ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM

摘要

ORB-SLAM3，第一个能够用单目、双目和RGB-D，针孔和鱼眼镜头模型进行视觉、视觉惯性和多地图SLAM的系统。

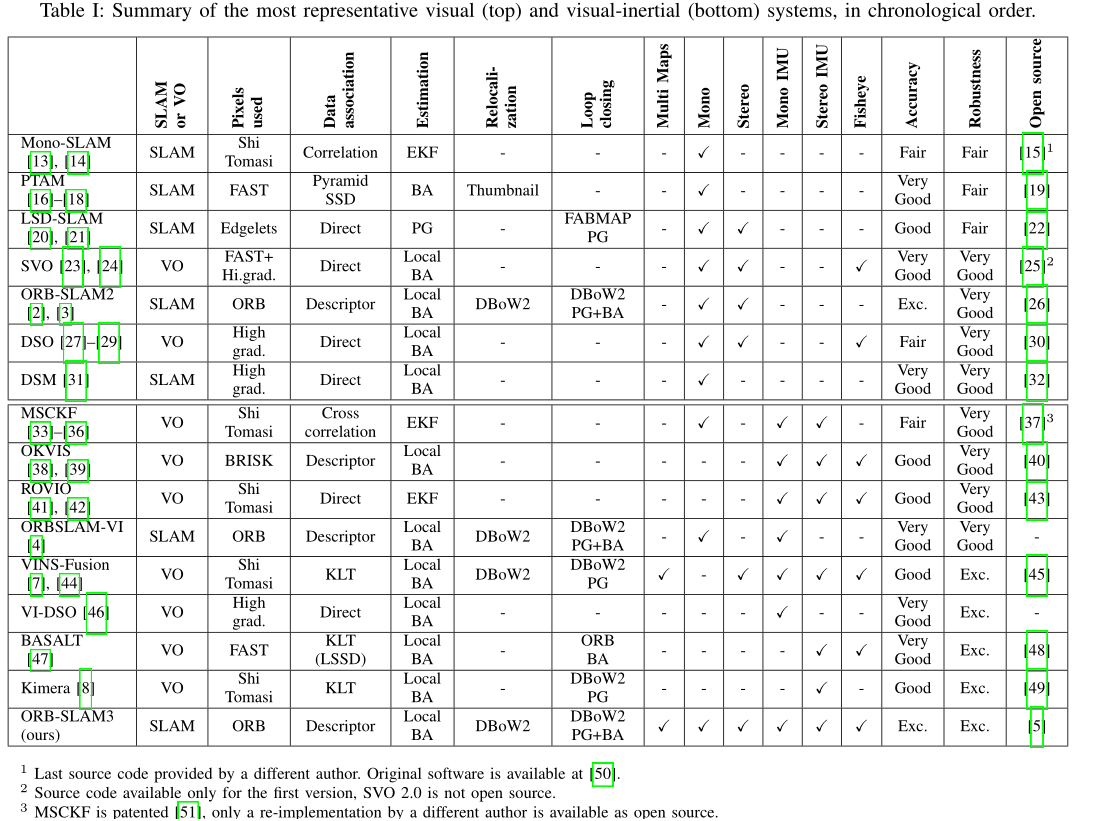
第一个创新点是基于特征的紧密集成视觉惯性SLAM系统，完全依赖于最大后验概率（MAP）估计，甚至在IMU初始化阶段也是如此。在小型和大型，室内和室外环境中都能稳定地运行，并且比以前的方法精确两到十倍。

第二个创新点是一个多地图系统，依赖于一种新的位置识别方法和改进的召回。能够在长时间的不良视觉信息下生存：定位失败情况下，当重新访问已经建图的区域时，会启动一个新的地图来和以前的地图无缝合并。与只使用最后几秒信息的视觉里程计系统相比，ORBSLAM3是第一个能够在所有算法阶段重用所有先前信息的系统。能BA优化共视关键帧，即使它们在时间上相隔甚远，或者它们来自以前的建图会话。

我们的实验表明，在所有传感器配置中，ORBSLAM3与文献中可用的最佳系统一样健壮，并且更精确。值得注意的是，我们的双目惯性SLAM在EuRoC无人机上的平均精度为3.6cm,在TUM-VI数据集(AR/VR场景的一个典型的代表),快速手持移动时的平均精度为9mm。为了社区的利益，我们公开了源代码。

