题目

ORB-SLAM2: An Open-Source SLAM System form Monocular, Stereo, and RGB-D Cameras

摘要

本文提出了ORB-SLAM2，它是基于单目、双目或RGB-D相机的一个完整的SLAM系统，其中包括地图重用、回环检测和重定位功能。这个系统可以适用于多种环境，无论是室内小型手持设备，还是工厂环境中飞行的无人机和城市中行驶的车辆，其都可以在标准CPU上实时运行。该系统的后端使用基于单目和双目观测的光束法平差法（bundle adjustment），这使得其可以精确估计轨迹的尺度。该系统包含一个轻量级的定位模式，它使用视觉里程计追踪未建图区域并匹配地图点，实现零漂移定位。在29个常用公开数据集上的实验评估显示本文方法在大多数情况下是精度最高的SLAM解决方案。我们公开了该系统的源代码，从而促进SLAM领域发展，同时也给其他领域的研究者提供一套能够开箱即用的SLAM解决方案。

1 引言

同时定位建图（SLAM）在过去二十年中一直是计算机视觉和机器人领域的研究热点，最近它也吸引了高科技企业的关注。SLAM技术对未知环境进行建图，同时实时地在地图中定位传感器的位置。在不同传感器中，相机相对便宜，同时能够提供鲁棒准确的位置识别所需的丰富的环境信息，所以以摄像头为主要传感器的视觉SLAM解决方案，是当前最受青睐的研究热点。位置识别是SLAM系统中实现回环检测（当检测到传感器回到已建图区域时，修正探索过程中的累积误差）的关键模块，它能够在因遮挡或剧烈运动导致追踪失败后以及系统重新初始化时，重定位相机的位置。

视觉SLAM只需一个单目相机即可实现，这是最便宜也是最小的传感器设备。但是，仅靠一个单目相机无法观测到深度信息，所以地图和估计轨迹的尺度是未知的。另外，由于单目视觉SLAM系统无法仅根据第一帧图像进行三角化测量（triangulate），所以系统启动时需要多个视角或者滤波技术来生成一个初始地图。同时，单目SLAM会造成尺度漂移，在纯旋转的探索过程中可能会失败。不过，通过使用双目或者RGB-D相机，这些问题都可以被解决，从而实现更可靠的视觉SLAM解决方案。

本文中，在我们之前提出的单目ORB-SLAM的基础上，我们进一步提出了ORB-SLAM2，它有以下贡献：

- 第一个开源的基于单目、双目、RGB-D相机的SLAM系统，其中包括回环检测、重定位、以及地图重用功能。
- 我们的RGB-D结果显示，相比较目前最好的基于迭代最近点法（ICP）或广度和深度误差最小法，我们通过使用光束法平差法（BA），可以达到更高的精度。
- 通过使用远近匹配双目点和单目观测，我们的双目结果比目前最好的直接双目SLAM的精度更高。
- 提出了一个轻量级的重定位模式，它可以在无法建图时，有效地重新使用地图。

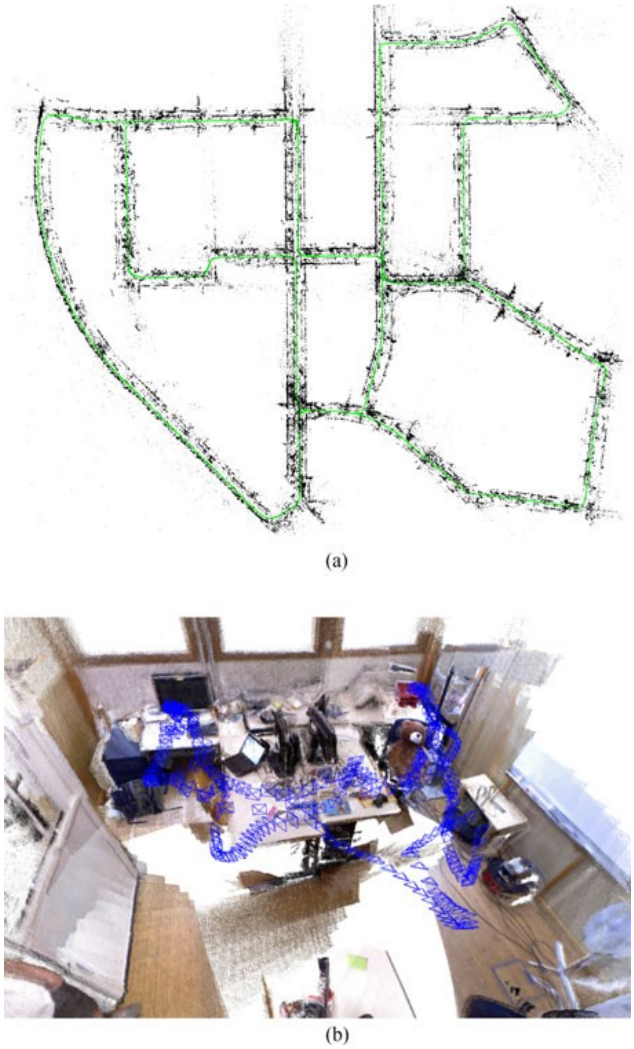

图 1 ORB-SLAM2 对双目和RGB-D输入进行处理，估计相机轨迹，并建立环境的地图。该系统能够该系统能够实时地在标准CPU上进行闭合回环，重定位，重用地图，并具有高精度和高鲁棒性。

