

## 基于事件相机的双目视觉 SLAM

刘 勇<sup>1</sup>, 刘智聪<sup>1</sup>, 邵津津<sup>2</sup>, 李 程<sup>1</sup>, 蓝德劭<sup>1</sup>

(1. 株洲中车时代电气股份有限公司, 湖南 株洲 412000; 2. 国家能源集团陕西神延煤炭有限责任公司, 陕西 榆林 719000)

**摘要:** 现有的基于事件相机的同步定位与地图构建 (SLAM) 方法大多依赖事件流长时间积累, 无法有效利用事件流之间的关联, 导致建图与定位精度欠佳等问题。为此, 本文探索了一种基于事件相机的双目视觉 SLAM 系统, 通过匹配左右时间曲面上的最近时间戳事件, 计算特征深度信息, 通过 Luacas-Kanede 算法匹配深度信息与时间曲面信息来计算位姿的微小增量, 以实现跟踪; 然后基于 Students't 概率模型, 使用 IRLS (迭代重加权最小二乘) 算法加速对深度信息的融合, 得到半稠密地图。使用开源数据集模拟极端环境对算法进行性能测试, 并对建图与跟踪的结果进行误差分析。试验结果表明, 相比于现有的最优的方法, 本文方法拥有更好的建图效果, 在不同数据集上的位姿误差降低 48.3%, 具备更高的定位精度; 同时在光线不足、高动态场景下均能鲁棒地工作。

**关键词:** 事件流; 同步定位与地图构建; 时间曲面; 高动态场景

## Binocular Vision SLAM Based on Event Camera

LIU Yong<sup>1</sup>, LIU Zhicong<sup>1</sup>, SHAO Jinjin<sup>2</sup>, LI Cheng<sup>1</sup>, LAN Deshao<sup>1</sup>

(1. Zhuzhou CRRC Times Electric Co., Ltd, Zhuzhou 412000, China;

2. Shaanxi Shenyang Coal Industry Limited Liability Company, CHN Energy, Yulin 719000, China)

**Abstract:** Most of the existing simultaneous localization and mapping (SLAM) methods based on event camera rely on the accumulation of event streams for a long time, and cannot effectively use the correlation between event streams, resulting in poor mapping and positioning accuracy. For this problem, a binocular vision SLAM system based on an event camera is explored in this paper. By matching the nearest timestamp events on the left and right time surfaces, the feature depth information is calculated, and the depth information and time surface information are matched by Luacas-Kanade algorithm to calculate the small increment of pose to achieve tracking. Then, the IRLS (iterative reweighted least squares) algorithm is used to accelerate the fusion of depth information based on the Students't probability model, to obtain a semi-dense map. The open source dataset is used to simulate the extreme environment to test the performance of the algorithm, and the error analysis on the mapping and tracking results is carried out. The experimental results show that compared with the state-of-the-art methods, the proposed method has better mapping effect, and the pose error on different datasets is reduced by 48.3%, demonstrating higher positioning accuracy. Meanwhile, it can work robustly in low light and high dynamic scenes.

**Keywords:** event stream; simultaneous localization and mapping; time surface; high dynamic scene

同步定位与地图构建 (SLAM) 作为机器人的核心技术, 使机器人在未知环境下具备获取场景信息、构建地图和估计自身运动的功能<sup>[1]</sup>, 进而在无人驾驶、室内导航和 underwater 作业等场景中完成自主建图和精确定位, 保证工作过程的安全和高效。

基于视觉传感器的 SLAM 系统由于设备结构轻便、便于安装等优势而得到广泛应用, 但是传统的视觉 SLAM 方法在一些特殊环境下的应用十分受限<sup>[2]</sup>。例如在环境光线高动态变化场景下, 传统视觉 SLAM 方法因图像模糊、曝光失衡, 难以实现特征的稳定提取与匹配, 严重影响系统定位与建图的精度与鲁棒性。

事件相机基于光照强度变化触发的异步成像原理, 所采集图像不受环境光照波动的影响, 有助于解决上述非理想环境下的 SLAM 系统性能下降问题。

文 [3] 基于概率滤波器实现对相机的位姿跟踪。文 [4] 利用事件帧作为目标函数对事件光流进行估计, 最终实现建图与定位。文 [5] 将连续的事件流进行叠加, 变成稠密事件流再进行非线性优化与追踪。文 [6] 提出了基于事件相机的实时 3 维重建与 6 自由度位姿跟踪的方法。文 [7] 提出了一个用于轨迹估计的连续时间模型, 该方法基于此连续时间模型, 直接使用事件流的时间戳与坐标进行轨迹

优化。但是该方法需要 GPU (图形处理器) 进行加速, 因此难以在算力一般的平台实现。文 [8] 在事件相机的基础上, 引入了传统相机、IMU (惯性测量单元), 首次将三者进行紧耦合, 构建 Ultimate SLAM 系统, 并实现高速和高动态等复杂场景下的定位与建图, 建图精度也得到了提高。文 [9] 提出了一种基于状态约束滤波算法融合图像、事件流、IMU 的方法, 对事件、视觉、IMU 实现了紧耦合, 并计算场景的深度。然而, 现有基于事件相机的 SLAM 方法依赖事件流长时间积累, 无法有效利用事件流之间的关联信息, 导致建图与定位精度欠佳等问题, 难以取得令人满意的效果。

