

介绍英特尔®RealSense™视觉SLAM和T265跟踪摄像机

版本1.0

Anders Grunnet-Jepsen, Michael Harville, Brian Fulkerson, Daniel Piro, Shirit Brook, Jim Radford

1. 介绍

位置感知可以说是在3D世界中生存和导航所需的最基本和最重要的感官之一。四处走动、觅食以及返回家园或巢穴的能力在所有动物身上都根深蒂固，因此机器人技术面临的挑战变成了“我们如何模仿甚至可能通过电子方式增强这种天生的能力？”在计算机科学中，这一研究领域通常被称为“同步定位和绘图”或SLAM。它包括对仪器和算法的研究，这些仪器和算法试图解决代理(又名机器人或自动驾驶汽车)在3D空间中的位置和方向，同时创建和更新未知周围环境的“心理”地图。有时这也被称为环境的“由内而外跟踪”，以确定代理自身的运动，以区别于“由外而内”的动作捕捉(又名mo-cap)或通过围绕被跟踪对象的固定摄像机阵列实现的对象跟踪。对于许多应用来说，SLAM最重要的方面之一是能够完全跟踪所有6个自由度(6DoF)，这意味着一个人自己的空间位置(X, Y, Z)以及角方向(俯仰，偏航，滚转)。

对于一些受限的映射问题，目前市场上确实存在成熟的解决方案。例如，全球定位系统(GPS)和辅助导航现在是如此广泛，以至于几乎每一部手机都嵌入了一个接收器，基本上不再需要学习如何阅读和导航地图。然而，GPS的精度目前被限制在大约10米。此外，更新速度在~10Hz相当慢，它们只能在室外工作，在那里它们能够接收来自GPS卫星的电子三角测量信号，并且由于山脉和建筑物的反射，它们受到多路径干扰。另一个近年来发展惊人的传感器是惯性测量单元，简称IMU。它由陀螺仪和加速度计的组合组成，通常由微型集成MEMs(微机电系统)组件组成，能够快速测量自身方向和加速度的变化。这些现在也已经进入了基本上每一部手机，并且可以以大约1美元的价格单独购买。在价格、功率和尺寸光谱的另一个范围内，光探测和测距系统(激光雷达)在开发全自动驾驶汽车的竞赛中变得非常受欢迎。激光雷达通常由旋转/扫描激光束组成，采用短光脉冲或调频，以毫米级精度测量与周围物体的距离。

我们将在这里重点介绍的技术被称为视觉惯性里程计(VIO)，它实际上是与大多数动物感知世界的方式最接近的电子等效物——使用CMOS传感器作为眼睛来观察周围环境;IMU充当内耳，感知平衡和方向;计算机充当大脑，将信息融合成即时定位和地图。VIO有

成为近年来出现的最有前途的SLAM技术之一，算法和变体继续以高速发展和改进。例如，今天大多数增强/虚拟现实耳机(包括微软HoloLens, Oculus Quest或Magic Leap One)和手机(苹果的AR Kit和谷歌的AR Core)都使用自己的VIO版本来准确和低延迟地跟踪设备的完整6DoF。此外，VIO现在也开始在机器人领域得到更广泛的应用。与现有的激光雷达解决方案相比，VIO系统体积更小、成本更低、功耗更低，并且可以根据视觉特征(而不仅仅是测距和几何形状)在大地图上识别和重新定位位置，从而提供更好的重新定位潜力。

在接下来的部分中，我们将探索更多的VIO是如何工作的，我们将介绍英特尔RealSense跟踪摄像机T265和跟踪模块T261，旨在作为完整的交钥匙嵌入式SLAM传感器解决方案，用于AR/VR头戴式设备，机器人和无人机。我们将解释一些关键的SLAM系统性能指标，我们将展示使用英特尔RealSense查看器开始使用T265相机是多么容易，并使用开源英特尔RealSense SDK以编程方式与之接口。本白皮书旨在作为快速介绍和概述，将更深入的处理留给其他技术文档。

