

ORB-SLAM 2: 用于单目、双目和 RGB-D 相机的开源 SLAM 系统

摘要: 本文介绍了一种用于单目、双目和 RGB-D 相机的完整的同步定位和映射(SLAM)系统 ORB-SLAM2, 包括地图重用、闭环关闭和重新定位功能。该系统可在多种环境下对标准中央处理器进行实时处理, 从小型手持室内序列, 到在工业环境中飞行的无人机, 以及在城市中行驶的汽车。我们的后端, 基于束调整与单眼和立体观测, 允许精确的轨道估计与公制尺度。我们的系统包括一个轻量级的定位模式, 它利用视觉里程表跟踪未映射的区域, 并匹配允许零漂移定位的地图点。对 29 个常用公共序列的评价表明, 我们的方法达到了最先进的精度, 在大多数情况下是最准确的 SLAM 解决方案。我们发布源代码, 不仅是为了 SLAM 社区的利益, 也是为了为其他领域的研究人员提供一个开箱即用的 SLAM 解决方案

一、引言

同步定位与绘图 (SLAM) 是近二十年来计算机视觉和机器人界的一个热门研究课题, 近年来引起了高科技公司的关注, SLAM 技术公司建立了一个未知的环境, 并将传感器定位在地图上, 重点放在实时操作上。在不同的传感器模式中, 摄像头价格低廉, 能够提供丰富的环境信息, 从而实现可靠和准确的位置识别。因此, 视觉 SLAM 解决方案, 其中的主要传感器是一个摄像头, 是主要的兴趣。位置识别是 SLAM 系统的一个关键模块, 用于闭合环路 (即, 检测传感器何时返回到映射区域并纠正勘探过程中累积的错误), 并在跟踪失败后 (由于遮挡或攻击性运动或系统重新初始化) 重新定位相机

2016 年 10 月 20 日收到的手稿; 2017 年 3 月 3 日修订; 2017 年 4 月 19 日接受。出版日期: 2017 年 6 月 12 日; 现行版本日期: 2017 年 10 月 2 日。本文由副主编 J.Piater 和编辑 T.Murphey 根据评审者的评价推荐出版。这项工作部分得到了西班牙政府在项目 DPI2015-67275 下的支持, 部分得到了阿拉贡地区政府在项目 DGA T04-FSE 下的支持, 部分得到了教育部奖学金 FPU13/04175 的支持。(通讯作者: Raul Mur Artal.) R.Mur Artal 在西班牙萨拉戈萨 50018 号萨拉戈萨大学的 ingenier´a de Aragon´ (I3a) 研究所任职, 直至 2017 年 1 月。他目前在美国华盛顿州雷德蒙的 Oculus Research 工作 (邮箱: raul.muratal@oculus.com)。J.D.Tardos 是西班牙萨拉戈萨大学 (地址: 50018 Zaragoza, Universidad d e Zaragoza) 的 ingenier´a d e Aragon 研究所 (I3a) (电子邮件: ´tardos@unizar.es) 本文提供了可下载的补充资料, 网址为 <http://ieeexplore.ieee.org>。本文中一个或多个图形的颜色版本可在 <http://ieeexplore.ieee.org> 上在线获取。

视觉 SLAM 可以只使用单眼相机, 这是最便宜和最小的传感器设置。然而, 由于深度不能只从一个相机观测到, 地图的比例和估计的轨迹是未知的。此外, 系统引导需要多视图或过滤技术来生成初始映射, 因为它不能从第一帧进行三角定位。最后, 但并非最不重要的是, 单眼撞击会产生尺度漂移, 如果在探索中进行纯旋转, 可能会失败。通过使用双目或 RGB-D 相机, 所有这些问题都得到了解决, 并提供了最可靠的视觉 SLAM 解决方案。

在本文中, 我们基于我们的单目 ORB-SLAM[1]提出了 ORB-SLAM 2, 其贡献如下:

- 1) 用于单眼、双目和 RGB-D 相机的第一个开源[1]SLAM 系统, 包括环路关闭、重新定位和地图重用;
 - 2) 我们的 rgb-d 结果表明, 通过使用束调整 (ba), 我们比基于迭代最近点 (icp) 或光度和深度误差的最先进方法获得更高的精度。
- 最小化;

3) 通过近、远立体点和单目观测，我们的立体效果比最先进的直接立体撞击更准确；

4) 轻量级本地化模式，可以在禁用映射的情况下有效地重用映射。

图 1 显示了双目和 RGB-D 输入的 ORB-SLAM2 输出示例。立体案例显示了基蒂数据集[2]中序列 00 的最终轨迹和稀疏重建。这是 ORB-SLAM2 能够成功检测到的具有多个环路闭合的城市序列。rgb-d 案例显示了从 tum rgb-d 数据集[3]按 fr1_room 序列估计的关键帧姿势，以及由从估计的关键帧姿势反投影传感器深度映射渲染的密集点云。请注意，我们的 SLAM 不会执行任何融合，如 Kinect Fusion[4]或类似，但良好的定义表明关键帧姿势的准确性。附加视频中显示了更多示例。