相关工作

表1概述了最具代表性的视觉和视觉惯性系统，显示了用于估计和数据关联的主要技术。表中的定性准确度和稳健性评级基于第七节中给出的结果，以及[2]中报告的PTAM、LSD-SLAM和ORB-SLAM之间的比较。



1. Visual SLAM

在MonoSLAM[13]、[14]、[52]中，首先使用扩展卡尔曼滤波器（EKF）和ShiTomasi点解决了单目SLAM问题，这些点在后续图像中通过相关引导搜索进行跟踪。中期数据关联得到了显著改善，使用的技术保证了所使用的特征匹配的一致性，实现了手持视觉SLAM[53]，[54]。

相比之下，基于关键帧的方法仅使用几个选定帧来估计地图，而丢弃来自中间帧的信息。这允许以关键帧速率执行成本更高但更精确的BA优化。最具代表性的系统是PTAM[16]，该系统将相机分为跟踪和建图分为两个线程。在计算成本相同的情况下，基于关键帧的技术比滤波更精确[55]，成为视觉SLAM和VO的黄金标准。在[56]中使用滑动窗口BA实现了大规模单目SLAM，在[57]中使用双窗口优化和共视图实现了大规模单目SLAM。

基于这些思想，ORB-SLAM[2]，[3]使用ORB特性，其描述子提供短期和中期数据关联，构建共视图以限制跟踪和建图的复杂性，并使用词包库DBoW2[9]执行回环和重定位，实现长期数据关联。到目前为止，它是唯一一个集成了三种数据关联的视觉SLAM系统，我们相信这是其卓越准确性的关键。在这项工作中，我们使用新的Atlas系统提高了其在纯视觉SLAM中的鲁棒性，该系统在跟踪丢失时启动新的地图，并使用具有改进召回率的新位置识别方法提高了其在回环场景中的准确性。

直接法使用图像中的像素强度，通过最小化光度误差来估计运动和空间结构。LSD-SLAM[20]能够使用高梯度像素构建大比例的半稠密地图（位姿图），其精度低于PTAM和ORB-SLAM[2]。混合系统SVO[23]，[24]从帧到帧提取FAST特征，使用直接方法跟踪特征和具有非零强度梯度的像素，并使用重投影误差优化相机轨迹和结构，只执行短期的数据关联，限制了准确性。直接稀疏里程计DSO[27]能够在特征点探测不佳的情况下计算出精确的相机位姿，增强了低纹理区域或模糊图像的鲁棒性。它引入了局部光度BA，同时优化了7个最近关键帧的窗口和点的逆深度，扩展到了双目[29]、使用特征和DBoW2[59][60]的回环闭合，以及视觉惯性里程计[46]。直接稀疏建图DSM[31]在直接方法中引入了地图重用的思想，说明了中期数据关联的重要性。在所有情况下，缺乏短期、中期和长期数据关联的整合导致精度低于我们的提案（见第七节）。

1. Visual-Inertial SLAM

视觉传感器和惯性传感器的结合提供了对不良纹理、运动模糊和遮挡的鲁棒性，并且在单目系统的情况下，使尺度可观察。

紧耦合方法的研究可以追溯到MSCKF[33]，特征边缘化避免了特征数量上的EKF二次方代价计算，系统在[34]中得到完善，在[35]、[36]中扩展到双目。第一个基于关键帧和光束平差的紧耦合视觉里程计系统是OKVIS[38]，[39]，能够使用单目和双目视觉。ROVIO[41]，[42]使用直接数据关联向EFK提供了光度误差。

ORB-SLAM-VI[4]首次提出了一种视觉惯性SLAM系统，能够重用三种数据关联的地图，并在基于IMU预积分的精确局部视觉惯性BA中使用它们[61]，[62]。然而它的IMU初始化技术太慢，耗时15秒，影响了系统的鲁棒性和准确性。[63]，[64]中提出了更快的初始化技术，是闭合形式的解决方案，联合了尺度恢复、重力、加速度计偏差、初始速度和特征深度。关键的是，忽略了IMU的噪声特性，将空间点的3D误差降到最小，而不是重投影误差，这是基于特征的计算机视觉的黄金标准。我们之前的工作[64]表明，这会导致大量不可预测的错误。

