

基于视觉的虚拟现实与增强现实融合技术

宁瑞忻, 朱尊杰, 邵碧尧, 龚冰剑, 颜成钢

杭州电子科技大学自动化学院, 智能信息处理实验室, 杭州 310018

摘要 基于视觉的虚拟现实(VR)和增强现实(AR)的融合发展是不可避免的趋势, 定位与地图构建(SLAM)是虚拟现实与增强现实融合应用的主要组成部分, 但在鲁棒性等方面仍存在很多具有挑战性的问题。本文提出了一种基于视觉的虚拟现实与增强现实融合技术, 总结了室内场景三维重建过程中设备选择、追踪和运动干扰、平面识别等问题, 并分析了当前SLAM中存在的5个具有挑战性的问题。

关键词 虚拟现实; 增强现实; SLAM; 三维重建

将现实场景与虚拟物体进行融合显示将会带来更加新颖的体验, 但融合的过程充满挑战。由于受到定位与建图技术的限制, 目前大多数虚拟现实(virtual reality, VR)设备都不能与真实场景进行交互, 为解决这种问题, 视觉SLAM组采用基于视觉的虚拟现实和增强现实(augmented reality, AR)融合技术, 并搭建了一套系统, 利用同时定位与地图构建(simultaneous localization and mapping, SLAM)技术进行定位和场景重建, 通过边缘检测和对角线判定等一系列方法对场景中的地板、桌子、墙壁等平面结构进行识别。该系统以相机作为传感器(一个深度相机、一个彩色相机), 利用深度图像(RGB-D)序列还原出真实场景的三维模型, 在此基础上识别出现实场景布局, 并利用增强现实技术为场景添加虚拟的辅助信息; 通过识别现实中的平面, 将虚拟物体和平面信息结合放置到现实场景中, 实现虚拟物体和真实场景的融合; 最后利用双目视图的生成办法产生人眼可观测的三维视觉效果, 通过头戴式虚拟现实设备实现虚拟和现实场景的交互。

1 VR与AR技术概况

虚拟现实(VR)技术能够创造出虚拟世界, 目前有大量的计算机应用^[1-2], 将多个信息源数据融合, 形成人机交互系统, 模拟三维场景和真实运动, 使用户身临其境。增强现实(AR)技术能够实时识别出相机图像的位置和角度, 并在双目视觉成像中将虚拟的物体和现实相结合。目前, 许多国际大公司都意识到了虚拟现实技术和增强现实技术巨大的市场潜力与价值, Oculus公司推出了Oculus Rift头戴设备, 谷歌公司不惜重金打造Google Glass, 微软公司也在积极开发Microsoft HoloLens(图1)。在早期, 虚拟现实技术和增强现实技术依赖计算机和大型工作站来构建系统操作平台, 随着互联网时代的到来, 尤其是智能手机的快速发展和普及, 虚拟现实技术和增强现实技术的应用在移动终端设备上得到了很好的推广。基于视觉的虚拟现实与增强现实融合技术在将RGB-D序列转换为3D模型的过程中, 帧率可以达到30帧/s。

收稿日期: 2018-04-14; 修回日期: 2018-04-27

作者简介: 宁瑞忻, 硕士研究生, 研究方向为视觉SLAM, 电子信箱: ningruixin_work@163.com; 颜成钢(通信作者), 教授, 研究方向为智能信息处理, 电子信箱: cgyan@hdu.edu.cn

引用格式: 宁瑞忻, 朱尊杰, 邵碧尧, 等. 基于视觉的虚拟现实与增强现实融合技术[J]. 科技导报, 2018, 36(9): 25-31; doi: 10.3981/j.issn.1000-7857.2018.09.003

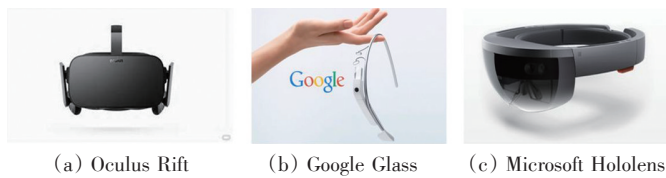


图1 3种虚拟现实和增强现实显示设备

Fig. 1 Three display devices of virtual reality and augmented reality

2 相机标定和图像数据获取

利用相机采集到的数据是二维平面图像,但在光线投影到成像平面的过程中,会受到相机镜头上透镜的影响,使结果产生畸变。为了防止畸变带来的影响,并将二维平面图像中的点准确映射到三维空间当中,就必需测得相机的参数。

