翻译 ORB-SLAM: a Verdatile and Accurate Monocular SLAM System

ORB-SLAM：一种通用、精确的单目 SLAM 系统

摘要

本文介绍了 ORB-SLAM，一种基于特征的单目 SLAM 系统。它能够在室内外、大小场景下实时运行。该系统对剧烈运动杂波具有鲁棒性，允许宽基线闭环及重定位，包含全自动初始化。基于近年的优秀算法，我们从零设计了一个新的系统。使用相同的特征（feature）来完成 SLAM 所有的功能：追踪（tracking）、建图（mapping）、重定位（relocation）、以及闭环（loop closing）。一种用于选择点（point）和重建关键帧（keyframe）的适者生存的策略带来了良好的鲁棒性，并生成精简、可追踪的地图，（这一地图）只有在场景内容发生变化时才会增长，允许终身运行。我们在最流行的数据集中取了 27 组图像序列，进行了详尽的评估。相比于其他的先进 SLAM 方案，ORB-SLAM 达到了前所未有的性能。为了社区的利益，我们将源代码开放。

关键词

终身地图（Lifelong Mapping），定位（Localization），单目视觉（Monocular Vision），识别（Recognition），SLAM

I. 简介

光束平差法（Bundle Adjustment，BA）在给出强大的匹配网络及合适的初值的条件下，为相机定位和稀疏几何重建提供精确估计[1][2]，由此广为人知。长久以来，这一方案被认为是实时应用例如视觉同步定位与建图（Visual Simultaneous Localization and Mapping，Visual SLAM）所难以负担的。视觉 SLAM 的目标是在重建环境的同时估计相机轨迹。现在我们得知，为了在不限制计算成本的情况下得到精确结果，一个实时 SLAM 算法必须为 BA 提供以下条件:

- 与选中的帧（关键帧，keyframe）的子集的场景特征（地图点，map point）相对应的观测值。
- 随关键帧的数目而增长的复杂度，其（关键帧）的选择要避免不必要的冗余。
- 一个强大的关键帧和点的网络配置会产生精确的结果，也就是说，一个良好的有着较大视差和足够闭环（loop closure）匹配的关键帧观测点的传播集。
- 用于非线性最优化的关键帧的位姿及点的位置的初值估计。
- 一个检测中的局部地图的最优化工作的重点在于实现其可扩展性。
- 执行快速全局最优化的能力（例如，位姿图，pose graph）以求实时闭环。

首个基于 BA 的实时应用是 Mouragon 等人的视觉里程计[3]，随后是 Klein 和 Murray 的开创性的 SLAM 工作[4]，称作 PTAM（Parallel Tracking and Mapping）。该算法仅限于小尺度运行，提供了简单而高效的关键帧选取、特征匹配、点的三角化、每一帧的相机定位及追

踪失败后的重定位的方案。不幸的是，一些因素严重地限制了它的应用：缺乏闭环和足够的阻塞处理，重定位视角的低不变性，以及地图启动时所需的人工干预。

在本课题中，我们基于 PTAM 的主体思想，G´alvez-L´opez 和 Tard´os 的场景重建工作[5]，Strasdat 等人的尺度感知闭环，以及使用大尺度运行的公共可见信息[7][8]，来从零开始设计 ORB-SLAM，一种新的单目 SLAM 系统。其主要贡献为：

- 在所有的功能中使用相同的特征：追踪，建图，重定位及闭环。这样使得我们的系统更加高效、简捷、可靠。我们使用 ORB 特征[9]，无需 GPU 即可实时运行，提供良好的视角不变性和光照不变性。
- 大环境下的实时运行。由于使用了共视图（covisibility graph），追踪和建图专注于局部可视区域，与全局地图的大小无关。
- 基于位姿图（pose graph）最优化的实时闭环，我们称之为本质图（essential graph）。它由一个生成树构建，由系统、闭环连接、共视图的一些强壮的边（edge）来维护。
- 视角、光照的不变性下，相机的实时重定位。它允许了追踪失败的恢复，并增强了地图的复用性。
- 一种新的基于模型选择的自动鲁棒初始化过程，允许创建平面和非平面场景的初始地图。
- 一种地图点和关键帧选择的适者生存方案，这一方案在生成关键帧时很大方，但是在剔除时却有很多限制。这一策略使得追踪更具有鲁棒性，并由于冗余关键帧被抛弃从而加强了终身运行能力。

我们在常见的公共数据集中，对系统在室内外环境的运行进行了广泛评估，包括手提式设备、汽车及机器人的图像序列。引人注目的是，我们做到了比目前先进的直接法[10]更好的相机定位精度。直接法即直接从像素灰度而非特征的重投影误差进行优化的方法。我们在章节IX-B 中对导致特征点法精度优于直接法的可能的原因进行了探讨。

本文中提出的闭环和重定位方案基于我们以往的工作[11]，即在文献[12]中提出的该系统的一个初始版本。在本文中，我们添加了初始化方案，本质图，以及对涉及的所有方案进行了完善。我们也详细介绍了构建的所有模块，并进行了详尽的实验验证。

