doi:10.6041/j.issn.1000-1298.2024.12.037

基于卷积神经网络的移动机器人自适应光照增强单目视觉 SLAM 算法

陈久朋[1]　陈治帆[1]　伞红军[1,2]　赵龙云[1]　彭　真[1]

(1.昆明理工大学机电工程学院,昆明 650500;2.云南省先进装备智能制造技术重点实验室,昆明 650500)

摘要:移动机器人视觉 SLAM 技术能够在一定条件下实时估计自身在环境中的位置,并构建和更新环境稀疏或稠密三维地图,这些信息可以帮助机器人提高对未知复杂环境的准确感知和适应能力,以执行更复杂的任务。但使用相机作为传感器的视觉 SLAM 在定位和建图的精度和稳定性方面在很大程度上依赖于采集到的图像质量,在弱光照环境中,现有的视觉 SLAM 算法难以有效地工作。针对视觉 SLAM 在弱光照环境中定位精度降低和跟踪丢失的问题。本文提出了一种适应弱光照环境的 RLMV-SLAM 算法,该算法使用一个轻量化的神经网络对输入图像进行预处理,增强其亮度、对比度、色彩和去噪,同时,该算法使用地图点补充策略、Sparse BA 和一种实时增量闭环检测方法提高了定位和建图精度和鲁棒性。在公开数据集和自采数据集上对该算法进行了实验验证,并与其他主流视觉 SLAM 方法进行了对比,结果表明本文提出的方法将弱光照环境中有效跟踪时长提升 30% 以上,且在公开数据集上估计位姿的误差也有明显降低,证明了所提算法的有效性,为弱光照环境中同步定位和建图提供了一定参考。

关键词:移动机器人;视觉 SLAM;弱光照环境;卷积神经网络

中图分类号:TP460　　**文献标识码**:A　　**文章编号**:1000-1298(2024)12-0383-09　　**OSID**:

Adaptive Illumination Enhanced Monocular Vision SLAM Algorithm for Mobile Robots Based on Convolutional Neural Networks

CHEN Jiupeng[1]　CHEN Zhifan[1]　SAN Hongjun[1,2]　ZHAO Longyun[1]　PENG Zhen[1]

(1. *Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming* 650500, *China*
2. *Key Laboratory of Advanced Equipment Intelligent Manufacturing Technology of Yunman Province, Kunming* 650500, *China*)

Abstract: The visual SLAM technology of mobile robots can estimate their position in the environment in real time under certain conditions, and build and update sparse or dense 3D maps of the environment. This information can help robots improve their accurate perception and adaptability to unknown complex environments, and perform more complex tasks. However, the accuracy and stability of localization and mapping of visual SLAM using cameras as sensors largely depend on the quality of the collected images. In low-light environments, existing visual SLAM algorithms have difficulty working effectively. In response to the problems of reduced positioning accuracy and lost tracking faced by visual SLAM in low-light environments, a visual SLAM algorithm suitable for low-light environments, RLMV-SLAM was proposed. This algorithm used a lightweight neural network to preprocess the input images, enhancing their brightness, contrast, color, and denoising. At the same time, the algorithm applied a map point supplement strategy, Sparse BA, and a real-time incremental loop closure detection method to improve the accuracy and robustness of localization and mapping. The research experimentally verified this algorithm on public datasets and self-collected datasets, and compared it with other mainstream visual SLAM methods. The results showed that the method proposed can increase the effective tracking time in low-light environments by more than 30% and significantly reduce the pose error of pose estimation on public datasets, proving the effectiveness of the proposed algorithm and providing a reference for simultaneous localization and mapping in low-light environments.

Key words: mobile robots; visual SLAM; low-light environment; convolutional neural network

收稿日期:2024-06-19　修回日期:2024-08-30

基金项目:云南省基础研究计划项目(202301AU070059)和昆明理工大学人才培养项目(KKZ320230104)

作者简介:陈久朋(1993—),男,讲师,博士,主要从事机器人技术及应用研究,E-mail:18314490225@163.com

通信作者:伞红军(1976—),男,副教授,博士,主要从事机器人技术及应用研究,E-mail:sanhjun@163.com

0 引言

同步定位和建图(Simultaneous localization and mapping,SLAM)对自主移动系统至关重要。在基于视觉的 SLAM 中,仅使用相机作为传感器,成本低且能提供丰富的语义与纹理信息。这些信息有助于机器人在复杂环境中精确定位和建图,提升其执行复杂任务的能力。语义视觉 SLAM 通过语义约束提高定位精确度和鲁棒性,并构建出详细的语义地图[1-2],对机器人在真实环境中交互与理解至关重要。