如何利用事件流异步特征, 设计出一种鲁棒的高精度 SLAM 算法, 对部署基于双目事件相机的 SLAM 法十分重要。本文利用事件相机的异步成像特性, 设计了一套基于事件相机的定位与建图框架。

### 1 基于事件相机的 SLAM 算法关键技术 (Key technologies for SLAM based on event camera)

如图 1 所示, 一个完整的基于事件相机的 SLAM 算法框架包含事件预处理模块、跟踪模块以及建图模块。原始事件流为事件相机获取的二进制异步信息, 其中包含时间戳信息与事件像素点, 事件流经过事件预处理模块处理后, 形成可供后续跟踪与建图模块处理的结构化图形信息。然后将该信息传递到跟踪模块以计算相机位姿, 根据计算得到的当前位姿结合原始事件流建立 3D 地图。

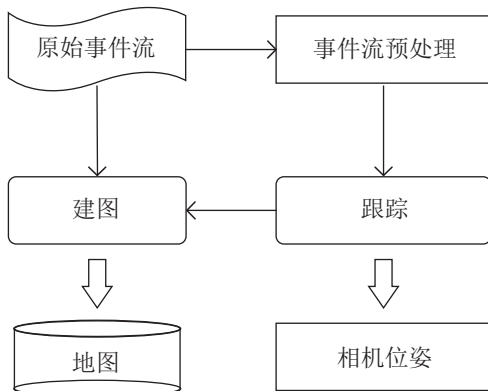


图 1 基于事件相机的 SLAM 算法框架

Fig.1 Framework of the SLAM algorithm based on event camera

单一事件所包含的信息十分稀疏, 难以反映其他事件的关系, 需要对事件进行积累以满足一定的信息密度, 供后续处理。主流的基于事件相机

的 SLAM 算法的预处理过程如下: 对同一坐标的像素产生的离散事件, 在“固定时间窗口”或“固定事件数”约束下, 通过统计事件数量、计算极性加权和、记录最新时间戳等方式, 将零散的事件转化为单个量化值, 为后续构建“事件伪帧”、支撑 SLAM 的特征提取与运动估计提供结构化输入。

假定事件流包含像素坐标  $(u, v)$  与时间戳  $t$ , 则单个事件可使用脉冲函数表示, 即:

$$e_0(u_0, v_0, t_0) = \delta(u - u_0, v - v_0, t - t_0) \quad (1)$$

事件流由大量离散的事件组成, 事件个数为  $N$ 。一般的方法是采用累加事件的脉冲函数方式表示事件流  $E$ :

$$E = \sum_{i=1}^N \delta(u - u_i, v - v_i, t - t_i) \quad (2)$$

图 2 为时空邻域示意图,  $L$  为空间邻域尺寸,  $\Delta t$  为时间邻域大小, 表示为  $\Omega_{\Delta t}^L$ , 即:

$$\begin{cases} u_0 - \frac{L}{2} \leq u_i \leq u_0 + \frac{L}{2} \\ v_0 - \frac{L}{2} \leq v_i \leq v_0 + \frac{L}{2}, & e_i(u_i, v_i, t_i) \in \Omega_{\Delta t}^L \\ t_i \leq t_0 - \Delta t \end{cases} \quad (3)$$

其中,  $e_0(u_0, v_0, t_0)$  为新到达事件, 该时间邻域的中心时间为新到达事件  $e_0$  的时间戳  $t_0$ 。

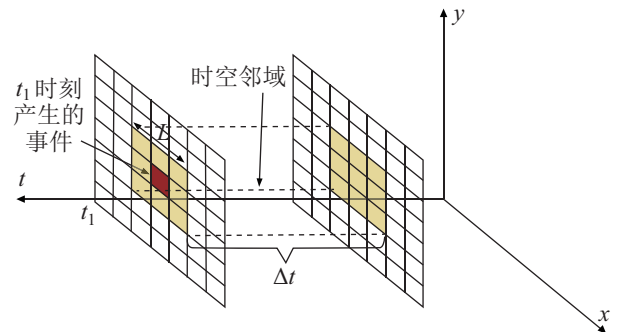


图 2 时空邻域示意图

Fig.2 Schematic diagram of temporal and spatial neighbourhood

单一事件所包含的信息难以反映与其他事件的关系, 因此需要使用上述时间邻域的概念提取事件  $e_0(u_0, v_0, t_0)$  对应的时空邻域  $\Omega_{\Delta t}^L$  内的事件流  $E_{sp}$ , 并对其降噪。