据我们所知，这是最完整、最可靠的单目 SLAM 方案。为了社区的利益，我们将源代码公开。演示视频和源代码可以在我们的主页上找到。

II. 相关工作

（这段先不翻）

III. 系统概述

III-A. 特征选择

我们的系统的主要设计思想之一是在建图、追踪、场景识别、基于帧率的重定位以及回环检测等功能中，使用相同的特征。这使得我们的系统更加高效，且避免了像从前的工作[6]，[7]那样，需要从邻近的 SLAM 特征中插入识别特征的深度。我们要求每幅图像提取特征所需时间远远小于 33ms，那么就排除了常用的 SIFT(~300ms)[19]、SURF(~300ms)[18]或最近的 A-KAZE(~100ms)[35]。为了获得综合位置识别能力，我们需要旋转不变性，那么就排除了

BRIEF[16]和 LDB[36]。

我们选择了 ORB[9]，它是带有方向的多尺度 FAST 角点（corner），具有 256 位相关描述子。它在具有良好的视角不变性的同时，有着极快的计算和匹配速度。这样就可以在宽基线对其进行匹配，提高了 BA 的精度。我们已经在文献[11]中展示了 ORB 在位置识别方面的良好性能。虽然我们当前的实现使用了 ORB，但是所建议的技术并不局限于这些特征。

III-B. 三个线程：追踪，局部地图和闭环

我们的系统如图 1 所示，包含三个并行的线程：追踪、局部地图和闭环。追踪负责在每一帧定位相机，并决定在何时插入新的关键帧。我们首先执行一个对于前一帧的初始特征匹配，并采用纯运动 BA 来优化位姿。若追踪失败(例如，由于遮挡或突然移动)，则使用位置识别模块执行全局重新定位。一旦相机姿态和特征匹配有了初始估计，则利用关键帧的共视图来恢复局部可见图（local visible map），该共视图由系统维护，见图 2(a)和图 2(b)。随后由重投影搜索局部地图点的匹配，并利用所有的匹配再次优化相机位姿。最终，追踪线程决定是否插入一个新的关键帧。追踪的所有步骤在章节 V 中有详细解释。章节IV中提出了创建初始地图的新过程。

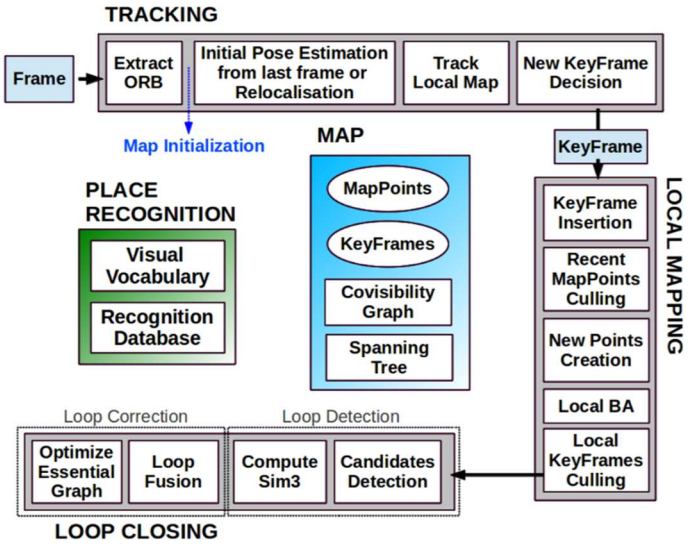
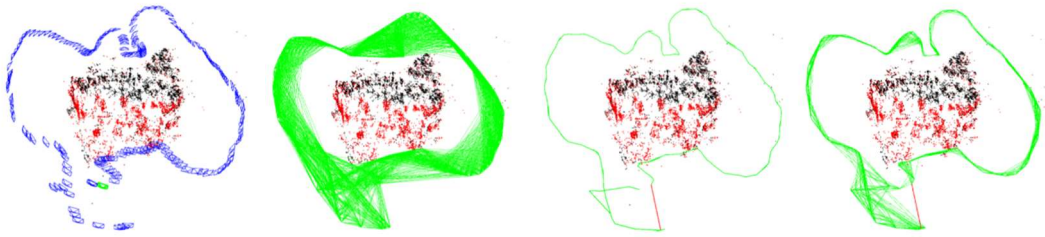


图 1 ORB-SLAM 的系统概述，显示了追踪、局部建图和闭环线程的所有执行步骤。
并给出了位置识别模块以及地图的主要组成部分。

局部地图处理新的关键帧，并执行局部 BA，来实现相机位姿环境下的最优重建。在共视图中相连的关键帧中搜索新关键帧中未匹配 ORB 的新的匹配，来三角化新的点。在创建后的一段时间内，基于追踪过程中收集到的信息，为了只保留高质量的点，使用紧急点剔除策略。局部地图还负责剔除冗余关键帧。我们将在章节VI中对局部建图的所有步骤进行详细解释。



