

## 人机共融的远程态势智能感知系统\*

牛文龙<sup>1,2</sup>, 樊铭瑞<sup>1,2</sup>, 李 运<sup>1</sup>, 彭晓东<sup>1</sup>, 谢文明<sup>1</sup>, 任敬义<sup>1</sup>, 杨 震<sup>1</sup>

(1. 中国科学院国家空间科学中心 复杂航天系统电子信息技术重点实验室, 北京 100190;

2. 中国科学院大学, 北京 100049)

**摘要:**针对人与远程无人设备协同精准配合的迫切需求,以机器人操作系统为基础构建了一种人机共融的远程态势感知系统,并开展了实验与分析。该系统以视觉定位技术为基础,以人机感知共融为切入点,通过实时三维场景重建技术与场景一致性融合方法,将无人设备探测到的环境及目标信息进行三维重构,并通过增强现实设备进行显示,与人的视觉信息进行一致性融合,实现无GPS条件下远程无人设备与人所佩戴的增强现实设备之间的协同定位。实验结果表明,系统在近距离时具有较好的人机协同定位准确度,定位精度随着距离的增加而逐渐降低。所构建的系统使无人设备成为人眼的延伸,在不干扰人员正常行动的情况下实现了穿障碍、跨视距的感知能力,在未来信息化作战中可发挥重要作用。

**关键词:**人机共融;远程态势感知;三维场景重建;协同定位;增强现实

**中图分类号:**TP39 **文献标志码:**A **文章编号:**1001-2486(2021)06-085-10

## Remote situational intelligent sensing system for human-machine integration

NIU Wenlong<sup>1,2</sup>, FAN Mingrui<sup>1,2</sup>, LI Yun<sup>1</sup>, PENG Xiaodong<sup>1</sup>, XIE Wenming<sup>1</sup>, REN Jingyi<sup>1</sup>, YANG Zhen<sup>1</sup>

(1. Key Laboratory of Electronics and Information Technology for Space Systems, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Aiming at the urgent need for human and remote unmanned devices to cooperate precisely and accurately, a remote situational awareness system of human-machine integration was proposed based on the robot operating system, and experiments and analysis were carried out. Based on visual positioning technology, with the integration of human-machine perception as the breakthrough point, through real-time 3D scene reconstruction technology and scene consistency fusion method, the environment and target information detected by unmanned equipment were 3D reconstructed. The result was consistent and fused with the human visual information, and displayed by the augmented reality device, realizing the coordinated positioning between the remote unmanned device and the augmented reality device worn by the person without GPS. The experimental results show that the system has better performance at close range. The accuracy of human-machine coordinated positioning is gradually reduced as the distance increases. The proposed system makes the unmanned device an extension of the human eye, realizing the ability to penetrate obstacles, and crosses the sight distance without interfering with the normal movement of personnel. It can play an important role in future information operations.

**Keywords:** human-machine integration; remote situational awareness; 3D scene reconstruction; collaborative positioning; augmented reality

随着人工智能与机器人技术的快速发展,智能机器人应用范围已经从搬运、安装等生产活动,扩展到引导、陪护、交流等生活活动,再到排雷、侦察、打击等军事活动,渗透到了人类的方方面面。因此,人与智能机器人的共生协同将成为未来工作生活的新常态,在未来相当长一段时间内,在现有的技术体系基础上构建“以人为中心的有人/无人协作智能”就成为一种必然且可行的方向选择。不论是美军2015年提出的“半人马”模式、

美国国防部高级研究计划局(Defense Advanced Research Projects Agency, DARPA)从“AI”到“IA”技术布局转变,还是人机技术领域提出的“Human-Center Computing”“Human-Engage Computing”等人机有机协同新概念,以及我国2018年“智能机器人”重点专项中的对“人-机器人智能融合技术”方向的布局,都反映出该人机协同技术的重要性与迫切性。但是,从当前的工业、服务、军事应用机器人的使用情况来看,在控

\* 收稿日期:2020-05-08

基金项目:中科院基础前沿科学研究计划资助项目(22E0223301)

作者简介:牛文龙(1988—),男,辽宁营口人,助理研究员,博士,E-mail:niuwenlong@nssc.ac.cn;

彭晓东(通信作者),男,研究员,博士,博士生导师,E-mail:pxd@nssc.ac.cn

制方式上大多还属于遥操作,不利于人员快速与多个无人系统进行协同配合,在信息交互手段上多以手持终端为主,不利于危险或应急情况下人员对局势的快速掌控与决策,因此需要一种新的人机协作信息应用模式与方法。

现代化军事作战中,多样化的侦察与反侦察手段使得战场变得更加复杂。如何在复杂环境下快速、准确地感知己方及敌方的战场态势显得十分重要<sup>[1-2]</sup>。一方面,作战人员在高度对抗状态下,除要完成自身的战术动作外,还要对战斗环境中的敌我部署、火力布置、威胁来源等态势信息进行现场收集分析并完成决策。另一方面,战场环境复杂、形式多变,作战个体多数情况下无法在视距范围内直接观察敌方行为,作战计划往往陷入被动。这对作战人员的作战能力提出了新的挑战。

针对复杂战场下作战人员对态势全方面感知的迫切需求,为进一步增强作战人员对远程态势的感知能力以及协同作战能力,本文构建了一种人机共融的远程态势感知系统,该系统在室内等无 GPS 的条件下,利用灵活、易于操作的无人机/无人车等无人平台<sup>[3]</sup>,搭载场景智能感知设备,来实时获取远程的环境及目标信息,将所获取的信息通过无线网络实时传输到服务器端进行融合处理,实现对远程态势的三维重构,并将重构结果发送到作战人员所佩戴的增强现实设备中进行显示,与人的视觉信息进行一致性融合。本文介绍了系统所涉及的基于视觉的人机协同定位技术、实时三维场景重建技术和场景一致性融合方法等关键技术和流程,并构建了以机器人操作系统(Robot Operating System, ROS)为基础的系统实现架构,对定位精度进行了实际实验分析。

