

王 皓, 曹 静, 胡佳楠, 等. 基于 SfM-MVS 的作物三维重建: 挑战与创新[J]. 江苏农业学报, 2024, 40(9): 1768-1776.
doi: 10.3969/j.issn.1000-4440.2024.09.021

基于 SfM-MVS 的作物三维重建: 挑战与创新

王 皓^{1,2}, 曹 静², 胡佳楠², 潘嫒嫒², 毕 然², 孙 挺², 张文宇^{1,2}

(1. 江苏大学农业工程学院, 江苏 镇江 212013; 2. 江苏省农业科学院无锡分院, 江苏 无锡 214174)

摘要: 三维重建技术可提供更全面、准确的作物表型信息, 已成为智慧农业和智慧育种领域的研究热点。基于多目视觉法的三维重建技术因其低成本、高精度的特性而被广泛应用于作物三维重建, 其中运动恢复结构-多视角立体视觉(SfM-MVS)流程是其主流实现方法。本文对不断发展的 SfM-MVS 流程进行了技术体系梳理, 并从当前作物三维重建的角度出发, 总结了 SfM-MVS 所面临的挑战和已有创新之处。本研究可为基于 SfM-MVS 的作物三维重建技术集成提供有力支撑, 并为关键技术突破提供重要参考。

关键词: 三维重建; 运动恢复结构; 多视角立体视觉; 作物; 表型

中图分类号: S126 **文献标识码:** A **文章编号:** 1000-4440(2024)09-1768-09

Three-dimensional crop reconstruction based on structure from motion-multiple view stereo (SfM-MVS): challenges and innovations

WANG Hao^{1,2}, CAO Jing², HU Jianan², PAN Yuanyuan², BI Ran², SUN Ting², ZHANG Wenyu^{1,2}

(1. College of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China; 2. Jiangsu Academy of Agricultural Sciences Wuxi Branch, Wuxi 214174, China)

Abstract: Three-dimensional (3D) reconstruction technology can provide more comprehensive and accurate crop phenotypic information, which is a research hotspot in the field of intelligent farming and intelligent breeding. The 3D reconstruction technology based on multi-view vision is widely adopted for its low cost and high accuracy. The structure from motion-multiple view stereo (SfM-MVS) pipeline is the primary method used for this technology. In this paper, the technical system of the evolving SfM-MVS process was sorted out, and the challenges and innovations of SfM-MVS were summarized from the perspective of current crop three-dimensional reconstruction. This study can provide robust support for the integration of crop 3D reconstruction technology based on SfM-MVS, and provide an important reference for key technology breakthroughs.

Key words: three-dimensional reconstruction; structure from motion; multiple view stereo; crop; phenotype

近年来, 作物表型在农业研究和生产中扮演着

日益重要的角色。精确的表型数据能够反映作物的生理生长状态, 为农业生产智慧决策提供关键信息。同时, 高通量作物表型技术可用于高产、优质种质的筛选以及表型-基因型关联分析, 从而推动智慧育种进程^[1]。然而, 传统的表型获取、分析耗时费力, 已无法满足高通量需求。随着计算机视觉技术的迅速发展, 基于视觉技术的作物表型平台成为一种有效的解决方案。这些平台可以分为利用可见光成像技术进行二维测量和依赖三维重建技术进行三维测量

收稿日期: 2024-07-09

基金项目: 国家重点研发计划子课题(2022YFD2001001、2022YFD200-1005); 国家自然科学基金项目(32201664、31871522); 江苏省重点研发计划项目(BE2023323); 江苏省农业自主创新基金项目[CX(21)2042]; 无锡市财政项目(33212303)

作者简介: 王 皓(2000-), 男, 江苏徐州人, 硕士研究生, 研究方向为作物三维重建。(E-mail) 454618365@qq.com

通讯作者: 孙 挺, (E-mail) drsunting@163.com

2类^[2]。在作物结构复杂且叶片遮挡严重的情况下,在二维平台上只能获得简单的表型特征,如叶长、叶面积。二维投影存在限制性,相比之下,在三维平台上能够准确地重建点云模型并获取更全面、准确的表型特征。

基于视觉的三维重建技术主要通过视觉传感器获取场景真实信息,经过视觉处理技术或投影模型获得物体的三维信息,是一种利用二维投影恢复三维信息的计算机技术^[3]。根据数据获取方式不同,基于视觉的三维重建技术可分为主动视觉法和被动视觉法2类^[4]。主动视觉法需要向场景中发射结构光源,并计算提取光源在场景中的投影信息,从而检测目标位置并进行测量。常见的技术包括激光扫描法、结构光法、飞行时间(ToF)技术和体感(Kinect)技术等。激光扫描法主要使用激光扫描仪和激光雷达发射束对目标表面进行扫描,测量反射或散射激光,并计算物体表面距离。该方法可以生成与实际大小相当的密集点云,具有测量速度快、准确性高、距离远等优势;然而边缘检测效果相对较差且设备价格昂贵,限制了其在农业生产中的大规模应用^[5]。结构光法通过在物体表面和参考平面上投射具有一定规律的结构图案,并采集图像以利用三角测量原理等方法获取物体表面深度信息。该方法的精度和分辨率都较高,但易受到环境照明条件及作物表面纹理的影响,在室内静态环境下较适用^[6]。ToF技术基于飞行时间原理,通过向观测对象连续发射光脉冲来测量光脉冲的飞行(往返)时间从而确定目标距离,进而获得物体深度信息。ToF技术的主要优势是只使用单一视点来计算深度,保证了遮挡和阴影下的鲁棒性,并保留了锐利的深度边缘;缺点是分辨率低,在强光下容易受到干扰,且测量距离短^[7]。Kinect技术最初是由微软公司开发的基于结构光和ToF技术的三维扫描技术,其优点是成本低、操作简便和实时性强,但Kinect技术的深度测量精度较低,生成的点云数据较为粗糙,对环境光照条件敏感,在强光、反射面或黑暗环境中性能受限^[8]。