(a) 关键帧（蓝色），当前相机（绿色），地图点（黑色，红色），当前局部地图点（红色）
(b) 共视图 (c) 生成树（绿色），闭环（红色）(d) 本质图

图 2 TUM RGB-D Benchmark 数据集中 fr3_long_office_household 的重建和图

闭环在每一个新的关键帧中进行回环搜索。如果检测到一个回环，我们计算一个相似变换，描述出回环中的漂移累积。随后将回环的两端对齐，并将重复的点融合。最终执行一个相似约束条件[6]下的位姿图最优化，以实现全局一致性。主要的创新点在于，我们通过本质图（essential graph）进行最优化。本质图是共视图的一个稀疏子图，在章节 III-D 中有所解释。回环检测及校正的步骤在章节VII中有详细解释。

我们使用在 g2o 中实现的 L-M 算法进行所有的优化。在附录中，我们阐述了每次优化所涉及的误差项、代价函数和变量。

III-C. 地图点，关键帧及其选取

每个地图点 p_i 存储:

- 它在世界坐标系下的 3D 位置 $\boldsymbol{X}_{w,i}$。
- 视角方向 $\boldsymbol{n}_i$，它是所有视角方向的平均单位向量(将该点与观察它的关键帧的光心连接起来的光线)。
- 一个典型的 ORB 描述子 $\boldsymbol{D}_i$，它是所有相关的 ORB 描述子中，与观察该点的关键帧中所有其它相关描述子的 Hamming 距离最小者。
- 根据 ORB 特征的尺度不变性极限，能够观察到该点的最大距离 $d_{\max}$ 及最小距离 $d_{\min}$。

每个关键帧 K_i 存储:

- 相机位姿 $\boldsymbol{T}_{iw}$，它是从世界坐标系到相机坐标系的刚体变换。
- 相机内参，包括焦距和主点。
- 所有在图像中提取出的 ORB 特征，不论其是否与地图点相关联。若提供了畸变模型，其坐标为校正后的坐标。

地图点和关键帧是由一个宽容的策略创建的，而随后有非常紧急的筛选机制，来负责检测冗余关键帧、以及误匹配或不可追踪的地图点。这样允许在探索过程中灵活地扩展地图，从而增强了在困难条件下(如旋转、快速移动)追踪的鲁棒性，同时其大小受到不断重复访问相同环境的限制，即，终身运行。此外，相比于 PTAM，我们的地图只有少量的异常值，这是以包含更少的点为代价的。地图点和关键帧的剔除过程分别在章节VI-B 和章节VI-E 中进行了说明。

III-D. 共视图和本质图

关键帧间的公共可视信息在系统的许多任务中都是十分有用的。并如文献[7]一般，以无向加权图的形式表示。每个节点（node）都是一个关键帧，如果两个关键帧共享同一地图点的观测值（至少 15 个），那么它们之间的存在边（edge），边的权重θ就是共同地图点的数目。

为了校正一个回环，我们执行了一个位姿图最优化[6]，将闭环误差分配到整个图中。为了避免将共视图提供的所有的边都包含在内（可能会非常稠密），我们决定构建一个保留所有节点（关键帧），但只有少量的边的本质图（essential graph），其仍保留有生成精确结果的强大网络。系统从初始关键帧开始，逐步构建一个生成树，提供共视图的一个具有最小边数的连通子图。当插入一个关键帧时，它被包含在与共享最多观测点的关键帧相连的树中；而当一个关键帧被剔除策略删除时，系统更新了受该关键帧影响的链接。本质图包含了生成树，由共视图中高共视性的边组成的子集（$\theta_{\min}=100$），以及闭环边，形成了一个相机的强大网络。图 2 展示了一个由共视图、生成树及相关的本质图构成的样例。如章节Ⅷ-E 中实验所示，当执行位姿图最优化时，其解的精度很高，以至于额外的完全 BA 优化几乎没有提升结果精度。本质图的高效性和$\theta_{\min}$的影响在章节Ⅷ-E 的结尾有所显示。

III-E. 词袋位置识别

系统嵌入了一个基于 DBoW2[5]的词袋位置识别模块，来实现回环检测和重定位。视觉词汇只是描述子空间的离散化，被称作视觉词表。词表由从大量图像中提取的 ORB 描述子离线创建。若图片足够通用，如我们之前的工作[11]所展示的那样，相同的词表可以用于不同的环境，以获得良好的性能。系统增量式地构建一个包含逆序目录的数据库，其为词表中的每个视觉词汇存储了它被看到的关键帧，因此查询这一数据库可以十分高效。该数据库同样可以在关键帧被剔除程序删除时更新。

由于关键帧之间存在视觉上的重叠，当查询数据库时，高分关键帧并不是唯一的。最初的 DBoW2 考虑了这一重叠，将时间上足够接近的图片相加。这里有一个限制，即不包括那些观察相同位置但插入时间不同的关键帧。相反，我们将那些在共视图中相连的图分组。此外，我们的数据库返回所有分数大于最高分 75%的关键帧匹配。