2.1 相机标定

为了确定空间物体表面某点的三维几何位置与其在图像中对应点之间的相互关系,就需要建立相机成像的几何模型,这些几何模型参数就是相机参数。在大多数条件下这些参数必须通过实验与计算才能得到,这个求解参数的过程就称之为相机标定。相机标定的结果将直接影响映射关系的准确性。因此,相机标定是必不可少的环节,通过相机标定可以获得外参矩阵、内参矩阵、畸变矩阵,并可知世界坐标和相机坐标的关系,最终映射出理想的三维视觉效果^[3-4]。Sinha等^[5]提出了较为简单的校正装置,Ben-Artzi等^[6]提出了较为有效的描述子。

2.2 数据获取

用棋盘格图片实现相机的标定,首先要在不同角度下对棋盘格图片进行拍照,然后从照片中提取角点(图2),最后利用约束条件和最小二乘法对参数进行估算。对直接采集到的RGB图和深度图像序列进行预处理,将深度相机和彩色相机采集到的数据合成四通道RGB-D图像。

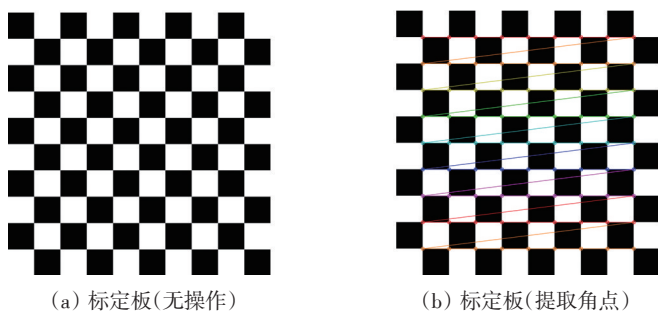


图2 标定板的数据获取

Fig. 2 Data extraction from the calibration board

2.3 相机位姿

刚体在三维空间中的位置和姿态由旋转和平移构成,为了用虚拟现实技术在虚拟世界中创造良好顺畅的体验效果,系统必须具有良好的实时性。因此必须获取相机的实时位置变化和姿态变化,即要实时掌握相机位姿以便进行下一步工作。首先要提取RGB-D图像序列中相邻两帧的特征点;然后对相邻两帧进行特征点匹配(图3),获取位姿变换矩阵;最后引入检测机制来考察匹配效果。

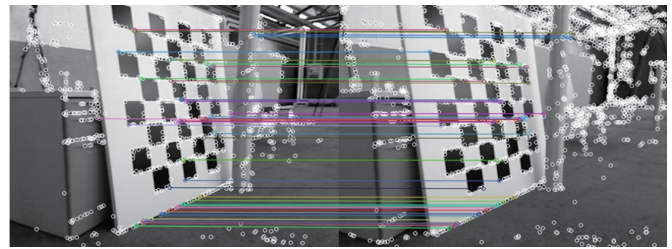


图3 特征点匹配

Fig. 3 Feature point matching

3 搭载传感器的主体SLAM

搭载传感器的主体使用SLAM技术在未知环境中对自身实时定位和地图重建,是虚拟现实与增强现实融合中的重要组成部分。SLAM自1988年提出至今,系统逐渐模块化、功能日渐完善,从开始的军用潜艇激光测距定位到走进千家万户的扫地机器人,以及如今耳熟能详的无人驾驶技术研究,SLAM正在逐步贴近人们的生活。

SLAM在发展的过程中,使用的数据逐渐多源,模块功能划分也更加明确。根据SLAM系统使用的传感器和主要功能模块的不同,可划分为不同分支,目前主流的两个方向为:基于激光雷达的SLAM(lidar SLAM)和基于相机的视觉SLAM(visual SLAM)。

3.1 激光SLAM

激光SLAM由最早出现在军用潜艇上的红外测距定位装置演变而来,使用激光雷达获取数据,相比红外测距装置之前其测距的距离远、数据精准度高(图4)。激光雷达获取的数据是一幅幅的点云,这些点云是外界环境中物体外观表面的点数据集合,且包含角度和深度(距离),有助于导航的路径规划。激光SLAM系统往往通过对相邻两幅点云进行特征的提取和匹配,解算传感器自身运动相对环境的位姿变换,可以实现传感器自身的定位。