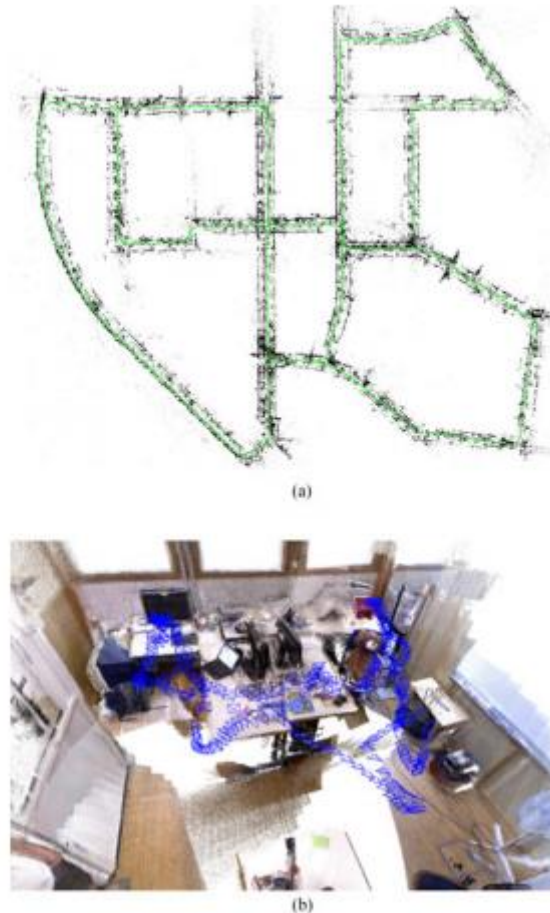


图 1 ORB-SLAM2 处理双目和 RGB-D 输入来估计相机的轨迹并构建环境地图。该系统能够在标准 cpu 上实时闭环、重定位和重用映射，具有较高的精度和鲁棒性。(a)立体输入:多回路闭合的城市环境轨迹和稀疏重建。(b) RGB-D 输入:一个房间场景的关键帧和密集的点云，一个循环闭合。点云是通过从估计的关键帧姿态反向投影传感器深度映射来呈现的

在本文的其余部分，我们将讨论第二节中的相关工作，在第三节中描述我们的系统，然后在第四节中给出评估结果，最后在第五节中给出结论。

二 相关工作

在本节中，我们讨论了有关 TereoandGB DSLAM 的相关工作。我们的讨论以及第四节中的评估只关注 SLAM 方法。

A. 双目 SLAM

比当时的其他方法。最重要的是，这是第一次利用近点和远点（即，由于立体摄像机中的小视差而无法可靠估计深度的点）进行立体撞击，后者使用反向深度参数化[6]。他们的经验表明，点的三方误差小于双目基线的 ~ 40 倍。在本文中，我们遵循这种以不同方式处理近点和远点的策略，如第 III-A 节所述。

大多数现代立体撞击系统都是基于关键帧的[7]系统，并在局部区域执行 BA 优化以实现可扩展性。斯特拉斯达等人的工作。[8]对关键帧的内部窗口中的 ba（点姿势约束）和外部窗口中的姿势图（姿势姿势约束）执行关节优化。通过限制这些窗口的大小，该方法以不保证全局一致性为代价，实现了恒定的时间复杂性。Mei 等人的相对同步定位和绘图系统（RSLAM）。[9]使用地标和姿势的相对表示，并在活动区域中执行相对 BA，该区域可以被限制为恒定时间。RSLAM 能够关闭循环，这允许在循环的两侧扩展活动区域，但不强制实现全局一致性。Pire 等人最近的 S-PTAM。[10]执行本地 BA，但缺少大循环关闭。与这些方法类似，我们在一组局部关键帧中执行 BA，以便复杂性与地图大小无关，我们可以在大型环境中操作。然而，我们的目标是建立一个全球一致的地图。当关闭一个循环时，我们的系统首先将两边对齐，类似于 RSLAM，以便跟踪能够继续使用旧地图进行定位，然后执行姿势图优化，以最小化循环中累积的漂移，然后执行完整的 BA。

Engel 等人最近的立体大尺度直接撞击（LSD-Slam）。[11]是一种半密集的直接方法，最大限度地减少高梯度图像区域的光度误差。该方法不依赖于特征，预计对运动模糊或纹理较差的环境更为稳健。然而，作为一种直接的方法，卷帘门或非朗伯反射等未建模效应会严重影响其性能。

B. RGB-D SLAM

最早也是最著名的 rgb-d slam 系统之一是 Newcombe 等人的 Kinect 融合。（4）。该方法将来自传感器的所有深度数据融合到一个体积密度模型中，该模型用于使用 ICP 跟踪摄像机姿态。该系统仅限于小工作空间，这是由于体积测量显示和回路闭合的缘故。Whelan 等人[12]能够使用滚动循环缓冲区在大型环境中操作，并使用位置识别和姿势图优化包括循环关闭。

可能第一个流行的开源系统是 Endres 等人的 rgb-d slam。（13）。这是一个基于特征的系统，其前端通过特征匹配和 ICP 计算帧到帧的运动。后端使用来自搜索的循环关闭约束执行姿势图优化。同样，Kerl 等人的 backenddvo slam。[14]优化一个姿势图，其中关键帧到关键帧的约束是根据视觉里程计计算的，该视觉里程计可最大限度地减少光度和深度误差。dvo-slam 还以启发式方式搜索所有之前帧的循环候选，而不是依赖于位置识别。

最近惠兰等人的弹性融合。[15]构建一个基于苏菲尔的环境地图。这是一种以地图为中心的方法，它可以忽略姿势并对地图应用非刚性变形来执行循环闭合，而不是标准的姿势图优化。该系统的详细重构和定位精度令人印象深刻，但目前的实现仅限于房间大小的地图，因为复杂度随地图中 SURFEL 的数量而变化。

如 Strasdt 等人提出的。[8]我们的 ORB-SLAM2 使用深度信息合成立体坐标，以提取图像上的特征。这样，我们的系统就不知道输入是双目还是 RGB-D。与上述所有方法不同，我们的后端是基于束调整，并建立一个全球一致的稀疏重建。因此，我们的方法是轻量级的，并且与标准中央处理器（CPU）一起工作。我们的目标是长期和全球一致的本地化，而不是建立最详细的密集重建。然而，从高精度的关键帧姿态，可以融合深度图，在局部区域进行实时精确重建，或者在一个完整的 BA 后对所有关键帧的深度图进行后处理，得到整个场景的精确三维模型。