文献[5]中还报告了代表特征匹配的词袋的另一重好处。当我们希望计算两个 ORB 特征集间的对应关系时，我们可以将暴力匹配限制在属于词汇树的仅一定层次内（在 6 层中选取 2 层），以加快搜索速度。在三角化新的点、回环检测及重定位时，我们使用这一方法对匹配进行搜索。我们还通过方向一致性检测，对对应点对进行了细化，具体可见文献[11]，抛弃异常值以确保所有的对应点都有一致的旋转。

IV. 地图自动初始化

地图初始化的目标是计算两帧间的相对位姿，，以三角化一组初始地图点。这一方法应独立于场景（平面或普通），且不需要人工干涉即可选取一个好的两视图配置，即，有着较大视差的配置。我们打算并行地计算两个几何模型，一个假设平面场景下的单应性矩阵，一个假设非平面场景下的基础矩阵（fundamental matrix）。随后我们使用启发式的方法来选择

模型，以使用合适的方法来尝试恢复相关的位姿。我们的方法仅在确定双视角配置是安全的情况下进行初始化；检测低视差情况及众所周知的双重平面模糊[27]，以避免得出一个有缺陷的地图。我们的算法步骤为：

1) 检测初始对应点

在当前帧 F_c 提取 ORB 特征(仅在最优尺度下)，并在参考帧 F_r 中寻找匹配 $\boldsymbol{x}_c \leftrightarrow \boldsymbol{x}_r$。如果没有发现足够的匹配，则重置参考帧。

2) 并行计算两个模型

并行计算单应性矩阵 $\boldsymbol{H}_{cr}$ 及基础矩阵 $\boldsymbol{F}_{cr}$：

$$\begin{aligned} \boldsymbol{x}_c &= \boldsymbol{H}_{cr}\boldsymbol{x}_r \\ \boldsymbol{x}_c^T \boldsymbol{F}_{cr}\boldsymbol{x}_r &= 0 \end{aligned} \tag{1}$$

在文献[2]中，分别解释了标准 DLT 和八点法算法，使用 RANSAC 方法计算。为了使两个模型的计算过程一致，两个模型迭代次数的前缀相同，且连同每次迭代使用的点的数目都相同（单应矩阵 4 个点，基础矩阵 8 个点)。在每一次迭代过程中，我们为每一个模型 M （单应矩阵为 H，基础矩阵为 F）计算一个评分 S_M：

$$\begin{aligned} S_M &= \sum_i \{\rho_M[d^2_{cr,M}(\boldsymbol{x}_c^i, \boldsymbol{x}_r^i)] + \rho_M[d^2_{rc,M}(\boldsymbol{x}_c^i, \boldsymbol{x}_r^i)]\} \\ \rho_M\left(d^2\right) &= \begin{cases} \Gamma - d^2 & \text{if } d^2 < T_M \\ 0 & \text{if } d^2 \geq T_M \end{cases} \end{aligned} \tag{2}$$

其中 d^2_{cr} 和 d^2_{rc} 是两帧间对称的传递误差[2]。T_M 是基于 χ^2 检测值的 95%（$T_H = 5.99$，$T_F = 3.84$，假设在量测误差中有 1 像素的标准偏差）确定的排除异常数据的阈值。Γ 被定义等于 T_H，因而对于内点中相同的 d，两个模型的得分相同，进而使得过程保持一致。

我们保持单应矩阵和基础矩阵的得分最高。若无法代入任何模型(没有足够的内点)，我们从 step 1 重启算法流程。

3) 模型选择

若场景为平面、接近平面或低视差，就可以用单应矩阵解释。若一个基础矩阵同样能够被找到，但是问题中并没有很好的约束[2]，那么从基础矩阵恢复运动的任何尝试都会得到错误的结果。(此时）我们应当选择单应矩阵，因为重建方案将会从一个平面开始正确初始化；否则将会检测到低视差情形，并拒绝初始化。另一方面，一个有着足够视差的非平面场景仅能由基础矩阵解释。但是若匹配点处于同一平面，或有着较低视差（距离很远)，也可以通过单应矩阵来解释匹配的子集。在这种情况下，我们应当选择基础矩阵。我们用如下具有鲁棒性的公式来计算

$$R_H = \frac{S_H}{S_H + S_F} \tag{3}$$

若 $R_H > 0.45$，则充分表明处于平面、低视差情形，选择单应矩阵；另一方面，若 $R_H \leq 0.45$，我们选择基础矩阵。

4) 运动和运动结构重构

一旦选定一种模型，我们恢复相关的运动假设。在单应矩阵的情况下，我们使用 Faugeras 等的方法[23]来恢复 8 个运动假设。这一方法通过测试来选取有效的解。然而，如果视差很小，这一测试将会失败，因为点很容易在相机前后移动，这样会导致选取一个错误的结果。我们打算直接三角化 8 个解，检测是否有一个解，它对于在具有视差的两视图看到的大多数点而言，在两相机前面且重投影误差很低。若没有明确的胜出的解，我们不会（继续）初始化并从 step 1 开始执行。这一消除解的歧义的技术使得我们的初始化在低视差、双重模糊配置之下具有鲁棒性，而且可以认作时我们的方案在鲁棒性上的关键。