被动视觉法通过获取外部光源的反射来进行三维测量,相比主动视觉法,被动视觉法不受环境和设备因素的限制,并且成本较低、应用范围更广。根据摄像头数量可将被动视觉法分为单目视觉法、双目视觉法和多目视觉法^[2]。单目视觉法仅使用单个

视觉传感器捕捉图像,并通过各种计算技术推断深度信息,具有系统结构简单、设置校准容易等优点,但其深度信息来自推断,导致精度较低且难以解决深度模糊和遮挡问题。双目视觉法使用2个相隔一定距离的摄像机捕捉图像,通过捕获待测场景或物体的不同视角的图像,分析图像中匹配特征点的视差来计算物体的深度^[9]。与单目视觉法相比,双目视觉法的深度感知更佳,通过比较不同视角的图像可以解决一些遮挡问题,立体匹配算法比单目深度估计算法更简单且研究较为成熟,但是有效的深度测量范围受到摄像头基线距离和视差测量分辨率的限制,此外对纹理的要求也较高。多目视觉是双目视觉的延伸,是在其基础上增加一台或多台摄像机辅助测量^[10],从而获得不同角度下同一物体的多视角图像。多目视觉法具有更好的精度和鲁棒性,能够更好地处理遮挡,适合高精度应用,但成本、复杂度和数据管理方面的挑战也更大^[11]。

近年来,随着工业级相机成本的不断下降、计算机算力的不断提升,基于多目视觉法的作物三维重建表型平台已成为主流形式之一。许多商业化三维重建表型平台基于多目视觉法建立,如MVS-Pheno^[12-13]、3DPhenoMV^[14]等。此外,得益于较为简单的系统架构,一些自搭建三维重建表型平台已被广泛应用于玉米^[12,15-16]、小麦^[17]、水稻^[18]、高粱^[19-20]、大豆^[21-23]、鹰嘴豆^[24]、马铃薯^[25]、甜菜^[26]、番茄^[27]、豇豆^[28]等作物的三维表型研究中,大幅降低了技术门槛。

目前,运动恢复结构-多视角立体视觉(SfM-MVS)流程是基于多目视觉法的三维重建主流解决方案。与近年来新兴的神经辐射场(NeRF)方案相比,SfM-MVS具有计算成本低、计算效率高等优点,同时重建出的三维点云密度高,能较好地还原场景信息。由于SfM-MVS在处理光照效果时有一定局限性,所以SfM-MVS更适合于室内三维重建,该流程包括基于运动恢复结构(SfM)流程的稀疏重建和基于多视角立体视觉(MVS)流程的密集重建。目前,一些成熟的商业化三维重建软件(如Agisoft Photoscan/Metashape、3DF Zephyr、ContextCapture、Pix4Dmapper)以及开源软件(COLMAP、OpenMVG、VisualSfM)都采用了SfM-MVS这一流程。随着计算机视觉技术和人工智能技术的不断发展,SfM-MVS流程也在持续迭代更新。本文拟从作物三维重建的

角度对当前主流 SfM-MVS 流程进行技术体系梳理,总结其面临的挑战和已有创新之处,以期为基于 SfM-MVS 的作物三维重建技术集成提供有力支持,并为关键技术突破提供重要参考。

1 SfM-MVS 流程

SfM 是通过多张二维图片来重建三维场景的过程^[29]。通过 SfM 流程可以得到稀疏三维点云模型,

这种点云模型密度较低,还需对其进行稠密估计,恢复稠密的三维点云。稠密重建是在相机位姿已知的前提下,逐像素地计算图像中每一个像素点对应的三维点,得到场景物体表面密集的三维点云的过程,可以通过 MVS 流程实现。通过结合 SfM 和 MVS 这 2 种方法,可以利用 SfM 流程获取的相机位姿信息以及稀疏点云信息,提高 MVS 流程获得的稠密点云的精度^[30],从而实现一个完整的三维重建过程(图 1)。

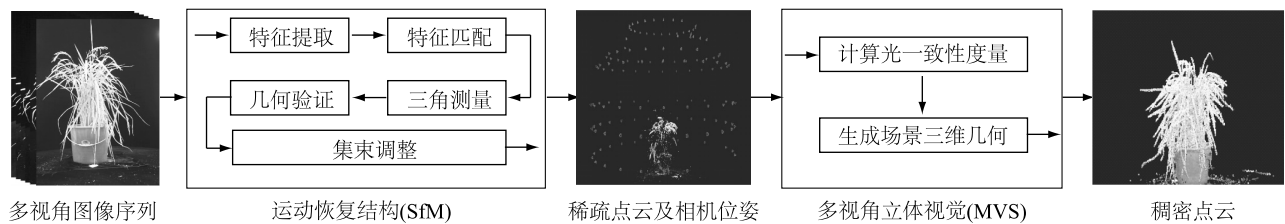


图 1 基于 SfM-MVS 流程的三维重建过程

Fig.1 Three-dimensional reconstruction process based on structure from motion-multiple view stereo (SfM-MVS) pipeline