三 ORB-SLAM2

双目和 RGB-D 相机的 ORB-SLAM 2 是基于我们基于单目特征的 ORB-SLAM[1]构建的，其主要组件在这里进行了总结，以方便读者阅读。系统的概述如图 2 所示。该系统具有三个主要的并行线程：1) 通过寻找与局部映射匹配的特征，并将只应用运动 BA 的重投影误差最小化，从而实现每帧相机的定位跟踪；2) 通过局部映射来管理局部映射并对其进行优化，从而实现局部 BA；3) 通过闭合循环来检测大循环，并对其进行校正。他通过执行姿势图优化来积累漂移。该线程在姿势图优化后启动第四个线程来执行完整的 BA，以计算最优结构和运动解。

系统嵌入了一个基于 dbow2[16]的位置识别模块，用于在跟踪失败（如遮挡）时重新定位，或在已映射的场景中重新初始化，以及用于循环检测。该系统保持一个共可见性图[8]链接观察公共点的任意两个关键帧，以及连接所有关键帧的最小生成树。这些图形结构允许检索关键帧的本地窗口，以便跟踪和本地映射在本地操作，允许在大型环境中工作，并作为关闭循环时执行的姿势图优化的结构。

系统使用相同的 ORB 功能[17]跟踪、映射和位置识别任务。这些特性对旋转和缩放都很稳健，对相机自动增益和自动曝光以及照明变化都具有很好的不变性。此外，它们提取和匹配速度快，允许实时操作，并且在字位识别包中显示出良好的精度/回忆性能[18]。

在本节的其余部分，我们将介绍如何利用立体/深度信息，以及系统的哪些元素受到影响。有关每个系统块的详细描述，我们请读者参阅我们的单目出版物[1]。

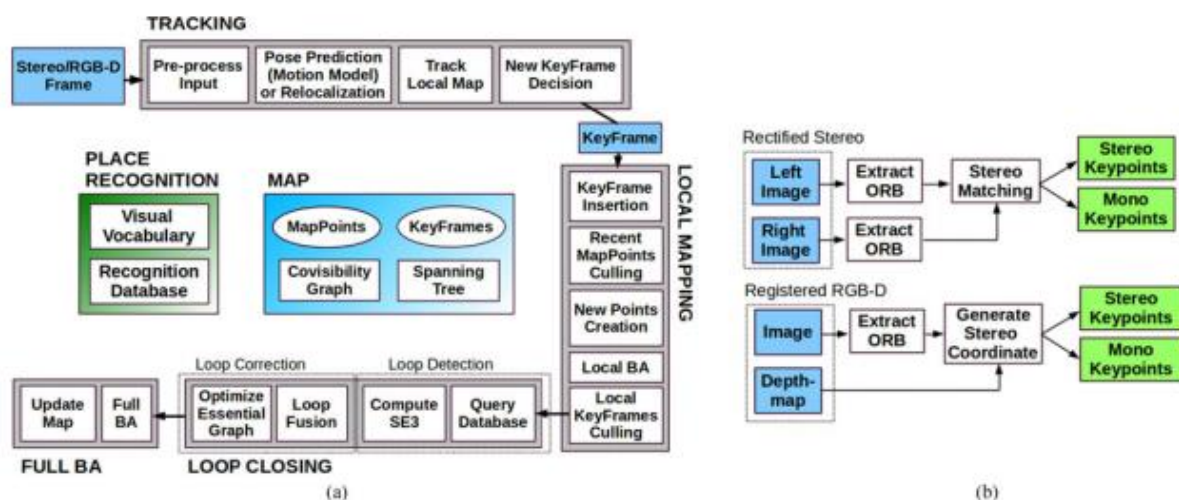


图 2 ORB-SLAM2 由三个主要并行线程组成:跟踪、本地映射和循环关闭，循环关闭后可以创建第四个线程来执行完整的 BA。跟踪线程对立体声或 RGB-D 输入进行预处理，使系统的其余部分独立于输入传感器工作。虽然没有在图中显示，但 ORB-SLAM2 也可以像[1]中那样使用单目输入。(a)系统线程和模块。(b)输入预处理。

A. 单眼、近双目和远双目关键点

ORB-SLAM2 作为一种基于特征的方法，对输入进行预处理，以在突出的关键点位置提取特征，如图 2 (b) 所示。然后丢弃输入图像，所有系统操作都基于这些功能，因此系统独立于双目或 RGB-D 传感器。我们的系统处理单眼和双目关键点，这些关键点进一步分类为近或远。

双目关键点定义为三个坐标 $s = (u_l, v_l, u_r)$ ，即 (u_l, v_l) 左图像上的坐标和右图像中的水平坐标。对于立体相机，我们提取两个图像中的圆珠，对于每一个左边的圆珠，我们在右边的图像中搜索匹配的圆珠。这可以非常有效地假设立体校正图像，使极线是水平的。然后，我们用左圆球坐标和右匹配的水平坐标生成立体关键点，这是由面片相关细化的亚像素。对于 RGB-D 相机，我们提取了 RGB 图像上的 ORB 特征，如 Strasdt 等人提出的。[8]对于具有坐标 (u_l, v_l) 的每个特征，我们将其深度值转换为虚拟右坐标。

$$u_R = u_L - \frac{f_x b}{d} \quad 1)$$

其中， f_x 是水平焦距，是结构光投影仪和红外摄像机之间的基线，对于 Kinect 和 Asus Xtion，我们大约为 8 厘米。深度传感器的不确定度由虚拟直角坐标的不确定度表示。通过这种方式，来自双目和 RGB-D 输入的功能被系统的其他部分平等地处理。外汇乙

