

虚实融合场景下移动端 AR 三维模型跟踪注册技术实现

赵丹

南宁师范大学, 广西 南宁 530001

DOI:10.61369/EDTR.2025120027

摘 要 : 基于移动设备的 AR 三维模型跟踪与配准技术是将数字三维模型正确地投射于真实世界的图像中, 并实现可靠的高精度定位的关键技术, 在教学、医疗、工业维修以及娱乐等领域有着广泛的应用。本文主要针对移动终端 AR 系统中的 3D 模型跟踪及配准算法进行研究。针对现有主流方案, 在系统层面对其进行分析总结, 包括基于 VIO 的方案、基于特征点跟踪的方法以及基于深度学习提高定位精度的方案, 并对其中存在的将该方案应用到移动端遇到的问题及解决办法进行讨论; 同时我们对其中模型配准的关键问题进行了深入探讨: 如坐标转换、尺度归一化、光照变化及物遮挡等问题, 并利用 SLAM 算法及多传感器信息融合来增强系统鲁棒性, 工作还提出部分未来发展方向的线索, 包括适用于轻量化模型、多平台适配性的设计思路, 基于 5G 边缘计算协同处理等新观点。

关 键 词 : 增强现实; 三维模型跟踪; 模型注册; 移动端 AR; SLAM; 位姿估计

Implementation of Mobile AR 3D Model Tracking and Registration Technology in Virtual-Real Integration Scenarios

Zhao Dan

Nanning Normal University, Nanning, Guangxi 530001

Abstract : AR 3D model tracking and registration technology based on mobile devices is a key technology for accurately projecting digital 3D models into real-world images and achieving reliable high-precision positioning, with extensive applications in fields such as education, healthcare, industrial maintenance, and entertainment. This paper primarily focuses on researching 3D model tracking and registration algorithms within mobile AR systems. It analyzes and summarizes existing mainstream solutions at the system level, including approaches based on Visual-Inertial Odometry (VIO), feature point tracking, and schemes leveraging deep learning to enhance positioning accuracy. Furthermore, it discusses the challenges encountered when applying these solutions to mobile devices and proposes corresponding solutions. Additionally, the paper delves into critical issues related to model registration, such as coordinate transformation, scale normalization, illumination variations, and object occlusion. It utilizes Simultaneous Localization and Mapping (SLAM) algorithms and multi-sensor information fusion to enhance system robustness. The work also provides clues for future development directions, including design considerations for lightweight models and multi-platform compatibility, as well as new perspectives such as collaborative processing based on 5G edge computing.

Keywords : augmented reality; 3D model tracking; model registration; mobile AR; SLAM; pose estimation

引言

随着智能硬件的发展和摄像机拍摄技术的进步, 手机 AR 功能发展为虚拟空间与真实世界的连接器, 在这个过程中, 3D 模型定位匹配是 AR 技术的重要环节之一, 影响着虚拟信息能否准确流畅地投射到人眼所见的真实世界中。近年来没有标记物同时能实时定位及构建地图的方法 (SLAM), 显著提高了 AR 的普及度, 然而在移动平台受限条件下兼顾实时性及提高跟踪精度与鲁棒性依然是一个热点也是难点问题, 本文旨在对现有方法进行系统归纳总结并分析其实现机理及其局限性以及可能的有效改进措施。

课题信息: 2023 年度广西高校中青年教師科研基础能力提升项目 (项目编号: 2023KY2016)。

作者简介: 赵丹 (1983.04—), 女, 满族, 广西壮族自治区南宁市人, 研究生学历, 毕业于东北石油大学, 南宁师范大学就职, 高级工程师。

一、三维模型跟踪技术原理与实现方式

移动终端 AR 中基于 3D 模型的跟踪主要采用的是视觉跟踪方法与惯性测量单元相结合的方法，而视觉跟踪方法又可分为基于特征点的方法如 SIFT、ORB 和直接估计相机姿态的方法。基于特征点的方法能够提取并匹配图像中的关键点来估计相机姿态，具有良好的鲁棒性及抗噪能力，但易受弱纹理区域干扰。利用直接法则来定位姿势时，会试图减小光学信号之间的亮度差异，这种方法对于具有丰富纹理的环境是有效的，但是它容易受到光线条件的影响。因此，许多现代 AR 设备选择使用视觉和惯性测量单元(IMU)相结合的方法，以弥补视觉系统中的滞后和偏移问题。基于深度学习的方法为跟踪提供了更多可能，如利用 CNN 构建端到端位置识别，在复杂场景中更具泛化能力。