图 1 展示了双目和RGB-D输入下ORB-SLAM2系统的输出。其中双目的例子展示了KITTI数据集 00序列的最终估计轨迹和稀疏重建结果。这是一个具有多个回环闭合的城市场景数据序列，ORB-SLAM2系统成功检测到了这些回环。RGB-D的例子展示了TUM RGB-D数据集中的frl-room序列的关键帧位姿估计结果及所得到的稠密点云，其中稠密点云是根据关键帧位姿，将传感器深度图反向投影所得到的。需要注意的是，ORB-SLAM2系统并没有像KinectFusion之类的系统一样进行任何融合，但是却能够精确地估计关键帧位姿。附件视频中将会展示更多的例子。

本文余下章节中，我们将会在第2节中讨论相关工作，在第3节中介绍我们的系统，在第4节中给出实验评估结果，最后在第5节中进行总结。

2 相关工作

在本节中，我们将会讨论双目和RGB-D SLAM的相关工作。本节中的讨论和第四节中的评估只针对SLAM方法。

2.1 双目SLAM

Paz等人曾做出了一个早期的卓越的双目SLAM系统[5]，他们基于条件独立分治的扩展卡尔曼滤波SLAM（EKF-SLAM），使该系统在那个年代相比较其他方式可以在更大的场景中运行。最重要的是，它是第一个同时使用近特征点和远特征点（由于该点在双目相机中的视差较小，使其深度无法得到可靠估计），并对后者使用逆深度参数估计[6]。他们经验性地指出，当特征点的深度小于双目相机基线长度的40倍时，特征点可以被可靠地三角化。本文工作中我们延续了这种用不同方式处理近特征点和远特征点的策略，这部分内容将在3.1节中进行解释。

大多数现代的双目SLAM系统是基于关键帧[7]和局部BA优化来实现可伸缩性（scalability）。Strasdat等人的工作[8]在关键帧窗内采用BA优化（点-位姿约束），在关键帧窗外采用位姿图优化（位姿-位姿优化）。通过限制窗的大小，该方法可以实现恒定的时间复杂度，但无法保证全局一致性。Mei等人提出了RSLAM方法[9]，其使用了地标和位姿的相对位置表示法，并在限制时间复杂度的条件下，在激活区域内采用相对BA。RSLAM可以实现回环的闭合，这可以扩展回环两端的激活区域，但并不能增强全局一致性。最近Pire等人提出的S-PTAM [10]采用了局部BA，但它缺少闭合大回环的功能。与这些方法类似，我们也在局部关键帧集合中采用了BA，因此该方法的复杂度不受地图尺寸影响，我们可以在大场景中实施该方法。但是，我们的目标是建立一个全局一致的地图。与RSLAM类似，当闭合一个回环时，我们的系统会首先将两端对齐，因此追踪模块可以使用旧地图继续定位，之后采用位姿图优化来最小化回环中的累积漂移，再之后进行全局BA。

最近Engel等人提出的双目LSD-SLAM [11]是一种半稠密的方法，它最小化图像梯度较大区域的光度误差。该方法希望在不依赖特征的条件下，在运动模糊或纹理较弱的环境下获得更好的鲁棒性。但是，作为直接法，该方法的性能会由于未建模因素而显著下降，例如卷帘快门或非朗伯反射。

2.2 RGB-D SLAM

Newcombe等人提出的KinectFusion [4]是最早也是最著名的RGB-D SLAM系统之一。该方法将传感器得到的所有深度数据融合至一个稠密的体积模型，并使用ICP来追踪相机位姿。由于该系统使用体积表示形式并且缺少回环检测，它只能应用于小规模的工作空间。Whelan等人提出的Kintinuous [12]使用了一个滚动循环缓冲区，并且包括了一个使用位置识别和位姿图优化的回环检测模块，从而能够在大规模场景运行。

Endres等人提出的RGB-D SLAM [13]可能是最早流行的开源系统。它是一个基于特征的系统，它的前端通过特征匹配和ICP来计算帧间的运动，它的后端使用位姿图优化，其回环检测约束条件由启发式搜索得到。与之相似，Kerl等人提出的DVO-SLAM [14]的后端也采用位姿图优化，其中关键帧之间的约束是由一个最小化光度和深度误差的视觉里程计计算得到。同时，DVO-SLAM在以往所有帧中启发式地搜索回环的候选者，而不依赖于位置识别。

最近Whelan等人提出的ElasticFusion [15]建立了环境的surfel地图，这是一种忽略位姿，而以地图为核心的方法，它采用对地图进行非刚性变形的方式来实现回环闭合，而不是采用位姿图优化的方法。该系统细节重建和定位精度是非常优秀的，但是由于地图中面元数量所带来的复杂度，目前它仍局限于建立房间大小的地图。

我们的ORB-SLAM2系统使用了一种Strasdat等人提出的方法[8]，该方法使用深度信息来为图像中提取的特征合成立体坐标。通过这种方法，我们的系统可以处理来自双目或者RGB-D的输入。与上述所有方法不同的是，我们方法的后端基于BA，并且能够得到一个全局一致的稀疏重建。因此我们的方法是轻

