人形机器人应用于精细农业的思考

**摘要**

目前，机器人在精细农业上的应用主要集中在平地和大规模的农场。轮式、履带式机器人和无人机等容易在平原地带使用，进行采摘、喷洒等作业。但是一些地区是梯田式山地地形。在这种地形上，轮式、履带式机器人难以应用。无人机存在续航时间短的问题，频繁升降，容易发生碰撞，且无人机适合喷洒，不适合采摘等作业。腿式的人形机器人能和人一样适合各种地形，可以在山地自由行走和作业。本文主要关注适用于人形机器人山地行走的SLAM技术。

1. **山地精细农业思考**

目前，采用机械化作业的农业主要集中在平原地区，几乎没有能在山地作业的机器人。主要原因是山地地形复杂，除了腿式机器人，大多数机器人不能在山地自由行走，而且腿式机器人在山地行走也需要出色的动力和平衡能力。在具有喀斯特地貌的桂林长大的我，偶尔会思考像龙胜梯田（图1-1）这种地形能否使用机器人作业。目前来看，比较适合在这种地形行走的，主要是腿式的人形机器人。在山地行走的人形机器人需要强劲的动力和优秀的平衡能力。波士顿动力公司之前发布的一款能在山地行走的人形机器人Atlas（图1-2）展示出了人形机器人应用于山地的可能。



图1- 1 龙胜梯田图

在山地行走的机器人除了具有出色的动力和平衡能力，还需具备自主定位和建图的功能。GPS和北斗等卫星定位受精度限制，不适合于这种情况。故机器人需要具备自主定位和建图的能力。SLAM（Simultaneous localization and mapping）可以实现在未知环境的定位和建图。接下主要讨论一下自己研究方向SLAM 技术，其有助于解决人形机器人在山地地形的定位和建图问题。SLAM主要分为视觉SLAM和激光SLAM。在山地行走颠簸比较厉害，不适合旋转式激光雷达。另外室外光照较强，不适合主动发射光源的传感器。单目视觉没有尺度信息，经过简单分析发现，使用单目相机或双目相机和IMU的融合是比较好的选择。



图1- 2 Atlas山地行走图

1. **SLAM简介**

SLAM是为了让机器人绘制地图而出现的技术，是在机器人移动到未知空间时通过探测周围环境来估计当前位置并同时绘制地图的方法[1],其在机器人运动中的作用如图2-1。SLAM是移动机器人的底层技术，服务于上层定位、导航、避障、重建和可视化等，稀疏地图可以用于机器人定位，但导航、避障和重建需要用到稠密地图。视觉SLAM可以用于建立稀疏、半稠密和稠密地图，激光SLAM可用于建立栅格地图。那么，精确的定位和可靠的地图对移动机器人来说至关重要。目前来说，单纯的使用激光或视觉进行定位和建图，还存在很多缺陷。多传感器融合技术可以有效弥补单传感器的缺陷，提供精确的位置和可靠的地图给上层算法使用。从而，推动移动机器人的发展，使其更好地应用于更多的室内外场景，如大型商超、大型场馆、家庭和公路等，更好地为人类服务。



图2- 1 SLAM和导航流程图

SLAM技术对于机器人或其他智能体的行动和交互能力至为关键，因为它代表了这种能力的基础：知道自己在哪里，知道周围环境如何，进而知道下一步该如何自主行动。它在自动驾驶、服务型机器人、无人机、AR/VR等领域有着广泛的应用，可以说凡是拥有一定行动能力的智能体都拥有某种形式的SLAM系统。

一般来讲，SLAM系统通常都包含多种传感器和多种功能模块。而按照核心的功能模块来区分，目前常见的机器人SLAM系统一般具有两种形式：基于激光雷达的SLAM(激光SLAM)和基于视觉的SLAM(Visual SLAM或VSLAM)。另外，按使用的方法可分为几何和学习两大类如图2-2。



图2- 2 SLAM分类图

现实中的激光与视觉SLAM系统几乎都会配备惯性元件、轮机里程计、卫星定位系统、室内基站定位系统等辅助定位工具，而近年来SLAM系统与其他传感器的融合成为了一大热点，也是未来的发展方向。不同于以往基于卡尔曼滤波的松耦合融合方法，现在学界的热点是基于非线性优化的紧耦合融合。例如与IMU的融合和实时相互标定，使得激光或视觉模块在猛烈加减速和旋转时可以保持一定的定位精度，防止跟踪丢失，极大的提高定位与地图构建的稳定性。

激光SLAM和视觉SLAM各有所长，单独使用都有其局限性，而融合使用则可能具有巨大的取长补短的潜力。例如，视觉在纹理丰富的动态环境中稳定工作，并能为激光SLAM提供非常准确的点云匹配，而激光雷达提供的精确方向和距离信息在正确匹配的点云上会发挥更大的威力。而在光照严重不足或纹理缺失的环境中，激光SLAM的定位工作使得视觉可以借助不多的信息进行场景记录。