事件流为二进制信息, 且含有大量传感器自身产生的噪声, 无法直接用于计算位姿, 需要先进行预处理降噪, 生成能够处理的图形信息。

事件相机产生的噪声分为两类: 固定位置的热像素噪声和时空相关性低的背景活动噪声。本文使用主流的时空相关滤波器 [10] 对事件流进行过滤处

理。降噪后的事件流可以利用时间邻域积累信息密度，用于位姿计算与建图。

现有方法依赖长时间积累事件流，无法有效利用事件流间的关联，导致真实事件在同一像素多次触发，被误判为噪声，进而导致系统定位与建图精度降低。

本文使用时空感知事件 (STAE) 框架将事件流创建为时间曲面，以此进行特征跟踪与关联，并使用基于 Student's  $t$  概率模型的 IRLS (迭代重加权

最小二乘) 方法进行时间曲面的逆深度融合，能够最大程度避免真实事件被错误滤除，保证了建图的稠密性与定位的准确性，以及建图的实时性。

## 2 鲁棒的基于事件相机的双目视觉 SLAM (Robust binocular vision SLAM based on event camera)

本文对现有基于事件相机的 SLAM 算法作出改进，提出如图 3 所示的方法。

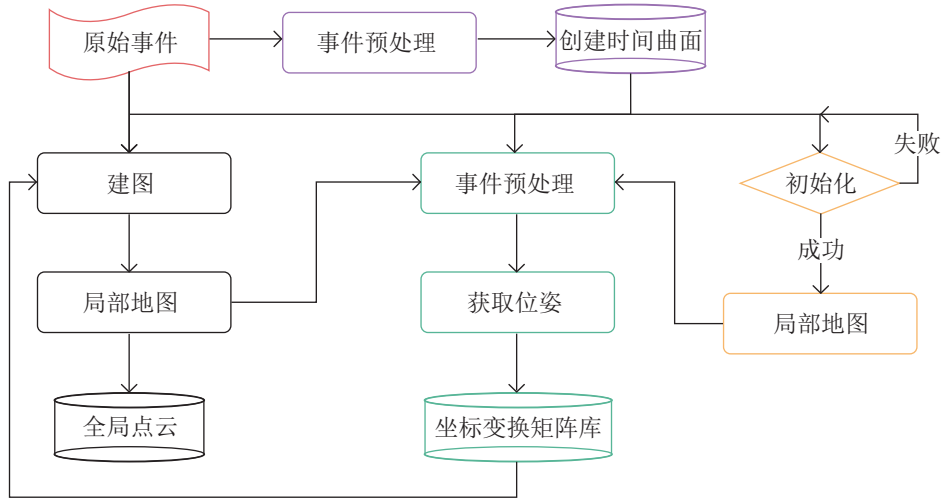


图3 快速场景下基于事件相机的双目视觉 SLAM 框架

Fig.3 The binocular vision SLAM framework based on event camera in fast scene

### 2.1 生成时间曲面与基于 Student's $t$ 模型的半稠密重建

采用一种便于处理的事件流表示方法，即时间曲面 (TS)。首先，针对每个像素所存储的事件，采用  $e = (u, v, t, p)$  的形式进行描述：其中  $\mathbf{x} = (u, v)$  表示像素坐标， $t$  为事件产生的时间戳，事件极性  $p \in \{+1, -1\}$  分别对应像素亮度上升、下降两种变化类型。

用下述公式定义时间曲面：

$$\mathcal{T}(\mathbf{x}, t) = \exp\left(-\frac{t - t_{\text{last}}(\mathbf{x})}{\eta}\right) \quad (4)$$

其中， $t$  为当前事件相机帧对应的时间戳， $t_{\text{last}}$  为上一事件相机帧对应的时间戳。 $\eta$  为事件相机衰减率参数，通常取 30 ms。

将事件流处理为逐帧的时间曲面以供后续在像素坐标上处理事件信息。

使用逆深度估计法还原事件像素点的 3D 位置，使用左事件相机来参数化逆深度。如图 4 所示，左相机观测到的事件可表示为  $e_{t-\varepsilon} \equiv (\mathbf{x}, t - \varepsilon, p)$ ： $t - \varepsilon$  为事件的时间戳 ( $\varepsilon \in [0, \delta t]$ )， $\delta t$  为时间窗口范围，

体现事件发生时刻的微小不确定性)；像素点的逆深度  $\rho^* \approx 1/Z^*$  ( $Z^*$  为各像素点在左相机坐标系下的真实深度，逆深度与深度成倒数关系，更利于处理远距离场景的不确定性)，相机的运动轨迹表示为  $\mathbf{T}_{t-\delta t:t}$ 。逆深度信息通过优化目标函数来估计：

$$\rho^* = \arg \min \mathbf{C}(\mathbf{x}, \rho, \mathbf{\Gamma}_{\text{left}}(\cdot, t), \mathbf{\Gamma}_{\text{right}}(\cdot, t), \mathbf{T}_{t-\delta t:t}) \quad (5)$$

$$\mathbf{C} = \sum_{\mathbf{x}_{1,i} \in W_1, \mathbf{x}_{2,i} \in W_2} \mathbf{r}_i^2(\rho_i) \quad (6)$$