目前,激光SLAM相比视觉SLAM,其理论和技术更加成熟,在成果和应用上也更加广泛,美国谷歌公司“Google X”实验室研发中的自动驾驶汽车就是利用激光雷达对道路状况进行重建(图4)。激光雷达探测距离较远,且能准确获取物体的三维点云信息,数据简单,误差模型简单,鲁棒性和稳定性好,即使在昏暗和光线变化的环境下也依然可以很好地发挥作用。

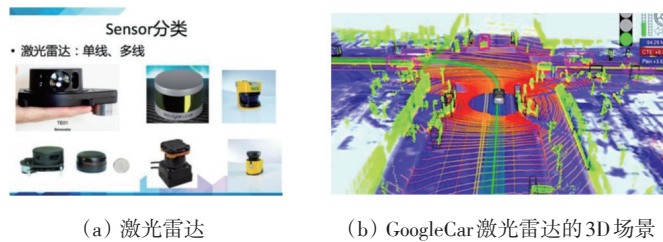


图4 激光传感器及应用效果

Fig. 4 Laser sensor and its application effect

3.2 视觉SLAM

视觉SLAM以相机作为主要传感器获取周围环境的图像序列,利用丰富的纹理信息,实现对自身的追踪。噪声的存在会对运动方程和观测方程的解算结果与现实位姿之间产生误差,为了在有噪声的数据中进行准确的状态估计,提出了基于滤波的SLAM系统和基于优化的SLAM系统。

一个视觉SLAM系统主要由前端、后端、回环检测3个部分组成。前端的任务是估算相邻图像间相机的运动和局部地图的信息。回环检测的目的是判断自身是否在之前已经到达过相同的位置。后端的任务是接受不同时刻前段估算的相机位姿,以及回环检测信息,对它们进行整合优化,最终得到全局一致的轨迹和地图,微软研究院在2011年提出的Kinect Fusion开启了使用RGB-D相机实时三维重建的序幕(图5)。

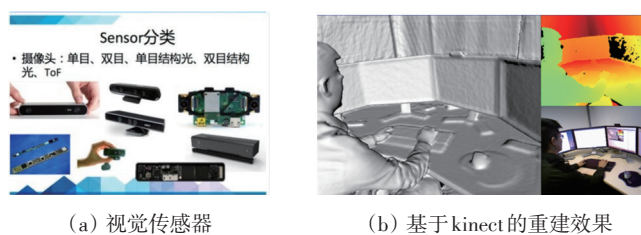


图5 视觉传感器及应用效果

Fig. 5 Visual sensor and its application effect

视觉SLAM相比激光SLAM具有丰富的纹理信息,颜色等重要属性是激光测距采集到的点云数据所不具备的,结合这些独有的纹理信息,视觉SLAM可以有效地识别和追踪场景中的动态变化,较为容易地辨别场

景,实现可靠的回环检测,完成定位和全局优化。

激光SLAM虽然具有相对完善的技术和理论支持,但根据数据精度不同,传感器的价格从数千元到数十万元不等,即使是最便宜的室内近距离平面激光雷达,价格也远高于视觉传感器,难以推广和量产,因此该系统中选择相机作为系统传感器。

3.3 运动模糊处理

在重建的过程中,人物的移动最自然的一件事情,如果视野的移动和身体的移动不结合起来,就必然会导致晕动症的产生。然而运动太快可能导致相机图像出现运动模糊,成像质量下降。获得的两个图像的重叠区不够,导致特征无法匹配,难以追踪相机的位置,也就无法进行准确连续的三维场景重建,所以采用慢速平移的策略。