在基础矩阵的情况下，我们使用标定矩阵 $\boldsymbol{K}$，将其转换为本质矩阵

$$\boldsymbol{E}_{rc} = \boldsymbol{K}^T \boldsymbol{F}_{rc} \boldsymbol{K} \tag{4}$$

然后根据文献[2]中的 SVD 方法，恢复 4 种运动假设。我们三角化四个解，如同单应矩阵时所做的那样，选取进行重建。

5) 光束平差

最终我们执行一个全局 BA，详细说明见附录，来优化初始重建。

如图 3 所示，NewCollege 机器人序列是一个初始化很有挑战性的室外实例。能够看出，PTAM 和 LSD-SLAM 如何将所有的点初始化到同一平面上；而我们的方法会等到有足够的视差后，通过基础矩阵，正确地进行初始化。

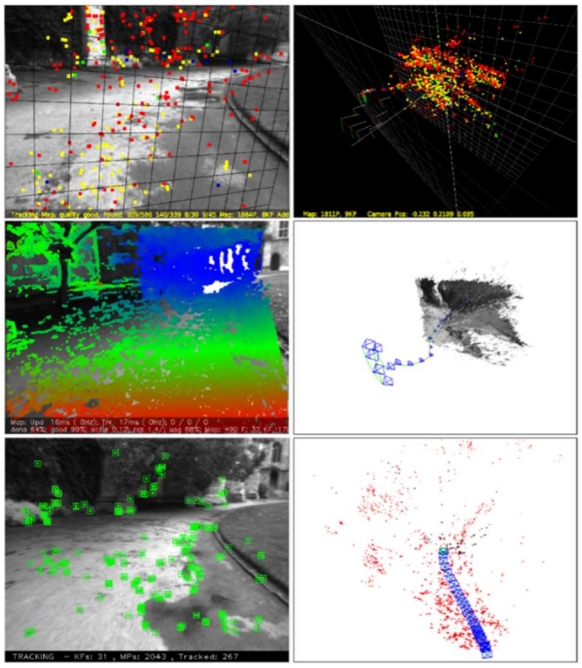


图 3 上方：PTAM；中部：LSD-SLAM；下方：ORB-SLAM；在 NewCollege 序列中[39]初始化一段时间之后。PTAM 和 LSD-SLAM 初始化一个错误的平面解，而我们的方法在检测到足够的视差时，通过基本矩阵自动初始化。根据手动选择的关键帧，PTAM 也能够很好地初始化。

V. 追踪

在这一章节中，我们描述追踪线程的步骤，这一线程在相机的每一帧中执行。相机位姿的最优化（在一些步骤中提及）由纯运动 BA 构成，在附录中介绍。

V-A. ORB 提取

我们将尺度因子定为 1.2，在 8 个尺度中提取 FAST 角点。对于图像分辨率从 512×384 到 752×480 像素，我们发现提取 1000 个角点较为合适。而对于 KITTI 数据集[40]中 1241×376 像素的分辨率，我们提取了 2000 个角点。为确保其为均匀分布，我们将每个尺度按网格进行拆分，尝试在每个网格中至少提取 5 个角点。随后我们在每个网格中检测角点，若未发现足够的角点，则改变其检测阈值。若某些网格中没有角点（无纹理或低对比度），那么也需要修改每个网格中保留角点数目的阈值。随后在保留的 FAST 角点上计算其方向和描述子。ORB 描述子用于所有的特征匹配，与 PTAM 中通过图像区块关联性进行搜索形成对比。

V-B. 由前一帧做初始位姿估计

若前一帧追踪成功，我们使用匀速运动模型，来预测相机位姿并对前一帧中观察到的地图点执行一个引导搜索。若未能发现足够的匹配（即，明确地违反了运动模型），我们对上一帧它们位置周围的地图点使用一个更加广泛的搜索。然后围绕所得的对应关系，进行位姿优化。

V-C. 经由全局重定位的初始位姿估计

若追踪失败，我们将这一帧转化为词袋，并且查询候选关键帧的识别数据库，以进行全局重定位。如章节 III-E 中阐述，我们计算 ORB 与每个关键帧中地图点之间的对应关系。随后，我们对每一个关键帧执行 RANSAC 迭代，尝试使用 PnP 算法[41]找出相机位姿。我们优化这一位姿，并执行一个导向搜索以寻找对于候选关键帧中地图点的更多匹配。最终，再次优化相机位姿，并且若有足够的内点支持，追踪线程会继续。