如果立体关键点的相关深度小于 40 倍，则将其分类为“近”，否则将其分类为“远”。当深度被精确估计并提供比例、平移和旋转信息时，闭合关键点可以从一个帧安全地三角定位。另一方面，远点提供了准确的旋转信息，但尺度和平移信息较弱。在多个视图支持的情况下，我们对远点进行三角测量。

单眼关键点由左侧图像上的两个坐标 $m = (u_l, v_l)$ 定义，并对应于无法找到立体匹配或在 rgb-d 情况下深度值无效的所有 ORB。这些点只从多个视图进行三角测量，不提供比例信息，但有助于旋转和平移估计。

B. 系统引导

使用立体或 RGB-D 相机的一个主要好处是，通过只从一帧获得深度信息，我们不需要像单目相机那样从运动初始化得到特定的结构。

在系统启动时，我们用第一帧创建一个关键帧，将其姿势设置为原点，并从所有立体关键点创建一个初始映射。

c. 单目和立体约束的束调整

我们的系统执行 BA 以优化跟踪线程中的相机姿势（仅运动 BA），优化本地映射线程（本地 BA）中关键帧和点的本地窗口，并在循环关闭后优化所有关键帧和点（完整 BA）。我们使用 g2o 中实现的 levenberg-marquardt 方法[19]。

仅运动 BA 优化摄像机方位 $\in so(3)$ 和位置 $\in \mathbb{R}^3$ ，最大限度地减小匹配三维点之间的重投影误差 $\in \mathbb{R}^{RT3}$ 我 3 在世界坐标和关键点中，无论是单眼的还是立体的，我所有匹配的集合

$$\{\mathbf{R}, \mathbf{t}\} = \underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \sum_{i \in \mathcal{X}} \rho \left(\left\| \mathbf{x}_{(\cdot)}^i - \pi_{(\cdot)}(\mathbf{R} \mathbf{X}^i + \mathbf{t}) \right\|_{\Sigma}^2 \right) \quad (2)$$

其中是鲁棒 Huber 成本函数和 Σ 与关键点规模相关的协方差矩阵。投影函数 $\pi_{(\cdot)}$ ，单眼 \mathbf{M} 和校正立体 \mathbf{S} 定义如下：

$$\pi_m \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix}, \quad \pi_s \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \\ f_x \frac{X-b}{Z} + c_x \end{bmatrix} \quad (3)$$

其中 (f_x, f_y) 是焦距， (c_x, c_y) 是主点和基线，所有这些都是通过校准知道的。乙

本地 BA 优化一组可重叠的关键帧和这些关键帧中的所有点 P。所有其他关键帧 K，而不是 K，观察点 P 对成本函数的贡献，但在优化中保持不变。定义 xas 关键帧中 pand 关键点之间的匹配集，优化问题如下：

$$\{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l | i \in \mathcal{P}_L, l \in \mathcal{K}_L\} = \underset{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l}{\operatorname{argmin}} \sum_{k \in \mathcal{K}_L \cup \mathcal{K}_F} \sum_{j \in \mathcal{X}_k} \rho(E(k, j))$$

$$E(k, j) = \left\| \mathbf{x}_{(\cdot)}^j - \pi_{(\cdot)}(\mathbf{R}_k \mathbf{X}^j + \mathbf{t}_k) \right\|_{\Sigma}^2. \quad (4)$$

全局 BA 是局部 BA 的特定情况，其中地图中的所有关键帧和点都进行了优化，但固定的原点关键帧除外，以消除仪表自由度。

d. 回路闭合和满 BA

环路阻塞分两种方式进行，第一种方式是检测并验证环路阻塞，第二种方式是对环路进行校正，优化姿态图。与单眼 ORB-SLAM 相比，在这种情况下可能会发生尺度漂移[20]，立体/深度信息使尺度可观测，几何验证和姿态图优化不再需要处理尺度漂移，而是基于刚体变换而不是相似性。

在 ORB-SLAM2 中，我们在姿态图之后加入了一个完整的 BA 优化，以获得最优解。这种优化可能非常昂贵，因此，我们在单独的线程中执行它，允许系统继续创建映射和检测循环。然而，这带来了将捆绑调整输出与地图当前状态合并的挑战。如果在优化运行时检测到新的循环，我们将中止优化并继续关闭循环，这将再次启动完整的 BA 优化。当完整的 BA 完成时，我们需要合并更新的关键帧子集。

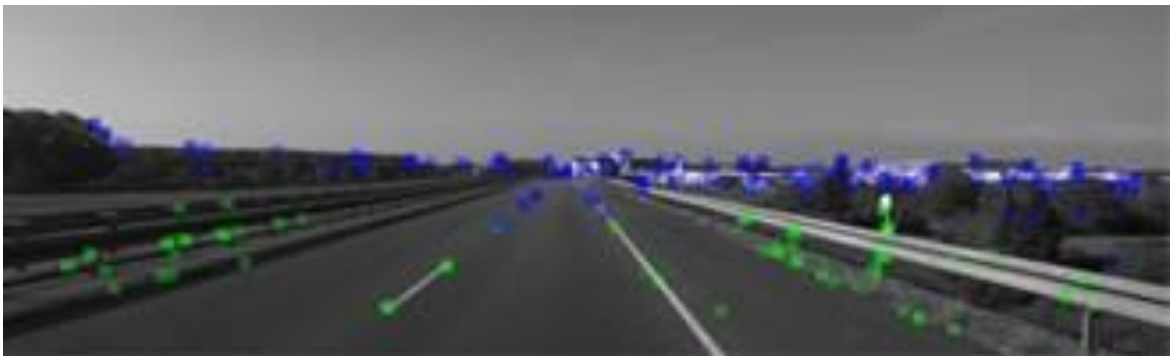


图 3 基蒂 01[2]中的跟踪点。绿色点的深度小于双目基线的 40 倍，而蓝色点则更远。在这种序列中，经常插入关键帧是很重要的，这样，闭合点的数量就可以进行精确的平移估计。远点有助于估计方向，但为翻译和尺度提供了薄弱的信息。