## 1 单兵态势感知系统发展现状

早在 20 世纪 80 年代,美国就率先提出了单兵信息化系统的概念,并于 90 年代启动“陆地勇士”计划,该计划借助微型计算机、传感器、通信导航设备等硬件装备嵌入至单兵可穿戴系统中<sup>[4-5]</sup>。随后,英、俄、法、德、日等国也相继推出各自的“未来士兵”计划<sup>[6-7]</sup>:如英国的“重拳”系统、法国的 FELIN 系统、德国的 Idz-es 系统及日本的“下一代近战信息共享研究”单兵作战系统等。这些系统使用不同功能的显示设备以获取战场态势信息,但皆存在与真实场景分离的缺陷:显示器的二维信息和真实的三维场景不尽相同,士兵需要在真实环境和手持/头戴设备之间频繁切换,造

成了士兵注意力的分散和实时反应能力难度的增加。同时,以上系统均只能探测可视范围,无法感知更远距离的威胁。

增强现实(Augmented Reality, AR)技术能让士兵视觉感知到经过处理后的态势信息与真实场景一致性融合的画面,使得作战变得更加高效和便捷<sup>[8-9]</sup>。增强现实眼镜可以将战场态势信息自然叠加在士兵的前方视野,并结合语音和手势等交互方式进行控制。2000 年,美国推出了第一个战场增强现实原型系统(Battlefield Augmented Reality System, BARS)<sup>[10]</sup>,并在此基础上启动了“ULTRA-Vis (urban leader tactical response, awareness & Visualization)”项目<sup>[11]</sup>,旨在研发适用于战场环境的 AR 软硬件。该项目设想不用离开当前视野,系统就可以在 AR 眼镜上为指挥员实时显示队员位置<sup>[12]</sup>,并能为作战人员标记目标区域。ULTRA-Vis 项目研发的“军用增强现实软件系统”(Augmented Reality Command Control Communicate Coordinate, ARC4)<sup>[13]</sup>,主要负责完成士兵户外位姿跟踪、地理配准与可视化以及网络通信管理等工作。2017 年美国陆军发布了“TAR(tactical augmented reality)”,一款更侧重于室内环境的单兵战术增强现实概念系统<sup>[14]</sup>。2019 年美国陆军公开了一款基于增强现实眼镜设备 HoloLens 2 的军用集成视觉增强系统“IVAS (integrated visual augmentation system)”,其透过 AR 眼镜能看到叠加在真实环境上显示敌我位置的场景三维地图,并能实现随眼镜的位置和姿态实时更新地图的功能<sup>[15]</sup>。综上所述,现有的基于增强现实的作战手段主要侧重于通过增强现实眼镜在真实环境上通过加标签等方式叠加信息层,并不具备对远程场景的三维感知能力以及协同定位能力。

作为单兵态势感知系统的重要组成,无人机/无人车等无人平台具有体积小、行动灵活、易于操控、隐蔽性好等特点,通过搭载场景感知设备,广泛应用在单兵协同作战以及作战侦察中<sup>[15-16]</sup>。以指针(Pointer)、大乌鸦(Raven)、龙眼(Dragon Eye)等为代表的小型无人机,可由单兵随身携带<sup>[17]</sup>,起飞后可自主导航至目标区域,并靠机身的光学、红外等传感器进行探测,再将获取的信息进行回传。

未来信息化战场必将是单兵结合各类传感器与其他作战单元进行全方位、高时效、高动态的协同作战。目前,美军正在加紧研发利用先进传感器和无人设备,并结合人机协同、增强现

实、态势感知等技术的“SXCT (squad X core technologies)”项目<sup>[18]</sup>,但其具体作战能力和战术指标并未公开。随着各国单兵作战能力的快速提升,研发我国自主的人机共融远程态势感知系统是面向未来信息化战场的必然发展趋势。

## 2 人机共融的远程态势智能感知系统架构

针对远程态势感知中无人设备与作战人员协同精准配合及态势融合显示的迫切需求,设计人机共融远程态势感知系统总体架构图如图1所示。

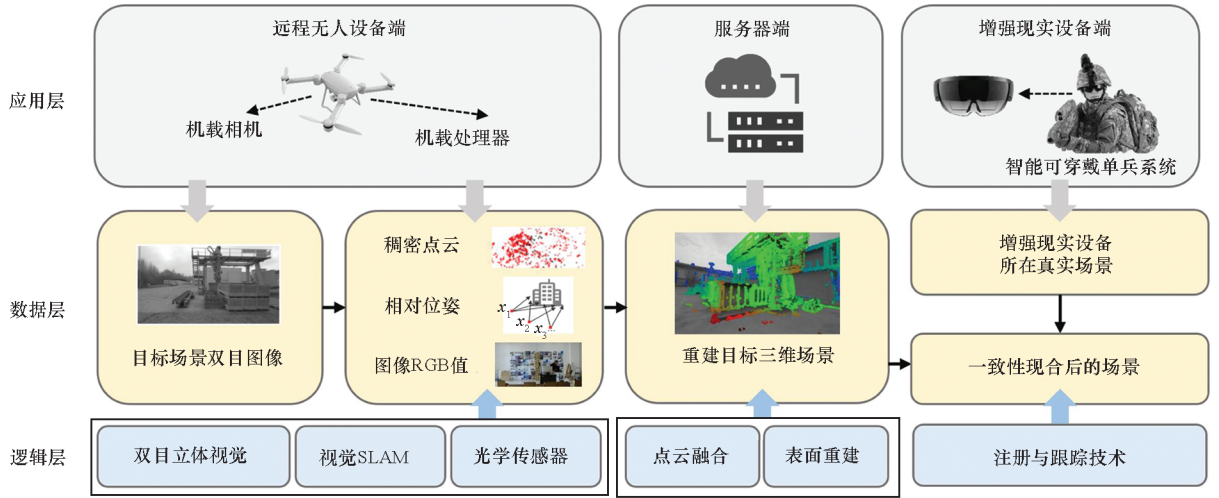


图1 人机共融的远程态势感知系统架构

Fig. 1 Architecture of remote situational awareness system based on human-machine integration