V-D. 局部地图追踪

一旦我们获得了关于相机位姿的估计，和一组初始特征匹配，我们就可以将地图投影到帧中，以寻找更多的地图点对应。为了限制其在大型地图中的复杂性，我们仅投影一个局部地图。这一局部地图包含了关键帧集 $\mathcal{K}_1$，它们与当前帧共享地图点；以及集合 $\mathcal{K}_2$，他们与共视图中的关键帧集 $\mathcal{K}_1$ 有邻点（neighbors）。局部地图中也有一个参考关键帧 $K_{ref}\in\mathcal{K}_1$，它与当前帧共享大部分地图点。现在在 $\mathcal{K}_1$ 和 $\mathcal{K}_2$ 中看到的每个地图点都通过如下步骤在当前帧中搜索：

1) 在当前帧中计算地图点的映射 $\boldsymbol{x}$。若其位于图像边缘之外，则丢弃之。
2) 计算当前视角光线 $\boldsymbol{v}$ 与地图点平均视角方向 $\boldsymbol{n}$ 之间的夹角。若 $\boldsymbol{v}\cdot\boldsymbol{n}<\cos 60°$，则丢弃之。
3) 计算地图点到相机中心的距离 d。若 d 不在地图点的尺度不变区间，即 $d\notin[d_{\min},d_{\max}]$，则丢弃之。
4) 通过比率 $d/d_{\min}$ 计算帧中的尺度。
5) 将地图点中代表性描述子 $\boldsymbol{D}$，与（当前）帧中预测尺度内、与 $\boldsymbol{x}$ 接近、且仍未匹配的 ORB 特征进行对比，并将地图点与其最优匹配进行关联。

最终相机位姿通过（当前）帧中所有的地图点进行最优化。

V-E. 新关键帧判别标准

最后一步时决定当前帧是否派生为新的关键帧。由于局部地图中具有剔除冗余关键帧的机制，我们会尝试尽快插入关键帧，因为这使得相机在有挑战性的运动中更加具有鲁棒性，典型的如旋转。插入一个新的关键帧必须具备以下条件：

1) 从上一次全局重定位起，至少经过 20 帧。
2) 局部建图处于空闲中，或者从插入上一个关键帧起，已超过 20 帧。
3) 当前帧追踪超过 50 个点。
4) 当前帧追踪少于 K_{ref} 点的 90%。

与 PTAM 等不同，我们没有使用到其他关键帧的一个距离标准，而是施加了一个最小视觉变化（条件 4）。条件 1 确保了一个良好的重定位，而条件 3 确保了一个良好的追踪。若在局部地图繁忙（条件 2 的第二部分）时插入关键帧，则会发送一个信号以暂停局部 BA，这样就可以尽快处理新的关键帧。

VI. 局部建图

这一章节中我们将介绍由每个新的关键帧 K_i 进行局部建图的步骤。

VI-A. 插入关键帧

首先我们更新共视图，为 K_i 添加一个新的节点（node），且更新那些与其他关键帧共享地图点而得到的边（edge）。随后我们更新生成树，连接 K_i 和与之有着最多共同点的关键帧。再之后，我们计算表现关键帧的词袋，这样将会在数据关联方面有所帮助，以三角化新的点。

VI-B. 近期地图点筛选

为了在地图中保留，地图点必须在创建后最初的三帧内通过一个严格的测试，以确保它们是可追踪的，且没有错误的三角化，即，没有虚假的数据关联。一个点必须满足以下两个条件：

1) 在该点预计可视的帧中，追踪线程必须在超过 25%的帧中发现该点。
2) 若有超过一个关键帧完成地图点的创建过程，它必须至少要被三个关键帧观察到。

一旦一个点通过了这一测试，它只有在被少于三个关键帧观测到的情况下，才与可能被删除。这可能在关键帧被剔除或局部 BA 抛弃异常观测时发生。这一策略使得我们的地图只包含少量的异常点。

VI-C. 构造新地图点

对来自共视图中相连的关键帧集 $\mathcal{K}_c$ 中的 ORB 特征进行三角化，可构造新地图点。对于 K_i 中每一个未匹配的特征，我们查找它和其他关键帧中未匹配点的匹配。这一匹配按章节 III-E 中阐述的那样进行，并抛弃那些不满足对极约束的匹配。将 ORB 点对三角化后，就会获得新的点，检查其在每个相机的深度、视差、重投影误差、及尺度一致性。起初，一个地图点是由两个关键帧观测到的，但是也可在其他关键帧中匹配；因此它被投影到其余相连的关键帧中，并如同章节 V-D 中所阐述的那样，搜索对应点。

VI-D. 局部光束平差

局部 BA 优化了当前处理的关键帧 K_i，共视图中所有与之相连的关键帧 $\mathcal{K}_c$，以及从这些关键帧中看到的所有地图点。所有能够看到这些都但是并不与当前处理关键帧相连的其他关键帧同样包含于优化中，但是保持固定。被标记为异常的观测值在优化的中间和结束时被抛弃。更多有关优化的细节详见附录。