(一) 视觉特征点跟踪与惯性传感器融合

在移动端 AR 应用中，精确估计手机位置姿态是实现三维目标跟踪的基础。通过摄像机获得图像信息进行特征点匹配称为视觉跟踪法，可以依据连续两帧之间的联系计算相机位移。纯视觉方法存在由于速度快或者弱纹理导致的问题。所以，加入了惯性测量单元来得到更高频率的加速及角速度信息，以补足视觉系统可能存在的滞后性，并利用两种系统的优势：视觉系统着眼于长帧时间上的稳定性以及惯性系统可以很好地填补帧之间的空缺，减少运动模糊。

(二) 基于深度学习的端到端位姿估计

传统的方法主要是采用手工特征进行描述，而引入 DL 可以直接从图像中提取姿态信息，其网络模型使用模拟数据集作为训练样本建立输入图像到 6D 位姿输出间的函数映射关系，并可有效解决复杂光照条件下的部分遮挡等问题，比如噪声环境下的跟踪鲁棒性。

(三) 直接法跟踪与光度误差最小化

采用直接的方法去掉特征提取过程并基于改进图像之间的像素光度差异去定位姿态。该方法适用于包含有复杂纹理的场景如文档或城市景观中可直接对像素变化进行处理，主要过程就是不断地更新及减小光度误差即 Lucas-Kanade 方法。该方法实现过程中需要用到大量对梯度的高效运算，在移动端可以通过 GPU 加速其运算速度；优势在于对运动模糊具有良好的鲁棒性，但易受光照变化影响，一般会结合特征点算法组成混合方案使用，例如 Apple ARKit 中采用的是直接法快速建立初值再跟踪特征点的方式，并进行了部分优化，其中包括在线更新权重以兼顾准确度和速度。

(四) SLAM 技术集成与实时地图构建

即时定位与地图构建(SLAM)技术为 AR 提供了场景中的 3D 模型并且也可以实现无标记物跟踪的功能。SLAM 利用关键点配准创建稀疏图并估计摄像机姿态。但是 SLAM 需要对移动端的数据进行优化，所以运行的速度会慢一些，如 ORB-SLAM2 就使用了词包模型来提高匹配速度。主要流程有：特征提取、地图更新及闭环检测。闭环检测可对累积误差进行校正，如识别相同场景下的环境。

二、三维模型注册的关键技术与挑战

关键是要让三维建模和真实环境实现空间上的对齐，包括诸多难点：坐标系对齐、标定缩放、去遮挡以及增量更新等。将视觉 SLAM 系统中的相对位姿切换到全局坐标系进行注册；考虑到不同来源数据可能存在的尺度差异，尤其是单目相机更为严重，

所以要用到循环观测或是利用标定系数去修正这种偏差。

对于物理世界物体-虚拟模型之间遮挡处理的问题，主要难点是如何快速优化渲染序列保证画面一致性。一些方法使用了预测深度信息以及语义分割来将前景和背景分开，使得生成更加合理自然的遮挡效果。在混合现实环境中，采用手机端 AR 三维跟踪技术从真实场景提取特征并用于手机定位，在一定程度上保证了虚拟模型与物理环境间的匹配精度。该方法主要分为以下几步：首先进行场景特征提取；其次为实时特征匹配跟踪；再次是基于 SLAM 推算定位；最后则使用多传感器融合优化。为兼顾效率和精度需考虑以下因素，例如实时光照和复杂场景。这一技术可应用于室内导航及产品展览等领域，利用 Unity 或 ARKit / ARCore 平台实现，提升用户体验度；发展方向将聚焦在边缘计算优化和基于 AI 的智能渲染领域。从而更加深入地实现虚实之间的交互融合。