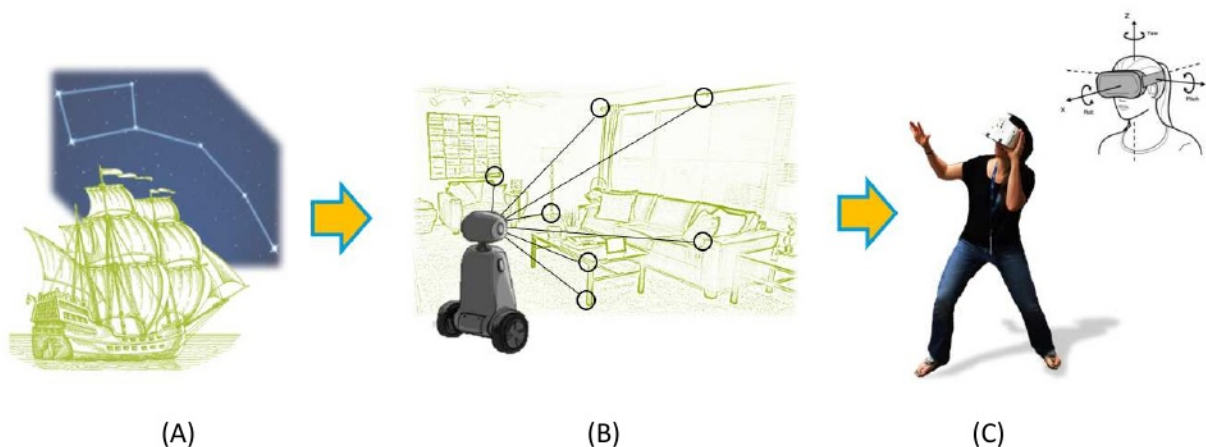


图1所示。熟练的水手习惯于根据星星导航(A)。视觉惯性里程计试图实现低延迟，高速，稳健和准确的全6DOF跟踪，可用于机器人(B)和AR/VR耳机(C)。

2. 视觉惯性里程计初学者指南

开始理解visual-inertial-odometry的最好方法是首先检查“视觉”部分。操作的核心概念可以追溯到指南针和gps出现之前的时代，那时人类会根据星星进行导航，如图1A所示。通过识别特殊的星座或特定的星星，人们可以找到北方(即方向)，也可以计算自己的纬度和经度(即位置)，通常使用查找表。

在计算机视觉中，我们力求以最快、最稳健的方式完成同样的任务。我们首先拍摄周围环境的电子照片。这通常由数百万个像素组成，所以我们使用“特征检测算法”来立即压缩信息内容，以至于我们基本上只剩下几百个命名点。换句话说，图像现在被替换成一张黑色的图片，上面有几个白点，每个白点都有一个身份，就像在晚上看天空一样。那么，什么是“特征”呢？特征可以很简单，比如说“找到图像中的所有角落”。实际上，这是一个非常活跃的研究领域，特征检测算法有很多名字，比如SIFT、LIFT、哈里斯角、Shi-Tomasi角、ORB、BLOB、SURF、KAZE、FAST等等。一旦检测到特征，它们都必须由一种算法“命名”，该算法为每个特征赋予一个“描述符”或“特征向量”。我们对接收到的每一个新图像帧重复这个过程，并使用特征跟踪算法对连续帧(或某些关键帧)中的特征进行比较和跟踪。我们强调，每个特征实际上都映射到3D空间，而不仅仅是二维空间。图2展示了一个检测到特征的图像示例。



图2所示。在图像中检测到“哈里斯角特征”的例子。转载自参考文献1。

这一切都运行得非常好，但最先进的解决方案正在不断发展以提高性能。例如，使用单个相机(又名。Mono-SLAM)的一个问题是它不能很好地解决缩放问题。这意味着如果相机向前移动1米，它没有办法将观察到的运动与1米联系起来，因为它也可能是1厘米。解决这个问题的两种最常见的方法是使用校准过的立体摄像机或IMU。另一个重大挑战是能够实现高帧率和低延迟跟踪。虽然在更高的帧率下运行相机传感器当然是可能的，但这些算法实际上是非常昂贵的计算成本(意味着计算和耗电)，而且你运行相机的速度越快，它能得到的光线就越少。事实证明，更好的设计方法是使用高速IMU，它在短时间内(10毫秒以内)跟踪速度和方向的能力出奇地好，但如果长时间运行，比如几秒钟，就会出现严重的漂移。IMU的增加实际上是VIO中“**I**”(“惯性”)的原因。表1很好地总结了Visual和IMU里程计如何很好地相互补充。此外，如果传感器连接到提供里程表读数(如速度)的车辆上，这也可以用来提高跟踪精度。最后，由于跟踪的主要前提是能够看到特征，因此大多数系统都会尝试使用光学传感器