VINS-Mono[7]是一个非常精确和健壮的单目惯性里程计系统，使用DBoW2和4-DoF位姿图优化和地图合并来检测回环。特征跟踪使用Lucas Kanade tracker，比匹配描述子稍微健壮一些。在VINS-Fusion [44]中扩展到双目和双目惯性。

最近，VI-DSO[66]将DSO扩展到视觉惯性里程计，提出了一种光束平差，将惯性观测和选定的高梯度像素的光度误差结合起来，从而获得非常好的精度。成功地利用了高梯度像素的信息，增强了对纹理较差的场景区域的鲁棒性。它们的初始化方法依赖于视觉惯性BA，需要20-30秒才能收敛到1%的尺度误差内。

最近的BASALT [47]是一种双目惯性里程计系统，它从视觉惯性里程计中提取非线性因素，用于BA，并闭合匹配ORB特征的环路，实现非常好到极好的精度。Kimera[8]是一种新颖的度量语义建图系统，度量部分包括双目惯性里程计加上DBoW2回环检测和位姿图优化，与VINS-Fusion的精度相似。

我们的工作建立在ORB-SLAM-VI上，并将其扩展到双目惯性SLAM。提出了一种新的基于最大后验概率（MAP）估计的快速初始化方法，考虑了视觉和惯性传感器的不确定性，并在2秒内估计到5%，15秒内收敛到1%的尺度误差。上面讨论的所有其他系统都是视觉惯性里程计方法，其中一些通过回环进行扩展，缺乏使用中期数据关联的能力。我们相信，这一点，加上我们快速而精确的初始化，是我们的系统获得更高精度的关键，即使在没有闭环也是如此。

1. Multi-Map SLAM

[65]在滤波方法中首次提出了通过地图创建和融合的方式，增强对探索过程中跟踪丢失的鲁棒性的想法。第一个基于关键帧的多地图系统之一[66]，但地图初始化是手动的，无法合并或关联不同的子地图。多地图作为协同制图系统的一个组成部分，由多个地图代理和一个只接收信息的中央服务器[67]或C2TAM[68]中的双向信息流组成。MOARSLAM[69]提出了一种用于协作多设备SLAM的健壮的无状态客户-服务架构，但主要关注的是软件架构，而没有准确的结果。

最近，CCM-SLAM[70]，[71]提出基于ORB-SLAM的分布式多地图系统，用于具有双向信息流的多架无人机。重点是克服有限带宽和分布式处理的挑战，我们的重点是准确性和稳健性，在EuRoC数据集上取得更好的结果。ORB-SLAMM[72]还提出了ORB-SLAM2的多层地图扩展，但将子地图作为独立的实体，我们执行无缝地图合并，从而构建更精确的全局地图。

对比VINS-Mono[7]，由于我们的方法能够使用中期数据关联，ORB-SLAM3在EuRoc数据集上的单目惯性单会话操作的精度是VINS-Mono的2.6倍。建图系统也建立在DBoW2上，但提出了一种新的更高召回率的位置识别技术，并使用局部BA进行更详细和更准确的地图合并，在EuRoC上多会话操作中，其精度比VINS-Mono提高了3.2倍。

**SYSTEMOVERVIEW**

ORB-SLAM3建立在ORB-SLAM2[3]和ORBSLAM-VI[4]的基础上。它是一个完整的多地图和多会话系统，能够在纯视觉环境下工作或者视觉惯性模式，使用针孔和鱼眼相机模型。图1显示了主要系统组件，这些组件与ORB-SLAM2的组件类似，具有一些显著的新颖性，下面对其进行总结：

ATLAS地图集由一系列断开的地图组成。活动地图中，跟踪线程定位传入的帧，并由局部建图线程不断优化和增加新的关键帧。其他地图称为非活动地图。该系统建立了一个独特的DBoW2关键帧数据库，用于重新定位、回环和地图合并。

TRACKING线程处理传感器信息，实时计算当前帧相对于活动地图的位姿，减少已匹配的地图特征的重投影误差，还决定当前帧是否成为关键帧。在视觉惯性模式下，在优化中加入惯性残差来估计物体速度和惯性测量单元偏差。当跟踪丢失时，跟踪线程将尝试重新定位所有地图集中的当前帧位姿，恢复跟踪，并在需要时转换换为活动地图。否则活动地图将被存储为非活动地图，从头开始初始化。