(一) 坐标系对齐与尺度一致性

其中最重要的是将虚拟对象正确映射到真实世界坐标系中，因为视觉 SLAM 的结果一般是以相对位姿的形式给出，因此需要转换为全局坐标系才能实现数字孪生体与现实场景的精准配准；其次还要考虑配比问题，在只有一组视角的视觉系统中，缺乏深度信息的输入会导致 3D 生成精度大打折扣。

例如在 AR 中，当虚拟房间无法与实际空间比例匹配的时候会产生视图误差，解决方案之一就是利用惯性测量装置辅助尺寸估计或基于多视角几何约束修正优化结果。但针对动态场景中的尺寸变化仍需进一步细化。例如将深度学习的方法应用其中来预测尺寸因子变化情况，在此过程中涉及复杂的数学变换，所以需要在移动端即时完成计算过程，避免因延时导致用户体验下降。

(二) 光照变化与动态环境适应

因为光线的变化让对齐模型变得很不稳定，为了让虚拟物体能够完全适应现实世界的照明效果(包括阴影以及高光)，如果不这样的话，就容易造成不真实的感觉；特别是当有动态的对象存在的场景中(例如：移动中的对象、人群等)。这种情况变得更为严重，因为系统需不断更新模型的位置和姿势参数以适应这些变量。传统方法是基于光流场的预测进行运算实现的，由于手机等移动端硬件性能的限制，在实时性方面无法得到很好的保障；而基于深度学习的方法，例如使用神经网络进行光学图像预测能获得更好的鲁棒性和通用性，但需要大量的训练集支撑。而最大的挑战在于如何平衡准确率和延迟的问题。特别是在户外移动的环境中，系统应该能够迅速调整以防止追踪失效的发生。

(三) 遮挡处理与渲染优化

AR 中常见的一种问题是遮挡(Occlusion)，即虚拟物体可能被物理世界部分甚至完全遮挡，导致沉浸感降低；一种解决方案是基于深度图像预测以及语义分割进行前景背景分离并动态调整渲染优先级，例如，当用户的手臂碰到虚拟物体时，则要求模型提前于手臂显示出来。

(四) 动态更新与实时性平衡

当场景中出现变化(例如用户的运动方式发生变化或者现实世界中的物体数量发生了变化等)，我们必须即时更新我们的模型配置来确保虚拟对象空间的位置正确性。定位跟踪是关键，因为如果响应太慢会丢失信息。由于移动终端的局限性，我们需要优化计算方法开销。例如 SLAM 加上 GPU 的支持可以实现每秒 1Hz 的速度，在高精度模型构建中可能超出硬件能力范围，如何平衡

更新频率不致于过度耗能是需要考虑的问题，未来发展方向将是边缘运算，即将部分工作交付云服务器完成，这样就能提升整体的系统工作效率了。

三、移动端 AR 系统优化策略与工程实践

虽然三维模型追踪和匹配技术得到了巨大的发展，但应用于移动设备过程中仍存在诸多挑战，例如硬件条件受限、续航能力不足以及散热等问题，因而如何在有限的资源环境中提高系统效率成为亟待解决的问题。对算法进行减法改造是一些常见做法，比如选用精度更高的特征描述子，适当的减小图像尺寸大小或是将浮点运算转换成整型运算等方式。在性能优化方面，合理的计算资源调度 (CPU / GPU memory)，避免重复劳动及数据搬运，能提升性能及响应速度；另外还要考虑跨端一致性问题，例如不同移动端 OS 的 API 标准差异、登录机制、绘图引擎等。以上都是需要有针对性地进行改善及优化之处。同时由于市面上智能设备存在很大的硬件差异性，因此也需要采用适应性的匹配策略来保障软件在低端设备上也能得以运行。最终方案是借助于云计算技术和边缘计算技术来解决设备端运算力不足的问题，在云端进行大规模的数据运算并返回给设备，使得追踪及登记系统更加准确可靠。