以及由完整的 BA 优化的点，带有未更新的关键帧和在优化运行时插入的点。这是通过在生成树中传播更新关键帧（即从非优化到优化姿势的转换）到非更新关键帧的校正来完成的。未更新的点根据应用于其参考关键帧的校正进行转换。

e. 关键帧插入

ORB-SLAM 2 遵循单目 ORBSLAM 中引入的策略，即经常插入关键帧，然后剔除多余的关键帧。近距和远距立体点之间的区别允许我们引入一种新的关键帧插入条件，这在场景的大部分远离立体传感器的具有挑战性的环境中至关重要，如图 3 所示。在这种情况下，我们需要有足够数量的近点来精确估计平移，因此，如果跟踪的近点数量降到下面，并且帧可以创建至少新的近立体点，系统将插入一个新的关键帧。我们根据经验发现 $n=100$ 和 $n=70$ 在我们所有的实验中都很有效。

f. 国产化模式

我们采用了一种定位模式，只要环境没有显著变化，这种模式对于在绘制良好的区域中进行轻型长期定位非常有用。在这种模式下，本地映射和循环关闭线程将被停用，并且如果需要，摄像机将通过使用重新定位的跟踪进行连续定位。在这种模式下，跟踪利用视觉里程计匹配和匹配地图点。视觉里程计匹配是当前帧中的 ORB 与上一帧中根据立体/深度信息创建的三维点之间的匹配。这些匹配使定位对未映射区域具有鲁棒性，但可以累积漂移。地图点匹配确保对现有地图进行无漂移定位。此模式在随附的视频中演示。

四、评价

我们在三个流行的数据集中对 ORB-SLAM2 进行了评估，并与其他最先进的 SLAM 系统进行了比较，使用的始终是原始作者发布的结果和文献中的标准评估指标。我们已经运行 ORB-SLAM2 在英特尔酷睿 i7-4790 台式机与 16 gb 内存。为了考虑多线程系统的不确定性，我们将每个序列运行 5 次，并给出估计轨迹精度的中值结果。我们的开源实现包括在所有这些数据集中运行系统的校准和说明。

表一 Kitti 数据集中精度的比较

Sequence	ORB-SLAM2 (stereo)			Stereo LSD-SLAM		
	t_{rel} (%)	r_{rel} (deg/100 m)	t_{abs} (m)	t_{rel} (%)	r_{rel} (deg/100 m)	t_{abs} (m)
00	0.70	0.25	1.3	0.63	0.26	1.0
01	1.39	0.21	10.4	2.36	0.36	9.0
02	0.76	0.23	5.7	0.79	0.23	2.6
03	0.71	0.18	0.6	1.01	0.28	1.2
04	0.48	0.13	0.2	0.38	0.31	0.2
05	0.40	0.16	0.8	0.64	0.18	1.5
06	0.51	0.15	0.8	0.71	0.18	1.3
07	0.50	0.28	0.5	0.56	0.29	0.5
08	1.05	0.32	3.6	1.11	0.31	3.9
09	0.87	0.27	3.2	1.14	0.25	5.6
10	0.60	0.27	1.0	0.72	0.33	1.5