VI-E. 局部关键帧筛选

为了维持一个简洁的重建，局部地图尝试检测冗余关键帧并删除它们。这是有益的，因为 BA 的复杂度增长相当于关键帧数量的立方；但是也由于它支持同一场景下的终身操作，其关键帧数量不会无限制增长，除非场景中的视觉内容发生改变。$\mathcal{K}_c$ 中部分关键帧的 90% 地图点在同一或过更精细尺度下，被其他至少三个关键帧看到，我们抛弃这部分的所有关键帧。尺度条件保证了地图点维持它最精确地被观测的关键帧。这一策略受 Tan 等的工作[24]影响，，其中关键帧在变化检测后被抛弃。

VII. 闭环

闭环线程获取局部建图处理的最后一个关键帧 K_i，并尝试检测及封闭回环。其步骤如下所述。

VII-A. 候选回环检测

首先我们计算 K_i 的词袋向量与它在共视图中所有的邻近图像（$\theta_{\min}=30$）之间的相似度，并保留最低分值 $s_{\min}$。随后我们查询识别数据库，并抛弃那些分数低于 $s_{\min}$ 的关键帧。这是一个与 DBoW2 中标准化得分以获得鲁棒性相类似的操作，它是由之前的图像计算的，但是这里我们使用共视图的信息。此外，所有与 K_i 直接相连的那些关键帧在结果中被抛弃。为获取一个候选回环，我们必须连续检测三个一致的候选回环（在共视图中相连的关键帧）。若有多处与 K_i 相似的地方，也可以有多个候选回环。

VII-B. 计算相似变换

在单目 SLAM 中，地图有七个自由度可以漂移，其中三个平移，三个旋转和一个比例因子[6]。因此，为了封闭回环，我们需要计算从当前关键帧 K_i 到到回环关键帧 K_l 的相似变换，它告诉了我们回环中的误差累积。这一相似计算同时也将作为回环的几何验证来服务。

我们首先计算与当前关键帧中地图点相关联的 ORB 特征与回环候选关键帧之间的对应关系，按章节 III-E 中所解释的那样执行。此时，对于每个候选回环，我们都有 3D-3D 的对应关系。我们对每个候选回环执行 RANSAC 迭代，使用 Horn 在文献[42]中的方案来尝试寻找一个相似变换。如果我们发现了一个相似变换 $\boldsymbol{S}_{il}$ 具有足够的内点，我们优化它（见附录），并执行一个导向搜索，以寻找更多的对应项。我们再度优化它，若 $\boldsymbol{S}_{il}$ 有足够的内点支持，那么对于 K_l 的回环将被采纳。

VII-C. 回环融合

回环校正的第一步是融合重复的对应点，并在共视图中插入附加闭环的新边。首先通过

相似变换 $\boldsymbol{S}_{il}$ 对当前帧位姿 $\boldsymbol{T}_{iw}$ 进行校正，这一校正应用于所有与 K_i 相邻的关键帧，将变换结果连接起来，使得回环两端能够对齐。将回环关键帧及其相邻关键帧看到的所有地图点投影到 K_i，并在投影的一个狭小区域内搜索它的近邻和匹配，如章节Ⅴ-D 所做的那般。将所有匹配的地图点和 $\boldsymbol{S}_{il}$ 计算过程中的内点进行融合。融合中包含的所有关键帧将更新它们在共视图中的边缘，有效地创建与封闭回环相关的边。

VII-D. 本质图最优化

为了有效地封闭回环，我们针对本质图执行了一个位姿图最优化，如章节Ⅲ-D 中介绍的那样，沿图分配闭环误差。该优化通过相似变换来修正尺度漂移[6]。误差项和代价函数在附录中有详细说明。优化之后，每个地图点根据观测到它的关键帧的校正进行变换。

VIII. 实验

（这段先不翻）

IX. 结论及讨论

（这段先不翻）

附录：非线性优化

- 光束平差法（BA）[1]:
  将地图点的 3D 位置 $\boldsymbol{X}_{w,j} \in \mathbb{R}^3$ 和关键帧位姿 $\boldsymbol{T}_{iw} \in \mathrm{SE}(3)$（其中 w 代表世界参考系）最优化以求其关于匹配的关键点 $\boldsymbol{x}_{i,j} \in \mathbb{R}^2$ 的重投影误差最小值。关键帧 i 中地图点 j 的观测值的误差项为:

$$\boldsymbol{e}_{i,j} = \boldsymbol{x}_{i,j} - \pi_i(\boldsymbol{T}_{iw}, \boldsymbol{X}_{w,j}) \tag{5}$$