1.1 SfM 流程

SfM 流程的实现一般包括特征点提取、特征匹配、几何验证、三角测量和集束调整等重要步骤^[31]。

1.1.1 特征点提取 特征点(也称为兴趣点或关键点)提取是整个 SfM 流程的基础,需要对每张图片提取出一些特征点,并计算它们的描述子。特征检测器是一种检测图像中特征点的算法,通常以角、斑点、边缘、连接点和线条等形式进行检测。经典的特征检测器有 HARRIS、SHI-TOMASI、尺度不变特征变换(SIFT)和加速稳健特征(SURF)等。近年来,一些更快的特征检测器已被开发出来,如加速分割测试特征(FAST)、定向 FAST 和旋转 BRIEF(ORB)、二进制鲁棒不变可扩关键点(BRISK)、加速分割测试(KAZE)和加速 KAZE 特征(AKAZE)。经过特征检测后,所观察到的特征会以一种逻辑上不同的方式描述其相邻像素所具有的独特模式,这个过程被称为特征描述。特征描述子算法是对检测到的特征点周围像素模式进行数学表示的一种方法,用于在不同图像中匹配相似的特征点。特征描述子算法主要分为基于字符串(浮点数)的描述子(SIFT、SURF、KAZE 等)算法和二进制描述子(AKAZE、ORB、BRISK 等)算法。基于字符串的描述子通常具有更高的描述性和准确性,但计算和存储成本较高,适用于精确的特征匹配和重建任务。二进制描述子由于其高效的生成和匹配速度,适合于实时应用和

资源受限的场景,如移动设备或嵌入式系统。

目前存在不同组合的特征点检测器和描述子。有些既是检测器同时也是描述子,例如 SIFT/SURF 既包括特征点检测器,也包括描述子;一些特征点检测器需要与它们指定的特征描述子算法一起使用,如 KAZE/AKAZE 描述子仅适用于 KAZE/AKAZE 特征点检测器;有些组合会出现问题,例如 SIFT 检测器和 ORB 描述子不能一起工作。目前主流的特征点检测器/描述子组合有 SIFT、SURF 和 ORB。SIFT 的主要优点是其对尺度、旋转、仿射变换等具有较好的不变性。由于 SIFT 基于浮点内核计算特征点,因此通常认为 SIFT 算法检测的特征在空间和尺度上定位更加精确,所以在要求匹配极度精准且不考虑匹配速度的场合可以考虑使用 SIFT 算法,但 SIFT 算法计算量较大,不适合实时应用。SURF 是 SIFT 的高效变种,也具有尺度不变性的特征,但 SURF 使用海森矩阵的行列式值进行特征点检测,并用积分图加速运算,速度相对 SIFT 提升明显,所以在一些实时视频流物体匹配上有较多的应用。ORB 是在著名的 FAST 特征点检测和二进制鲁棒独立基本特征(BRIEF)描述子算法的基础上提出的,能够在保持良好性能的同时实现较高的计算效率。相比 SIFT 和 SURF 等具有尺度不变性的算法,ORB 在处理图像中尺度变化时的表现不佳。

1.1.2 特征匹配 特征匹配需要计算 2 个描述子

之间的距离,其差异可以被转换成一个单一的数字,作为一个简单的相似性度量。曼哈顿距离(L1-norm)、欧几里德距离(L2-norm)和汉明距离(Hamming distance)是常见的距离度量方法。二进制描述子需要选择汉明距离,而字符串描述子需要选择欧几里德距离。寻找匹配对的方法有蛮力匹配(BFM)和快速最近邻搜索(FLANN)。蛮力匹配是最简单的直接匹配方法,对于给定的一组特征描述子,通过计算它与另一组特征描述子之间的距离来寻找最近邻或者最优匹配,适用于规模较小的特征集合,在局部特征描述子(如 SIFT、SURF、ORB)的匹配中较为常见。FLANN 则适合需要在大型数据集中进行特征匹配的场景。选择匹配对的方法有交叉检查、最近邻方法(NN)和 K 最近邻方法(K-NN)。交叉检查一般只用于蛮力匹配,K-NN 常用于局部特征描述符的匹配过程。

1.1.3 几何验证 通过特征点匹配,可以得到不同图片之间的几何变换关系。通过这些变换关系,可以估计得到相机的姿态,包括相机的位置和方向。如果一个有效的变换能够在图像之间映射足够数量的特征,那么这些变换就可以被视为几何验证。由于匹配的对应关系经常受到离群值的污染,因此需要剔除不符合几何约束的匹配对。最常用的几何验证算法是随机抽样一致性算法(RANSAC),属于迭代算法。它通过从上一步骤得到的匹配点中随机选择数据子集(例如使用归一化的八点法)来估算基础矩阵和本质矩阵的参数。几何验证可以显著提高匹配的质量,进而提高后续三维重建的准确性和鲁棒性。

1.1.4 三角测量 三角测量主要通过几何验证中基础矩阵以及本质矩阵分解得到的相机参数以及特征点在图像中的投影位置,计算出这些特征点在三维空间中的坐标。通常,三角测量包含 3 种算法策略,即增量式^[32]、层级式^[33]以及全局式^[34]。在增量式 SfM 中,算法会逐步处理每个新的图像,一次添加一个图像,并不断更新相机姿势和场景结构的估算结果^[32]。该算法通常从初始视觉数据集开始,然后逐步引入新的图像。每个新图像都会与先前处理的图像进行匹配,并通过三角测量或其他方法来估计新的相机姿势和场景结构。增量式方法可以在处理大型数据集时更好地利用内存和计算资源,是目前最常见的 SfM 类型。然而,增量式方法可能会受到误差累积的影响,并且对初始视觉数据集的选择和