LOCAL MAPPING线程将关键帧和点添加到活动地图，移除多余的，并用视觉或视觉惯性BA优化地图，操作在接近当前帧的关键帧局部窗口中进行。惯性情况下，利用我们新的最大后验概率MAP估计技术，通过建图线程初始化和优化IMU参数。

LOOP and MAP MERGING以关键帧率检测活动地图和整个地图集的公共区域，如果公共区域属于活动地图，则执行回环校正；否则，两个地图将无缝地合并为一个活动地图。在循环校正之后，在一个独立的线程中启动一个完整的BA，以便在不影响实时性能的情况下进一步优化地图。

**CAMERA MODEL**

ORB-SLAM在所有系统组件中均假设为针孔相机模型。我们的目标是通过将与相机模型相关的所有属性和函数（投影和非投影函数、雅可比矩阵等）提取到单独的模块中，从整个SLAM管道中提取相机模型。这允许我们的系统通过提供相应的摄像头模块来使用任何摄像头型号。在ORB-SLAM3库中，除了针孔模型外，我们还提供了Kannala Brandt[12]鱼眼模型。

由于大多数流行的计算机视觉算法都采用针孔相机模型，因此许多SLAM系统要么校正整个图像，要么校正特征坐标，以在理想的平面视网膜中工作。然而，这种方法对于鱼眼镜头是有问题的，鱼眼镜头可以达到或超过180度的视野（FOV）。图像校正不是一个选项，因为外围的对象被放大，而中心的对象分辨率变差，阻碍了特征匹配。校正特征坐标需要使用小于180度的FOV，这给许多计算机视觉算法带来了麻烦，这些算法假设图像上存在均匀的重投影误差，而这在校正后的鱼眼图像中远远不是真的。这迫使裁剪出图像的外部部分，失去了大视场的优势：更快地映射环境和更好地抵抗遮挡。接下来，我们讨论如何克服这些困难。

A. Relocalization  
 一个健壮的SLAM系统需要在跟踪失败时重新定位摄像机的能力。ORB-SLAM通过设置基于ePnP算法[73]的透视n点解算器来解决重新定位问题，该算法假设在所有公式中都有一个校准的针孔相机。为了跟进我们的方法，我们需要一个PnP算法，该算法独立于使用的摄像机模型。因此，我们采用了最大似然透视n点算法（MLPnP）[74]，该算法与相机模型完全解耦，因为它使用投影光线作为输入。相机模型只需要提供从像素到投影光线的非投影函数，就可以使用重新定位。

B. Non-rectified Stereo SLAM

大多数双目SLAM系统假设双目帧被校正，即两幅图像使用相同的焦距转换为针孔投影，图像平面共面，并与水平极线对齐，这样一幅图像中的特征可以通过查看另一幅图像中的同一行轻松匹配。然而，校正双目图像的假设非常有限，在许多应用中既不合适也不可行。比如，矫正发散的双目对或双目鱼眼相机需要严重的图像裁剪，从而失去大视场的优势。

我们的系统不依赖图像校正，将双目视为两个单目相机：

1. 它们之间有一个连续的相关变换SE（3）
2. 在观测相同场景的一部分中有一个共同区域。

这些约束允许我们在三角剖分新地标和BA优化时通过引入这些信息来有效地估计地图的比例。根据这一想法，我们的SLAM管道估计6自由度刚体姿态，其参考系统可以位于其中一个摄像机或IMU传感器中，并表示摄像机相对于刚体姿态的位置。

如果两台摄像机都有一个重叠区域，我们可以在第一次看到真实比例的地标时对其进行三角测量。这两幅图像的其余部分仍然有很多相关信息，在SLAM管道中用作单目信息。首先在这些区域中看到的特征是从多个视图进行三角剖分的，如单目情况。

**VISUAL-INERTIALSLAM**

ORB-SLAM-VI[4]是第一个能够重用地图的真正视觉惯性SLAM系统。然而，它仅限于针孔单目相机，初始化速度太慢，在一些具有挑战性的场景中失败。在这项工作中，我们以ORB-SLAM-VI为基础，提供了一种快速、准确的IMU初始化技术，以及一个开源的SLAM库，能够使用针孔和鱼眼相机进行单目惯性和双目惯性SLAM。

1. Fundamentals

在纯视觉SLAM中，估计状态仅包括当前相机姿态，在视觉惯性SLAM中，需要计算附加变量。这些是在世界坐标系下的相机位姿，速度，还有陀螺仪和加速度计的偏差和，假设它们按照布朗运动演化。有状态向量：