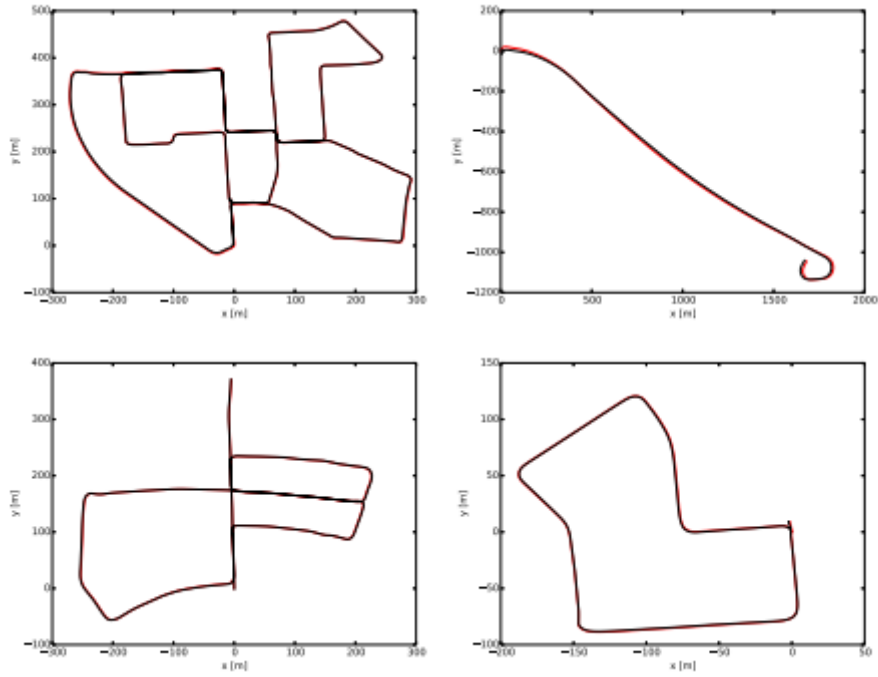


图 4 Kitti00、01、05 和 07 的估计轨迹（黑色）和地面实况（红色）。

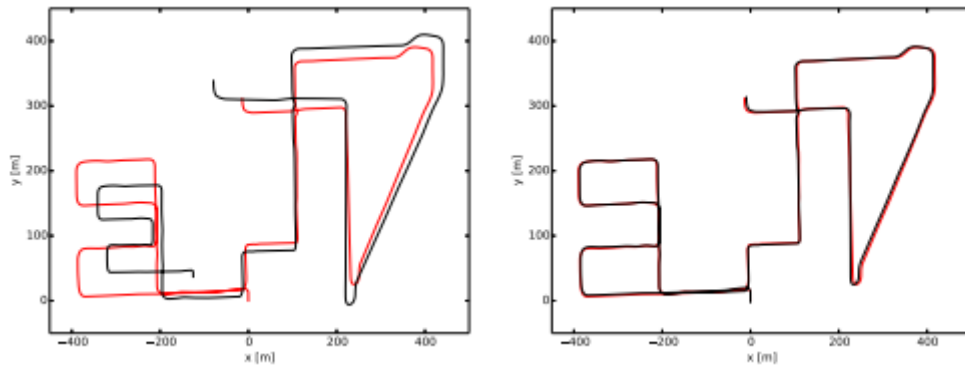


图 5 Kitti 08 的估计轨迹（黑色）和地面实况（红色）。左：单眼 ORB-SLAM[1]，右：ORB-SLAM 2（双目）。单眼大圆球在这个过程中，特别是在转弯处，会出现严重的尺度漂移。相比之下，所提出的立体版本能够估计轨道和地图的真实比例，而没有比例漂移。

A. 基蒂数据集

Kitti 数据集[2]包含在城市和公路环境中从汽车记录的双目序列。立体传感器有~54 厘米的基线，在校正 1240×376 像素后以 10 赫兹的分辨率工作。序列 00、02、05、06、07 和 09 包含循环。

我们的 ORB-SLAM2 检测到所有循环，并且能够在之后重用其映射，除了序列 09，其中循环发生在序列末尾的极少数帧中。表中显示了 11 训练的结果。

表二 欧洲委员会数据集

Sequence	ORB-SLAM2 (stereo)	Stereo LSD-SLAM
V1_01_easy	0.035	0.066
V1_02_medium	0.020	0.074
V1_03_difficult	0.048	0.089
V2_01_easy	0.037	-
V2_02_medium	0.035	-
V2_03_difficult	X	-
MH_01_easy	0.035	-
MH_02_easy	0.018	-
MH_03_medium	0.028	-
MH_04_difficult	0.119	-
MH_05_difficult	0.060	-

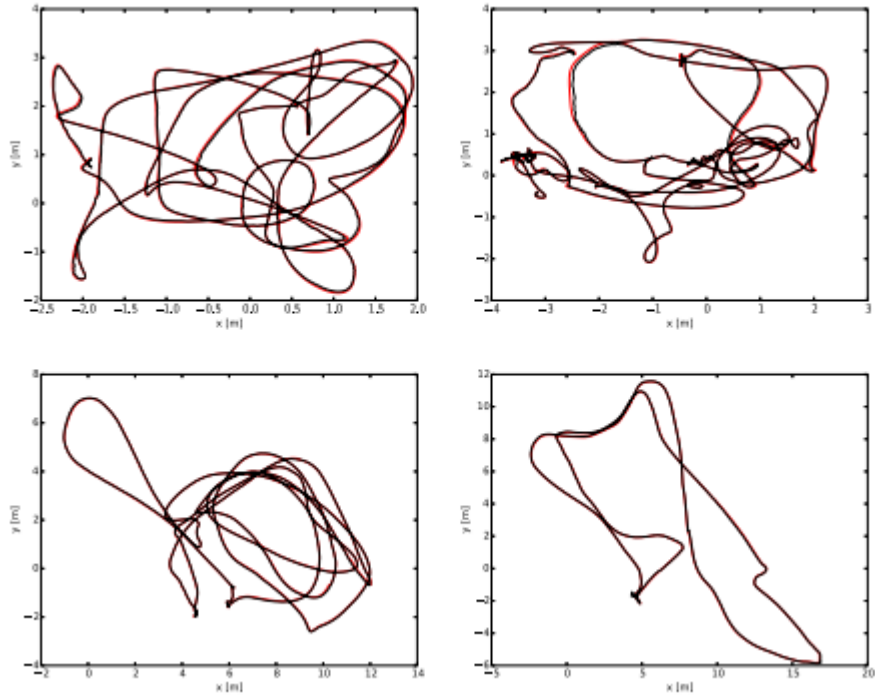


图 6 Euroc v1_02_Medium、v2_02_Medium、mh_03_Medium 和 mh_05_Diffible 中的估计轨迹（黑色）和地面真值（红色）。

表三 TUM RGB-D 数据集

Sequence	ORB-SLAM2	Elastic- (RGB-D)	Kintinuous Fusion	DVO SLAM	RGBD SLAM
fr1/desk	0.016	0.020	0.037	0.021	0.026
fr1/desk2	0.022	0.048	0.071	0.046	-
fr1/room	0.047	0.068	0.075	0.043	0.087
fr2/desk	0.009	0.071	0.034	0.017	0.057
fr2/xyz	0.004	0.011	0.029	0.018	-
fr3/office	0.010	0.017	0.030	0.035	-
fr3/nst	0.019	0.016	0.031	0.018	-

