et al. Robust and accurate monocular pose tracking for large pose shift [J]. IEEE Transactions on Industrial Electronics, 2023, 70(8): 8163–8173.
- [27] CHEN B, CAO J W, PARRA A, et al. Satellite pose estimation with deep landmark regression and nonlinear pose refinement [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2020: 2816–2824.
- [28] WANG Q F, ZHOU J X, LI Z, et al. Robust monocular object pose tracking for large pose shift using 2D tracking [J]. Visual Intelligence, 2023, 1(1): 22.
- [29] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization [J]. The International Journal of Robotics Research, 2015, 34(3): 314–334.
- [30] SIBLEY G, MATTHIES L, SUKHATME G. A sliding window filter for incremental SLAM [M]//KRAGIC D, KYRKI V. Unifying perspectives in computational and robot vision. Boston: Springer US, 2008: 103–112.
- [31] 黄丹丹, 郝文豪, 杨阳. 基于误匹配剔除和地面约束的视觉SLAM算法 [J]. 中国惯性技术学报, 2024, 32(2): 163–169, 179.
- HUANG D D, HAO W H, YANG Y. Visual SLAM algorithm based on mismatch rejection and ground constraints [J]. Journal of Chinese Inertial Technology, 2024, 32(2): 163–169, 179. (in Chinese)
- [32] 潘小鹏, 刘浩敏, 方铭, 等. 基于语义概率预测的动态场景单目视觉SLAM [J]. 中国图象图形学报, 2023, 28(7): 2151–2166.
- PAN X K, LIU H M, FANG M, et al. Dynamic 3D scenario-

- oriented monocular SLAM based on semantic probability prediction [J]. *Journal of Image and Graphics*, 2023, 28(7): 2151–2166. (in Chinese)
- [33] MOULON P, MONASSE P, MARLET R. Global fusion of relative motions for robust, accurate and scalable structure from motion[C]//*Proceedings of the 2013 IEEE International Conference on Computer Vision*, 2013: 3248–3255.
- [34] MATSUMOTO T, HANARI T, KAWABATA K, et al. Automatic system for sequential reconstruction from image sequences acquired from a camera by SfM-MVS [C]//*Proceedings of 2023 the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2023: 1–7.
- [35] 宋书学, 孙统领, 刘文锴, 等. 无人机序列影像快速三维重建方法研究[J]. *华北水利水电大学学报(自然科学版)*, 2024, 45(1): 31–36.  
SONG S X, SUN T L, LIU W K, et al. Research on fast 3D reconstruction method for UAV sequence images[J]. *Journal of North China University of Water Resources and Electric Power (Natural Science Edition)*, 2024, 45(1): 31–36. (in Chinese)
- [36] 姜三, 刘凯, 李清泉, 等. 融合深度特征的无人机影像 SfM 重建[J]. *测绘学报*, 2024, 53(2): 321–331.  
JIANG S, LIU K, LI Q Q, et al. Learned local features for SfM reconstruction of UAV images[J]. *Acta Geodaetica et Cartographica Sinica*, 2024, 53(2): 321–331. (in Chinese)
- [37] QIU K J, QIN T, GAO W L, et al. Tracking 3-D motion of dynamic objects using monocular visual-inertial sensing[J]. *IEEE Transactions on Robotics*, 2019, 35(4): 799–816.
- [38] CHEN J J, WANG R, WANG R M. Vision positioning method for autonomous precise landing of UAV based on square landing mark [J]. *Journal of Physics: Conference Series*, 2020, 1651(1): 012182.
- [39] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021 [EB/OL]. (2021–08–06) [2024–01–12]. <https://arxiv.org/abs/2107.08430>.
- [40] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 1401–1409.
- [41] JIANG T, LU P, ZHANG L, et al. RTMPose: real-time multi-person pose estimation based on MMPose [EB/OL]. (2023–07–03) [2024–01–12]. <https://arxiv.org/abs/2303.07399>.
- [42] DENNINGER M, SUNDERMEYER M, WINKELBAUER D, et al. BlenderProc: reducing the reality gap with photorealistic rendering [C]//*Proceedings of the 16th Robotics: Science and Systems*, 2020.
- [43] SHIGURE N. PaddlePaddle/PaddleDetection [EB/OL]. [2024–01–12]. <https://gitee.com/paddlepaddle/PaddleDetection>.