所设计的系统架构被划分为三个层次,分别是应用层、数据层、逻辑层。应用层由远程无人设备端、服务器端、增强现实设备端构成,其中无人设备配备机载双目相机和便携式的处理器,服务器端完成数据处理和场景重构,增强现实设备端主要装备智能可穿戴单兵系统,并将服务器端重构的场景由AR眼镜进行显示。服务器与远程无人设备、AR眼镜与服务器之间均通过WiFi连接,数据使用传输控制协议(Transmission Control Protocol, TCP)传输。系统主要实现过程如下:

1) 远程无人设备端:无人设备端的便携计算机完成初始化相对定位后,前往未知环境中获取前方态势信息。当无人设备进入目标区域,其使用机载相机获取目标场景双目图像和图像RGB值,利用视觉同时定位与建图(Simultaneous Localization And Mapping, SLAM)技术计算自身与环境的相对定位实现自主导航和避障,并通过机载处理器实时生成环境稠密点云。所获取的场景三维结构和目标状态,能协助用户完成目标区域的侦察。

2) 服务器端:以上数据传输到服务器端,由服务器端提供算力辅助,结合无人机的位姿和场景深度信息,以粗精度重构目标场景,并检测和标注目标物体,进而帮助作战人员在复杂战场环境中快速获取前方环境及人员情况。

3) 增强现实设备端:AR设备在初始化定位后,显示由服务器所提供的三维场景重建和目标识别结果,与AR设备所在真实场景和人的视觉信息进行一致性融合,呈现出“穿障碍”的虚实效果。

由于整个系统的数据交互比较复杂,故通过ROS的消息发布订阅机制简化数据交互的逻辑,降低各模块间的耦合度。ROS是一个开源的机器人操作系统,为软件开发人员提供了许多用于机器人实际应用的软件包和算法库,例如导航、定位、路径规划等功能。ROS最大的特点就是复用性,它能够设置不同功能的节点,多个节点之间以统一的消息格式来订阅和发布话题,从而降低机器人开发过程中硬件及接口的不一致性对其造成的影响。

系统构建了以ROS为基础的系统实现架构,所构建的ROS节点关系如图2所示,表1列举了各节点发布和订阅的话题。图像接收节点的作用是从无人机基站获取双目图像数据;特征点计算节点的作用是从图像中提取特征点;位姿估计局部建图回环检测节点计算无人机的位置和姿态变化矩阵;深度估计节点用于计算深度图和点云数据;实时建模节点使用深度图、点云数据、RGB图片和相机内参对点云进行网格化,并且实时输出整个场景的三维场景。

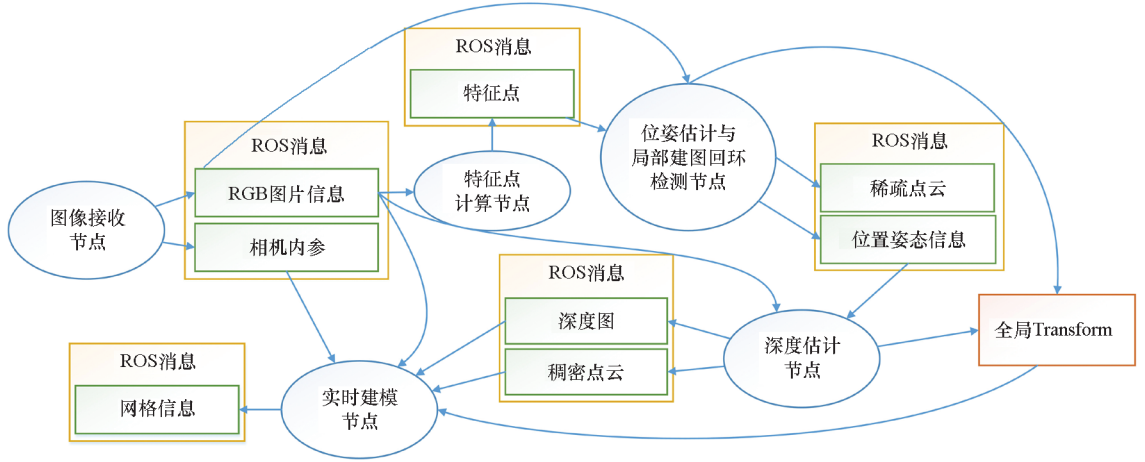


图 2 重构后 ROS 节点关系图  
Fig.2 Relationship of ROS nodes

表 1 ROS 节点发布与订阅话题

Tab.1 ROS nodes publish and subscribe topics

| 节点名称          | 发布话题                         | 订阅话题   |
|---------------|------------------------------|--|
| 图像接收          | 双目 RGB 图像<br>相机内参            | 无  |
| 特征点计算         | 图像特征点                        | 双目 RGB 图像  |
| 位姿估计与局部建图回环检测 | 稀疏点云<br>位置姿态<br>全局 Transform | 双目 RGB 图像、<br>图像特征点                              |
| 深度估计          | 深度图<br>稠密点云<br>全局 Transform  | 双目 RGB 图像、<br>位置姿态                               |
| 实时建模          | 网格信息                         | 相机内参<br>双目 RGB 图像<br>深度图<br>稠密点云<br>全局 Transform |

服务器端。服务器端接收到无人机端发送的数据后,进行解压处理并通过截断符号距离函数<sup>[19]</sup>(Truncated Signed Distance Function, TSDF)进行点云快速融合,然后使用体素化网格的方法实现点云的快速网格化生成网格模型,形成便于观察的可视化地图。最后在 AR 设备端完成实时显示,通过增强现实跟踪和注册技术实现一致性融合等功能。