量级的，可以在标准CPU上运行。我们的目标是实现长期并且全局一致的定位，而不是进行具有更多细节的稠密重建。但是，我们的方法也可以通过精度很高的关键帧位姿，进行深度图融合来实时地对局部环境进行准确重建，或者在全局BA后对所有关键帧的深度图进行处理从而得到整个场景的精准三维模型。

3 ORB-SLAM2

基于双目和RGB-D相机的ORB-SLAM2是建立在我们的基于特征的单目ORB-SLAM [1]的基础上的。为读者方便，我们在这里总结一下单目ORB-SLAM的基本组成部分。总体概述了该系统基于双目和RGB-D相机的ORB-SLAM2是建立在我们的基于特征的单目ORB-SLAM的基础上的。为读者方便，我们在这里总结一下单目ORB-SLAM的基本组成部分。图 2 ORB-SLAM2由三个主要的并行线程组成：追踪、局部建图和回环检测。在回环检测后会执行第四个线程，进行全局BA。追踪线程会对双目和RGB-D输入进行预处理，从而使得系统其它部分可以独立于输入传感器运行。虽然这张图没有展示，但ORB-SLAM2也可以基于单目输入运行。图 2总体概览了该系统。该系统具有三个主要的并行线程：1）追踪线程是用来在每一帧中定位相机的位置，通过匹配特征和局部地图并且进行运动BA（motion-only BA）最小化重投影误差；2）局部建图线程是用来管理和优化局部地图；3）回环检测线程是用来检测大回环，并通过执行位姿图优化来修正累积误差。该线程在位姿图优化后会启动第四个线程来执行全局BA，计算最优的结构和运动结果。

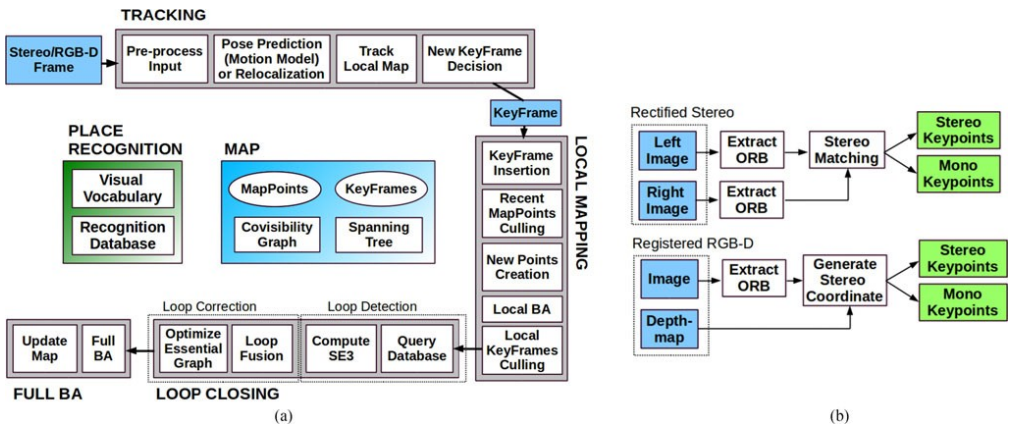

图 2 ORB-SLAM2由三个主要的并行线程组成：追踪、局部建图和回环检测。在回环检测后会执行第四个线程，进行全局BA。追踪线程会对双目和RGB-D输入进行预处理，从而使得系统其它部分可以独立于输入传感器运行。虽然这张图没有展示，但ORB-SLAM2也可以基于单目输入运行[1]。(a)系统的线程和模块 (b)输入的预处理

该系统嵌入了一个基于DBoW2 [16]的位置识别模块，在发生追踪失败（例如：碰撞）或者在建好图的场景中重新初始化时用来进行重定位，以及用来进行回环检测。该系统维护了一个关联可见地图（convisible map）[8]，此地图将每两个观察到相同地图点的关键帧连接到一起；同时该系统也维护了一个最小生成树，它连接了所有的关键帧。这种地图结构实现了对于关键帧局部窗的检索，因此追踪线程和局部建图线程可以局部地运行，使其可以在大场景中工作；同时该结构在回环闭合进行位姿图优化时，也可作为优化的图结构。

该系统在追踪、建图和位置识别任务中，都使用相同的ORB特征[17]。这些特征对于旋转和尺度变化具有很好的鲁棒性，同时对于相机的自动增益、自动曝光和光线变化也具有不变性。另外，提取和匹配ORB特征的速度很快，使其可以实时运行，并且在词袋模型位置识别任务上表现出良好的查准率/查重率（precision/recall）[18]。

在本节的余下部分中，我们会展示如何使用双目/深度信息，以及系统的哪些部分将会被影响。关于该系统每一部分更详尽的描述，请参考我们的单目ORB-SLAM论文[1]。

3.1 单目、近处立体和远处立体关键点

作为一种基于特征的方法，ORB-SLAM2会对输入进行预处理，在显著关键点位置提取特征，如图 2b所示。接下来，输入的图片会被丢弃，系统的全部运算会基于这些特征，因此无论是双目还是RGB-D输入，本系统都可以工作。我们的系统会处理单目和双目的关键点，这些点又会被分为近处点和远处点。