激光点云信息本身也仍有潜力可挖。在高端的远距离多线激光雷达上，返回的点云除了包含方向和距离信息，还可以加入目标点的反射率信息。当线数较多较密时，由反射率信息构成的数据可以视为一种纹理信息，因此可以在一定程度上享受视觉算法和纹理信息带来的重定位等方面的优势。这些信息一旦融入到高精度地图中，高精度地图就可以在点云\纹理两种形式间无缝切换，使得利用高精度地图的定位可以被只拥有廉价摄像头的自动驾驶汽车分享。这也是目前国外一些团队的研究方向。

同时，视觉所依赖的投影模型，蕴含着非常丰富的不同相机间的融合方法。长、短基线的单双目结合，可以在保证大尺度定位水平的同时提高中近距离的障碍探测和地图构建精度；广角鱼眼和360度全向摄像头与标准单双目的结合，使得VSLAM的覆盖范围可以进一步提升，特别适合对场景按照距离的远近进行不同精度不同速度的定位。被动视觉与深度相机的结合，催生了RGB-D SLAM，而深度相机量程的逐步扩大，将给这种特殊VSLAM带来更大的应用空间。

VSLAM的常用的开源SLAM方案如表2-1所示。除了这些开源方案之外，还能在openslam.org之类的网站上找到许多其他的工作，例如VINS、DVO-SLAM，RGBD-SLAM-V2，DSO以及一些 Kinect Fusion 相关的工作等。相机优点是结构简单，安装的方式多元化，无传感器探测距离的限制，成本低，图像信息丰富，可提取语义信息；缺点是环境光影响大、暗处或无纹理区域无法工作，运算量大，构建地图本身难以直接用于路径规划和导航，传感器动态性能还需提高，地图构建时会存在累积误差。激光SLAM优点是可靠性高、技术成熟，建立的地图直观、精度高、不存在累积误差，地图可直接用于路径规划和导航；缺点是受雷达探测范围限制，安装有结构要求，地图缺乏语义信息。激光SLAM是目前比较成熟的定位导航方案，视觉SLAM是未来研究的主流方向。

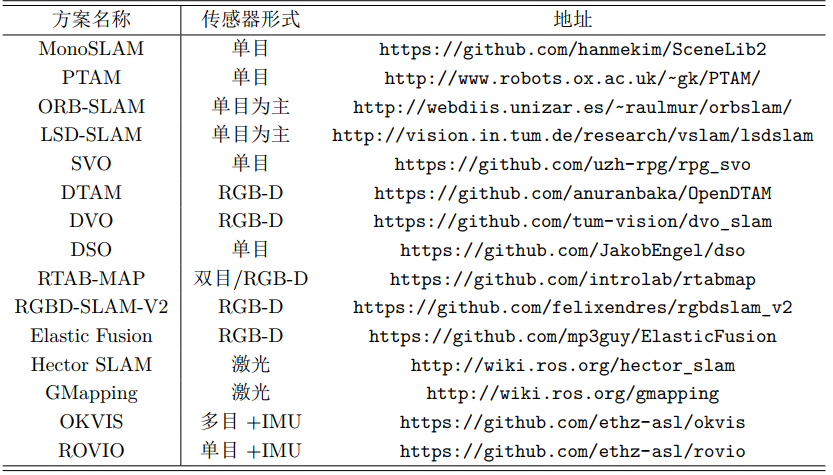


表2- 1 常用开源SLAM方案

另外，惯性传感器（ IMU）能够测量传感器本体的角速度和加速度，被认为与相机传感器具有明显的互补性，而且十分有潜力在融合之后得到更完善的 SLAM 系统。IMU 虽然可以测得角速度和加速度，但这些量都存在明显的漂移，使得积分两次得到的位姿数据非常不可靠。当运动过快时，（卷帘快门的）相机会出现运动模糊，或者两帧之间重叠区域太少以至于无法进行特征匹配，所以纯视觉 SLAM 非常害怕快速的运动。而有了 IMU，即使在相机数据无效的那段时间内，我们也能保持一个较好的位姿估计，这是纯视觉 SLAM 无法做到的。相比于 IMU，相机数据基本不会有漂移。如果相机放在原地固定不动，那么（在静态场景下）视觉 SLAM 的位姿估计也是固定不动的。所以，相机数据可以有效地估计并修正IMU读数中的漂移，使得在慢速运动后的位姿估计依然有效。当图像发生变化时，本质上我们没法知道是相机自身发生了运动，还是外界条件发生了变化，所以纯视觉 SLAM 难以处理动态的障碍物。而 IMU 能够感受到自己的运动信息，从某种程度上减轻动态物体的影响。