近年来,视觉位姿估计方法在学术与商业领域取得显著进展[3-6],但性能受图像质量影响。主流算法分为直接法和基于特征的方法,直接法[7]对光照变化敏感,光照差异显著时易失效,增加了图像处理难度。基于特征的方法使用点[3-4]、线[8-11]、面[12]和语义信息[13]作为特征。尽管在稳定光照和丰富纹理环境中表现稳定,但在复杂光照下鲁棒性不足,且易受视角快速变化影响。研究显示 U-SIFT 具有较高光照不变性[14],但未涵盖广泛使用的 ORB 特征提取器[15]。目前,也有一些基于学习的特征描述方法,如 SuperPoint[16]、SuperGlue[17],或者基于神经网络的端到端 SLAM 方法[18-19]。但由于这些算法计算效率较低,尚未融入现有 SLAM 系统。

实际应用中,光照不足和急剧变化会显著降低视觉 SLAM 位姿估计的精度,尤其在室外和夜间环境,这成为视觉 SLAM 的主要挑战[20-22]。同时,在光照和纹理缺乏的室内环境,如无人值守的自动化场所,增加照明设备成本较高。

针对弱光照环境中的视觉 SLAM 方法,目前只有少数研究。LMVI-SLAM[14]方法通过伽马矫正和直方图均衡化技术提升了图像亮度和对比度,然而,在去噪和颜色恢复方面仍面临挑战,并且该算法尚未在户外环境中进行测试。DarkSLAM[23]采用生成对抗网络(GAN)对图像进行预处理,但 GAN 在色彩矫正和去噪方面的局限性影响了其稳定性,且其复杂的网络结构不适合实时 SLAM 系统。SINGH 等[24]对多种基于深度学习的图像增强方法和自动曝光技术在弱光照环境下对视觉 SLAM 性能提升进行了评估,发现即便使用这些技术,视觉 SLAM 在弱光照条件下仍会出现跟踪丢失的问题,且未能证明这些图像增强模块能够实时集成到 SLAM 系统中。

HUANG 等[25]提出了一种结合多种图像特征的 SLAM 方法,以增强弱光照下的鲁棒性,虽保持实时性能,但对建图任务无效。LIU 等[26]开发了基于物理 Retinex 模型的网络以增强视觉 SLAM 图像序列,但仅用 KITTI 数据集[27]评估,该数据集不含弱光照场景,故其有效性存疑。现有方法在图像预处理上取得了一定成果,但处理速度与效果间存在权衡,尚无实时稳定的视觉 SLAM 方法适用于弱光照环境。测试了 ORB-SLAM3[3]和 OV2-SLAM[4]在弱光照下性能,发现易出现姿态漂移或跟踪失败,表明主流的视觉 SLAM 方法对光照变化敏感。

本文基于 ORB-SLAM3 提出一种 RLMV-SLAM(Real time low-light monocular visual SLAM)算法。RLMV-SLAM 采用一个轻量级卷积神经网络,该网络能够有效增强图像亮度和对比度,同时去除噪声并校正色彩偏差,克服原始 SLAM 方法对输入图像光照条件的依赖,使得 RLMV-SLAM 能够在弱光照条件下稳定运行。

1 面向弱光照环境的单目视觉 SLAM 算法

RLMV-SLAM 是在 ORB-SLAM3[3]的基础上进行扩展的,RLMV-SLAM 系统架构如图 1 所示,包含图像增强模块和 SLAM 定位建图模块,采用分布式计算的方法,图像增强由 GPU 或 NPU 芯片负责计算,SLAM 定位和建图模块使用 CPU 进行计算。图像增强模块用于估计给定输入图像的 1 组最佳映射曲线系数,该组系数被用来迭代调整 V 通道图像以实现亮度增强。之后将原始的三通道图像与增强后的 V 通道图像沿着通道维度进行拼接,将拼接后的张量送入后续的卷积层,最终输出增强后图像。增强后图像由 SLAM 模块的特征提取器提取特征点,后依次经过跟踪线程、局部建图线程和闭环线程后输出相机估计位姿和稀疏点云。