立体（双目）关键点通过三维坐标 $x_s = (u_L, v_L, u_R)$xs=(uL,vL,uR) 来定义，(u_L, v_L)(uL,vL) 是关键点在左图的坐标，u_RuR 是关键点在右图的水平坐标。对于双目相机，我们在左右两张图片中同时提取ORB特征。对于左图中的每个ORB特征，我们在右图中搜索一个相应的匹配。对于校正后的双目图像来说，极线是水平的，所以上述任务可以很高效地完成。之后我们根据左图ORB特征坐标和右图相匹配的特征水平坐标来生成立体关键点。对于RGB-D相机，正如Strasdat等人所言[8]，我们在RGB图像上提取ORB特征，对于每个坐标为 (u_L, v_L)(uL,vL) 的特征，我们根据它的深度值d计算出一个虚拟的右图坐标：

$$u_R = u_L - (f_x b)/d$$
uR=uL−(fxb)/d

其中 $f_x$fx 是水平焦距；bb 是结果光投影机和红外相机之间的基线长度，在Kinect和Asus Xtion相机中我们将其大概设定为8厘米。深度传感器的不确定性由虚拟的右坐标表示。通过这种方式，系统余下部分可以以相同的方法处理来自双目或者RGB-D输入的特征。

正如文献[5]所述，如果一个立体关键点的深度值小于双目/RGB-D的基线长度的40倍，则认为它是近处点，否则认为它是远处点。近处关键点可以被安全地三角化，因为它的深度可被精确估计，且提供了尺度、平移和旋转的信息。另一方面，远处关键点虽然提供了精确的旋转信息，但不能提供精确的尺度和平移信息。所以当远处关键点在多个视图中存在时，我们才对其进行三角化。

单目关键点通过左图中的二维坐标 $x_m = (u_L, v_L)$xm=(uL,vL) 定义，若ORB特征的双目匹配失效或者RGB-D相机无法得到其有效深度值，则采用此方式。这些点只会在多视图时进行三角化，且不会提供尺度信息，但它们可用于旋转和平移估计。

3.2 系统启动

使用双目或者RGB-D相机的最主要的好处之一是，我们可以直接获得单帧图像的深度信息，不用像在单目SLAM中一样需要使用特定的SFM（structure from motion）初始化。在系统启动时，我们将第一帧设为关键帧，将其位姿设置为初始位姿，并且根据所有的立体关键点来建立一个初始地图。

3.3 单目和双目约束下的光束优化法（BA）

我们的系统在追踪线程中使用BA来优化相机位姿（纯运动BA），在局部建图线程中优化关键帧和点的局部窗（局部BA），在回环检测后优化所有的关键帧和点（全局BA）。我们使用g2o [19]中的实现的Levenberg-Marquardt方法来进行优化。

纯运动BA（motion-only BA）优化相机的旋转矩阵 $\mathbf{R} \in SO(3)$R∈SO(3) 和位置 $\mathbf{t} \in \mathbb{R}^3$t∈R3，最小化相匹配的世界坐标系下的三维点 $\mathbf{X}^i \in \mathbb{R}^3$Xi∈R3 和关键点 $\mathbf{x}^i_{(\cdot)}$x(·)i 之间的重投影误差（单目点

$\mathbf{x}_{\mathrm{m}}^{i} \in \mathbb{R}^{2}$xmi∈R2 或者双目点 $\mathbf{x}_{\mathrm{s}}^{i} \in \mathbb{R}^{3}$xsi∈R3，对于所有匹配对 $i \in \mathcal{X}$i∈X）：

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

{R,t}=argminR,t∑i∈Xρ(‖x(·)i−π(·)(RXi+t)‖Σ2)

其中ρρ是鲁棒Huber代价函数，∑∑是关键点尺度的协方差矩阵。其中投影函数$\pi_{(.)}$π(·)，单目投影函数π_mπm，校正双目投影函数π_sπs如下定义：

$$\pi_{\mathrm{m}}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix}, \pi_{\mathrm{s}}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \\ f_x \frac{X-b}{Z} + c_x \end{bmatrix}$$

πm([XYZ])=[fxXZ+cxfyYZ+cy],πs([XYZ])=[fxXZ+cxfyYZ+cyfxX−bZ+cx]

其中$\left(f_x, f_y\right)$(fx,fy)是焦距，$\left(c_x, c_y\right)$(cx,cy)是光心点，bb是基线长度，这些值都通过标定得到。