视觉惯性里程计（VIO，Visual Inertial Odometry）的融合都是相当复杂的。其复杂性主要来源于 IMU 测量加速度和角速度这两个量的事实，所以不得不引入运动学计算。

目前 VIO 的框架已经定型为两大类：松耦合和紧耦合[2]。松耦合是指IMU 和相机分别进行自身的运动估计，然后对其位姿估计结果进行融合。紧耦合是指把IMU 的状态与相机的状态合并在一起，共同构建运动方程和观测方程，然后进行状态估计。我们可以预见，紧耦合理论也必将分为基于滤波和基于优化两个方向。在滤波方面，传统的 EKF（Extended Kalman Filter）以及改进的 MSCKF（ Multi-State Constraint KF）都取得了一定的成果，研究者对EKF也进行了深入的讨论；优化方面亦有相应的方案。值得一提的是，尽管在纯视觉 SLAM 中优化方法已经占了主流，但在VIO中，由于IMU的数据频率非常高，对状态进行优化需要的计算量就更大，因此目前仍处于滤波与优化并存的阶段。

总之，我们看到IMU为快速运动提供了较好的解决方式，而相机又能在慢速运动下解决IMU 的漂移问题。在这个意义下，它们二者是互补的。

SLAM技术将赋予为机器人和智能体前所未有的行动能力。作为当前SLAM框架的主要类型，激光SLAM与视觉SLAM必融合其他传感器获得更高精度的定位和建立更高可靠性的地图，必将带来机器人技术和人工智能技术的真正革命，也将使得机器人从实验室和展示厅中走出来，真正服务和解放人类。

综上所述，多传感器信息融合技术在容错性、互补性和实时性等方面表现出的优势非常突出，其研究成果已经展现出巨大的研究价值和应用潜力。信息融合技术横跨多门学科， 处于所有新研究方向和新成果的交汇点。因此，延伸和扩展智能的融合理论与方法，是解决信息融合技术中融合推理和融合损失等关键问题的有效途径。

与此同时，人工智能的发展趋势是走向融合: 传统机器学习、深度学习、强化学习、知识推理、智能决策，这一趋势将伴随着智能机器人的产业化发展。可以预见，在未来利用人工智能的各种方法，以知识为基础构成多传感器信息融合将继续成为信息融合技术重要研究方向之一。多传感器信息融合技术的研究已成为近年来十分热门的课题，它在机器人系统智能化发展中起着无可比拟的作用，应用前景十分广阔。多传感器融合的slam技术也适合于山地行走的人形机器人。

1. **传感器简介**

对于山地中行走的人形机器人，室外和颠簸的条件对传感器要求较高。下面简单分析一下常见的几种传感器，以便从中选择合适的传感器。

激光的测量精度、范围和价格都高于超声波。超声波测距仪的测量精度是厘米级的，激光测距传感器的测量精度是毫米级的。超声波测距仪的测量范围通常在80米以内，而手持式激光测距传感器的测量范围最高可到200米，激光测距望远镜的测量范围更是多达几百几千米，甚至更远。激光测距精度与光强关系较大，只在一定光强范围测得距离是准确的，如激光在玻璃、反光很强或很弱的材质测得距离不准确，但超声波可以测得正确距离。

相机可分为单目相机、双目相机和深度相机。

日常生活中使用的一般是单目相机。其优点是结构简单，成本低，便于标定和识别。缺点是在单张图片里，无法确定一个物体的真实大小。它可能是一个很大但很远的物体，也可能是一个很近很小的物体。通过相机的运动形成视差，可以测量物体相对深度。但是单目SLAM估计的轨迹和地图将与真实的轨迹和地图相差一个因子，也就是尺度（scale），单凭图像无法确定这个真实尺度，所以称尺度不确定性。

双目相机代表性的产品包括Point Grey的Bumblebee相机和韩国InRobot公司的OjOcamStereo，以及小觅摄像头、Intel实感跟踪摄像头T265（适合室内）。其优点是基线距离越大，能够测量的距离就越远；并且可以运用到室内和室外。缺点是配置与标定较为复杂，深度量程和精度受到双目基线与分辨率限制，视差计算非常消耗计算资源，需要GPU/FPGA设备加速。

深度相机（Depth Camera）有多种名称，在类似LDS（laser distance sensor，激光距离传感器）的范畴内被称为Depth sensor，可以获得彩色图像时也被称为RGB-D camera，而微软公司成功普及的深度相机被称为Kinect Camera。根据获取信息的方法，Depth camera可以被分成多种类型，诸如ToF（Time of flight，飞行时间）、结构光（Structured Light）等。其优点是通过结构光或ToF的物理方法直接测量物体深度信息。缺点是测量范围窄，噪声大，视野小，易受日光干扰（红外对阳光敏感），无法测量透射材质等问题，主要用在室内，室外很难应用。