顺序敏感。层级式 SfM 通过分层的方式处理图像数据和场景结构,以便在不同的分辨率和精度级别上进行处理^[33]。这种方法通常是先在低分辨率或粗略的级别上处理图像数据和相机位姿,然后逐步提高分辨率或精度以提升估算结果的精确度。层级式方法能够加快算法的收敛速度,并且对噪声和不确定性具有一定的鲁棒性。此外,它还能够有效减少处理大型数据集时所需的计算资源。全局式 SfM 是一种以整个数据集为基础的方法,它在整个数据集上同时优化相机姿势和场景结构。这种方法通常会建立一个全局的优化问题,其中相机姿势和场景结构之间的关系被建模为一个大型优化策略,例如通过最小化重投影误差来优化。全局式方法在完成所有两两视图的重建后,统一进行集束调整(BA)的优化,效率高,但是重建精度一般^[34]。

1.1.5 集束调整 在 SfM 流程中,当从多张图片中提取出特征点并估计出初始的相机位置和三维点时,这些估计通常是不精确的。误差主要来源于图像中点定位不精确、匹配噪声和标定不精确,这些误差随着视图数量的增加而累积,这种效应称为漂移。减少漂移的一种方法是改进相机姿势和三维点位置,可以通过集束调整进行细化。集束调整的目标是通过最小化重投影误差,对所有相机参数和三维点位置进行调整,以使得在所有图像中观察到的特征点的投影误差最小化。作为多视图几何系统中的核心组件之一,集束调整通常被用于近似初始场景估计和消除增量重建过程中的漂移。此优化模型通常是非线性的,并且 LM(Levenberg-Marquardt)算法已被证明是解决此模型较好的方法。

1.2 MVS 流程

多视角立体视觉(MVS)假设每张图像的相机参数已知,通过利用一组图像及其对应的相机参数来重建场景的三维几何结构,从而恢复密集点云。根据场景表示不同,MVS 算法通常可分为基于体素的 MVS 算法、基于点云扩散的 MVS 算法和基于深度图融合的 MVS 算法。

1.2.1 基于体素的 MVS 算法 体素是三维计算机图形学中的基本单位,类似于二维图形中的像素。在基于体素的 MVS 算法中,首先需要在对象周围初始化一个 3D 体素网格。网格的大小和分辨率取决于对象的预期大小和所需的细节要求。对每个体素进行光一致性度量时,将其投影到每张图像上并比

较像素值以计算成本。通过成本来表示该体素被对象占据的可能性,并使用成本体积将其分类为占用或空置状态。基于体素的 MVS 方法可以生成规则点云,并易于提取 Mesh 网格;但精度取决于体素粒度,难以应用于遮挡较多以及较大的场景。

1.2.2 基于点云扩散的 MVS 算法 基于点云扩散的 MVS 算法通过对特征点进行三维重建,随后将结果迭代传播到相邻区域,从而重建出三维形状^[35]。初始种子点可以由 SfM 流程生成,然后对稀疏特征点进行深度和法向量的重建,在扩张过程中对已重建三维点的邻域进行匹配,最后通过光度一致性约束和可视性约束去除噪点。其中,最为典型的的就是基于补丁(Patch)的 MVS(PMVS)^[36]。PMVS 基于输入图像中检测到的特征点生成补丁,补丁是以特征点为中心的局部图像块,用于表示每个特征点周围的局部视觉信息。对于每个补丁,PMVS 在不同视角的图像间进行匹配。这包括对每对图像中的相似补丁进行特征匹配和深度一致性验证,以确定每个特征点精确的三维位置。通过多视角几何和像素一致性约束,PMVS 将匹配的补丁中的信息合并,生成高密度的三维点云。基于点云扩散的 MVS 具有较高的精度和稠密性,能够处理视角变化大和几何形状复杂的场景,但由于算法的复杂性导致计算成本较大。

1.2.3 基于深度图融合的 MVS 算法 基于深度图融合的 MVS 算法从多视角图像中估算每个视角的深度图,然后对其进行集成以重建目标的三维形状。首先,为每一幅图选择领域图像构成立体图像组,常用方法有全局视图选择和局部视图选择。全局视图选择根据视差、场景内容和尺度等寻找一组用于立体匹配的候选相邻视图。局部视图选择则在全局视图选择结果的基础上再次进行筛选,通过计算图像之间的相似度进行评估优化。其次,对于选定的视图,通过平方差之和(SSD)、绝对差之和(SAD)、归一化互相关(NCC)等实现立体配准和融合。基于深度图融合的 MVS 算法具有并行计算能力,适用于处理大场景海量图像,并且得到的点云数量较多。因此,在当前开源软件和商业软件中普遍采用这类方法。

2 作物三维重建过程的难点以及 SfM-MVS 技术的革新

尽管 SfM-MVS 技术在文物保护、城市规划、地形测绘、影视、游戏和工业品控等多个领域中的应用

取得了成功,但在作物三维重建中的应用仍然面临诸多挑战。作物具有复杂而多样的形态结构,作物之间以及器官和组织之间的相互遮挡会导致部分细节丢失,从而对重建效果产生影响;某些作物表面纹理特征不足,使得图像匹配和对齐更加困难,进而导致重建结果不准确;风等环境因素的干扰会增加图像对齐和匹配的难度;光照在植物冠层中的反射和折射会产生阴影和反光现象,进一步导致图像质量不稳定。此外,在大规模作物三维重建过程中需要处理大量图像数据,并且对计算与存储资源需求极高,这对硬件设备与算法优化提出了更高要求。这些挑战使得将 SfM-MVS 技术应用于作物三维重建比其他领域更为复杂、困难,需要进行进一步研究和技术改进来解决这些问题。

2.1 图像预处理

目前,典型的多视角作物三维表型平台主要部署在室内环境中,由相机、支撑臂、旋转平台及其驱动部件等构成。在作物多视角图像的采集过程中,环境、转盘和栽培容器等冗余的背景信息是不可避免的。尽管这些无效信息可以通过不同方法在三维点云重建完成后去除,但可能会降低作物三维点云重建的效率和精度。因此,在 SfM-MVS 工作流程中进行图像预处理是非常必要的。