1.1 弱光照环境中图像预处理

文献[28]提出应用图像特定曲线映射(Image specific curve mapping)对弱光照图像进行增强,将输入像素映射到新的输出像素,这种映射关系由一个特定的函数或曲线定义,可以根据特定的图像进行调整。这种策略可以增强图像,并且不会出现不现实的伪影。考虑到 SLAM 任务中实时性至关重要,本文使用 HSV 颜色空间中表示图像亮度 V 通道进行曲线映射,并且对同一幅图像中的所有像素使用 1 条全局映射曲线,而不是对输入图像的每个像素都给定 1 条映射曲线。迭代调整低光照图像计算式为

$$\boldsymbol{I}_{k+1}=\boldsymbol{I}_k+\boldsymbol{X}_k\boldsymbol{I}_k(\boldsymbol{I}-\boldsymbol{I}_k) \tag{1}$$

其中 $\boldsymbol{X}=[x_1 \quad x_2 \quad \cdots \quad x_n]$

式中 $\boldsymbol{X}$——1 组映射曲线系数

$\boldsymbol{I}$——输入的 V 通道图像

k——迭代次数

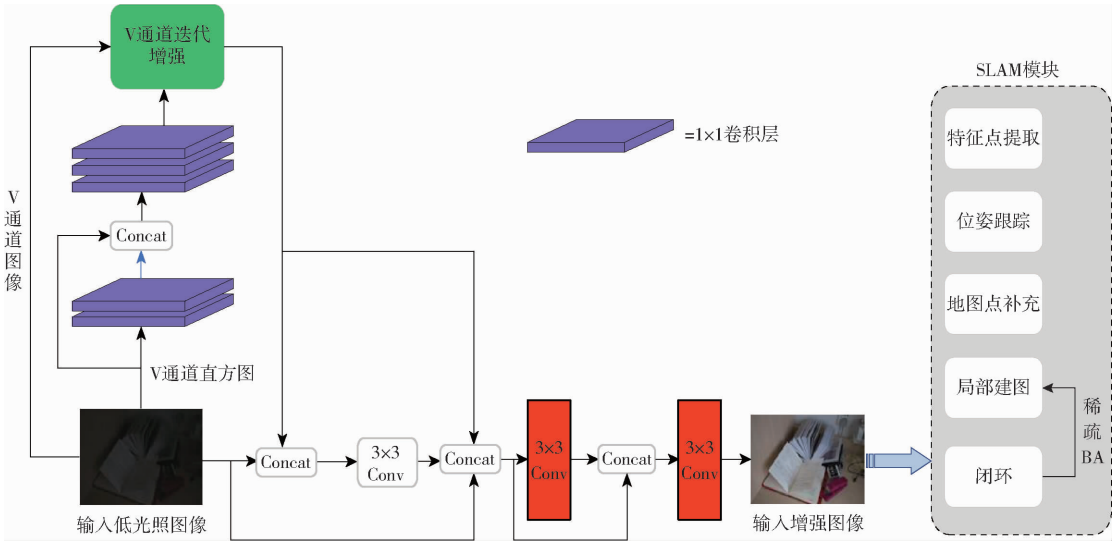


图 1　RLMV－SLAM 系统流程图

Fig. 1　Pipeline of RLMV－SLAM

为了得到全局高阶映射曲线系数并进行后续的增强，设计了图 1 所示的卷积神经网络，前向传播流程为：首先提取输入图像 V 通道直方图作为输入送入后续 5 个 1×1 的卷积层，输出 1 组映射曲线系数

$$\boldsymbol{X}=[x_0 \quad x_1 \quad \cdots \quad x_n]=F(H(\boldsymbol{I}^v)) \tag{2}$$

式中　$\boldsymbol{I}^v$——输入的 V 通道图像

$H(\boldsymbol{I}^v)$——获取 V 通道图像的直方图

F——5 个 1×1 的卷积层

得到映射曲线系数后使用式(1)对原始 V 通道图像进行迭代增强。因为卷积层能够提取局部信息并产生高频细节，将迭代增强后 V 通道图像与原始的三通道弱光照图像沿着通道维度进行拼接，将拼接后张量送入几个具有局部感受野的卷积层，并进一步将其与中间层拼接进行特征融合，卷积层输出得到最终的增强图像。使用对应的参考图像端到端训练整个网络以优化网络整体性能。使用 LOL (Low light paired dataset) 数据集[29] 对网络进行训练，并将其应用到 SLAM 任务中。