ToF方法是发送红外线后利用返回所需的时间测量距离。通常，IR发光部和收光部是成对的，并读取由每个像素测量的距离。ToF方法比后面将介绍的利用相干辐射模式的结构光方式更昂贵的原因是这种结构方面的原因提高了硬件的价格（最近，引入了使用相位差的距离计算方法，因此价格在下降）。采用ToF方式的传感器有Panasonic的D-IMAGER、MESA Imaging的SwissRanger、Fotonic的FOTONIC-B70、pmdtechnologies的CamCube和CamBoard、SoftKinectic的DepthSense DS系列以及微软最新发布的Kinect 2。

结构光方式的代表性产品是微软的Kinect和华硕的Xtion，它们使用相干辐射模式（pattern of coherent radiation，利用US20100225746专利）。此外，还有PrimeSense的Carmine和Capri以及最近的Occipital的Structure Sensor。这些传感器的共同点是都使用PrimeSense公司的PrimeSense片上系统（SoC）。使用PrimeSense公司的 PrimeSense SoC的Depth Camera是一款由一个红外投影仪和一个红外相机组成的传感器，它使用了现有的ToF方法中从未使用的相干辐射模式。该技术解决了现有ToF方式的硬件昂贵的问题和外部干扰等问题，因此备受关注。但是，苹果在2013年12月收购PrimeSense时出现了问题。PrimeSense的Carmine和Capri产品已经不再可用，而且微软的Kinect也停产，而华硕的Xtion也即将停产（库存除外）。Occipital公司的Structure Sensor是采用PrimeSense SoC的最后一款产品，目前是将此产品作为苹果的附件出售，但无法知道未来会发生什么。以低价流行的产品已经隐藏在历史中。

激光距离传感器（Laser Distance Sensor，LDS）有多种名称，比如激光雷达（LIDAR）、激光测距仪（Laser Range Finder，LRF）和激光扫描仪（Laser Scanner）。LDS是利用激光光源来测量与物体的距离的传感器。LDS传感器具有高性能、高速度和实时数据采集的优点，因此在距离测量方面有着广泛的应用。由于这些优点，它是在机器人领域被广泛使用的传感器，比如用于使用距离传感器的SLAM或用于识别人或物体识别。由于其优越的实时性能，最近还被广泛用于无人驾驶车辆。典型的产品是在室内广泛使用的Hokuyo的URG系列。多用于室外的产品有SICK和Velodyne的配有多个激光传感器的HDL系列。这些传感器最大的问题是价格。一般来说，不同产品的价格不尽相同，但大多是几千美元左右，而其中Velodyne的HDL系列是几十万元的产品。弥补这些缺点的中国产品（如RPLIDAR）以2800元左右的低价进入了市场，而近期则出现了韩国的一家公司推出的一款700多元的LDS（HLS-LFCD2）。国产的EAI产品更便宜，它的YDLIDAR X4只需499元。

IMU可靠性高，很少因为环境问题出故障。

1. **总结**

为了能在山地地形进行机械化作业，人形机器人是不错的选择。在室外和颠簸的条件下。因为室外有强光干扰，会对RGB-D相机、激光雷达这种主动发射光源的传感器造成很大的扰，因而被动式相机比较合适。而颠簸的条件，不适合旋转式激光雷达，然而，颠簸往往伴随着强烈的加减速，特别适合利用IMU与单目视觉的融合恢复单目视觉的尺度，从而使人形机器人即使在单目的情况下，也具有尺度信息。综上，使用单目相机或双目相机和IMU的融合的SLAM技术比较适合山地行走的人形机器人，香港科技大学发布的开源项目VINS中实现了这两种方案，具体实现见VINS相关论文[3]-[6]及其相关文献和源码。

参考文献

1. Cadena C , Carlone L , Carrillo H , et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age[J]. IEEE Transactions on Robotics, 2016, 32(6):1309-1332.
2. A Review of Visual-Inertial Simultaneous Localization and Mapping from Filtering-Based and Optimization-Based Perspectives.
3. Tong Q , Peiliang L , Shaojie S . VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2018:1-17.
4. Robust Initialization of Monocular Visual-Inertial Estimation on Aerial Robots.
5. Li P , Qin T , Hu B , et al. Monocular Visual-Inertial State Estimation for Mobile Augmented Reality[C]// 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE Computer Society, 2017.
6. Qin T , Li P , Shen S . Relocalization, Global Optimization and Map Merging for Monocular Visual-Inertial SLAM[J]. 2018.