经典的图像预处理方法包括噪声去除和背景分割。Sandhu 等^[18]采用色调-饱和度-亮度(HSV)颜色空间对水稻红-绿-蓝(RGB)图像进行转换,并通过颜色阈值去除背景,从而减少了三维重建过程中需要处理的特征数量和计算时间。Zhu 等^[23]利用小波变换阈值去噪法对大豆植株图像序列进行降噪,然后运用蓝屏匹配技术实现背景扣除,以便将预处理后的图像应用于后续 SfM-MVS 三维重建流程中。Xiao 等^[26]使用 3DF Zephyr Aerial 软件自带的掩膜技术,能够精确地分离出甜菜根和黑色背景作为重建对象,只有白色部分被选为 SfM-MVS 输入。Sun 等^[21]对大豆植株多角度图像进行掩膜处理,消除与大豆植株无关的所有背景和校准垫区域,仅保留完整的大豆植株信息供后续三维重建使用。Yang 等^[37]通过在 U²-net 网络上训练模型,将图像中的每个像素分类为植物或非植物,并成功开发出适用于多分蘖、形态复杂的结球甘蓝和羽衣甘蓝植物的图像分割模型,在最终点云重建过程中只选择了含有植物信息的像素。

此外,He 等^[22]提出了一种改进的检测和匹配策略。通过裁剪原始图像来增加感兴趣区域(ROI)的比例,有效降低了图像分辨率,并增加了大豆植物在整个图像中的比例,从而减少了特征检测的计算量。通过对大豆多视角图像序列原图像的 ROI 进行提取后,使用 SfM-MVS 流程进行大豆植物的三维重建,稠密点云生成率提高了 81.62%。Tian 等^[38]将包含小麦籽粒的核心区域定位为图像中心,并自动裁剪以获得固定像素大小的图像,从而提高了后续小麦籽粒三维重建任务的准确性和效率。此外,在解决拍摄过程中对焦不准导致纹理模糊和细节不清晰问题时,研究人员采用基于 T-transformer 网络的超分辨率技术来提升图像分辨率。该方法能够恢复小麦籽粒纹理信息且不显著增加捕获时间和工作量。与单纯基于原始图像获得的结果相比,采用超分辨率重建可以使结果结构相似性(Structural similarity, SSIM)提高 38.96%。

2.2 特征检测优化

目前,大部分多视角三维重建平台具有固定的作物位置、相机与相机之间的相对位置以及固定的旋转角度。在三维重建过程开始时,研究者可以向 SfM-MVS 流程提供一组植物图像序列、相机内部参数和畸变系数的初始估计值,以减少计算误差、加速算法收敛并提高重建质量。Yang 等^[37]利用多视角相机检测了一个黑白相间棋盘格的角点,并获取其像素坐标,然后根据已知网格大小和世界坐标系统原点来计算它们的物理坐标,从而确定树莓派相机内部参数矩阵。Wang 等^[25]使用 Matlab 中的立体摄像机校准工具箱和类似棋盘格实现了立体校准,并得到相机内部方向和相机之间的相对方位 2 组参数,用于后续三维重建。

针对不同作物特性选择或改进特征检测器和描述符算法可提高特征检测准确性和效率。Yang 等^[37]在 OpenMVG 框架基础上改进了 SfM 流程的算法,并提出了 IVOP&AKAZE-SfM 算法。增量式视觉里程计与建图(IVOP)能够通过增量方式优化视点,在每次增加新视点时最大化地利用现有数据,从而提高三维重建精度。处理图像序列时,IVOP 不断优化视点位置和姿态以减少累积误差。与 SIFT 和 SURF 使用高斯平滑不同,AKAZE 使用非线性扩散方程构建尺度空间,在捕获形态复杂的甘蓝的精细特征方面表现出卓越效果。IVOP&AKAZE-SfM 较

OpenSfM 误差降低 58.05%,时间消耗降低 44.77%;IVOP&AKAZE-SfM 较 Colmap 误差降低 55.94%,时间消耗降低 48.40%;IVOP&AKAZE-SfM 较 OpenMVG 误差降低 60.20%,时间消耗降低 38.42%。

随着人工智能技术的发展,越来越多的深度学习方法被集成到 SfM 流程中。Detone 等^[39]设计了 1 个名为 SuperPoint 的全卷积神经网络架构,在全尺寸图像上进行单次前向传递以实现关键点检测和描述子提取。相较于传统特征点检测方法,SuperPoint 具备更高的计算效率,适用于实时应用场景。Sarlin 等^[40]提出了一种名为 SuperGlue 的深度学习方法用于特征匹配,并与 SuperPoint 结合使用以提高关键点匹配的准确性和鲁棒性。SuperGlue 利用图注意力机制和图神经网络(GNN)建立 2 个图像之间的匹配关系,并通过图神经网络进行全局优化,从而提升匹配精度,在纹理相似或重复场景中表现出色,并且对视角变化、光照变化等复杂条件有良好的适应能力。通过端到端训练方式,该模型能够更好地适应实际应用中的变化。Liu 等^[17]将 Pixel-perfect SfM 优化组件作为 SfM 重建基础模型,并采用 SuperPoint 进行特征点检测和描述;同时使用 SuperGlue 进行图像对之间的匹配,在小麦植物特征识别任务中成功找到正确匹配并排除错误匹配。