由于图像增强任务通常会包含多个参考图像，因此较为常用的 L_1 和 SSIM 等损失在图像增强方面的性能可能不如在其他低级图像处理任务(如图像去噪、图像去模糊和图像修复)方面表现好，即这些常用的损失函数更适合学习输入的训练图像与绝对参考图像之间的一对一的映射关系，而在图像增强任务中并不存在绝对的参考图像。因此，为了更好地完成弱光照图像增强任务，参照文献[30]引入几种损失函数，包括颜色损失、亮度损失和结构损失，所有损失函数都使用余弦相似度，这是因为余弦相似性自然地对像素值向量进行了标准化，即使向量长度改变，只要方向不变其余弦相似性就不会改变，而 L_1 或 L_2 范数对向量长度敏感，当数据维度较高时，余弦相似性比 L_1 或 L_2 范数更稳定，因为在高维空间中，所有点都几乎等距，使得 L_1 或 L_2 范数在高维空间中失去了区分能力，而余弦相似性仅仅依赖于角度，因此不受维度影响。

颜色损失用于评估增强图像与参考图像之间的颜色相似性，计算式为

$$L_{\text{color}}=1-\sum_{i=1,\,j=1}^{m,n} S(E(\boldsymbol{I}_{i,j}),\boldsymbol{G}_{i,j})) \tag{3}$$

式中　$\boldsymbol{G}_{i,j}$——正常光照像素

$\boldsymbol{I}_{i,j}$——输入的弱光照图像像素

$S(\cdot)$——取余弦相似度

$E(\boldsymbol{I}_{i,j})$——弱光照图像像素经过增强网络 E 后的像素

为了使得增强后图像与参考图像具有相同的亮度顺序(如果参考图像 A 比参考图像 B 亮，则增强后的图像 A′应该也比 B′更亮)，将输入图像沿对称轴划分为左上、右上、坐下、右下 4 个部分，分别计算 4 个部分的损失函数，可以对同一图像中的不同区域实现不同程度增强，避免因为图像局部亮度过高或过低导致图像整体亮度的变化，亮度损失计算式为

$$L_{\text{brightness}}=1-\sum_{c\in H,S,V}\sum_{i=1,\,j=1}^{m,n} S(B(E(\boldsymbol{I}_{i,j}^c)),B(\boldsymbol{G}_{i,j}^c)) \tag{4}$$

式中　$B(E(\boldsymbol{I}_{i,j}^c))$——以像素 $E(\boldsymbol{I}_{i,j})$ 为中心的图像块

$B(\boldsymbol{G}_{i,j}^c)$——以像素 $G_{i,j}$ 为中心的图像块

使用图像块像素梯度代替像素值构成结构损失，即

$$L_{\text{struchure}}=1-\sum_{c\in H,S,V}\sum_{i=1,\,j=1}^{m,n} S(B(\nabla E(\boldsymbol{I}_{i,j}^c)),B(\nabla \boldsymbol{G}_{i,j}^c)) \tag{5}$$

所有这些损失函数,包括 L1 和 SSIM 损失,都被结合起来形成训练这个网络所使用的总损失函数。

1.2 弱光照环境中位姿估计

目前主流视觉 SLAM 算法主要是基于在图像帧中提取的特征,通过相邻帧之间匹配好的特征来估计相邻帧之间相机的运动和位姿,过程中最关键的部分就是特征提取和相邻帧之间特征匹配,特征提取和匹配数量和质量取决于采集到的图像质量,在昏暗光照下的图像中只能提取到极少的特征,因此会导致位姿估计的失效,同时同步建图也会失效。针对此问题,本文提出三点改进措施在以下部分详细说明。

1.2.1 地图点补充

为了提高 SLAM 系统在弱光照环境中的跟踪鲁棒性和实时性,在局部建图线程中,首先对关键帧进行筛选剔除,如在局部地图中至少有 95% 的地图点被 3 个以上的关键帧观测到,只保留其中共视关系最高的 3 个关键帧和它们对应的地图点。将筛选出的关键帧分别与其共视程度最高的 10 个共视关键帧两两进行重投影特征匹配,在此处引入反向匹配方案来剔除误匹配的特征,具体来说,当某一关键帧特征点重投影到另一关键帧中,在投影区域附近进行描述子匹配找出匹配的特征点后,反向将特征点投影回原来的关键帧,只有投影后点与原来的特征点误差不超过 3 个像素时,才认为两个特征点相互匹配。匹配后进行三角化生成新的地图点。这些新的地图点将会与原本的局部地图点一起参与后续的如图 2 所示的局部 BA 优化即