3.4 动态场景处理

人物的快速移动会导致重建的失败,那么当场景中存在移动的物体,也会影响到重建的效果。通常的SLAM都是以静态场景为基础,假定场景中的物体长时间静止不变,但在实际应用中,最好的情况也会面临一些少量的动态变化。在过去的几年里,逐渐有人提出了大量的运动去除方法来解决动态环境中场景重建的问题。Namder等^[7]提出了一种基于视觉的在线的方法,采用多几何约束和密集光流算法对运动目标进行分割,能够滤除移动物体所带来的干扰。报告的结果表明,三维场景重建的结果得到了有效的改善。Teichman等^[8]提出了一种基于RGB-D数据运动场景重建方法,并描述出了条件随机场(CRF)模型。分割线索包括光流、颜色和深度数据不连续等。设计一个处理动态场景的系统是重建质量、重建速度、空间尺度和场景环境假设之间的一个复杂的平衡。现有的在线方法牺牲空间尺度可以高质量地重建小物件和场景,牺牲实时性能或质量可以处理大范围的场景。视觉SLAM组采用重建静态部分,检测并滤除动态部分的方法(图6),直接使用深度传感器采集的数据,基于点云进行融合,能够提高速度和内存效率。

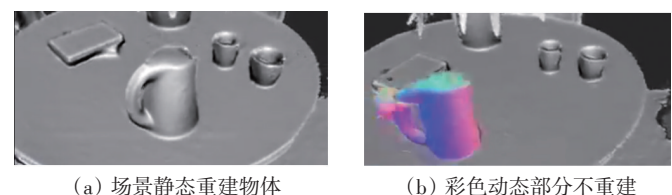


图6 动态场景处理系统

Fig. 6 Dynamic scene processing system

因此,基于视觉的虚拟现实与增强现实融合选用深度相机和彩色相机,在慢速移动下实现虚拟和现实场景的交互。

4 室内场景的三维重建

4.1 场景平面追踪

实现对平面结构的跟踪及重建是三维重建的难点。传统的相机追踪方法是使用迭代最近点(iterative closest point, ICP)^[9]估计两个深度帧获得的点云之间的对应关系,再通过估计的相机运动将点云进行融合。然而,在基于迭代最近点的方法中,需要丰富的纹理特征,通过场景中的几何特征,鲁棒地估计正确的相机运动。对于3D空间中的两个纯2D平面,最接近的点可能不是正确的对应关系,因此,迭代最近点可能会产生错误的摄像机运动。此外,迭代最近点需要大量采样点,需要迭代才能收敛到最终的对应关系,这意味着相对较重的计算成本。即使一些基于迭代最近点的系统使用GPU实现实时性能,但是因为GPU可能被其他任务占用仍然不适用于许多实际应用。

因此,视觉SLAM组提出了一种新的解决方案:基于曼哈顿假设^[10]的实时相机运动估计方案。当记录的帧包含足够的正交平面时,说明可能存在较大的平面墙壁,因为平面墙壁通常具有一致的颜色,所以只能提取很少的特征。在这种情况下,基于曼哈顿假设的方案给出了精确的运动估计,首先估计记录帧的所有3D点的法线方向,然后通过主成分分析(principal component analysis, PCA)估计3个正交平面主要方向;由于从所有深度点估计主要方向,随机噪声被大量滤除,然后进一步估计相应的主轴(即法线),确定每帧的平面坐标;最后,通过由平面坐标和主轴获取的变换矩阵来估计摄像机姿态。

该方案可在几何和纹理特征较少情况下重建室内场景,这对于基于图像特征的方法和基于迭代最近点的方法都是重要且具有挑战性的。其主要思想涉及曼哈顿假设,有效地估计主导轴和主平面,并通过对齐连续输入帧的轴和平面重建相机运动。由于根据传感器记录的深度输入估计的轴和平面具有少量自由度,所以估计对于输入噪声具有很好的鲁棒性。

4.2 场景平面识别

若要在真实场景中生成虚拟物体,需要了解真实

场景中各部分的结构,否则生成虚拟物体时会产生误差,导致虚拟物体与真实场景不融合等问题。例如实际应用中,需要在墙壁上生成一块虚拟屏幕,就需要基于场景中的平面检测机制识别出场景中的墙壁,从而生成虚拟屏幕。完成平面识别阶段之后,通过面与面之间的交叉线识别检测出场景的边缘位置,即边缘检测。要将得到的RGB-D图像序列输入到场景平面识别模块中,首先,检测场景中的直线。这里采用边缘检测方法并循环执行,直到识别且判定出直线;其次,确定平面,通过对角线判定框定直线;最后,判断获取到的平面的深度信息是否符合三维平面函数,如果符合,就认为识别出的是正确的平面,从而输出平面的边缘坐标,便于在片面上生成虚拟场景。通过边缘检测和对角线判定等一系列方法可对场景中的地板、桌子、墙壁等平面结构进行识别。