2.3 深度图融合优化

近年来,基于深度学习的多视角立体视觉取得了快速发展。MVSNet 及其改进版是 MVS 流程中突出的优化案例。MVSNet 采用卷积神经网络(CNN)结构,以端到端的方式学习从多个视角获取的图像中提取场景的深度信息,并通过学习视图之间的对应关系和光度一致性来提高深度图的精确性和稳定性。Liu 等^[17]将 RepC-MVSNet 算法用于小麦三维点云重建,该算法考虑了由光度不一致引起的光度损失误差,还通过引入渲染一致性替代原始无监督方法减少对光度一致性依赖,并结合神经身体渲染作为参考。此外,还引入了视角合成损失来构建 RGB 监督信号,以消除因不同方向视图而导致的光度损失误差,并增强模型泛化能力。Tian 等^[38]使用 Vis-MVSNet 对小麦籽粒进行三维重建,这是一种端到端网络结构,考虑了像素可见性信息,在多视图立体匹配网络中明确推断并整合像素级遮挡信息,并利用成对不确定性地图和深度地图联合推断作为多视图代价体积融合中的加权指导,有效抑制遮挡像素带来的负面影响。此

外,MVSNet 等深度学习方法利用图形处理器(GPU)进行大规模像素级深度估计和立体匹配,在处理大规模图像数据时具备高效能力。

2.4 SfM-MVS 整合优化

基于 SfM-MVS 流程框架,已经出现了整合多个算法库的开源软件并用于作物三维重建。其中最为常见的有 VisualSfM^[16,27]、OpenMVG+OpenMVS^[13]以及 Colmap^[41]。VisualSfM 整合了基于 GPU 的 SIFT (SiftGPU)、多核集束调整以及线性时间增量式运动结构。利用多核并行计算进行特征检测、匹配和集束调整,运行速度较快。该程序同时集成了 PMVS/CMVS 工具链用于稠密重建。OpenMVG 和 OpenMVS 是 2 个常用的开源库,提供了 1 套强大的接口,通常结合使用进行从图像到三维模型的重建。OpenMVG 负责特征检测、匹配和初步的三维结构重建,而 OpenMVS 则负责密集点云生成、网格重建和纹理映射。MVS-Pheno V2 三维表型平台采用 OpenMVG+OpenMVS^[13]。Colmap 支持图形用户界面和命令行界面,使用 SIFT 算法进行特征提取,并提供中央处理器(CPU)和 GPU 选项,并且还提供广泛功能匹配选项,在无人机田块及更大尺度上被广泛应用于重建任务^[41-42]。Liu 等^[43]比较 COLMAP、COLMAP+PMVS、VisualSfM、Meshroom 和 OpenMVG+MVE 5 种 SfM-MVS 开源整合框架重建 12 个玉米基因型根系的质量,结果发现 COLMAP 和 COLMAP+PMVS 获得了很好的视觉效果,在模型完整性方面,没有出现明显的根断连或根系部分缺失现象。与 VisualSfM 相比,OpenMVG+MVE 能显示更精细的细节,但不提供每个点的颜色信息。在作物三维重建过程中,应根据作物冠层特征、算力以及关注性状来选择适合的 SfM-MVS 整合框架。

2.5 SfM-MVS 流程评估

无论是 SfM-MVS 的整合优化还是单独算法的优化,不同算法间的比较都需要建立在相同的评估标准上。通常情况下,MVS 流程往往基于 DTU 数据集进行评价,DTU 数据集是一种公开数据集,由丹麦技术大学提供,并且其稠密结构真值通常基于激光雷达或深度相机获取整个场景点云。点云重建评估指标包括准确性、完整性、精度和召回率等,用于衡量重建结果与真值之间的差距以及重建算法的整体性能。

对于 SfM-MVS 流程综合评估体系而言,并没有固定指标。Yang 等^[37]提出了一种联合评估方案,

通过点云数量、平均轨迹长度、配准图像数、平均重投影误差和时间效率等指标对点云模型的重建质量进行定量分析。观察到更多数量的点云则表明算法提取到更丰富的三维结构特征和细节信息;较小的平均轨迹长度说明算法提取到更丰富的特征;配准图像数量表示在重建过程中利用信息的多少;较小的平均重投影误差意味着高精度点云;时间效率作为一个关键指标通常代表完成任务所需持续时间,持续时间越短则表示算法越有效。Liu 等^[43]使用 CloudCompare 软件计算 3D 点云模型中的点数,并通过计算每个半径为 R 球体内邻居数来估计点云表面密度。Tian 等^[38]在三维重建时采用超分辨率技术,并使用峰值信噪比 (PSNR) 和结构相似性指数 (SSIM) 作为量化评估指标。

许多作物表型研究并未对 SfM-MVS 流程进行评估,仅对后续提取的表型参数进行整体评价。例如,Xu 等^[44]通过决定系数 (R^2)、均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 来量化三维表型与实际测量值之间的差异。Wu 等^[13]则通过 MAPE 评估三维表型是否能够保留叶尖和边缘特征,并利用 R^2 和 RMSE 评估三维表型与手动测量得到的株高以及数字化仪获取的叶长、叶宽测量值之间的差异。

3 展 望

目前,基于 SfM-MVS 流程的作物三维重建主要应用 RGB 图像领域,面临作物形态结构复杂、环境与背景干扰大、计算成本高等一系列挑战,已有研究在图像预处理、特征检测优化、深度图融合优化等方面开展了技术创新,同时对 SfM-MVS 流程进行了整合优化,并进行了系统评估。

随着传感技术的迅速发展,SfM-MVS 流程将从 RGB 图像领域逐渐扩展到高光谱、热红外、荧光等其他光学成像领域,这也带来了新的挑战。传统的 SfM-MVS 流程在处理高光谱图像时面临困难,因为这些图像包含数十个甚至上百个波段。不同波段图像中提取的特征点会生成不同的三维模型,其尺度和点集不一致,这些模型不能直接合并形成一个完整的三维模型。如果尝试合并从不同波段捕获的所有关键点进行三维重建,嵌入在图像中的精细结构和光谱信息可能会丢失^[45]。优化当前的 SfM-MVS 流程,使其能够实现不同波段下特征点的检测、描述