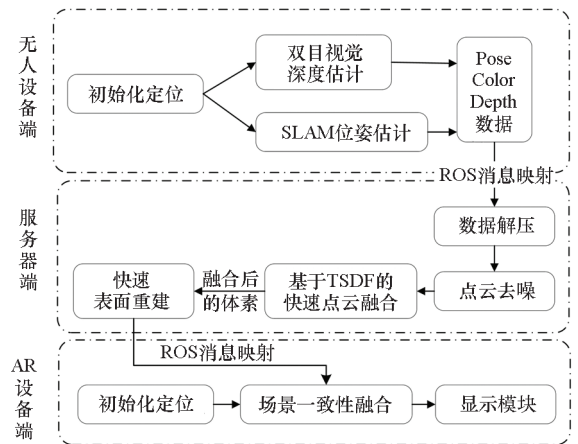


图 3 人机共融的远程态势智能感知技术流程图  
Fig.3 Flowchart of human-machine integration remote situation intelligent perception technology

### 3 人机共融的远程态势智能感知系统实现

人机共融的远程态势感知系统主要包括远程无人设备端的基于视觉的人机协同定位、服务器端的实时三维场景重建和增强现实设备端的场景一致性融合。

系统使用指定基准,并通过基于视觉特征的初始化相对定位的方式对无人设备和增强现实设备进行初始协同定位。如图 3 所示,无人设备在未知环境中使用 SLAM 进行实时定位,并利用双目立体视觉估计深度数据,通过 ROS 消息映射到

#### 3.1 基于视觉的人机协同定位

基于视觉的人机协同定位技术采用计算机视觉算法,可在无 GPS 的室内环境中实现无人机与 AR 设备间的实时相对定位。这使得无人机拍摄到的场景随着二者的移动仍能精确地注册到佩戴者所看到的真实场景中,实现“穿障碍,跨视距”的融合效果。



### 3.1.1 基于视觉特征的初始化相对定位

系统采用已知位置和大小视觉基准码对设备进行初始化相对定位。具体流程如图 4 所示,在初始化相对定位前需标定和校准无人机与 AR 眼镜上用于定位的机载相机内参  $H_1$ 、 $H_2$ ,并测量机载相机在设备上的具体位置,即相对于设备质心的位置。随后,确定世界坐标系,测量已知基准码大小和其在世界坐标系下的位置,并分别求得无人机和 AR 眼镜相对于基准码的相对位姿,进一步可得二者的相对定位。

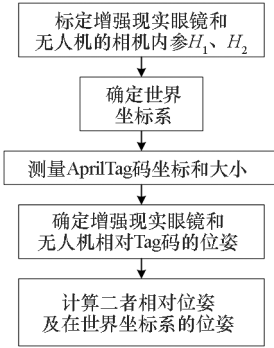


图 4 AR 眼镜和无人机的初始相对定位流程

Fig. 4 Initial relative positioning process of AR glasses and drones

常见的基准有 ARToolkit<sup>[20]</sup>、ARTag<sup>[21]</sup>、AprilTag<sup>[22]</sup>等。考虑到速率与鲁棒性,系统采用 AprilTag 码对无人机与 AR 设备进行初始相对位置定位。AprilTag 视觉定位方法包括检测线段、检测方形、单应性矩阵估计三个步骤。

1) 检测线段:计算图像中每个像素梯度的大小和方向,再将梯度方向和大小相近的像素聚集到一起。聚类时首先创建一个图,将图像中的每个像素表示成图中的一个节点,相邻节点间建立一条边,边的权重为沿梯度方向的两像素之间的差异。对于每个边,判断相应像素间是否应当连接在一起。将每个边按权重排序,以判断连通分支是否可以合并。

2) 检测方形:采用深度优先策略进行四轮搜索。第一轮搜索时,找到每一条线段,后面每一轮都搜索起始位置与上一轮搜索到线段的结束位置接近的线段。

3) 单应性矩阵估计:最后利用直接线性变换(Direct Linear Transform, DLT)算法计算单应性矩阵。

本实验中,无人机和增强现实眼镜位置与真实位置精度误差控制在 1 cm 以内,基本可以满足定位精度要求。

### 3.1.2 基于双目视觉的实时定位

未知环境中运动的相机或者机器人可依赖 SLAM 算法通过对环境的观测确定自身的运动轨迹,并同时构建连续地图。

视觉 SLAM 主要分为两个流派:特征点法和直接法。基于特征点法的 ORB-SLAM2<sup>[23]</sup>算法流程如图 5 所示,其采用 ORB 特征作为视觉特征,匹配相邻帧的对应特征点,根据图像的变化反向计算出相机的运动,实现室内场景下以高精度实现设备自身的定位和建图功能,具有性能高且适用于计算资源小的设备等优点。

ORB-SLAM2 在 TX2 可以通过 ROS 和 C++ 两种方式编译运行,并实现厘米级别的精确定位。整个系统分为追踪、局部建图、回环检测 3 个并行处理的线程。

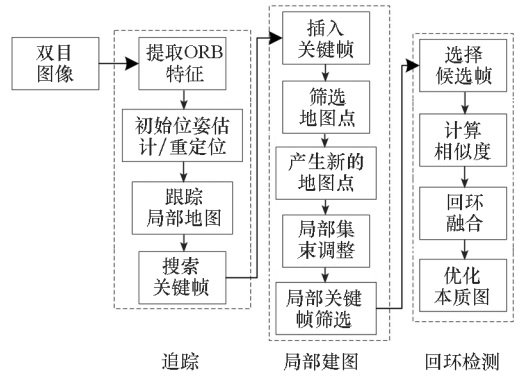


图 5 ORB-SLAM2 算法流程

Fig. 5 ORB-SLAM2 algorithm flowchart

## 3.2 实时三维场景重建

### 3.2.1 基于双目立体视觉的深度估计

稠密的三维场景重建需要计算场景中像素点的深度信息,仅由特征点法得到的稀疏地图是远远不够的。通过结构光或飞行时间(Time-of-Fly, ToF)等直接测量物体深度的方法易受到光照影响,且探测范围有限。故本系统采用基于双目立体视觉的深度估计算法,这部分在无人机端进行,其中无人机携带的双目摄像头为 ZED 立体相机,随后将估计得到的深度信息传给服务器端进行进一步的点云融合和三维重建。