全局BA是局部BA的一种特殊情况，在全局BA中，除了初始关键帧因用来消除计算自由度而被固定之外，所有关键帧和地图点都会被优化。

局部BA对一个关联可见的关键帧$\mathcal{K}_L$KL集合和这些关键帧中所有可见的点$\mathcal{P}_L$PL。所有不在$\mathcal{K}_L$KL中，但也观测到$\mathcal{P}_L$PL中的点的其它关键帧$\mathcal{K}_F$KF，也会参与到代价函数的计算中，但是不会被优化。我们将$\mathcal{P}_L$PL中的点与关键帧k中的关键点之间的匹配对的集合定义为$\mathcal{X}_k$Xk，将优化问题进行如下定义：

$$\left\{\mathbf{X}^{i}, \mathbf{R}_{l}, \mathbf{t}_{l} | i \in \mathcal{P}_{L}, l \in \mathcal{K}_{L}\right\} = \underset{\mathbf{X}^{i},\mathbf{R}_{l},\mathbf{t}_{l}}{\operatorname{argmin}} \sum_{k \in \mathcal{K}_{L} \cup \mathcal{K}_{F}} \sum_{j \in \mathcal{X}_{k}} \rho\left(E_{kj}\right)$$

$$E_{kj} = \left\|\mathbf{x}_{(\cdot)}^{j} - \pi_{(\cdot)}\left(\mathbf{R}_{k}\mathbf{X}^{j} + \mathbf{t}_{k}\right)\right\|_{\Sigma}^{2}$$

{Xi,Rl,tl|i∈PL,l∈KL}=argminXi,Rl,tl∑k∈KL∪KF∑j∈Xkρ(Ekj)Ekj=‖x(·)j−π(·)(RkXj+tk)‖Σ2

全局BA是局部BA的一种特殊情况，在全局BA中，除了初始关键帧因用来消除计算自由度而被固定之外，所有关键帧和地图点都会被优化。

3.4 回环检测和全局BA

回环检测分两步进行：第一步是检测和确认回环，第二步是通过优化位姿图来修正回环。相较于单目ORB-SLAM可能会发生尺度漂移[20]，双目/深度信息会使尺度变得可以观测，所以几何验证和位姿图优化不再需要处理尺度漂移；同时它是基于刚体变换，而不是基于相似性。

在ORB-SLAM2中，我们在位姿图优化后，采用全局BA优化来得到最优解。这个优化过程可能开销会很大，所以我们将其放在一个独立的线程中，从而使得系统可以持续建立地图、检测回环。但这样的话，将BA输出与当前地图状态之间进行融合就会产生困难。如果在优化运行的同时发现了新的回环，那么我们就停止优化，转而去闭合回环，这将再次启动全局BA优化。当全局BA完成时，就需要将全局BA优化更新后的关键帧和点的集合，与在优化过程中插入的未更新的关键帧和点，进行融合。这通过将更新的关键帧的修正（未优化位姿至优化位姿的变换）沿生成树传递至未更新的关键帧来完成。未更新的点依据它们的参考帧的修正来进行变换。

3.5 关键帧的插入

ORB-SLAM2沿用了单目ORB-SLAM中介绍的策略：频繁插入关键帧，之后再剔除冗余的关键帧。近处立体点和远处立体点之间的区别使我们在插入关键帧时可以引入一个新的条件，当环境中存在很大一块场景远离双目传感器时，这是非常重要的，如图 3所示。在这样的环境中，我们需要有足够多的的近处点来精确地估计平移量，因此，当追踪的近处点数目低于τ_t τt并且此帧能够创建至少τ_c τc个新的近处立体点时，系统就会插将此帧作为一个新的关键帧插入。根据经验，根据经验，在我们的实验中，$\tau_t = 100$τt=100 $\tau_c = 70$τc=70 的效果较好。

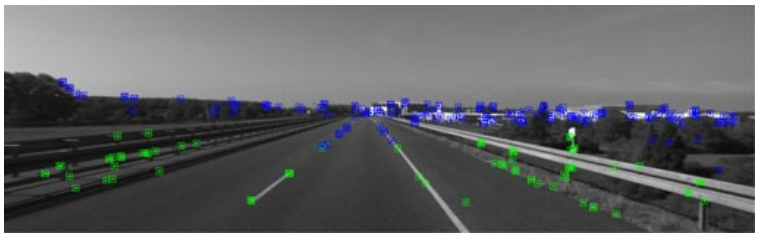

图 3 KITTI 01中的追踪点。绿色点表示深度值小于40倍双目基线长度的点，蓝色点表示更远的点。在这种视频序列中，需要频繁地插入关键帧，使近处点的总数满足精确估计平移量的要求。远处点可用于估计旋转量，但对于估计平移量和尺度帮助不大。

3.6 定位模式

我们的系统包括了一个定位模式，只要环境没有大的变化，该模式就可以在建图良好的区域中有效地进行轻量级的长期定位。在此模式中，局部建图线程和回环检测线程是停用的，如果需要的话，相机会持续通过追踪线程的重定位功能进行定位。在此模式中，追踪线程会使用视觉里程计中的匹配对，并将其与地图点进行匹配。视觉里程计中的匹配对是指当前帧中的ORB特征和之前帧根据双目/深度信息所创建的三维点之间的匹配对。这些匹配对使得定位功能在未建图区域更加鲁棒，但是会产生累积漂移。与地图点的匹配保证了在已建图区域的定位是零漂移的。此模式会在附带的视频中进行展示说明。

4 评价

我们评价了ORB-SLAM2在三个常用数据集中的表现，并将其与其它当前最好的SLAM系统进行比较。我们使用的其它SLAM系统的评价结果都是来自原作者的文章中的标准评价指标。我们在一台Intel核心i7-4790、16GB内存的台式电脑上运行ORB-SLAM2。为了防止多线程系统的不确定性对评价结果产生影响，我们对每个数据序列运行5次，最终展示轨迹估计精度结果的中值。我们的开源实现中包括了相机标定，以及如何在这些这些数据集中运行ORB-SLAM2系统的用法说明。