和匹配以及后续的融合,是基于 SfM-MVS 的作物三维重建面临的挑战和创新的方向之一。

与此同时,多源传感器融合也为基于 SfM-MVS 的作物三维重建提供了新的机遇。融合更加丰富的传感器信息,可以提高三维重建的精度。例如,将 RGB 图像与深度传感器数据结合,可以得到更准确的深度信息,从而提高重建模型的细节水平。多源传感技术还可以提高系统对单一传感器失效或噪声干扰的鲁棒性。例如,结合视觉和 LiDAR 数据可以在光照条件不佳的情况下仍保持良好的性能。尽管多源传感技术在 SfM-MVS 流程中带来了许多创新,但也需要克服诸多挑战。为了解决多源数据融合中的问题,新的算法和模型,如深度学习模型、多模态数据融合算法等,需要进一步耦合到 SfM-MVS 流程。未来,随着多源传感技术和 SfM-MVS 流程算法的不断发展,作物三维重建将变得更加精准和高效,能够更全面地捕捉和理解作物的复杂结构和动态变化,进而推动智慧农业和智慧育种的进一步发展。

参考文献:

- [1] 张建华,姚 琼,周国民,等. 作物农艺性状与形态结构表型智能识别技术综述[J]. 智慧农业(中英文),2024,6(2):14-27.
- [2] 周静静,郭新宇,吴 升,等. 基于多视角图像的植物三维重建研究进展[J]. 中国农业科技导报,2019,21(2):9-18.
- [3] EBRAHIMNEZHAD H, GHASSEMIAN H. Robust motion from space curves and 3D reconstruction from multiviews using perpendicular double stereo rigs[J]. Image and Vision Computing,2008,26(10):1397-1420.
- [4] ISGRO F, ODOE F, VERRI A. An open system for 3D data acquisition from multiple sensor[C]//Institute of Electrical and Electronics Engineers. Proceedings of the Seventh International Workshop on Computer Architectures for Machine Perception, July 4-6,2005. Palermo, Italy:IEEE,2005:120-125.
- [5] GOEBEL W, KAMPA B M, HELMCHEN F. Imaging cellular network dynamics in three dimensions using fast 3D laser scanning[J]. Nature Methods,2007,4(1):73-79.
- [6] ROCCHINI C, CIGNONI P, MONTANI C, et al. A low cost 3D scanner based on structured light[J]. Computer Graphics Forum,2001,20(3):299-308.
- [7] PARK J, KIM H, TAI Y W, et al. High quality depth map up-sampling for 3D-TOF cameras[C]//Institute of Electrical and Electronics Engineers. Proceedings of the IEEE International Conference on Computer Vision, November 06-13, 2011. Barcelona, Spain:IEEE,2011:1623-1630.
- [8] KHOSHELHAM K, ELBERINK S O. Accuracy and resolution of kinect depth data for indoor mapping applications[J]. Sensors,2012,12(2):1437-1454.
- [9] 张文明,刘 彬,李海滨. 基于双目视觉的三维重建中特征点提取及匹配算法的研究[J]. 光学技术,2008,34(2):181-185.
- [10] BAILLARD C, ZISSERMAN A. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images[C]//International Society for Photogrammetry and Remote Sensing. Proceedings of the International Archives of Photogrammetry and Remote Sensing, 1999. Amsterdam, Netherlands:ISPRS,2000:56-62.
- [11] PARK H, LEE H, SULL S. Efficient viewer-centric depth adjustment based on virtual fronto-parallel planar projection in stereo 3D images[J]. Ieee Transactions on Multimedia,2014,16(2):326-336.
- [12] WU S, WEN W L, WANG Y J, et al. MVS-Pheno:a portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction[J]. Plant Phenomics,2020,2020:1848437.
- [13] WU S, WEN W, GOU W, et al. A miniaturized phenotyping platform for individual plants using multi-view stereo 3D reconstruction[J]. Frontiers in Plant Science,2022,13:897746.
- [14] DAS CHOUDHURY S, MATURU S, SAMAL A, et al. Leveraging image analysis to compute 3D plant phenotypes based on voxel-grid plant reconstruction[J]. Frontiers in Plant Science,2020,11:521431.
- [15] LI Y L, WEN W L, MIAO T, et al. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning[J]. Computers and Electronics in Agriculture,2022,193:106702.
- [16] LI Y C, LIU J Y, ZHANG B, et al. Three-dimensional reconstruction and phenotype measurement of maize seedlings based on multi-view image sequences[J]. Frontiers in Plant Science,2022,13:974339.
- [17] LIU H, XIN C, LAI M Z, et al. RepC-MVSNet:a reparameterized self-supervised 3D reconstruction algorithm for wheat 3D reconstruction[J]. Agronomy-Basel,2023,13(8):13081975.
- [18] SANDHU J, ZHU F Y, PAUL P, et al. PI-Plat:a high-resolution image-based 3D reconstruction method to estimate growth dynamics of rice inflorescence traits[J]. Plant Methods,2019,15(1):162.
- [19] TROSS M C, GAILLARD M, ZWEINER M, et al. 3D reconstruction identifies loci linked to variation in angle of individual sorghum leaves[J]. PeerJ,2021,9:12628.
- [20] GAILLARD M, MIAO C Y, SCHNABLE J C, et al. Voxel carving-based 3D reconstruction of sorghum identifies genetic determinants of light interception efficiency[J]. Plant Direct,2020,4(10). DOI:10.1002/pld3.255.
- [21] SUN Y Z, ZHANG Z X, SUN K, et al. Soybean-MVS:annotated three-dimensional model dataset of whole growth period soybeans for 3D plant organ segmentation[J]. Agriculture-Basel,2023,13(7):13071321.
- [22] HE W, YE Z H, LI M S, et al. Extraction of soybean plant trait parameters based on SfM-MVS algorithm combined with GRNN[J]. Frontiers in Plant Science,2023,14:1181322.
- [23] ZHU R S, SUN K, YAN Z Z, et al. Analysing the phenotype de-