其中， $\mathbf{\Gamma}_{\text{left}}(\cdot, t)$ 、 $\mathbf{\Gamma}_{\text{right}}(\cdot, t)$  分别为左、右事件相机的时间曲面， $\mathbf{T}_{t-\delta t:t} \in \text{SE}(3)$  表示左相机在时刻  $t - \delta t$  的位姿 ( $\text{SE}(3)$  为特殊欧氏群，描述刚体的旋转与平移)。代价函数  $\mathbf{C}$  定义为残差的平方和，用于聚合邻域内多组像素的时空偏差，其中  $W_1$ 、 $W_2$  是左、右相机图像平面上的像素邻域，第  $i$  个像素对应残差  $\mathbf{r}_i(\rho_i)$ ：

$$\mathbf{r}_i(\rho_i) \approx \mathbf{\Gamma}_{\text{left}}(\mathbf{x}_{1,i}, t) - \mathbf{\Gamma}_{\text{right}}(\mathbf{x}_{2,i}, t) \quad (7)$$

残差表示邻域内的左右相机产生的时间曲面相应的像素  $\mathbf{x}_{1,i}$  与  $\mathbf{x}_{2,i}$  之间的时间差。事件相机内参、外参已知，左事件相机在  $[t - \delta t, t]$  内任何时间戳的位姿

可以使用坐标变换矩阵库  $SE(3)$  的内置矩阵表示。像素点  $\mathbf{x}_1$  与  $\mathbf{x}_2$  由下述公式给出:

$$\mathbf{x}_1 = \boldsymbol{\pi}(\mathbf{T}_{c_{i-\varepsilon}} \cdot \boldsymbol{\pi}^{-1}(\mathbf{x}, \boldsymbol{\rho})) \quad (8)$$

$$\mathbf{x}_2 = \boldsymbol{\pi}(\mathbf{T}_{\text{left}} \cdot \mathbf{T}_{c_{i-\varepsilon}} \cdot \boldsymbol{\pi}^{-1}(\mathbf{x}, \boldsymbol{\rho})) \quad (9)$$

上述每个事件都使用其时间戳对应的相机位姿进行坐标变换。函数  $\boldsymbol{\pi}: \mathbf{R}^3 \rightarrow \mathbf{R}^2$  将一个 3 维点投影到相机的像素平面上, 而其反函数  $\boldsymbol{\pi}^{-1}: \mathbf{R}^2 \rightarrow \mathbf{R}^3$  将一个像素反投影到 3 维空间中, 给出了逆深度  $\boldsymbol{\rho}$ 。 $\mathbf{T}_{\text{left}}$  表示从左到右事件摄像机的位姿的变换, 该变换是恒定的。所有事件坐标都是未失真的, 并且使用事件相机进行了校正。

同一真实事件在左右事件相机上同时激发事件像素点, 理论上它们的时间残差最小, 借助这个特点估计同一个真实事件在左右两个相机下的像素坐标, 然后去估计其逆深度。逆深度估计算法框架如算法 1 所示。目标函数使用高斯-牛顿法求解非线性最小二乘问题, 迭代得到最优解, 公式如下:

$$\frac{\partial \mathcal{C}}{\partial \boldsymbol{\rho}} = 2\mathbf{J}^T \mathbf{r} = 0 \quad (10)$$

其中,  $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N^2})^T$ ,  $N^2$  表示使用 STAE 框架时每次搜索的单元像素框的大小,  $\mathbf{J} = \partial \mathbf{r} / \partial \boldsymbol{\rho}$ 。将  $\mathbf{r}$  的线性化表达式代入到泰勒公式中得到:  $\mathbf{r}(\boldsymbol{\rho} + \Delta \boldsymbol{\rho}) \approx \mathbf{r}(\boldsymbol{\rho}) + \mathbf{J}(\boldsymbol{\rho}) \Delta \boldsymbol{\rho}$ , 即  $\mathbf{J}^T \cdot \mathbf{J} \Delta \boldsymbol{\rho} = -\mathbf{J}^T \mathbf{r}$ , 且又因为  $\mathbf{J}^T \cdot \mathbf{J} = \|\mathbf{J}\|^2$ , 因此公式最终简化为

$$\boldsymbol{\rho} \leftarrow \boldsymbol{\rho} + \Delta \boldsymbol{\rho}, \quad \Delta \boldsymbol{\rho} = -(\mathbf{J}^T \mathbf{r}) / \|\mathbf{J}\|^2 \quad (11)$$

#### 算法 1 逆深度估计算法框架

输入: 事件  $e_{i-\varepsilon}$ , 左右相机时间曲面  $\Gamma_{\text{left}}(\cdot, t), \Gamma_{\text{right}}(\cdot, t)$ , 双目系统位姿变化矩阵  $\mathbf{T}_{c_{i-\varepsilon}}$