 (1)

对于视觉惯性SLAM，我们按照[60]中开发的理论，在连续的视觉帧i和i+1之间预积分IMU测量，并在[61]中的流形上进行公式化。我们获得了预积分的旋转、速度和位置测量，表示为：，以及整个测量矢量的协方差矩阵。给定这些预先整合的术语和状态,我们采用[61]惯性残差的定义：

 （2）

从李群映射到向量空间。与惯性残差一起，我们在帧i和位置 的3D点j还使用重投影误差：

 （3）

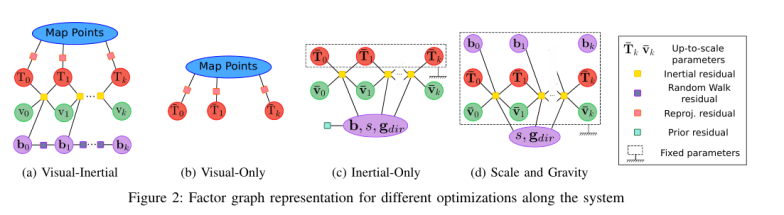
是对应相机模型的投影函数，是在图片i的观察点j，具有协方差矩阵的，表示从车身IMU到摄像头（左或右）的刚体变换，通过校准已知，

是组的变换操作。

结合惯性项和视觉残差项，视觉惯性SLAM可以作为基于关键帧的最小化问题[39]。给定一系列的k+1关键帧和它的状态 ，和个的3D点以及他们的状态****，可视惯性优化问题可以表述为：

 （4）

是观测到3D点j的一系列关键帧，该优化可概括为图2a所示的因子图，注意，对于重投影误差，我们使用稳健的Huber核来减少虚假匹配的影响，而对于惯性残差，则不需要它，因为不存在未命中关联。这种优化需要在跟踪和映射过程中进行调整以提高效率，但更重要的是，它需要良好的初始种子来收敛到精确解。



1. IMU Initialization

该步骤的目标是获得惯性变量的良好初始值：速度、重力方向和IMU偏差。一些系统，如VI-DSO[46]试图从头开始解决视觉惯性BA，避开特定的初始化过程，获得惯性参数的缓慢收敛（长达30秒）。

在这项工作中，我们提出了一种基于三个关键见解的快速准确的初始化方法：

* 纯单目SLAM可以提供非常精确的初始地图[2]，其主要问题是比例未知。首先解决仅视觉问题将增强IMU初始化。
* 如[56]所示，当scale显式表示为优化变量时，它的收敛速度要快得多，而不是使用BA的隐式表示。
* 在IMU初始化过程中忽略传感器的不确定性会产生大量不可预测的错误[64]

因此，在适当考虑传感器不确定性的情况下，我们将IMU初始化描述为MAP估计问题，分为三个步骤：

1. Vision-only MAP Estimation：我们初始化纯单目SLAM[2]，并在2秒内运行，在4Hz处插入关键帧。在这段时间之后，我们有了一张由K=10个摄影机姿势和数百个点组成的放大地图，该地图仅使用视觉BA进行了优化（图2b）。这些姿势转换为身体参考，获得轨迹，其中条形表示单目情况下的最大比例变量。
2. Inertial-only MAP Estimation: 在这一步中，我们的目标是在地图估计的意义上，仅使用这些关键帧之间的惯性测量，获得惯性变量的最佳估计。这些惯性变量可叠加在仅惯性状态向量中：

 （5）

是纯视觉的比例尺度，是旋转矩阵，用于计算世界参考中的重力矢量g,为加速度计和陀螺仪偏差假设在初始化时恒定；是从第一个关键帧到最后一个关键帧的按尺度速度，初始估计从开始估计。此时，我们只考虑惯性测量值。因此，最大化的后验分布是：

 （6）

代表似然，代表先验。考虑到测量的独立性，纯惯性地图估计问题可以写成：

 （7）

采用负对数并假设IMU预积分和先验分布的高斯误差，最终得到优化问题：

 （8）

这种优化，如图2c所示，与等式4的不同之处在于不包括视觉残差，多了先验残差来使IMU偏差接近于零，其协方差由IMU的特性给出。

当我们在流形中进行优化时，定义一个retraction[61]，以便在优化过程中更新重力方向估计，由于围绕重力方向旋转不会假设重力发生变化，因此此更新使用两个角度（，）进行参数化：