具有非常宽的视场，因此它们可以看到尽可能多的特征。当传感器移动到靠近白墙时，这一点特别有用，因为白墙本身没有任何特征。使用鱼眼镜头，即使在非常接近无特征物体的情况下，它也应该仍然可以跟踪视觉外围的特征。

	Visual Odometry	IMU Odometry
Update Frequency	Low	High
Stability over Time	High	Low
Scene Dependent	Yes	No
Re-localization	Yes	No
CPU Utilization	High	Low

表1。视觉和IMU里程计技术相辅相成。特别是，IMU数据的频率比视觉里程计高。另一方面，随着时间的推移，IMU数据本身就存在大量的累积漂移，这可以通过视觉里程计的不频繁更新来纠正。此外，视觉里程计依赖于场景中存在的特征，而IMU相对独立于环境。

总之，目前的最佳解决方案往往是将多个鱼眼成像仪的信息与IMU和/或车辆里程计的传感器读数相结合。然后，它将实时融合所有信息，并以极低的延迟跟踪传感器拥有的6DoF。图1B和1C分别是使用这种嵌入式传感器对机器人和AR/VR头戴式设备进行以自我为中心的6DOF跟踪的示例。



图3所示。英特尔RealSense跟踪相机T265是一个完整的嵌入式SLAM解决方案，使用视觉惯性里程计(VIO)在3D空间中跟踪其自身的方向和位置(6DoF)。

如图3所示，英特尔RealSense跟踪摄像头T265是一个完整的独立解决方案，它利用最先进的算法输出基于VIO的6DoF跟踪。硬件设计包括一套宽视场立体鱼眼相机，由直径约800像素的单色全局快门相机捕获~165度的圆形FOV。它还集成了一个硬件同步博世BMI055 200Hz陀螺仪和62.5Hz加速度计以及一个强大的英特尔®Movidius™Myriad™2视觉处理单元(VPU)。嵌入式处理器运行整个SLAM算法，分析立体图像并将所有传感器信息融合在一起进行6DoF跟踪，同时功耗低于1.5W。相机姿态信息以200Hz的速率通过USB提供，可以很容易地使用开源英特尔RealSense SDK接口到几乎任何主机平台，详见第4节。

该设备是为客户寻找一个无忧无虑的独立SLAM系统。对于那些想要开发或部署自己的跟踪算法的人，他们可以从T265获得原始数据(相机和IMU)，或者构建在我们的英特尔RealSense深度相机D400系列之上，该系列提供类似的传感器功能(尽管具有较小的FOV)和输出深度图。但是，请注意，任何SLAM算法都需要在主机平台上运行，而不是嵌入，这将增加目标平台的重量、功率预算和成本。

3. 性能指标

为了对6DoF跟踪器的性能进行基准测试，最广泛使用的关键性能指标(kpi)是漂移和地板位置误差。漂移是在“未知”环境中未激活重新定位时的累积误差。它被计算为每经过的路径长度的误差百分比。地板位置误差是在已知环境下，当重新定位活动时，地板报告位置与实际位置的均方根误差。其测量单位为cm。