初始化: 在相机左右时间曲面上匹配初始化逆深度

- 1: **while** 匹配的值没有收敛使用式 (7) 计算残差 **do**
- 2:   计算雅可比行列式  $\mathbf{J}(\boldsymbol{\rho})$
- 3:   更新计算得到的深度信息
- 4: **end while**
- 5: **return** 计算得到逆深度  $\boldsymbol{\rho}^*$

计算好的逆深度信息需要进行深度融合, 以提高地图构建质量。通过不同场景下的数据集, 利用事件、真实深度和位姿信息, 计算了逆深度的残差分布  $\mathbf{r}_i(\boldsymbol{\rho})$ , 实验表明, Student's t 概率模型可以很好地拟合直方分布图。因此, 得到  $\mathbf{r}$  的概率模型, 表示为  $\mathbf{r} \sim St(\boldsymbol{\mu}_r, s_r^2, \nu_r)$ , 其中  $\boldsymbol{\mu}_r$ 、 $s_r$ 、 $\nu_r$  分别表示  $\mathbf{r}$  的均值、标准差和自由度。

由于 Student's t 概率模型在仿射变换下具有封闭性, 因此随机坐标  $\mathbf{x} \sim St(\boldsymbol{\mu}, S, \nu)$  的仿射变换同样遵循 Student's t 概率模型。将此理论应用于式 (11), 且  $\mathbf{A} \equiv -\sum_i \mathbf{J}_i / \|\mathbf{J}\|^2$ ,  $\mathbf{b} \equiv 0$ , 可以得到:

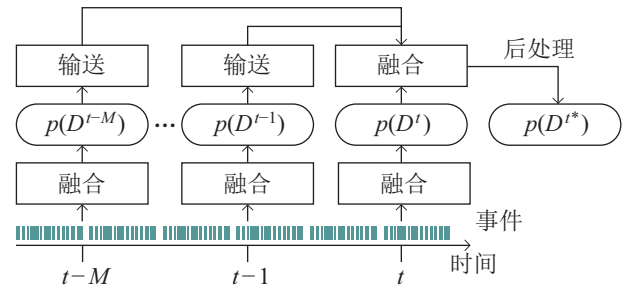
$$\Delta \boldsymbol{\rho} \sim St\left(-\frac{\sum \mathbf{J}_i}{\|\mathbf{J}\|^2} \boldsymbol{\mu}_r, \frac{s_r^2}{\|\mathbf{J}\|^2}, \nu_r\right) \quad (12)$$

最终方差表示为

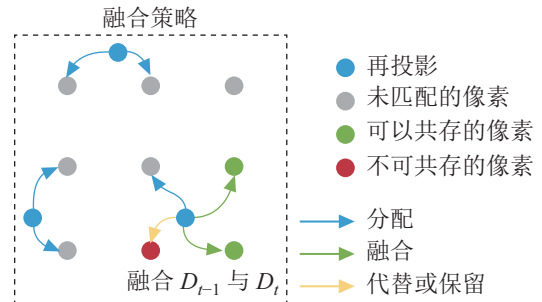
$$\sigma_{\boldsymbol{\rho}^*}^2 = \frac{\nu_r}{\nu_r - 2} \frac{s_r^2}{\|\mathbf{J}\|^2} \quad (13)$$

结合 IRLS 算法来融合事件逆深度估计以生成当前时间的半稠密深度图。该方法如图 4 所示, 采用概率深度融合框架可提高 3D 重建的密度, 同时提高建图结果的精度。

如图 4 所示, 逆深度计算过程将对选定的时间戳  $(t, \dots, t-M)$  的事件计算逆深度, 然后将计算得到的逆深度估计结果反馈给融合函数。使用融合函数对相机位姿、原始事件进行融合, 得到逆深度的真实值; 其中, 像素点与投影点的关系分为可以共存、不可共存、未匹配成功, 对应融合、代替或者保留当前像素坐标处的事件。



(a) 建图模块的事件融合框架



(b) 建图模块的事件融合策略

图 4 建图模块的事件融合框架及策略

Fig.4 Event fusion framework and strategy of the mapping module

融合算法使用 IRLS 算法来提升建图精度并提高计算速度。首先, 假设在像素坐标  $\mathbf{x}$  下的当前事件逆深度满足 Student's t 概率模型, 即  $St(\boldsymbol{\mu}_a, s_a^2, \nu_a)$ , 然后需要判断是否存在先前的事件逆深度  $St(\boldsymbol{\mu}_b, s_b^2,$

$v_b$ ), 如果不存在, 则利用当前的事件进行该像素坐标下的逆深度初始化; 如果当前时间戳的事件满足式 (14), 则进行事件逆深度的融合; 反之, 该像素坐标处的事件将会被认为是不可共存事件而过滤。

$$\mu_b - 2\sigma_b \leq \mu_a \leq \mu_b + 2\sigma_b \quad (14)$$

## 2.2 Luucas-Kanade 算法

使用向前组合 Luucas-Kanade 方法<sup>[10]</sup> 最小化迭代过程中的误差, 该方法的计算复杂度较小, 可以提高跟踪线程运算效率。Luucas-Kanade 方法通过迭代优化位姿参数, 逐步最小化产生的误差。使用下述公式描述位姿迭代的过程:

$$F(\Delta\theta) = \sum_{x \in S^{\text{ref}}} (\tilde{\Gamma}_{\text{left}}(\mathbf{W}(x, \rho; \theta), k))^2 \quad (15)$$