  其中 π_i 为投影函数:

$$\pi_i(\boldsymbol{T}_{iw}, \boldsymbol{X}_{w,j}) = \begin{bmatrix} f_{i,u}\frac{x_{i,j}}{z_{i,j}} + c_{i,u} \\ f_{i,v}\frac{y_{i,j}}{z_{i,j}} + c_{i,v} \end{bmatrix} \tag{6}$$

$$\begin{bmatrix} x_{i,j} & y_{i,j} & z_{i,j} \end{bmatrix}^T = \boldsymbol{R}_{iw}\boldsymbol{X}_{w,j} + \boldsymbol{t}_{iw}$$

  其中 $\boldsymbol{R}_{iw} \in \mathrm{SO}(3)$ 和 $\boldsymbol{t}_{iw} \in \mathbb{R}^3$ 分别表示 $\boldsymbol{T}_{iw}$ 中的旋转和平移部分；$(f_{i,u}, f_{i,v})$ 和 $(c_{i,u}, c_{i,v})$ 为焦距长度和关于相机 i 的主点。求其最小值的代价函数为:

$$C = \sum_{i,j} \rho_h(\boldsymbol{e}_{i,j}^T \boldsymbol{\Omega}_{i,j}^{-1} \boldsymbol{e}_{i,j}) \tag{7}$$

  其中 ρ_h 为 Huber 鲁棒代价函数，$\boldsymbol{\Omega}_{i,j} = \sigma^2 \boldsymbol{I}_{2\times 2}$ 为关于被检测关键点的尺度的协方差矩阵。在全局 BA（用于章节Ⅳ中解释的地图初始化和章节Ⅷ-E 中的实验）的情况下，我们优化所有的点和关键帧，除了第 1 帧作为原点保持固定。在局部 BA（见章节Ⅵ-D）中，包含在局部区域内所有的点都要进行优化，同时关键帧集的一个子集固定。在位姿优化，或纯运动 BA（见章节Ⅴ）中，所有的点都是固定的，只有相机的位姿进行优化。

- 基于$\mathrm{Sim}(3)$约束[6]的位姿图优化：（$\mathrm{Sim}(3)$表示 3 维相似变换）

  给定一个带有二值边（binary edge）的位姿图（见章节Ⅶ-D），我们将一条边中的误差项设定为：

$$\boldsymbol{e}_{i,j} = \log_{\mathrm{Sim}(3)}(\boldsymbol{S}_{ij}\boldsymbol{S}_{jw}\boldsymbol{S}_{iw}^{-1}) \tag{8}$$

  其中$\boldsymbol{S}_{ij}$是由进行位姿图优化并将尺度因子置 1 之前的$\mathrm{SE}(3)$位姿计算得出的两个关键帧之间的相对$\mathrm{Sim}(3)$变换。在回环闭合边（loop closure edge）的情况下，这一相对变换由 Horn 的方案[42]进行计算。该$\log_{\mathrm{Sim}(3)}$[48]函数为到正切空间的变换，因而误差项为一个$\mathbb{R}^7$内的 7 维向量。其目标是通过最小化代价函数来优化$\mathrm{Sim}(3)$关键帧位姿：

$$C = \sum_{i,j}(\boldsymbol{e}_{i,j}^T\boldsymbol{\Lambda}_{i,j}\boldsymbol{e}_{i,j}) \tag{9}$$

  其中$\boldsymbol{\Lambda}_{i,j}$为边的信息矩阵，其中，如同文献[48]中的那样，我们将其设置为恒等式。我们将闭环关键帧固定，以固定 7 测量自由度。即使这一方法是全局 BA 的一个粗略的近似，然而我们在章节Ⅷ-E 中以实验证实，它相较于 BA，在速度和收敛性方面有显著的提升。

- 相对$\mathrm{Sim}(3)$的最优化

  给定一个从关键帧 1 到关键帧 2 的n个匹配对组成的集合，其中匹配对为$i \Rightarrow j$（关键点及其相对的 3D 地图点）的匹配。我们希望通过最小化每幅图像中的重投影误差来优化其相对$\mathrm{Sim}(3)$变换$\boldsymbol{S}_{12}$（见章节Ⅶ-B）：

$$\begin{aligned}\boldsymbol{e}_1 &= \boldsymbol{x}_{1,i} - \pi_1(\boldsymbol{S}_{12}, \boldsymbol{X}_{2,j})\\ \boldsymbol{e}_2 &= \boldsymbol{x}_{2,j} - \pi_2(\boldsymbol{S}_{12}^{-1}, \boldsymbol{X}_{1,i})\end{aligned} \tag{10}$$

  而用来最小化的代价函数为

$$C = \sum_n[\rho_h(\boldsymbol{e}_1^T\boldsymbol{\Omega}_{1,i}^{-1}\boldsymbol{e}_1) + \rho_h(\boldsymbol{e}_2^T\boldsymbol{\Omega}_{2,j}^{-1}\boldsymbol{e}_2)] \tag{11}$$

  其中$\boldsymbol{\Omega}_{1,i}$和$\boldsymbol{\Omega}_{2,j}$为相对于检测到的图像 1 和图像 2 中关键点尺度的协方差矩阵。在这一优化中，点是固定的。