4.1 KITTI 数据集

KITTI数据集[2]包括了城市和高速公路环境中车辆采集的双目视频序列。其中双目传感器的基线长度为54厘米，工作频率为10Hz，矫正后的分辨率为1240×376像素。其中视频序列00、02、05、06、07和09包括回环。我们的ORB-SLAM2系统可以检测到所有的回环，并且可以在之后重新使用地图（除了视频序列09，因为回环只在序列快结束时的很少几帧出现）。表 1展示了在11段训练视频序列中的评价结果，这些序列都含有公开的对应真实值。据我们所知，只有双目SLAM算法可以在上述所有视频序列中都运行得到细致的结果，所以我们将我们的方法与当前最好的双目LSD-SLAM [11]进行比较。我们使用两个不同的指标，绝对平移均方根误差（absolute translation RMSE）t_abs [3]、相对平移平均误差t_rel、相对旋转平均误差r_rel [2]。我们的系统在大多数视频序列中都表现得比双目LSD-SLAM更好，通常情况下，相对误差不到1%。图 3所示视频序列是训练集中唯一一个高速公路上的视频序列，它的平

移误差稍稍差一些。在该序列中，平移量更难估计，因为高速和低帧率导致被追踪的点非常少。但是，估计的旋转量很准确，每100米的误差仅为0.21度，因为有很多可以被长期追踪的远处点。图 4 展示了一些估计轨迹的例子。

表 1 KITTI数据集中各SLAM系统精度的比较

	ORB-SLAM2 (stereo)			Stereo LSD-SLAM		
Sequence	t_{rel} (%)	r_{rel} (deg/100m)	t_{abs} (m)	t_{rel} (%)	r_{rel} (deg/100m)	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

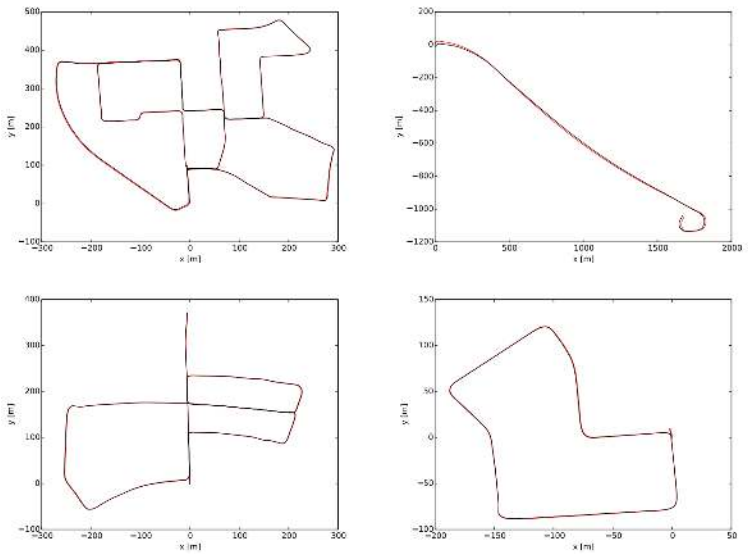

图 4 KITTI 00、01、05、07中的估计轨迹（黑色）和真实值（红色）：与[1]中单目ORB-SLAM结果相比，本文提出的双目版本可以处理单目系统处理不了的视频序列01。通过图 3可以看出，在该高速公路序列中，近处点只能持续出现在很少几帧中。双目版本可以根据一个双目关键帧来创建地图点，而不需要双目版本只需一个双目关键帧来创建地图点，而不需要像单目一样使用延迟初始化（在两个关键帧中寻找匹配对），这使其该序列中不易追踪丢失。另外，双目系统可以估计地图和轨迹的尺度，而不会产生尺度漂移。

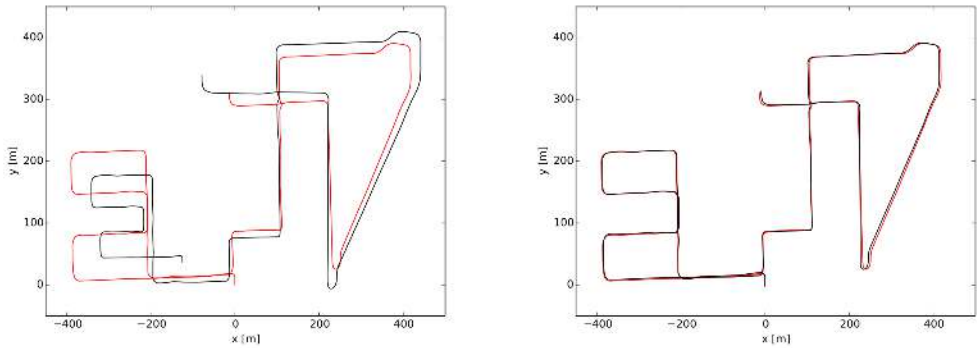

图 5 KITTI 08中的估计轨迹（黑色）和真实值（红色）。左图：单目ORB-SLAM [1]；右图：ORB-SLAM2（双目）。单目ORBSLAM在此序列中产生了严重的尺度漂移，特别是在转弯处。相比较而言，本文提出的双目版本可以很好地估计轨迹和地图的准确尺度，而不产生尺度漂移。