一般基于双目的深度估计主要包括极线搜索、立体匹配、视差输出和深度信息提取四个步骤。本系统的立体匹配算法采用半全局块匹配<sup>[24]</sup>(Semi-Global Block Matching, SGBM)算法,其优势是视差效果好且运算速度快。SGBM 选取匹配的像素点和其周围的像素点计算它们的视差,并构成视差图。建立视差图的能量函数,并求解其最小值。其中能量函数为:

$$E(D) = \sum_p \{C(p, D_p) + \sum_{q \in N_p} P_1 I[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 I[|D_p - D_q| > 1]\} \quad (1)$$

式中： $D$  是视差图； $E(D)$  为视差图的能量函数； $p$ 、 $q$  为像素点； $N_p$  为像素点  $p$  附近的像素点； $P_1$ 、 $P_2$  是两个惩罚系数； $I[\cdot]$  判断函数的真假，若为真，则返回 1，反之返回 0。两个正则项可以平滑视差图和控制图像边缘位置不连续性。

通过对每个像素点进行上述操作，得到像素点的视差图，结合双目相机的基线便可估计得到深度值。

### 3.2.2 基于 TSDF 的快速点云融合

TSDF 是一种对空间体素进行描述来计算空间中物体表面位置的方法，可以在 GPU 显存中并行计算，具有时间和空间的高效性。系统利用 TSDF 使用点云构造体素的方法快速融合点云，同时根据 TSDF 构建过程中更新权重的方法来实时更新融合后的点云分布结果，并据此构建体素模型。

符号距离函数 (Signed Distance Function, SDF) 用来重建三维空间模型，使用大小相同的体素来构成三维空间。如图 6 所示， $x$  表示某体素中心， $P$  为相机光心  $O$  经体素  $x$  的射线与障碍物表面的交点， $sdf(x)$  表示该体素中心与最近物体表面沿相机光轴的带符号距离。其中，物体外体素的距离值为正，物体内部体素的距离值为负； $i$  表示第  $i$  次的观测值。

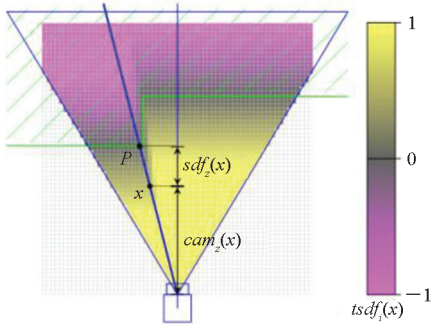


图 6 TSDF 示例  
Fig. 6 TSDF sample

根据图 6 中关系， $sdf_i(x)$  可表示为：

$$sdf_i(x) = depth_i(pic(x)) - cam_z(x) \quad (2)$$

其中， $pic(x)$  表示某体素中心  $x$  到深度图的投影， $depth_i(pic(x))$  为相机射出的光线经过  $x$  到达最近物体表面点  $P$  的测量深度， $cam_z(x)$  为该体素沿相机光轴到相机的距离。

由于在空间中，大部分体素并不在物体表面，故使用 TSDF，设定固定的截断距离  $t$ ，避免对

距离表面较远的体素进行计算带来的不必要时间和空间浪费。对于该体素的第  $i$  次观测可表示为：

$$tsdf_i(x) = \max\left(-1, \min\left(1, \frac{sdf_i(x)}{t}\right)\right) \quad (3)$$

因为单次计算准确度并不高，且视场有限，空间存在大量未被定义的体素，故需要随着相机的移动不断更新空间中体素的截断距离，更新后的截断距离表示为  $TSDF_i(x)$ 。从相机视点使用光线投射，找到 TSDF 符号变化附近的体素并对其进行空间中的三线性插值，即可得到精确的物体表面点的位置。

### 3.2.3 基于体素的实时表面重建

经过 TSDF 处理过后的三维场景只是离散的点云集合，不能显式地表达场景的表面信息，因此与正常可视化习惯存在较大差异。为了能将前方态势信息实时显示在用户视野，考虑到现有增强现实眼镜处理器的性能限制，本系统使用对点云空间快速划分的方法——对融合后的点云通过体素化网格的方式实现表面重建。该方法对全局的点云空间进行三维体素栅格处理（文中采用 0.05 m 大小的三维立方体），如图 7 所示，并用不同颜色区分不同深度区域，实现对空间的划分和表面信息的表达。在实时显示场景的同时，也能尽量拟合物体表面形状，达到提升效率和提高地图可视化能力的效果。



(a) 实验场景  
(a) Experimental scene



(b) 对场景三维重建  
(b) 3D reconstruction of the scene

图 7 实际的三维重建效果图

Fig. 7 Effect diagram of actual 3D reconstruction

### 3.3 场景一致性融合

基于头戴式 AR 眼镜的场景一致性融合可以将重建场景显示在 AR 眼镜视野的对应位置,便于用户的查看和交互,实现穿越障碍感知场景态势的目标。这其中包括增强现实的注册和跟踪技术。增强现实的注册技术能将重建场景由其所在的坐标系转换到眼镜坐标系。跟踪技术可以保障用户在运动过程中,跟踪观察者不断变化的视点并实时更新坐标系的对应关系,这就需要眼镜要根据周围环境和自带传感器的变化不断估计自身位姿。

场景一致性融合流程图如图 8 所示,无人机通过指定路径到达陌生的待重建环境之后,根据指令进行三维场景重建。通过 SLAM 自主定位算法,无人机便得知机载相机和重建的三维模型在每个时刻的实时相对位姿。由于在初始化定位部分完成了无人机和 AR 眼镜的相对定位并统一了二者的坐标系,重建后的真实场景便可以对应显示在 AR 眼镜视野中:如目标场景在用户左前方,观察左前方视野便能看到重建后的场景叠加在真实环境之上,达到场景的一致性显示效果。

过惯性传感器来获取 AR 系统中摄像机位姿的实时变化量,通过累积的方式求解每个时刻点的相机位姿。这种跟踪技术所需要的数据量较少,实时性能较好,但随着时间的累积,精确度会有所下降。为了弥补惯性传感器的累积误差,通过 SLAM 算法实时构图并解析摄像机位置对惯性传感器的结果进行修正。