(一) 算法轻量化与硬件加速优化

为应对移动端 AR 系统的运算能力不足的问题，算法的简化及优化是主要解决方案之一。如使用更加高效的特征描述子 (BRISK 替代 SIFT) 来替代传统的图像处理方式；再如降低图像处理的分辨率或是采用固定点数替代浮点数的方式有效减少 CPU 的运算量。应当充分利用移动端特有的硬件特性，例如 GPU 进行图像特征提取的并行计算、NPU 加速神经网络运算以及 DSP 对传感器数据融合等等。例如在特征点检测中，采用 GPU 的着色器进行编程实现并行运算，将计算时间降至亚微秒级；结合移动终端 SoC 结构特点，使用 NEON 指令集优化向量化操作提升数学密集运算性能。该改进方式在不影响跟踪精度的情况下，同时提升了器件寿命及温升水平。

(二) 跨平台兼容性与自适应降级机制

由于 Android 和 iOS 操作系统下 API 接口、授权机制以及图片处理流程的不同，我们需要设计统一通用的跨端实现方案。本文采用中间抽象的方法将各平台差异部分进行封装隔离，确保算法底层不变性。例如采用相同的摄像头访问方式屏蔽底层差异性以及选用 OpenGL ES 作为通用图像处理引擎。智能降级策略是针对不同级别的硬件分别采用的策略：高级别的硬件使用完整模

式；中等硬件使用平衡算法；低级别硬件则切换至简略模式。该自适应方法在运行阶段会根据硬件信息进行动态决策（如 CPU 频率、内存容量及 GPU 规模）。同时根据以上信息即时更改系统的工作状态，从而保证我们无论在任何类型的设备上都能获得流畅的 AR 体验。

(三) 多线程架构与内存管理优化

大多数 AR 移动设备都需要执行大量的计算任务，包括图像分析、传感器整合以及 3D 模型渲染等。为了保证系统响应速度的要求，我们必须使用高效的多线程方案进行并行处理。因此，本文设计了生产者-消费者架构来分配相机捕获模块、特征提取算法以及跟踪器等核心功能到不同的线程上进行并发运算。我们还使用了线程池优化线程资源，减少不断创建线程对象及销毁线程对象造成的影响；对一些临时对象进行对象池缓存处理，减少 GC 对系统的负担。

(四) 云端协同与边缘计算集成

为解决移动端的计算瓶颈，云计算和边缘计算成为一种提高性能的方法：将复杂的数据处理任务（如大范围地图构建或者深度学习中的模型推理）放置在云中，并使用 5G 技术以减少数据传输延迟；将边缘计算部署到用户附近。以满足对实时性的高需求，像是对迅速特征比对和位置调整等操作。举例来说，在工场中的 AR 运用里，机器部分需要持续追踪，由边缘节点完成 3D 重构，然后交由云端去运行更高级别的分析。该层级的计算模式采取混合模式 (分布式 + 负载均衡) 来优化整个系统性能，并增加了预取机制来根据用户特征进行资源预测，有效缓解了因网络延迟带来的用户体验问题；同时我们也对信息传递过程进行了加密保护，保证整个过程中数据安全不外泄；最终构建了高可靠性的云-端协同计算平台。

四、结束语

综上所述，目前，针对虚实混合环境下移动设备上的三维 AR 模型追踪技术处于蓬勃发展期，在其中具有代表性的方法有视惯融合算法、基于深度学习的姿态估计算法及 SLAM 新的应用形式等。但是，在面对各种复杂的、变化的环境和人们的多样需求方面还存在很多缺陷。接下来，轻量化体系架构设计、多平台集成及自然语言处理将是未来研究的重点方向之一，并结合边缘智能。随着第五代移动通信技术 (5G)、AI 芯片、新型显示技术的进步与发展，相信基于移动端 AR 将会在海量级的应用场景下扮演重要角色。并且促进人机交互新形态走向深入。

参考文献

- [1] 施端阳, 林强, 胡冰, 等. 深度学习在雷达目标检测中的应用综述 [J]. 雷达科学与技术, 2022, 20 (6): 589-605.
- [2] 朱洪峰, 熊伟, 崔亚奇, 等. 新的自适应转弯模型的 IMM 算法研究 [J]. 计算机工程与应用, 2019, 55 (17): 252-258.
- [3] 张娜, 王锐, 蔡炯. 基于机动检测的参数自适应跟踪算法 [J]. 信号处理, 2022, 38 (2): 367-374.