- velopment of soybean plants using low-cost 3D reconstruction[J]. Scientific Reports, 2020, 10(1): 7055.
- [24] SALTER W T, SHRESTHA A, BARBOUR M M. Open source 3D phenotyping of chickpea plant architecture across plant development[J]. Plant Methods, 2021, 17(1): 95.
- [25] WANG L Y, MIAO Y L, HAN Y X, et al. Extraction of 3D distribution of potato plant CWSI based on thermal infrared image and binocular stereovision system[J]. Frontiers in Plant Science, 2023, 13: 1104390.
- [26] XIAO S F, CHAI H H, WANG Q, et al. Estimating economic benefit of sugar beet based on three-dimensional computer vision: a case study in Inner Mongolia, China[J]. Eur J Agron, 2021, 130: 126378.
- [27] WANG Y H, HU S T, REN H, et al. 3DPhenoMVS: a low-cost 3D tomato phenotyping pipeline using 3D reconstruction point cloud based on multiview images[J]. Agronomy-Basel, 2022, 12(8): 12081865.
- [28] YU L, SUSSMAN H, KHMELNITSKY O, et al. Development of a mobile, high-throughput, and low-cost image-based plant growth phenotyping system[J]. Plant Physiology, 2024. DOI: 10.1093/plphys/kiae237.
- [29] SZELISKI R. Computer vision: algorithms and applications[M]. Cham: Springer, 2011.
- [30] LOU L, LIU Y, HAN J, et al. Accurate multi-view stereo 3D reconstruction for cost-effective plant phenotyping[C]//International Association for Pattern Recognition. Proceedings of the 11th International Conference on Image Analysis and Recognition, October 22-24, 2014. Porto: Springer, 2014: 467-474.
- [31] SCHÖNBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//Institute of Electrical and Electronics Engineers. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016. Las Vegas: IEEE, 2016: 4104-4113.
- [32] WU C. Towards linear-time incremental structure from motion[C]//Institute of Electrical and Electronics Engineers. Proceedings of the International Conference on 3D Vision. Seattle: IEEE, 2013: 127-134.
- [33] GHERARDI R, FARENZENA M, FUSIELLO A. Improving the efficiency of hierarchical structure-and-motion[C]//Institute of Electrical and Electronics Engineers. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010. San Francisco: IEEE, 2010: 1594-1600.
- [34] SWEENEY C, SATTLER T, HÖLLERER T, et al. Optimizing the viewing graph for structure-from-motion[C]//Institute of Electrical and Electronics Engineers. Proceedings of the IEEE International Conference on Computer Vision, December 7-13, 2015. Santiago, Chile: IEEE, 2015: 801-809.
- [35] LHUILLIER M, QUAN L. A quasi-dense approach to surface reconstruction from uncalibrated images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 418-433.
- [36] FURUKAWA Y, PONCE J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376.
- [37] YANG D N, YANG H J, LIU D F, et al. Research on automatic 3D reconstruction of plant phenotype based on multi-view images[J]. Computers and Electronics in Agriculture, 2024, 220. DOI: 10.1016/j.compag.2024.108866.
- [38] TIAN Y J, ZHANG J N, ZHANG Z J, et al. Research on super-resolution enhancement technology using improved transformer network and 3D reconstruction of wheat grains[J]. IEEE Access, 2024, 12: 62882-62898.
- [39] DETONE D, MALISIEWICZ T, RABINOVICH A, et al. Superpoint: self-supervised interest point detection and description[C]//Institute of Electrical and Electronics Engineers. CVF Conference on Computer Vision and Pattern Recognition, June 18-22, 2018. Salt Lake City: IEEE, 2018: 224-233.
- [40] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: learning feature matching with graph neural networks[C]//Institute of Electrical and Electronics Engineers. CVF Conference on Computer Vision and Pattern Recognition, June 14-19, 2020. Seattle: IEEE, 2020: 4938-4947.
- [41] POTENA C, KHANNA R, NIETO J, et al. AgriColMap: aerial-ground collaborative 3D mapping for precision farming[J]. IEEE Robotics and Automation Letters, 2019, 4(2): 1085-1092.
- [42] GAO J, LIU J, JI S P. A general deep learning based framework for 3D reconstruction from multi-view stereo satellite images[J]. Isprs Journal of Photogrammetry and Remote Sensing, 2023, 195: 446-461.
- [43] LIU S, BONELLI W P, PIETRZYK P, et al. Comparison of open-source three-dimensional reconstruction pipelines for maize-root phenotyping[J]. The Plant Phenome Journal, 2023, 6(1): 20068.
- [44] XU B, WAN X Y, YANG H, et al. TIPS: a three-dimensional phenotypic measurement system for individual maize tassel based on TreeQSM[J]. Computers and Electronics in Agriculture, 2023, 212. DOI: 10.1016/j.compag.2023.108150.
- [45] ZIA A, LIANG J, ZHOU J, et al. 3D reconstruction from hyperspectral images[C]//Institute of Electrical and Electronics Engineers. Proceedings of the IEEE Winter Conference on Applications of Computer Vision, December 7-13, 2015. Santiago, Chile: IEEE, 2015: 318-325.

(责任编辑:王 妮)