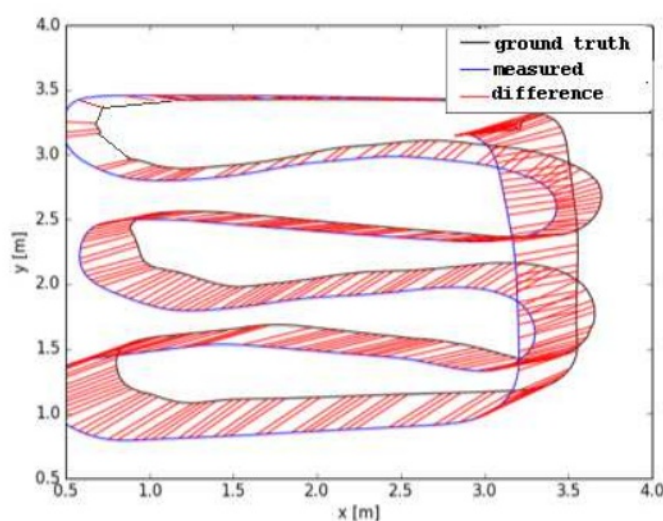


图4所示。SLAM运动轨迹误差示例，显示地面真实情况和估计位置(此处为2D)。像地板位置误差和漂移这样的kpi可以从这些类型的记录中测量出来。

一个好的SLAM系统的一个极其重要的属性是它能够重新定位到一个已知的或预先生成的地图。这有时被称为解决“被绑架的机器人”问题。基本上，当机器人最初通电，或者断电，或者在运动过程中视野长时间受阻时，传感器就有可能失去对它所在位置的追踪。如果它有内部地图，重新定位误差描述的是系统识别之前环境并在该地图上重新定位自己的能力。这对于大规模跟踪运动或穿越多个房间尤为重要。一个好的系统以厘米级的绝对精度重新定位，并且可以选择频繁地重新定位，并按照预定义的节奏进行重新定位。特征地图生成的另一个非常重要的方面是，地图是否可以在多个智能体之间共享。例如，任何机器人或自动驾驶汽车都可以从另一个机器人已经绘制出该区域的事实中受益，这是可取的。为了实现这一点，每个机器人不仅必须能够捕获，还必须能够导出和共享它们的地图，更重要的是，检测到的特征在每个机器人上看起来应该是一样的。T265允许这种协作映射。

另外两个性能标准是主机CPU利用率和内存使用率。由于T265 SLAM算法在设备本身上运行(即在嵌入式视觉处理器上)，因此它很少使用这两种算法，从而为主机上的应用程序留下空间和计算能力，即使在弱平台上运行也是如此。

目标kpi请参考T265数据表[参考文献2]。

4. T265入门

4.1 校准与初始化

T265跟踪摄像头已在工厂进行校准，这些参数可以从英特尔RealSense SDK中获得。当相机启动时，其中一些参数可能会改变，因此T265内部的算法会在启动时估计它们。它假设设备在启动时没有移动，只有当设备移动到足以估计其内部参数时，才会返回“高”置信度。非常慢和非常快的移动都会阻碍这种初始化，应该避免。

4.1a 安装注意事项

设备应牢固地固定在目标平台上，机壳背面有两个M3螺丝孔，方便安装。在传感器可能经历高振动(如无人机)或临时冲击(如机器人)的应用中，最好将T265连接到减振底座上。

T265还提供M2螺钉孔，以牢固地连接微型USB 3电缆，如果可能的话，应使用这些螺钉孔，以防止USB断开。

4.1b 坐标系

T265提供的姿态是相对于两个摄像头的中心提供的，如图4b所示。这个姿态就是从这个参照系到重力对齐参照系的转换

以起始位置为中心，Y向上，-Z相对于初始相机轴向前，X向右(见图4c)。

在应用程序中，您通常需要知道T265以外的东西的姿势。例如，在VR头显中，您可能想知道渲染中心的位置。T265与目标坐标系之间的变换应仔细测量和应用。

应用程序还可能希望在将来的某个时间知道位置(例如，在计划渲染VR场景的下一帧的时候)。英特尔RealSense SDK提供了一个名为**pose-predict**的示例，其中展示了如何使用T265的速度和加速度估计来预测设备未来的位置。

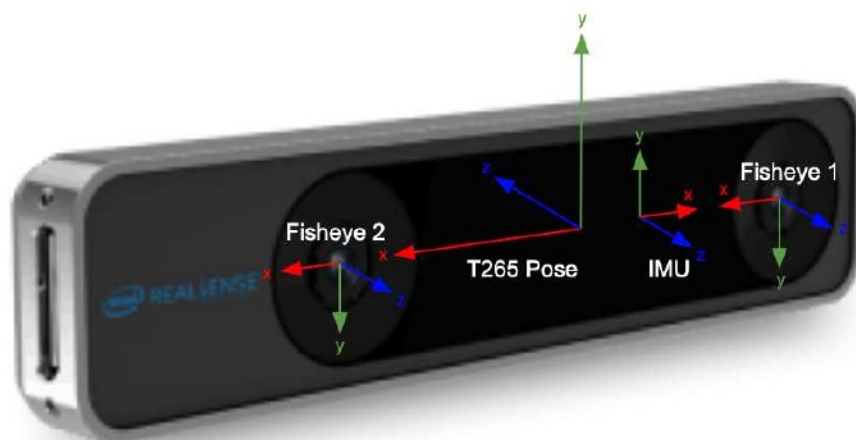


图4 b。T265的坐标系。传感器之间的精确变换可以从英特尔RealSense SDK。



图4 c。T265的原点坐标系。

4.1c时间戳

T265报告RS2_TIMESTAMP_DOMAIN_GLOBAL_TIME中所有传感器的时间。它使用主机和T265之间的时间同步机制，以毫秒级精确的主机时钟时间提供所有传感器数据时间戳。