### 4 实验与分析

真实应用场景中,便携式计算设备需要利用无人机进行搭载,而为了测试易于操作,实验采用三角支架代替无人机平台。为便于误差估计,三角支架与待重建目标距离为 1 m,支架所支撑的相机高度设置为 1.6 m——近似人眼高度。另外,真实应用场景中重建目标与 AR 眼镜佩戴者之间应存在障碍物遮挡,而为了测试的可行性,实验中使重建目标位于 AR 眼镜佩戴者视野范围内。便携式计算设备采用 NVIDIA Jetson TX2,双目相机采用 ZED 立体相机,AR 眼镜采用 Microsoft HoloLens,服务器和以上硬件的参数设置如表 2 所示。其中,TX2 和服务设备均为 Ubuntu16.04 操作系统,ROS 平台版本为 Kinetic。

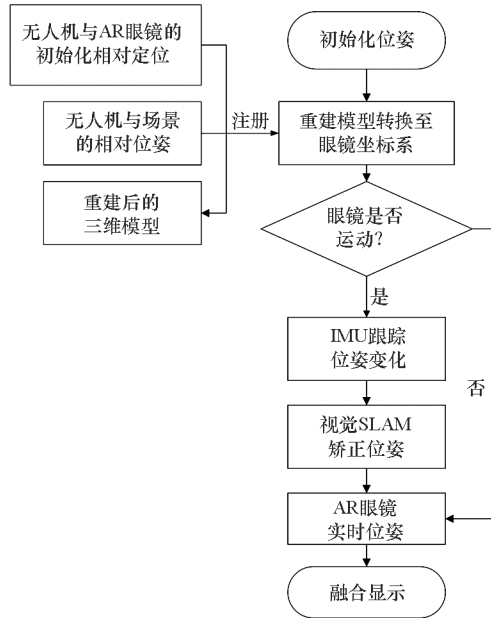


图 8 场景一致性融合流程图

Fig. 8 Flowchart of scene consistency fusion technology

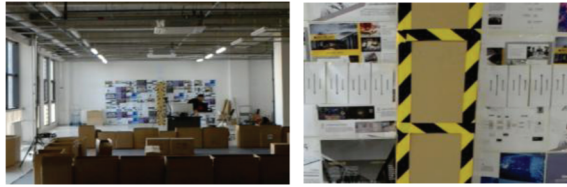
由于场景大小不定,用户可能需要转动头部甚至在周围活动才能完整地观察到整个模型的细节。系统使用的 AR 眼镜 Microsoft HoloLens 携带有 4 个环境感知摄像头、1 个深度摄像头、1 个惯性测量单元,这些传感器对 HoloLens 的自主定位起到主要作用。当相机在场景中移动时,需要实时获取变化后的相机位姿。由于 AR 系统带有惯性传感器,因此可以采用基于惯性的跟踪技术,通

表 2 实验硬件设备参数设置

Tab. 2 Parameter setting of hardware device

| 硬件        | 配置参数  |
|-----------|---|
| 低功耗移动计算设备 | GPU: NVIDIA Pascal™, 256 CUDA cores<br>CPU: HMP Dual Denver 2/2 MB L2 + Quad ARM® A57/2 MB L2<br>数据存储: 32 GB eMMC, SDIO, SATA<br>内存: 8 GB 128 位 LPDDR458.3 GB/s |
| 双目相机      | 深度感知距离: 70 cm ~ 20 m<br>视野范围: 110°广角视频和深度<br>帧率: 30 帧/s(1 080P 高清)和 100 帧/s(WVGA)<br>图像分辨率: 2 000 像素  |
| 服务器       | 处理器: AMD Ryzen 7 700 Eight-Core P<br>显卡: GeForce GTX 1 080 Ti/PCIe/SSE2<br>内存: 16 GB<br>硬盘: 1 TB SSD  |
| 增强现实眼镜    | 全息分辨率: 最高 230 万光学点<br>处理器: Intel32 位体系结构;<br>定制的 Microsoft 全息处理单元(HPU 1.0)<br>内存: 64 GB 闪存;2 GB RAW(2 GB CPU 和 1 GB HPU)<br>WiFi 输入: 802.11AC                   |

实验中,在 TX2 上连接 ZED 相机进行图像和深度数据的采集,并发送给服务器做信息融合处理,并对所处场景进行多分辨率重构及对自身进行高精度定位,之后将处理结果通过网络发送给增强显示设备进行一致性融合展示,原型系统的功能及性能指标通过展示结果来验证。结合实验场地的实际范围,本文搭建了三个不同测试距离的场景对无人机协同定位精度进行测试与分析,如图 9 所示。



(a) 场地一 (a) Site 1 (b) 测试立柱 (b) Test pillar



(c) 场地二远景 (c) Long shots of site 2 (d) 场地二近景 (d) Close shots of site 2

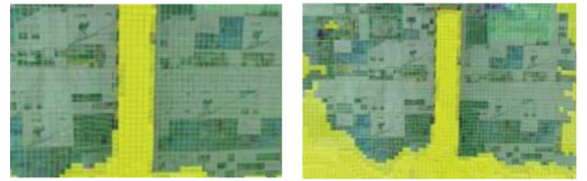


(e) 场地三 (e) Site 3

图 9 实验场景设置

Fig. 9 Experimental scenario

实验中,观察者佩戴 AR 眼镜观察作为目标的  $2\text{ m} \times 0.3\text{ m} \times 0.3\text{ m}$  黄色立柱,并在黄色立柱的各个棱线及中间张贴黑黄相间的条纹以对结果进行分析。白色墙壁特征点稀疏的问题可通过在立柱两侧的墙壁上随机张贴黑黄相间的条纹来解决。如图 10 所示,在场地一中测量观察者与目标距离分别为  $3\text{ m}$ 、 $5\text{ m}$ 、 $8\text{ m}$ 、 $10\text{ m}$  和  $20\text{ m}$ ,在场地二中测量观察者与目标距离分别为  $30\text{ m}$ 、 $45\text{ m}$ 、 $60\text{ m}$ ,场地三中为  $100\text{ m}$ 。对于每个距离,观察者在所处位置左右适当范围内移动,同时录制视频记录 AR 眼镜观察到的画面。对于每一个视频,随机提取 20 张不同视角的图片作为该位置点的采样图片用来计算定位误差。