 （9）

Exp从so(3)到SO(3)的指数映射。为了保证在优化过程中尺度因子为正，将其更新定义为：

（10）

一旦仅惯性估计优化完成，帧的位姿和速度以及3D地图点将按估计的比例缩放，旋转z轴与估计的重力方向对齐。偏差被更新，重复IMU预积分，来减少未来的线性化误差。

3）视觉惯性估计：一旦我们对惯性和视觉参数有了良好的估计，我们就可以执行联合视觉惯性优化，以进一步完善解决方案。该优化可表示为图2a，但所有关键帧都有共同的偏置，并且包括与仅惯性步骤中相同的偏置先验信息.

我们在EuRoC数据集[6]上进行的详尽初始化实验表明，该初始化非常有效，在2秒的轨迹上实现了5%的标度误差。为了改进初始估计，在初始化后5秒和15秒执行视觉惯性BA，收敛到1%的标度误差，如第七节所示。在这些BAs之后，我们说地图是自然的，这意味着比例尺、IMU参数和重力方向已经被精确估计。

我们的初始化比求解一组代数方程[62]–[64]的联合初始化方法要精确得多，并且比ORB-SLAMVI[4]中使用的初始化方法要快得多，ORB-SLAMVI[4]中使用的初始化方法需要15秒才能获得第一次标度估计，或VI-DSO中使用的初始化方法[46]，这从一个巨大的标度误差开始，需要20-30秒才能收敛到1%的误差。不同初始化方法之间的比较见[6]。

在某些特定情况下，当慢运动不能提供惯性参数的良好可观测性时，初始化可能无法在15秒内收敛到精确解。为了获得对这种情况的鲁棒性，我们提出了一种新的比例优化技术，基于改进的纯惯性优化，其中包括所有插入的关键帧，但比例和重力方向是唯一要估计的参数（图2d）。请注意，在这种情况下，恒定偏差的假设是不正确的。相反，我们使用从建图估计的值，并修复它们。此优化在计算上非常高效，每10秒在本地映射线程中执行一次，直到贴图具有100多个关键帧，或者自初始化以来已超过75秒。

最后，我们很容易地将单目惯性初始化扩展到双目惯性，方法是将比例因子固定为1，并将其从纯惯性优化变量中取出，从而增强其收敛性。

1. Tracking and Mapping

对于跟踪和映射，我们采用了[4]中提出的方案。跟踪解决了一个简化的视觉惯性优化问题，其中仅优化最后两帧的状态，而建图点保持不变。

对于建图，尝试从方程4中求解整个优化对于大型建图来说是困难的。我们使用关键帧及其点的滑动窗口作为可优化变量，还包括从可共视关键帧对这些点的观察，但保持其姿势不变。

1. Robustness to tracking loss

在纯视觉SLAM或VO系统中，时间相机遮挡和快速运动会导致视觉元素失去跟踪，导致系统丢失。ORB-SLAM率先使用了基于单词袋位置识别的快速再定位技术，但事实证明它们不足以解决EuRoC数据集中的困难序列[3]。当跟踪的点地图少于15个时，我们的视觉惯性系统进入视觉闭合状态，并在两个阶段实现鲁棒性：

* 短期丢失：根据IMU读数估计当前状态，并以估计的相机姿势投影地图点，并在大型图像窗口中搜索匹配。结果匹配包含在视觉惯性优化中。在大多数情况下，这允许恢复视觉跟踪。否则，5秒后，我们进入下一阶段。
* 长期丢失：如上所述初始化一个新的视觉惯性地图，它将成为活动地图。如果系统在IMU初始化后15秒内丢失，则地图将被丢弃。这样可以防止累积不准确和无意义的地图。

VI. MAPMERGING ANDLOOPCLOSING

跟踪和建图线程通过将地图点投影到估计的相机姿态并在仅几个像素的图像窗口中搜索匹配，常规地发现帧和活动贴图之间的短期和中期数据关联。为了实现再定位和循环检测的长期数据关联，ORB-SLAM使用DBoW2单词袋位置识别系统[9]，[75]。这种方法也被最新的VO和SLAM系统采用，这些系统实现了闭环（表1）。