4.2 EuRoC数据集

最近的EuRoc数据集[21]包含11段双目视频序列，其由一个微型无人机（MAV）在两间不同的屋子和一个很大的工业环境中拍摄。该双目传感器的基线长度为11厘米，其以20Hz的频率输出WVGA图像。这些视频序列根据微型无人机的速度、光照条件、场景的纹理被分成简单、中等、困难三等。在所有的视频序列中，微型无人机都会回到曾经到过的地方，这样ORB-SLAM2系统就可以在需要的时候重用地图或闭合回环。表 2展示了ORB-SLAM2与双目LSD-SLAM [11]在该数据集所有视频序列中的绝对平移量均方根误差。ORB-SLAM2的定位精度达到厘米级，比双目LSD-SLAM精度更高。由于严重的运动模糊，我们的追踪模块在V2_03_difficult序列的某些部分会丢失。如文献[22]所述，该序列可以通过添加惯性测量单元（IMU）信息进行处理。图 6展示了一些计算得到的轨迹估计与真实值之间的比较。

表 2 EuRoC数据集：平移量均方根误差（RMSE）的结果比较

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	-

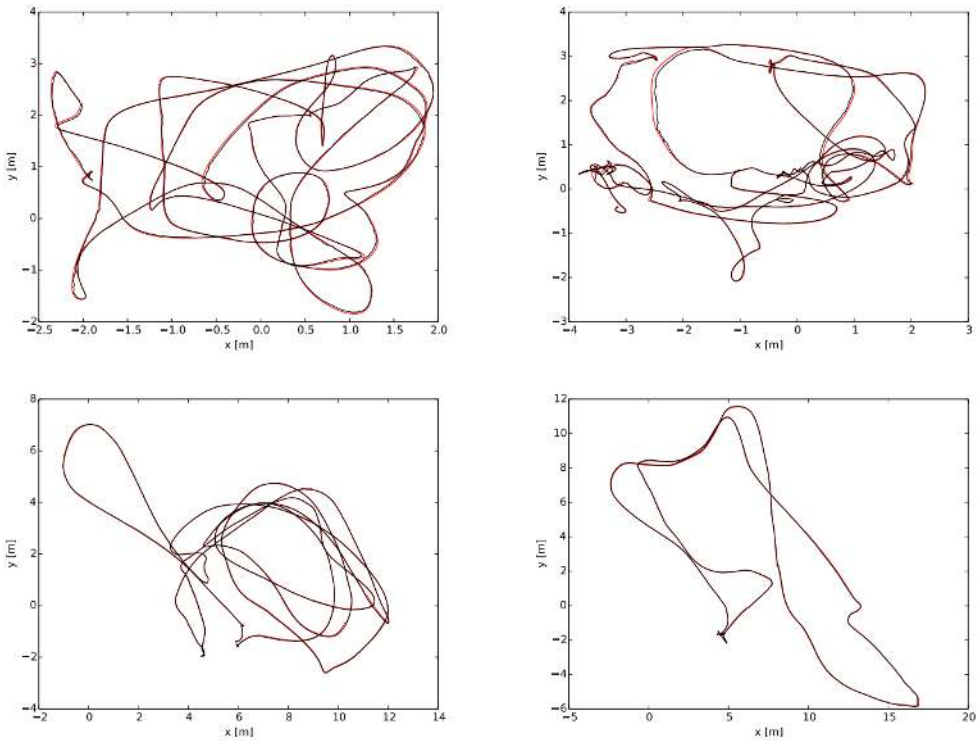

图 6 EuRoC V1 02 medium, V2 02 medium, MH 03 medium and MH 05 difficult序列中的估计轨迹（黑色）和真实值（红色）。

4.3 TUM RGB-D数据集

TUM RGB-D数据集[3]包括RGB-D传感器采集的室内视频序列，其被分组在几个分类中，用于在不同纹理、光照和结构条件下评价物体重建和SLAM/里程计方法的效果。我们展示了几个最常用于RGB-D方法

评价的视频序列的评价结果。在表 3中我们将我们的方法与当前最好的方法（ElasticFusion [15]、Kintinuous [12]、DVO-SLAM [14]、RGB-D SLAM [13]）的精度进行了对比，我们的方法是唯一一个基于BA的方法，且其效果在大多数序列中都好于其它方法。我们已经观察到文献[1]中的RGB-D SLAM，freiburg 2序列的深度图有4%的尺度偏差，这可能是错误的标定导致的，所以我们在运行中对此进行了补偿。这能部分解释我们取得很好结果的原因。图 7展示了四个视频序列中将传感器深度图根据计算所得的关键帧位姿反向投影得到的点云结果。实验结果显示，我们的方法很好地重建了桌子和海报的轮廓线，这说明我们的方法具有很高的定位精度。

表 3 TUM RGB-D数据集：平移量均方根误差的结果比较

Sequence	ORB-SLAM2 (RGB-D)	Elastic-Fusion	Kintinuous	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	-