4.3 双目视差图

最终要通过人眼观测并产生逼真效果,就必须将场景转化为人眼能直接观测到的三维画面。而获取立体显示的生理基础依赖于深度直觉,人类的眼睛和大脑可以感受到很多深度提示,当这些提示在现实世界中出现的时候,它们之间会互相加强,在人类的脑海中形成清晰的3D环境图像。为此,要采用双目视差图的方法来实现。VR头显利用两个微型显示器生成左右两幅图像(图7),左右眼观测到的画面中只有部分重合,相同物体图像的位置也具有差别,当左右眼的图片信息分别被人脑接收,大脑将会分析出二维视网膜图像在立体视觉的深度信息,产生3D效果,从而模拟真实场景的效果。用相机模型模拟人眼,估计人眼相机参数,将图像转换成双目视图。将双目视图图像序列呈现于屏幕,此时可以观察到实际场景中的三维模型,由于实现了场景中平面结构的识别跟踪及物体识别,用户可以通过增强现实技术在图像序列中的场景结构上将虚拟三维模型与平面结构融合。



图7 头显设备中的视差图

Fig. 7 Parallax diagram in head display device

5 SLAM 技术现有挑战

1) 视觉SLAM的鲁棒性。

视觉SLAM系统的鲁棒性是极具挑战性的问题。定位和重建的过程中会将特征和观测到的路标相关联^[11],将数据与状态关联到一起称为一致性问题,建立在不同时间、不同地点获得的传感器测量之间、传感器测量与地图特征之间的对应关系,目的是为了确定测量是否源于环境中同一物理实体。但是视觉SLAM依赖视觉特征,在昏暗、光线变化及动态的场景中容易出现数据关联错误的问题,如果特征和路标建立了错误的匹配关系,后端也会受到影响,执行错误的估计和优化,进而导致整个算法失效。因此,目前视觉SLAM研究大多都是建立在环境不变的基础上。如果是进行大范围、长时间的地图构建,就难免会遇到昼夜变化、物体变化,甚至是季节变化,在没有构建动态环境模型的情况下会带来更严重的长期数据关联错误,使SLAM系统失效。

在SLAM系统中,数据与状态关联的工作在前端中进行。短时间内的数据关联准确率高,易于处理,使用数据采样频率较高的传感器,采用基础的特征提取方法进行图像匹配,或者使用光流的方法,都可以对路标进行关联和追踪。在较高的采样频率下,相邻两组数据之间传感器(相机,激光)的位姿变化非常小,所以 t 和 $t+1$ 两个相邻时刻的特征非常接近^[12]。长时间的地图构建任务重中,数据关联容易出现错误,所以同时需要闭合回路的检测和校验,使用词袋模型可以量化特征空间并进行更有效的搜索来避免这种难题,将词袋模型设计成叉树形式的字典,保证了对数级别的查找效率,使用了词袋模型的SLAM系统在回环检测中的结果,十分可靠。但是,当图像剧烈运动,同时带有模糊效果的时候,词袋模型将无法实现有效的回环检测,因为图像无法与视觉单词进行正确匹配。

回环检测效果可以通过几何验证步骤进行检验。在视觉SLAM系统中,随机采样一致性算法^[13](random sample consensus, RANSAC)常用于几何校验和离群点去除工作。在激光SLAM系统中,可以通过分析当前激光扫描图像和已经扫描好的地图的匹配程度(匹配得有多好)来校验回环效果。

当回环检测环节出现问题,错误的回环检测结果仍会传递给后端,导致后端最大后验估计和优化的质

量被严重损害。针对这种情况,目前已有部分研究提出了改善方法,可以使系统后端对不合格的检测结果进行更有弹性的处理^[14-15]。这些方法基于回环检测的有效性进行推导,通过优化过程中的残差进行约束。