与最先进的双目 LSD-SLAM[11]相比，具有公共基础真实性的序列，据我们所知，只有双目 SLAM 显示所有序列的详细结果。我们使用两种不同的度量，即[3]中提出的绝对平移均方根误差（RMSE）和[2]中提出的平均相对平移和旋转误差。我们的系统在大多数序列中都优于双目 LSD-SLAM，并且通常可以实现低于 1% 的相对误差。序列 01（见图 3）是训练集中唯一的公路序列，翻译错误稍差。在这个序列中很难估计平移，因为由于高速和低帧速率，可以跟踪的闭合点很少。

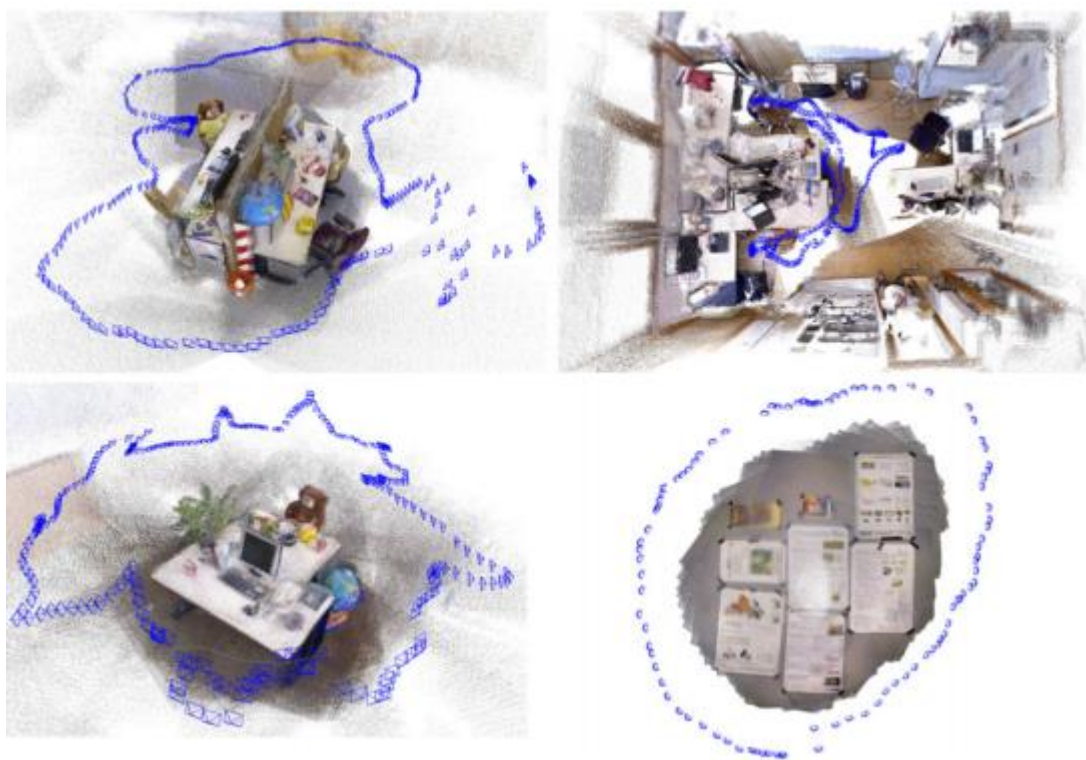


图 7 从 TUM rgb-d fr3_办公室、fr1_房间、fr2_书桌和 fr3_nst 中估计的关键帧姿态和传感器深度图进行密集点云重建。

表四
每根螺纹的计时结果（单位：毫秒）（平均值 \pm 2 标准偏差）

Settings	Sequence	V2_02	07	fr3_office
	Dataset	EuRoC	KITTI	TUM
	Sensor	Stereo	Stereo	RGB-D
	Resolution	752 \times 480	1226 \times 370	640 \times 480
	Camera FPS	20 Hz	10 Hz	30 Hz
	ORB Features	1000	2000	1000
Tracking	Stereo Rectification	3.43 \pm 1.10	-	-
	ORB Extraction	13.54 \pm 4.60	24.83 \pm 8.28	11.48 \pm 1.84
	Stereo Matching	11.26 \pm 6.64	15.51 \pm 4.12	0.02 \pm 0.00
	Pose Prediction	2.07 \pm 1.58	2.36 \pm 1.84	2.65 \pm 1.28
	Local Map Tracking	10.13 \pm 11.40	5.38 \pm 3.52	9.78 \pm 6.42
	New Keyframe Decision	1.40 \pm 1.14	1.91 \pm 1.06	1.58 \pm 0.92
	Total	41.66 \pm 18.90	49.47 \pm 12.10	25.58 \pm 9.76
Mapping	Keyframe Insertion	10.30 \pm 7.50	11.61 \pm 3.28	11.36 \pm 5.04
	Map Point Culling	0.28 \pm 0.20	0.45 \pm 0.38	0.25 \pm 0.10
	Map Point Creation	40.43 \pm 36.10	47.69 \pm 29.52	53.99 \pm 23.62
	Local BA	137.99 \pm 248.18	69.29 \pm 61.88	196.67 \pm 213.42
	Keyframe Culling	3.80 \pm 8.20	0.99 \pm 0.92	6.69 \pm 8.24
	Total	174.10 \pm 278.80	129.52 \pm 88.52	267.33 \pm 245.10
Loop	Database Query	3.57 \pm 5.86	4.13 \pm 3.54	2.63 \pm 2.26
	SE3 Estimation	0.69 \pm 1.82	1.02 \pm 3.68	0.66 \pm 1.68
	Loop Fusion	21.84	82.70	298.45
	Essential Graph Opt.	73.15	178.31	281.99
BA	Total	108.59	284.88	598.70
	Full BA	349.25	1144.06	1640.96
	Map Update	3.13	11.82	5.62
	Total	396.02	1205.78	1793.02
	Loop size (#keyframes)	82	248	225