4.2 英特尔RealSense浏览器

熟悉T265跟踪摄像机的最佳方法是运行英特尔®RealSense™查看器应用程序，该应用程序包含在英特尔®RealSense™SDK 2.0中。该应用程序通过图形用户界面可视化姿势输出。请参阅<https://www.intelrealsense.com/tracking-camera-t265/>和T265在GitHub上获得更多信息和代码示例。

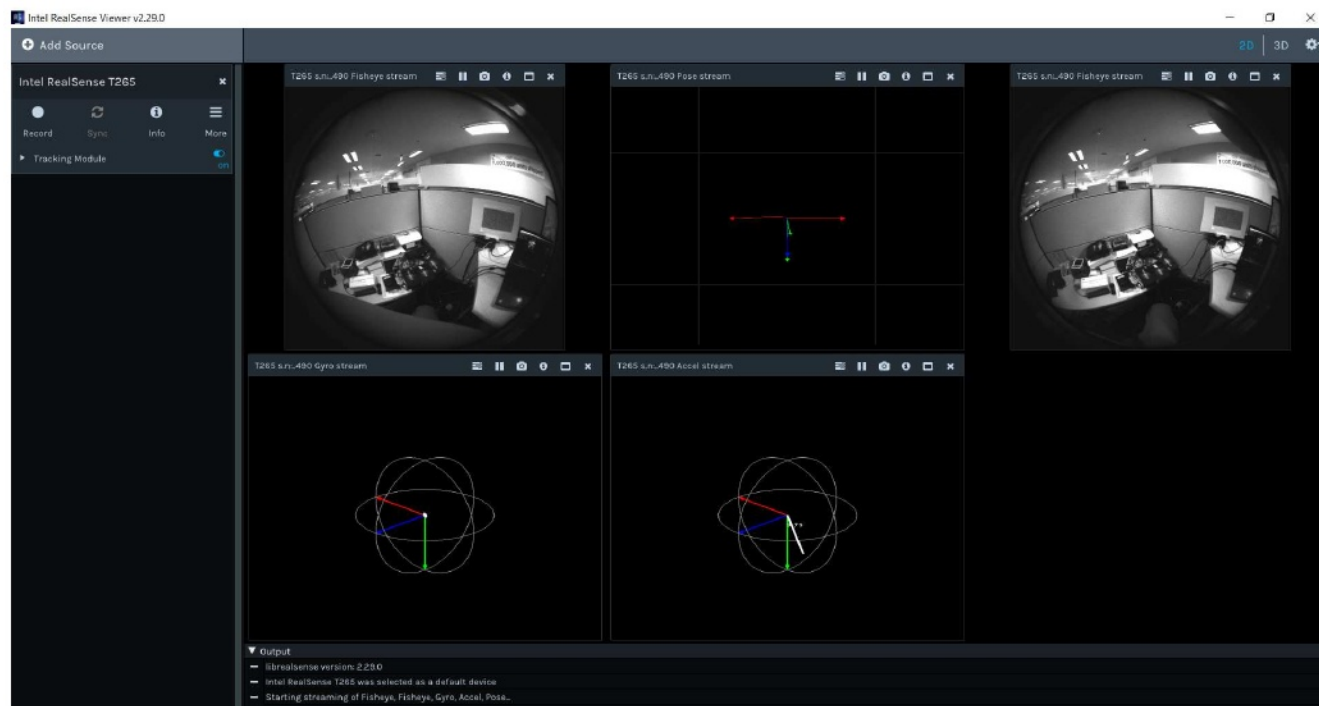


图5。英特尔RealSense浏览器，显示在这里，是一个伟大的起点，熟悉英特尔RealSense跟踪相机T265。在“2D”视图中，不同的原始传感器输入被可视化。