2) SLAM的失效保护和恢复机制。

离群点对于SLAM系统的鲁棒性具有很大的影响,这是因为一个成熟鲁棒的SLAM系统是建立在非凸集迭代优化之上的。当出现离群点的时候,任何一个离群点都会降低初始估计的质量,而系统去除离群点的能力又取决于估计的质量,所以将会导致识别和去除离群点的能力逐渐下降,最终导致算法失效。成熟完善的系统应当能够及时发现即将由离群点引起的算法失效问题,并采取相应的失效保护和恢复机制恢复之前的操作任务。目前SLAM系统还不具备这种功能。

3) 几何度量重定位。

与基于特征的方法相比,基于贴图的方法应用在包含昼夜交替的图像序列上可以实现回环闭合,因此在包含季节变换的图像序列上实现回环闭合也是可以实现的。对于几何度量重定位,利用特征的办法可能更规范化,但无法应用到这种变化的复杂环境中。假设视觉变成了对于不同环境的传感器的选择,回环检测变成了传感器位置匹配问题,那么也可以对SLAM问题所需的其他信息和传感器进行研究。例如,Brubaker^[16]提出了轨迹匹配中相机的劣势。先用一个相机构建地图,再用另一个相机在构建的地图中进行定位,会有比较好的效果,Wolcott和Forster的工作在这个方向上有所突破^[17]。Forster^[18]研究了在激光地图中的视觉定位问题,Majdik^[19]研究了如何在Google街景照片的3D纹理地形图上定位无人机。Behzadin^[20]演示了如何在手绘地图用激光扫描仪,Winterhatter^[12]演示了如何在2D户型图上用RGB-D定位。

4) 时间变化和非固定地图。

目前主流的两种SLAM方法是激光SLAM和视觉SLAM,对此相关的研究大部分建立在现实世界固定不变的条件之上。但是随着时间变化、昼夜更替、季节变换、物体变化,真实世界也是在不断变化的。完善的SLAM系统应该具有处理环境中动态因素,包括非固定部分的能力,在大范围、长时间的地图构建中,也能实时更新地图。在计算机视觉方向上,人们始终没有放弃尝试在有限的应用中从非固定的物体上恢复形状,比如Pertlend^[21]指出需要物体结构属性的信息。

Bregler^[22]要求限定物体的畸变,在人脸识别上做了一些演示。在SLAM领域,Newcombe^[15]针对小范围地图构建提出了非固定环境实例。然而,大范围非固定地图的构建仍然没有得到解决。

5) 参数自动调整。

在视觉SLAM系统中,为了更好地将特征和观测到的路标相关联,将数据与状态关联到一起,提高系统的鲁棒性,就需要进行大量准确快速的参数调整,其中包括采样一致性算法的参数,特征匹配的阈值等,进而可以在规定的场景中正确地工作。为了使系统能够在更多的场景中正确工作,适应更多的场景变化,让大量的参数调整工作可以自动在系统内部完成,需要考虑一种参数自动调整的方法。

6 结论

目前,SLAM是虚拟现实与增强现实融合应用的主要组成部分,与人类自身建立环境模型的能力相比,现有SLAM系统远远不能提供有效的、可用的、精简的环境模型。场景重建及虚拟现实与增强现实融合领域面临更多更新的挑战:失效保护SLAM系统的设计,参数自动调整算法,环境高级几何度量和语义表示等都极具挑战性^[23]。

虽然基于视觉的虚拟现实和增强现实当前看似各自发展,但结合三维场景重建,二者融合是未来不可避免的趋势。虚拟现实可以创建出虚拟世界,能够在短时间内给使用者带来愉悦美好的体验,但却将外部真实世界完全隔离,所以无法长时间沉浸其中。增强现实能够把有趣惊奇的虚拟物体融合到现实世界当中,没有将外部真实世界隔离开,而是与现实世界互动,但是增强现实沉浸感不够,使用者在现阶段明显能够区分何为虚拟,何为真实。两者结合,将会带来更精彩和美好的体验。

参考文献(References)