可以精确估计, 每 100 米误差为 0.21 度, 因为有许多远点可以长期跟踪。图 4 显示了一些估计轨迹的例子。

与[1]中的单眼结果相比, 所提出的立体版本能够处理单眼系统失效的序列 01。在这个公路序列中, 见图 3, 闭合点仅在几个帧中出现。立体版只从一个立体关键帧而不是单目镜的延迟初始化创建点的能力, 包括在两个关键帧之间找到匹配点, 这一序列不适合进行跟踪。此外, 立体版系统用度量尺度估计地图和轨迹, 不受尺度漂移的影响, 如见图 5。

B. 欧洲数据集

最新的 EuroC 数据集[21]包含 11 个立体序列, 这些立体序列是从一辆微型飞行器 (MAV) 上记录下来的, 该飞行器在两个不同的房间和一个大型工业环境中飞行。双目传感器有~11 厘米的基线, 提供 20 赫兹的 WVGA 图像。序列分为简单、中等和困难三类, 这取决于 Mav 的速度、照明度和场景纹理。在所有的序列中, MAV 重新访问环境, ORB-SLAM2 能够重用其映射, 必要时关闭循环。表 II 显示了与立体 LSD-SLAM 相比, 所有序列的 ORBSLAM2 的绝对平移 RMSE, 结果如[11]所示。ORB-SLAM 2 的定位精度达到几厘米, 比双目 LSD-SLAM 更精确。由于严重的运动模糊, 我们的跟踪在 V2-03 的某些部分会丢失。如[22]所示, 可以使用 IMU 信息处理此序列。图 6 显示了与地面真值比较的计算轨迹示例。

C. TUM RGB-D 数据集

tum rgb-d 数据集[3]包含来自不同类别的 rgb-d 传感器的室内序列, 用于评估不同纹理、照明和结构条件下的对象重建和 SLAM/Odometry 方法。我们将结果显示在序列的一个子集中, 其中大多数 RGB-D 方法通常都是被评估的。在表三中, 我们将我们的精确度与以下最先进的方法进行了比较: 弹性融合[15]、扭结法[12]、dvo-slam[14]和 rgb-d slam[13]。我们的方法是唯一一种基于束调整的方法, 在大多数序列中都优于其他方法。正如我们在[1]中已经注意到的, 对于 rgb-d slam 结果, freiburg2 序列的 depth maps 具有 4% 的比例偏差, 可能来自错误校准, 我们已经在运行中进行了补偿, 并且可以部分地解释我们的显著更好的结果。图 7 显示了点云: 从计算的关键帧姿势中反投影传感器深度图的结果 in 四个序列。良好的清晰度和书桌和海报的直线轮廓证明了我们方法的高精度定位。

D 计时结果

为了完成对所提议系统的评估, 我们在表四中给出了三个具有不同图像分辨率和传感器的序列的计时结果。显示每个线程任务的平均值和两个标准偏差范围。由于这些序列包含一个单循环, 因此整个 BA 和循环关闭线程的一些任务只执行一次, 并且只报告一次时间测量。每个帧的平均跟踪时间低于每个序列的摄像机帧速率的倒数, 这意味着我们的系统能够实时工作。由于立体图像中的 ORB 提取是并行的, 可见在 v2_02 的立体 wvga 图像中提取 1000 个 ORB 特征与在 fr3_office 的单个 vga 图像通道中提取相同数量的特征类似。

循环中关键帧的数量显示为与循环关闭相关的时间的参考。虽然 kitti 07 中的循环包含更多的关键帧, 但为室内 fr3_办公室构建的可互见性图更密集, 因此, 循环融合、姿势图优化和完整的 ba 任务更昂贵。由于共可见性图的密度越高, 使得局部地图包含的关键帧和点越多, 因此局部地图跟踪和局部 BA 也就越昂贵。

五、结论

我们已经为单眼、双目和 RGB-D 传感器提供了一个完整的系统, 能够在标准 CPU 上执行重新定位、循环关闭和实时重用地图。我们专注于构建全球一致的地图, 以便在各种环境中实现可靠和长期的本地化, 如实验所示。所提出的定位模式具有系统的定位能力, 在已知

的环境下产生了一种非常健壮、零漂移和轻量级的定位方法。这种模式对某些应用程序很有用，例如在一个映射良好的空间中跟踪虚拟现实中的用户视点。

与最新技术的比较表明，ORB-SLAM2 在大多数情况下达到了最高的精度。在基蒂视觉里程表基准 ORB-SLAM2 是目前最好的双目 SLAM 解决方案。至关重要的是，与近年来蓬勃发展的立体视觉里程计方法相比，ORB-SLAM2 在已绘制的区域实现了零漂移定位。

令人惊讶的是，`argb-dresultsdemonstrate` 相机定位的准确度是理想的，捆绑调整比直接方法或 ICP 执行得更好，其额外的优点是计算成本更低，不需要 GPU 处理实时运行。

我们已经发布了系统的源代码，以及示例和说明，以便其他研究人员可以轻松使用。据我们所知，ORB-SLAM 2 是第一个开源的可视 SLAM 系统，可以使用单目、双目和 RGB-D 输入。此外，我们的源代码还包含一个使用单目摄像机的增强现实应用程序的示例[2]来展示我们解决方案的潜力。

未来的扩展可能包括，举例来说，非重叠多镜头、鱼眼或全向摄像头支持、大规模密集融合、合作映射或增强的运动模糊鲁棒性。