其中,  $S^{\text{ref}}$  为时间曲面深度信息,  $\tilde{\Gamma}_{\text{left}}$  为左时间曲面,  $\mathbf{W}(x, \rho; \theta)$  为用于优化运动参数  $\theta$  的翘曲函数, 其中, 运动参数  $\theta = (c^T, d^T)$ ,  $c = (c_1, c_2, c_3)$  为凯莱参数<sup>[11]</sup>,  $d$  为平移向量。翘曲函数  $\mathbf{W}(x, \rho; \theta)$  由式 (16) 逐步更新:

$$\mathbf{W}(x, \rho; \theta) \leftarrow \mathbf{W}(x, \rho; \theta) + \mathbf{W}(x, \rho; \Delta\theta) \quad (16)$$

结合式 (15)(16), 可以计算出运动参数  $\theta$ , 进而得到事件相机的真实位姿。

此外, 为了兼顾效率, 应用 Huber 范数<sup>[12]</sup> 和 IRLS 方法来提升计算速度。在每次迭代中, 在参考系中随机选取一批  $N_p$  (通常取  $N_p = 300$ ) 个 3 维点用于评估式 (13) 计算得到的运动参数。

## 3 性能测试分析 (Performance test and analysis)

针对基于事件相机的双目视觉 SLAM 系统, 在开源数据集上进行模拟来评估其性能。在数据集 1<sup>[5]</sup> 上对比 IRLS 计算方法与 LS 计算方法在“建图模块融合程度”和“建图效果”上的差异, 以评估两种方法的性能。

在数据集 2<sup>[12]</sup> 上对比本文算法与现有基于事件相机的视觉 SLAM 算法的建图质量以及位姿估计误差, 验证本文算法的性能。数据集 3<sup>[13]</sup> 为弱光、高动态环境。在这种困难环境下进行测试, 通过分析系统建图效果以及位姿误差, 验证本文方法在困难环境下的可靠性。

### 3.1 IRLS 算法性能分析

使用基于事件的双目视觉 SLAM 框架, 在数据集 3 上使用两种计算优化方法进行测试, 得到表

1 中的数据。事件融合程度用来表示事件流与真实场景的融合程度, 其值越高说明建立的地图越接近真实场景。平移误差均值是估计位姿与真实位姿在平移分量上绝对差异值的平均值, 误差均值越大, 所估计的位姿越不准确, 建图效果越差。方差用来衡量建图误差的离散程度, 方差越小, 建图误差越小。

表 1 IRLS 算法与 LS 算法对比  
Tab.1 Comparison of IRLS and LS algorithms

方法	事件融合程度	平移误差均值 /cm	方差 /cm
LS	$3.33 \times 10^5$	2.76	2.94
IRLS	$5.07 \times 10^5$	2.15	1.29

从表 1 看出, 使用 IRLS 方法进行计算优化时建图结果的事件融合程度更高, 且误差均值更小, 方差更小, 这证明 IRLS 方法优化建图的效果优于 LS 方法。

采用 IRLS 方法的误差均值更小, 因此最终估计的位姿更加接近真实位姿。图 5 放大了地图的局部区域。对比两图, 采用 IRLS 方法所得到的点云更加集中, 采用 LS 方法所得到的点云较为离散, 这也与表 1 中 IRLS 方法的方差更小相对应。上述实验结果证明了 IRLS 方法能够提高建图质量和定位精度。

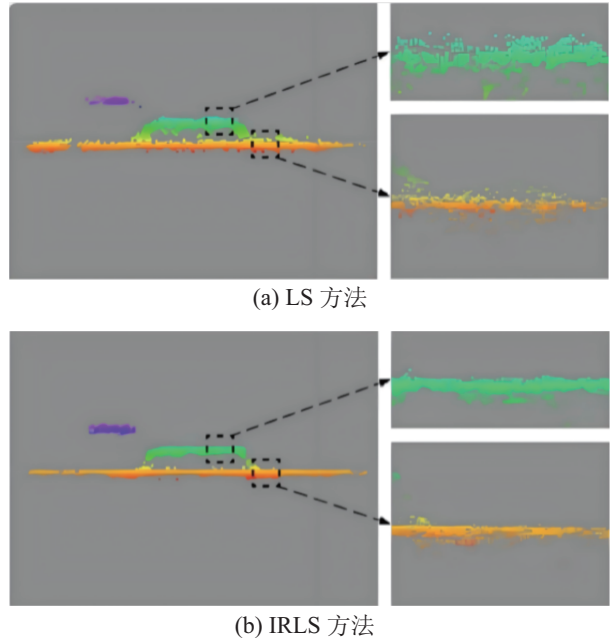


图 5 LS 方法与 IRLS 方法的建图效果

Fig.5 The mapping effect of LS method and IRLS method

### 3.2 Luucas-Kanade 算法性能分析

在数据集 1 上测试 Luucas-Kanade 算法的性能以及主流的边缘图对齐算法<sup>[6]</sup>, 结果如表 2 所示。

表 2 Luacas-Kanade 算法与边缘图对齐算法对比

Tab.2 Comparison between the Luacas-Kanade algorithm and the edge graph alignment algorithm