- [1] Yan C G, Zhang Y D, Xu J Z, et al. Efficient parallel framework for hevc motion estimation on manycore processors[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(12): 2077-2089.
- [2] Yan C G, Zhang Y D, Xu J Z, et al. A highly parallel framework for hevc coding unit partitioning tree decision on many-core processors[J]. IEEE Signal Processing letters, 2014, 24(5): 573-576.
- [3] Laurentini A. The visual hull concept for silhouette-based image understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(2): 150-162.
- [4] Miller G, Hilton A. Exact view-dependent visual hulls[J]. International Conference on Pattern Recognition, 2006, 1(1): 107-111.
- [5] Sinha S N, Pollefeys M. Camera network calibration and synchronization from silhouettes in archived video[J]. International Journal of Computer Vision, 2010, 87(3): 266-283.
- [6] Ben-Artzi G, Kasten Y, Peleg S, et al. Camera calibration from dynamic silhouettes using motion barcodes[J]. Computer Science, 2015: 4095-4103.
- [7] Namdev R K, Kundu A, Krishna K M, et al. Motion segmentation of multiple objects from a freely moving monocular camera [C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2012: 4092-4099.
- [8] Teichman A, Thrun S. Learning to segment and track in RGBD [J]. IEEE Transactions on Automation Science & Engineering, 2013, 10(4): 841-852.
- [9] Besl P J, McKay N D. A method for registration of 3-D shapes [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 14(2): 239-256.
- [10] Furukawa Y, Curless B, Seitz S M, et al. Manhattan-world stereo[J]. IEEE Computer Vision and Pattern Recognition, doi: 10.1109/CVPR. 2009.5206867.
- [11] Aragues R, Carlone L, Sagues C, et al. Distributed centroid estimation from noisy relative measurements[J]. Systems & Control Letters, 2012, 61(7): 773-779.
- [12] Bibby C, Reid I. A hybrid SLAM representation for dynamic marine environments[C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2010, 58(8): 257-264.
- [13] Agudo A, Morenonoguer F, Calvo B, et al. Sequential non-rigid structure from motion using physical priors[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(5): 979.
- [14] Bloomenthal J, Wyvill B. Introduction to implicit surfaces[M]. San Francisco: Morgan Kaufmann Publishers, 1997, 22(9): 213-234.
- [15] Bailey T, Bryson M, Mu H, et al. Decentralised cooperative localisation for heterogeneous teams of mobile robots[C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2011: 2859-2865.
- [16] Bardow P, Davison A J, Leutenegger S. Simultaneous optical flow and intensity estimation from an event camera[C]//IEEE Conference on Computer Vision and Pattern Recognition.

- IEEE Computer Society, 2016: 884–892.
- [17] Bosse M, Zlot R. Continuous 3D scan-matching with a spinning 2D laser[C]//IEEE International Conference on Robotics and Automation. Salt Lake City, Utah, USA: Piscataway, NJ: IEEE, 2009: 4244–4251.
- [18] Brand C, Schuster M J, Hirschmuller H, et al. Stereo-vision based obstacle mapping for indoor/outdoor SLAM[C]//International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2014: 1846–1853.
- [19] Benosman R, Ieng S H, Clercq C, et al. Asynchronous frameless event-based optical flow[J]. Neural Networks the Official Journal of the International Neural Network Society, 2012, 27(3): 32.
- [20] Anderson S, Dellaert F, Barfoot T D. A hierarchical wavelet decomposition for continuous-time SLAM[C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2014:373–380.
- [21] Aragues R, Carlone L, Calafiore G, et al. Multi-agent localization from noisy relative pose measurements[C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2011: 364–369.
- [22] Carlone L. A convergence analysis for pose graph optimization via Gauss-Newton methods[C]//IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2013: 965–972.
- [23] Cadena C, Carlone L, Carrillo H, et al. Simultaneous localization and mapping: Present, future, and the robust-perception age[J]. IEEE Transactions on Robotics, 2017, 32(6): 1309–1332.

Vision-based virtual reality and augmented reality fusion technology

NING Ruixin, ZHU Zunjie, SHAO Biyao, GONG Bingjian, YAN Chenggang

Laboratory of Intelligent Information Processing, College of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract Visual-based virtual reality and augmented reality have their own developments, but the fusion development trend of the two is unavoidable in the future. SLAM (simultaneous localization and mapping) is a major component of the application of virtual reality and augmented reality, but there are still many challenging issues in terms of robustness. This paper proposes a vision-based virtual reality and augmented reality fusion technology, which can analyze and solve the problems such as device selection, tracking, motion interference, and plane recognition in the 3D scene reconstruction process. Finally, the challenging issues in SLAM are discussed.

Keywords virtual reality; augmented reality; SLAM; 3D scene reconstruction ●



(责任编辑 卫夏雯)