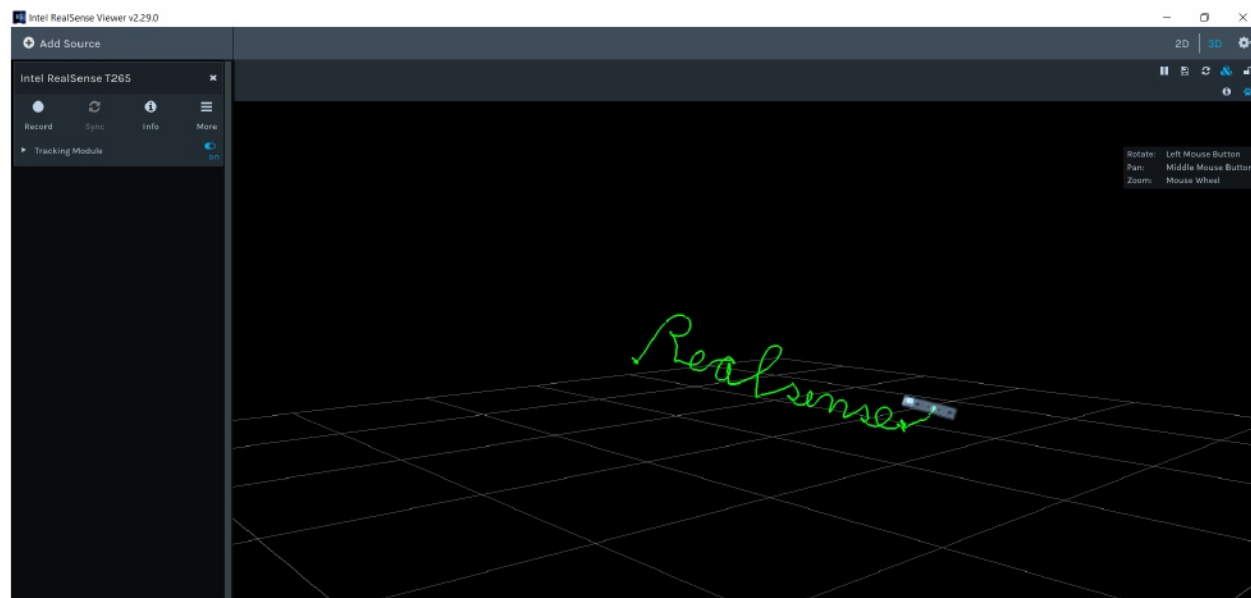


图5 b。在英特尔RealSense查看器中，切换到3D视图可以通过图形用户界面可视化姿势。这个点云视图可以在3D中旋转。



图5度。在英特尔RealSense查看器中，可以调整参数以查看其效果。此外，可以启用重新定位和映射。

4.3 英特尔RealSense SDK API

英特尔®RealSense™跟踪相机T265的主要软件接口是英特尔®RealSense™SDK 2.0，可从<http://intelrealsense.com/developers/>获取

英特尔®RealSense™SDK 2.0是跨平台和开源的，也可用于驱动其他英特尔®RealSense™产品，如D400系列深度相机。它在Windows和Linux平台上得到最好的支持，在Mac上的功能会有所减少。

SDK 2.0还包含各种其他示例应用程序(带有源代码和文档)，展示如何使用英特尔®RealSense™设备，包括T265跟踪摄像头。ROS(机器人操作系统)也可用于与英特尔®RealSense™设备进行交互。<https://github.com/IntelRealSense/realsense-ros/>包含基于英特尔®RealSense SDK 2.0构建的ROS集成、工具和示例应用程序。所有这些代码示例都可以直接用于测试，修改以适应测试目的，或作为用户构建新应用程序的灵感。

对于英特尔®RealSense™跟踪相机T265，最基本的示例应用程序位于“pose”和“pose-predict”文件夹中。这两个样例分别展示了如何通过轮询和回调接口获取姿态数据。在英特尔®RealSense™Software ROS存储库中，`T265RealsenseNode`类是一个很好的起点，可以在[这里找到如何通过ROS驱动T265的几个示例](#)。

4.4 挑战

我们正在不断改进T265的稳定性、鲁棒性和跟踪性能。您可以通过更新到最新版本的英特尔RealSense SDK来获取最新版本的固件。我们想强调一些适用于所有SLAM系统的挑战，包括T265。