与跟踪不同，位置识别并不是从摄像机姿势的初始猜测开始的。相反，DBoW2使用关键帧的词包向量构建了一个关键帧数据库，并且给定一个查询图像能够根据它们的词包有效地提供最相似的关键帧。仅使用第一个候选者，原始DBoW2查询就可以达到50-80%的精度和召回率[9]。为了避免误报损坏地图，DBoW2实施了时间和几何一致性检查，将工作点移动到100%精度和30-40%召回率[9]，[75]。关键的是，时间一致性检查至少在3个关键帧期间延迟位置识别。当我们尝试在Atlas系统中使用它时，我们发现这种延迟和低召回率导致在相同或不同的地图中经常出现重复区域。

在这项工作中，我们提出了一种新的地方识别算法，改进召回长期和多地图数据关联。每当贴图线程创建新的关键帧时，就会启动位置识别，尝试检测与Atlas中已存在的任何关键帧的匹配。如果找到的匹配关键帧属于活动地图，则执行循环闭合。否则，它是一个多地图数据关联，然后将活动地图和匹配地图合并。作为我们方法中的第二个新颖之处，一旦估计了新关键帧和匹配贴图之间的相对姿势，我们定义了一个局部窗口，其中包含匹配关键帧及其在共视图中的邻域。在这个窗口中，我们集中搜索中期数据关联，提高闭环和地图合并的准确性。这两个新特性解释了在EuRoC实验中，ORB-SLAM3比ORB-SLAM2获得更好的精度。下面将解释不同操作的细节。

A. Place Recognition

为了获得更高的召回率，对于每个新的活动关键帧，我们在DBoW2数据库中查询Atlas中的几个类似关键帧。为了达到100%的精度，每个候选对象都要经过几个几何验证步骤。所有几何验证步骤的基本操作包括检查图像窗口内是否存在ORB关键点，其描述符与地图点的ORB描述符匹配，使用它们之间的汉明距离阈值。如果搜索窗口中有多个候选项，要放弃不明确的匹配项，我们将选中与第二个最近匹配的距离比[76]。我们的位置识别算法的步骤如下：

（1）DBoW2 candidate keyframes。我们使用活动关键帧查询Atlas DBoW2数据库，检索三个最相似的关键帧，不包括与活动关键帧共视的关键帧。我们将每个匹配的候选位置识别称为。

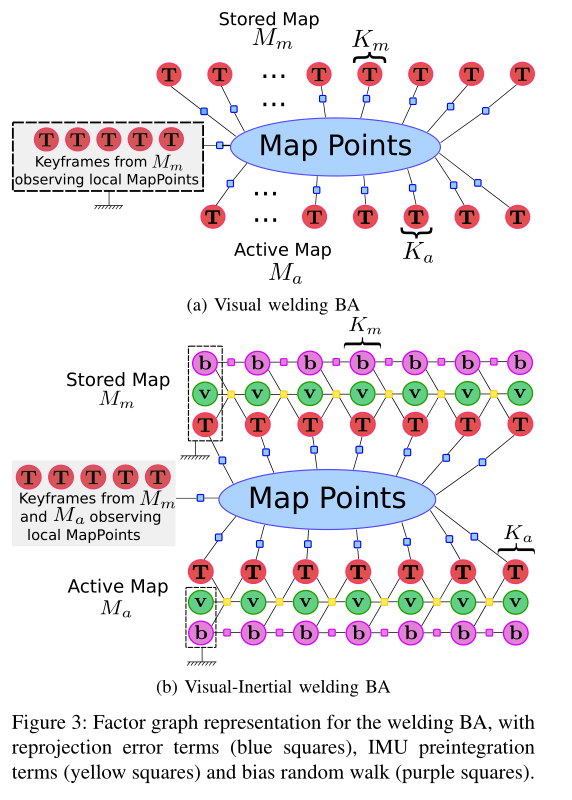
（2）Local window.对于每个，定义一个本地窗口，其中包括、其最佳共视关键帧以及所有这些关键帧观察到的地图点。DBoW2直接索引提供了一组假定的关键点匹配在本地窗口关键帧之间。对于这些2D-2D匹配中的每一个，我们还提供了其相应地图点之间的3D-3D匹配。

（3）3D aligning transformation.我们使用RANSAC计算转换，它可以更好地将局部窗口中的地图点与的地图点对齐.在纯单目或单目惯性中，当地图还未成熟时，我们计算∈Sim（3），其他方式∈SE（3）。在这两种情况下，我们使用Horn算法[77]，使用三个3D-3D匹配的最小集合来找到每个假设。假设的匹配项是，在将地图点变换后，实现了重投影误差，对假设投赞成票。如果票数超过阈值，则选择票数更多的假设。