(a) 场地一采样距离  $3\text{ m}$  (a) Sampling distance  $3\text{ m}$  in Site 1 (b) 场地一采样距离  $5\text{ m}$  (b) Sampling distance  $5\text{ m}$  in Site 1



(c) 场地一采样距离  $8\text{ m}$  (c) Sampling distance  $8\text{ m}$  in Site 1 (d) 场地一采样距离  $10\text{ m}$  (d) Sampling distance  $10\text{ m}$  in Site 1



(e) 场地一采样距离  $20\text{ m}$  (e) Sampling distance  $20\text{ m}$  in Site 1 (f) 场地二采样距离  $30\text{ m}$  (f) Sampling distance  $30\text{ m}$  in Site 2



(g) 场地二采样距离  $45\text{ m}$  (g) Sampling distance  $45\text{ m}$  in Site 2 (h) 场地二采样距离  $60\text{ m}$  (h) Sampling distance  $60\text{ m}$  in Site 2



(i) 场地三采样距离  $100\text{ m}$  (i) Sampling distance  $100\text{ m}$  in Site 3

图 10 场景重建的实验采样结果

Fig. 10 Sampling results of scene reconstruction

场地一近距离的采样图片如图 10 中的 (a) ~ (c) 所示,其采样距离分别为  $3\text{ m}$ 、 $5\text{ m}$ 、 $8\text{ m}$ ,由于在此距离下可清晰观察到黄色立柱旁边设定的最小刻度为  $2.5\text{ cm}$  的标尺,故利用标尺上的刻度求得虚实立柱之间的实际距离偏差为:



$$\delta_1 = 2.5 \times a/b \quad (4)$$

其中,  $a$  为虚实立柱之间偏差的像素数,  $b$  为刻度在图像中所占的像素数。

场地一中等距离的采样图片如图 10(d) ~ (e) 所示, 其采样距离分别为 10 m、20 m, 由于此距离下在 AR 眼镜中已无法清晰观察黄色立柱旁的标尺, 故利用虚拟场景中边长均为 5 cm 的立方体模块求得虚实立柱之间的实际距离偏差为:

$$\delta_2 = 5 \times a/c \quad (5)$$

其中,  $c$  为立方体水平边在图像中所占的像素数。

场地二、场地三中远距离的采样图片如图 10(f) ~ (i) 所示, 其采样距离分别为 30 m、45 m、60 m、100 m, 由于立柱旁门框的实测宽度为 1.5 m, 故可得虚实立柱之间的实际距离偏差为:

$$\delta_3 = 150 \times a/d \quad (6)$$

其中,  $d$  为门框的水平边在图像中所占的像素数。

通过以上计算可得到 AR 眼镜与立柱不同距离下的计算定位误差, 如表 3 所示。定位误差随距离变化趋势如图 11 所示。从图 11 和表 3 可以看出, AR 眼镜视场范围 20 m 以内所构建的虚拟场景与真实环境之间的空间位置误差小于 10 cm, 20 m 至 100 m 范围内所构建的虚拟场景与真实环境之间的空间位置误差小于 35 cm。

表 3 误差与距离关系

Tab. 3 Error and distance relationship

| 距离/m | 定位误差/cm |
|------|---------|
| 3    | 1.11    |
| 5    | 1.62    |
| 8    | 2.50    |
| 10   | 4.17    |
| 20   | 9.17    |
| 30   | 11.86   |
| 45   | 14.89   |
| 60   | 19.10   |
| 100  | 32.71   |

实验结果表明, 协同定位误差随着 AR 眼镜与前端感知设备距离的增加而逐渐增加, 主要原因在于当二者的距离变大时, 前端无人感知设备和 AR 眼镜之间的累计误差也会逐渐增大, 该误差一方面是由于惯性传感器等硬件造成的漂移, 另一方面是因为定位算法本身的累积误差。当重建好的三维场景不能按照坐标系的转换关系准确

地叠加在眼镜视野上, 就会出现匹配不准的情况。

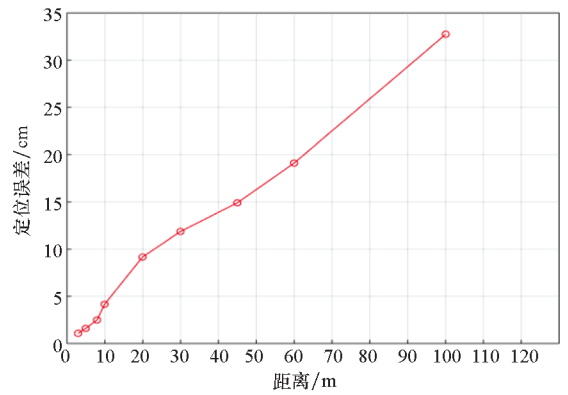


图 11 定位误差与测试距离关系

Fig. 11 Relationship between positioning error and test distance

上述实验中, 从前端的无人机定位到服务器端重建再到 AR 眼镜端显示这一过程的平均处理速度总计为 15 帧/s, 定位和建图效率基本可以满足实时性的要求。

## 5 结论

针对复杂战场环境下作战人员的态势感知需求, 本文设计了一种人机共融的远程态势感知系统, 并进行了初步的系统实现和关键技术验证。实验结果表明, 所构建的系统能够在近距离时有较好的协同定位精度, 定位精度随着距离的增加而减小。本文的工作能够为人机融合作战系统的设计提供一定的启示作用, 但是为了适应大场景作战的应用需求, 在未来的工作中将从以下三点展开更完善的系统研究工作: ①分布式多节点的态势融合, 实现多无人设备协同探测及信息融合处理; ②加入对环境的关键目标的识别能力, 并对系统延迟进行统一分析; ③本文演示系统中使用的是 WiFi 无线通信, 但是在真实战场环境中, 为了保证无线通信的安全性、即时性和可靠性, 则更多使用军用通信系统, 包括卫星通信、4 G 以及正在发展的 5 G 技术, 后续将对无线通信系统进行进一步的完善。

## 参考文献 (References)

[1] 任智源, 肖尧, 郭凯, 等. 适应低时延业务需求的分布式可穿戴单兵作战信息系统[J]. 国防科技大学学报, 2018, 40(4): 159-165.  
REN Zhiyuan, XIAO Yao, GUO Kai, et al. Distributed wearable individual soldier combat information system for low-latency business [J]. Journal of National University of Defense Technology, 2018, 40(4): 159-165. (in Chinese)

[2] 原大川, 王涛, 李丽君, 等. 单兵班组多平台协同作战应用研究[J]. 火力与指挥控制, 2019, 44(6): 148-152.