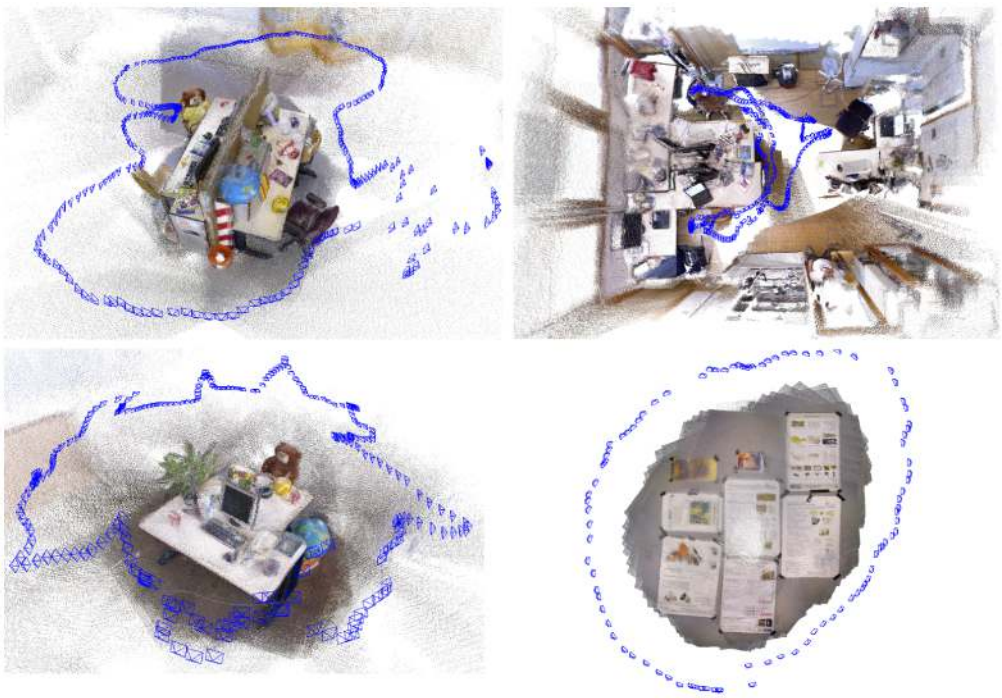

图 7 TUM RGB-D fr3_office、fr1_room、fr2_desk、fr3_nst序列的根据估计关键帧位姿和传感器深度图得到的稠密点云重建。

4.4 运算耗时结果（Timing results）

为了更全面地对我们提出的系统进行评价，我们在表 4中展示了三段视频序列中不同图像分辨率和传感器条件下的计算耗时。我们展示了每个线程的平均耗时及其两倍标准差范围。因为这些序列包括一个单回环，所以全局BA和回环检测线程的一些任务只需执行一次，因此我们只展示了单个时间测量。每个序列中的平均每帧追踪耗时都低于相机帧率的倒数，这意味着我们的系统可以实时运行。因为在双目左图和右图中的ORB特征提取是并行的，所以可以看出，在V2_02序列的双目WVGA图像中提取1000个ORB特征的速度与在fr3_office的单VGA图像通道中提取相同数量特征的耗时是差不多的。

表 4 每个线程的计算耗时/毫秒（平均值±2倍标准差）

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

我们展示了回环中的关键帧数量，用于给回环检测的耗时做参照。虽然KITTI 07序列中包含更多的关键帧，但fr3_office室内序列的关联可见图（convisibility graph）更加稠密，因此回环融合、位姿图优化和全局BA的开销更大。关联可见图（convisibility graph）越稠密，局部地图就含有越多关机帧和点，因此局部地图追踪和局部BA的开销会更大。

5 结论

我们提出了一个完整地基于单目、双目或RGB-D传感器的SLAM系统，其可以在标准CPU上实时实现重定位、回环检测和重用地图。我们的重点在于建立全局一致的地图，用于在实验中所介绍的大规模环境中进行长期定位。我们提出的包含重定位功能的定位模式，是一种可以在已知环境中进行鲁棒的、零漂移的、轻量级的定位方法。该模式可适用于特定应用，例如在环境建图良好的虚拟现实中追踪使用者的视点。

与当前最好的SLAM系统的对比，ORB在大多数情况下达到了最高的精度。在KITTI视觉里程计基准测试中，ORB-SLAM2是目前最好的双目SLAM解决方案。很重要的是，与最近流行的双目视觉里程计方法相比，ORB-SLAM2实现了在已建图区域的零漂移定位。

令人惊讶的是，我们的RGB-D结果显示，如果需要精度最高的相机定位，那么BA的表现比直接法或ICP更好，另外它的计算量也更小，不需要依赖GPU就可以实时运行。

我们发布了系统源代码、例子和使用说明，因此其他研究者可以很方便地使用本系统。据我们所知，ORB-SLAM2是第一个在单目、双目或RGB-D输入下都可以工作的开源视觉SLAM系统。另外，我们的源代码包括了一个增强现实的应用例子，其使用单目相机，用于展示我们的解决方案的可能性。

未来的研究方向可能包括：非重叠多幅相机、鱼眼相机、全景相机支持，大规模稠密融合、协作建图以及增强运动模糊的鲁棒性。