1. Guided matching refinement。使用变换本地窗口中的所有地图点，以查找与中的关键点的更多匹配项。搜索也会反转，在本地窗口的所有关键帧中查找匹配的点。利用找到的所有匹配，通过非线性优化对进行细化，其中目标函数为双向重投影误差，使用Huber影响函数提供对虚假匹配的鲁棒性。如果优化后的内联线数量超过阈值，则使用较小的图像搜索窗口启动引导匹配和非线性优化的第二次迭代。
2. Verification in three covisible keyframes.为了避免误报，DBoW2在三个连续的关键帧中等待位置识别触发，延迟或丢失位置识别。我们的关键洞察是，大多数时候，核查所需的信息已经在地图上。为了验证位置识别，我们在地图的活动部分搜索两个与共视关键帧，其中与局部窗口中的点匹配的数量超过阈值。如果未找到这些关键帧，则将使用新传入的关键帧进一步尝试验证，而无需再次触发单词包。验证将继续，直到三个关键帧或两个连续的新关键帧无法验证为止。
3. VI Gravity direction verification.在视觉惯性的情况下，如果活动地图是成熟的，我们估计了。我们进一步检查俯仰角和滚转角是否低于最终接受的阈值。

B. Visual Map Merging

当一个成功的位置识别产生多个地图数据关联时，在活动地图中的关键帧和存储在地图集中的不同地图的匹配关键帧之间，通过对齐变换，启动地图合并操作。在这个过程中，必须特别注意确保中的信息能够被跟踪线程及时重用，以避免地图重复。为此，将地图引入参考。由于可能包含许多元素，合并它们可能需要很长时间，因此合并分为两个步骤。首先，在和邻域定义的连接窗口中进行合并，然后通过位姿图优化将修正传播到合并后的其余部分。合并算法的具体步骤是：  
1）连接窗口组合。连接窗口包括及其共视关键帧，及其共视关键帧以及它们观察到的所有地图点。在将它们包含在连接窗口中之前，属于的关键帧和地图点由变换以使它们与对齐。  
2）合并地图。和融合在一起成为新的活动地图。若要删除重复点，将在关键帧中主动搜索匹配点。对于每个匹配，中的点被移除，中不断累积已移除点的所有观测值。共视和本质图[2]通过添加和关键帧相连的边来更新，这些边是由新的中期点关联找到的。  
3）连接窗口BA。在连接窗口（图3a）中，执行局部BA优化和中的所有关键帧。为了固定自由度，与中共视的关键帧保持固定。优化完成后，连接窗口区域中包含的所有关键帧都可以用于相机追踪，实现快速准确地地图重用。  
4）位姿图优化。在保持连接区域关键帧固定的前提下，利用融合后的整个地图的本质图进行位姿图优化。此优化将修正从连接窗口传播到地图的其余部分。



C. Visual-Inertial Map Merging

视觉惯性合并算法遵循与纯视觉情况类似的步骤。修改步骤1）和3）以更好地利用惯性信息：

1. VI welding window assembly。如果活动地图是成熟的，在被包含到连接窗口之前，用有效的变换到地图。如果不成熟，用有效的对齐。
2. VI welding bundle adjustment。关键帧及其最近时间5个关键帧的位姿、速度和偏差作为优化项。这些变量通过IMU预积分项进行关联。对于地图，我们进行了类似的处理，包括k及其5个时间邻域的姿态、速度和偏差，如图3b所示。对于，包含了紧挨着局部窗口的关键帧，但固定了；对于，包含了相似的关键帧，但其位姿仍是可优化的。所有这些关键帧看到的所有点，以及观察这些点的关键帧位姿也得到优化。所有的关键帧和点通过重投影误差来关联。

D. Loop Closing

回环校正类似于地图合并，但是在位置识别匹配的两个关键帧都属于活动地图的情况下。连接窗口由匹配的关键帧组合而成，重复点被检测和融合，在共视和本质图中创建新的连接。下一步是位姿图优化，将回环校正传播到地图的其余部分。最后一步是全局BA，在考虑回环的中期和长期匹配后找到MAP估计。在视觉惯性情况下，全局BA只在关键帧数目低于阈值的情况下进行，避免巨大的计算开销。