- **场景外观和几何形状:**SLAM算法通常会创建一个世界模型，该模型依赖于“地标”，或传感器检测到的相对特征配置，以便在该世界中导航。例如，英特尔®RealSense™跟踪相机T265部分依赖于其相机所看到的“有趣”视觉特征。如果场景的特征太少，例如当摄像头面对一堵白墙或一个表面没有纹理的空房间时，传感器可能很难创建地图。另一方面，视觉纹理过多或重复视觉图案的场景偶尔会造成地图和导航上的混乱。
- **动态场景:**如果场景中有许多物体(如人)在移动，SLAM系统可能难以在3D中创建静止的地标地图。
- **照明:**如果没有足够的光线，英特尔®RealSense™跟踪相机T265可能根本无法检测到任何视觉特征。在适度低光下，传感器噪声会影响图像质量，较长的曝光时间会导致图像模糊。非常强的全局光照，或者将相机对准光源，也会导致图像的部分或全部饱和。在饱和的图像区域内检测不到任何视觉特征，并且沿着这些区域的边缘可能会产生虚假的特征。
- **反射:**场景中的高光表面，如镜子或玻璃窗，可能会导致“幻影”地标的感知，这些地标似乎位于不正确的3D位置。
- **设备运动:**在高速下，运动模糊会导致图像数据变得基本无用。突然加速，无论是在位置还是方向上，都可能超出IMU的测量极限，或者打破视觉惯性传感器融合算法的假设基础。
- **时序和计算资源:**实时视觉惯性融合算法对其输入信号的相对时序非常敏感，因此如果涉及多个设备时钟，任何处理不当的异步都可能对结果产生巨大的负面影响。同样，计算资源的暂时短缺或中断可能导致破坏性的数据丢失或跳过关键的算法步骤。与许多其他SLAM系统相比，英特尔®RealSense™跟踪相机T265的架构解决了大多数这些问题，但客户在将其与主机上的应用程序集成时仍然必须注意。

4.5 超越SLAM

在开发自主导航系统时，至少还有两个重要的方面需要考虑。在3D世界中移动时，除了映射之外，下一个最重要的任务是避免碰撞。这通常需要连续测量到所有障碍物的精确3D距离

的路径。虽然VIO系统实际上通过跟踪特征点创建了一个“稀疏”深度图，但它通常不足以实现鲁棒避碰。出于这个原因，我们通常建议将T265与英特尔RealSense深度相机配对，如模型D435或D415，这将提供一个非常密集和准确的深度图，可用于3D重建，以建立一个完整的环境3D模型，或用于对象扫描。英特尔RealSense查看器是一个很好的地方，开始看看这些传感器如何一起工作。更多信息可以在<https://www.intelrealsense.com/depth-and-tracking-combined-get-started/>上找到。

另一个需要考虑的方面是如何在超大空间中导航。T265原生支持的一种方法是“四月标签”跟踪，如参考文献3所述。四月标签是一种开源的2D条形码(或基准标记)，可以在3D空间中很好地跟踪，以给出标签的位置和方向，或者相对于标签的代理的位置和方向。四月标签特别吸引人，因为它们是旋转、剪切、照明、缩放和平移不变的。由于有许多独特的代码，这为系统提供了基于与Visual SLAM不同的模态进行重新定位的选项。此外，它很容易为用户打印出四月标签和标记不同的区域或对象。使用四月标签的代码示例可以在这里找到：

<https://github.com/IntelRealSense/librealsense/tree/master/examples/pose-apriltag>。



图6所示。跟踪四月标签的示例。截图转载自Ref 4中的Video。四月标签跟踪包含在T265中。

5. 总结

通过对6DoF位置和方向进行精确、高速、低延迟的“由内而外”跟踪，许多新的令人兴奋的应用得以实现。跟踪设备具有小尺寸，低功耗和良好的开环跟踪，例如，可以安装在AR/VR用途的头盔中，或安装在机器人上进行位置跟踪。具有重新定位到地图的能力，大面积覆盖和召回并被启用，地图甚至可以在多个代理之间共享。T265跟踪摄像头是一个出色的开箱即用的解决方案，应该能够快速原型和熟悉性能和编程使用RealSense Viewer应用程序和SDK。T261相机模块可用于更多自定义集成。

参考文献

1. [https://en.wikipedia.org/wiki/Feature_detection_\(computer_vision\)](https://en.wikipedia.org/wiki/Feature_detection_(computer_vision))
2. <https://dev.intelrealsense.com/docs/tracking-camera-t265-datasheet>
3. <https://april.eecs.umich.edu/software/apriltag.html>

4. <https://www.youtube.com/watch?v=Y8WEGGbLWIA>