- YUAN Dachuan, WANG Tao, LI Lijun, et al. Cooperative engagement operation research on soldier squad and multiple platforms [J]. *Fire Control & Command Control*, 2019, 44(6): 148–152. (in Chinese)
- [3] 张代兵, 王勋, 钟志伟, 等. 融合地面多传感器信息引导无人机着陆[J]. *国防科技大学学报*, 2018, 40(1): 151–156.
- ZHANG Daibing, WANG Xun, ZHONG Zhiwei, et al. Guidance of unmanned aerial vehicles landing by ground-based multisensory fusion[J]. *Journal of National University of Defense Technology*, 2018, 40(1): 151–156. (in Chinese)
- [4] 邓威, 张德彬. 智能可穿戴设备军事应用与发展趋势[J]. *国防科技*, 2016, 37(1): 57–60.
- DENG Wei, ZHANG Debin. The research on military application and development tendency of intelligent wearable devices[J]. *National Defense Science & Technology*, 2016, 37(1): 57–60. (in Chinese)
- [5] MURRAY J. Wearable computers in battle: recent advances in the land warrior system [C]//*Proceedings of Fourth International Symposium on Wearable Computers*, 2000: 169–170.
- [6] 郭亮. 单兵信息化系统及应用研究[D]. 昆明: 云南大学, 2012.
- GUO Liang. Research on individual soldier information system and its application [D]. Kunming: Yunnan University, 2012. (in Chinese)
- [7] LE SUEUR P. The Felin soldier system; a tailored solution for networked operations [C]//*Proceedings of Defense and Security Symposium*, 2007, 6542: 65422X.
- [8] YOU X, ZHANG W W, MA M, et al. Survey on urban warfare augmented reality[J]. *ISPRS International Journal of Geo-Information*, 2018, 7(2): 1–16.
- [9] JALUI S, HAIT T, HATHI T, et al. Advanced military helmet aided with wireless live video transmission, sensor integration and augmented reality headset [C]//*Proceedings of International Conference on Communication and Electronics Systems*, 2019: 123–127.
- [10] JULIER S, BAILLOT Y, LANZAGORTA M, et al. Bars: battlefield augmented reality system [C]//*Proceedings of NATO Symposium on Information Processing Techniques for Military Systems*, 2000: 9–11.
- [11] ROBERTS D, MENOZZI A, COOK J, et al. Testing and evaluation of a wearable augmented reality system for natural outdoor environments [C]//*Proceedings of SPIE Defense, Security, and Sensing*, 2013, 8735: 87350A.
- [12] ROBERTS D C, SNARSKI S, SHERRILL T, et al. Soldier-worn augmented reality system for tactical icon visualization [C]//*Proceedings of SPIE Defense, Security, and Sensing*, 2012, 8383: 838305.
- [13] GANS E, ROBERTS D, BENNETT M, et al. Augmented reality technology for day/night situational awareness for the dismounted soldier [C]//*Proceedings of SPIE Defense + Security*, 2015, 9470: 947004.
- [14] 邓晨, 游雄, 张巍巍. 面向城市分队作战的增强现实研究[J]. *指挥控制与仿真*, 2019, 41(2): 1–10.
- DENG Chen, YOU Xiong, ZHANG Weiwei. Augmented reality in urban operations for unit dismounted warfighters [J]. *Command Control & Simulation*, 2019, 41(2): 1–10. (in Chinese)
- [15] KASE S E, PERRY V, ROY H, et al. Comparative analysis of mission planning and execution times between the microsoft HoloLens and the surface touch table [C]//*Proceedings of International Conference on Human-Computer Interaction*, 2020.
- [16] 李增彦, 李小民. 单兵无人机发展现状及关键技术分析[J]. *飞航导弹*, 2017(6): 23–29.
- LI Zengyan, LI Xiaomin. Development status and key technology analysis of individual soldier unmanned aerial vehicle [J]. *Aerodynamic Missile Journal*, 2017(6): 23–29. (in Chinese)
- [17] 汪浩洋, 杨梅枝. 美军无人机发展现状及趋势[J]. *飞航导弹*, 2020(2): 46–50.
- WANG Haoyang, YANG Meizhi. Development status and trend of U.S. military unmanned aerial vehicle [J]. *Aerodynamic Missile Journal*, 2020(2): 46–50. (in Chinese)
- [18] PARK S J, ALLEN S, SHIRES D. Urban reconnaissance planning: discovering new applications with tactical high performance computing [C]//*Proceedings of Military Communications Conference*, 2016: 1121–1124.
- [19] WERNER D, AL-HAMADI A, WERNER P. Truncated signed distance function: experiments on voxel size [C]//*Proceedings of International Conference Image Analysis and Recognition*, 2014: 357–364.
- [20] KATO H, BILLINGHURST M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system [C]//*Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality*, 1999: 85–94.
- [21] FIALA M. ARTag, a fiducial marker system using digital techniques [C]//*Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*, 2005: 590–596.
- [22] OLSON E. AprilTag: a robust and flexible visual fiducial system [C]//*Proceedings of International Conference on Robotics and Automation*, 2011.
- [23] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255–1262.
- [24] KIM K R, KIM C S. Adaptive smoothness constraints for efficient stereo matching using texture and edge information [C]//*Proceedings of International Conference on Image Processing*, 2016: 3429–3433.