序列名	Luacas-Kanade		边缘图对齐	
	旋转 误差 /( $^{\circ}$ )	平移误差 均值 /cm	旋转 误差 /( $^{\circ}$ )	平移 误差 /cm
rpg_bin	1.2	3.1	0.6	1.2
rpg_desk	3.4	7.2	1.8	5.1
rpg_monitor	3.1	4.5	2.4	3.3
upenn_flying1	5.4	16.2	1.2	6.3
upenn_flying3	3.0	20.2	1.2	7.1

由表 2 可知, 在 Tracking 模块, 使用 Luacas-Kanade 算法的策略相较于使用边缘图对齐的策略取得了更好的效果。这证明了本文的 Luacas-Kanade 算法能够显著提升定位精度。

### 3.3 整体性能分析

在数据集 2 的两个场景测试本文方法的效果, 并与当前最优的基于事件相机的方法对比。

#### 3.3.1 不同方法的定位建图效果对比

利用数据集 2 的 Hexacopter 序列和 Car 序列测试本文方法, 并与基于事件相机的其他优秀方法对比, 对比方法依次表示为 GTS<sup>[14]</sup>、SGM<sup>[12]</sup>。

如表 3 所示, 计算了不同方法估计的位姿与真实位姿之间的误差。本文方法拥有更高的定位精度, 提升了约 48.3%, 为后续建图提供了精确的位姿。

表 3 不同方法在数据集 2 上的效果

Tab.3 The effect of different methods on dataset 2

组别	Hexacopter 序列		Car 序列	
	平均误差 均值 /cm	相对 误差 /%	平均误差 均值 /cm	相对 误差 /%
GTS	31	5.64	44	7.26
SGM	29	5.58	21	3.28
本文方法	16	<b>3.05</b>	19	<b>3.13</b>

图 6 展示了本文方法与 GTS、SGM 两种方法的建图效果。得益于优异的跟踪策略以及建图优化算法, 本文方法建立的地图拥有更高的信噪比以及更好的细节还原度。

#### 3.3.2 极端场景性能测试

此外, 还测量了本文方法在数据集 3 这类极端场景下的定位与建图效果。如图 7 所示, 在弱光、高动态场景下测试了本文方法的建图效果, 结果表明它能够很好地还原环境中的细节。

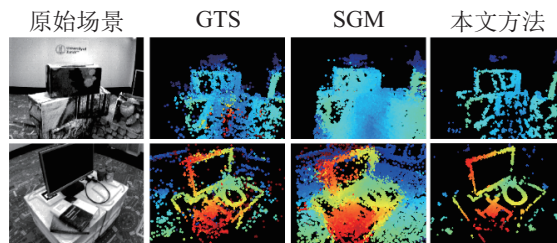
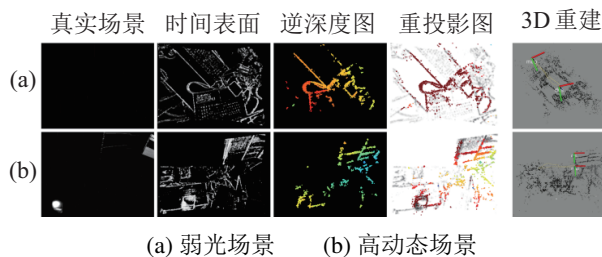


图 6 本文方法与其他方法建图效果对比

Fig.6 Comparison of the mapping effect between the proposed method and other methods



(a) 弱光场景 (b) 高动态场景

图 7 本文方法在极端场景下的建图效果

Fig.7 The mapping performance of the proposed method in extreme scenes

测量并计算了本文方法在两种困难场景下的位姿误差。如表 4 所示, 本文方法在极端场景下的相对误差能够控制在 3% 以内, 能够很好地满足定位建图精度的需求。

表 4 本文方法在极端场景下的位姿误差

Tab.4 The pose error of the proposed method in extreme scenes

场景	平均误差 /m	相对误差 /%
弱光	0.13	2.44
高动态	0.17	2.96

#### 3.3.3 实时性测试

为了验证本文方法的实时性, 测试了本文方法在数据集 1 以及数据集 2 上的平均耗时, 具体结果如表 5 所示。

表 5 本文方法的平均耗时

Tab.5 The average time consumption of the proposed method

模块	操作	耗时 /ms	总计 /ms
事件预处理	生成时间曲面	7	21
	逆深度估计	14	
建图	事件匹配	6	41
	逆深度融合	15	
	局部地图优化	20	
跟踪	位姿计算	10	10

如表 5 所示, 跟踪线程在不同数据集上的平均耗时为 10 ms, 即能够以 100 Hz 的频率实时估计位

姿。同时,事件预处理模块与建图模块分别耗时 21 和 41 ms,即分别以 50 和 25 Hz 的频率在场景中执行预处理与建图操作。

上述实验很好地证明了本文方法拥有优良的实时性,能够在不同场景中实时定位建图。

#### 4 结论 (Conclusion)

针对现有基于事件相机的 SLAM 算法定位建图精度欠佳的问题,提出了一套基于事件相机的双目视觉 SLAM 框架。对比了本文方法与现有先进方法的定位与建图精度,结果证明本文方法取得了更优的定位与建图效果。此外,在极端场景下测试了系统的性能,结果表明本文方法在极端场景下依然能够取得较好的效果。最后,测试了系统的实时性,结果表明所提方法拥有较好的实时性。

#### 参考文献 (References)

- [1] BURRI M, OLEYNIKOVA H, ACHELIK M W, et al. Real-time visual-inertial mapping, re-localization and planning onboard MAVs in unknown environments[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2015:1872-1878.
- [2] ZHONG F W, WANG S, ZHANG Z Q, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]//IEEE Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2018: 1001-1010.
- [3] CENSI A, SCARAMUZZA D. Low-latency event-based visual odometry[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 703-710.
- [4] GALLEGO G, SCARAMUZZA D. Accurate angular velocity estimation with an event camera[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 632-639.
- [5] ZHOU Y, GALLEGO G, SHEN S. Event-based stereo visual odometry[J]. IEEE Transactions on Robotics, 2021, 37(5): 1433-1450.
- [6] KIM H, LEUTENEGGER S, DAVISON A J. Real-time 3D reconstruction and 6-DoF tracking with an event camera[C]//

European Conference on Computer Vision. Cham, Switzerland: Springer, 2016: 349-364.

- [7] MUEGLER E, FORSTER C, BAUMLI N, et al. Lifetime estimation of events from dynamic vision sensors[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2015: 4874-4881.
- [8] VIDAL A R, REBECQ H, HORSTSCHAEFER T, et al. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 994-1001.
- [9] KOGLER J, HUMENBERGER M, SULZBACHNER C. Event-based stereo matching approaches for frameless address event stereo data[C]//International Symposium on Visual Computing. Berlin, Germany: Springer, 2011: 674-685.
- [10] LIU H, BRANDLI C, LI C, et al. Design of a spatiotemporal correlation filter for event-based sensors[C]//IEEE International Symposium on Circuits and Systems. Piscataway, USA: IEEE, 2015: 722-725.
- [11] REBECQ H, HORSTSCHAEFER T, GALLEGO G, et al. EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 593-600.
- [12] ZHOU Y, GALLEGO G, REBECQ H, et al. Semi-dense 3D reconstruction with a stereo event camera[C]//European Conference on Computer Vision. Cham, Switzerland: Springer, 2018: 242-258.
- [13] ZHU A Z, THAKUR D, OZASLAN T, et al. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception[J]. IEEE Robotics & Automation Letters, 2018, 3(3): 2032-2039.
- [14] HIRSCHMÜLLER H. Stereo processing by semiglobal matching and mutual information[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328-341.
- [15] IENG S-H, CARNEIRO J, OSSWALD M, et al. Neuromorphic event-based generalized time-based stereovision[J]. Frontiers in Neuroscience, 2018, 12. DOI: 10.3389/fnins.2018.00442.

#### 作者简介:

刘 勇 (1981-), 男, 博士。研究领域: 轨道交通牵引传动控制, 故障诊断等。

(上接第 47 页)

- [31] LI Y, XIE J X, LI Y, et al. SD-PDMD: Deep reinforcement learning for robotic trajectory imitation[C]//IEEE 34th International Conference on Tools with Artificial Intelligence. Piscataway, USA: IEEE, 2022: 1181-1186.
- [32] FARD M J, AMERI S, CHINNAM R B, et al. Soft boundary approach for unsupervised gesture segmentation in robotic-assisted surgery[J]. IEEE Robotics and Automation Letters, 2017, 2(1): 171-178.
- [33] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//34th International Conference on Machine Learning. Cambridge, USA: MIT Press, 2017: 214-223.

- [34] CUTURI M. Sinkhorn distances: Lightspeed computation of optimal transport[C]//Advances in Neural Information Processing Systems 26. Red Hook, USA: Curran Associates, Inc., 2013: 2292-2300.

#### 作者简介:

谢劼欣 (1990-), 男, 博士。研究领域: 护理机器人, 机器学习, 强化学习。

李 洋 (1992-), 男, 博士。研究领域: 机器人结构, 控制工程, 强化学习。

郭士杰 (1963-), 男, 博士, 教授。研究领域: 护理机器人, 机械工程, 控制工程。