$$\{\hat{\boldsymbol{T}}_{iw},\hat{\boldsymbol{P}}_k^w\}=\underset{\hat{\boldsymbol{T}}_{iw},\hat{\boldsymbol{P}}_k^w}{\operatorname{argmin}}\sum_i\sum_k\left\|\boldsymbol{u}_k^i-\frac{1}{s_k}M(\hat{\boldsymbol{T}}_{iw},\hat{\boldsymbol{P}}_k^w)\right\|_{\Sigma}^2 \tag{6}$$

$$\boldsymbol{P}_k^w=s_k\boldsymbol{T}_{wi}\boldsymbol{M}^{-1}(\boldsymbol{u}_k^i) \tag{7}$$

式中 $\boldsymbol{T}_{iw}$——第 i 帧图像在世界坐标系下位姿变换矩阵

$\boldsymbol{u}_k^i$——第 i 帧图像中第 k 个二维特征点

$\boldsymbol{\Sigma}$——二维特征点协方差矩阵

$\boldsymbol{P}_k^w$——世界坐标系下第 k 个三维地图点

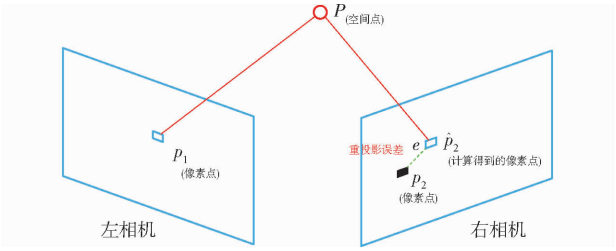


图 2 优化重投影误差示意图(BA)

Fig. 2 Optimizing reprojection error(bundle adjustment)

s_k——三维地图点 $\boldsymbol{P}_k^w$ 在 Z 轴上的深度

在三维向二维的映射中会丢失掉深度,所以在像素坐标中归一化 Z 轴坐标为 1,M 表示相机模型,定义为

$$M\begin{bmatrix}X\\Y\\Z\end{bmatrix}=\begin{bmatrix}f_x\dfrac{X}{Z}+c_x\\f_y\dfrac{Y}{Z}+c_y\end{bmatrix} \tag{8}$$

在闭环线程中,使用与当前关键帧共视程度最高的 10 个共视关键帧和闭环候选关键帧进行两两特征正反向匹配,三角化生成新的三维地图点,这些新的地图点将会参与后续的闭环矫正。采用上述的策略,RLMV-SLAM 可以极大地提高位姿估计过程中匹配到的特征点数量并消除误匹配,从而大幅提高跟踪稳定性和鲁棒性。其中,f_x、f_y、c_x、c_y 为相机内参,在相机标定后均已知。

1.2.2 Sparse BA

SLAM 系统通过同时优化相机位姿和三维地图点来进行 BA(Bundle adjustment),这种方法能够有效地解决大规模定位和建图问题。然而,随着 SLAM 系统长时间运行,关键帧和地图点积累会导致计算效率逐渐下降。如果在闭环矫正后使用全局的 BA 优化,可能会大幅降低系统实时性,甚至导致系统崩溃。因此,RLMV-SLAM 在闭环矫正后选择使用 Sparse BA,意味着只对受到闭环矫正影响的地图部分应用 BA 优化,即所有被修正的关键帧及其对应的 3D 地图点。通过优化一部分关键帧和 3D 地图点,避免了应用全局 BA(如 ORB-SLAM 所做)的负载。这种策略大大减少了涉及最近关键帧的闭环检测的运行时间。Sparse BA 优化可能仍需要几秒才能执行,同时添加的新关键帧和地图点也必须更新,以保持轨迹的一致性。使用 Sparse BA 之前的位姿传播关键帧位姿的修正,并通过式(8)优化三维地图点。

1.2.3 实时词袋模型

使用基于外观的闭环检测方法 iBoW-LCD[31],所提出的方法利用基于二进制描述符的增量词袋(BoW)方案来检索视频序列中与当前帧类似的帧。增量式词袋模型优势避免了经典 BoW 模型通常需要的提前训练词袋库,可以完全在线式检测回环,并且 iBoW-LCD 可以保证词袋模型始终适应于当前环境,可进一步提高闭环检测精确度。

为了检测闭环,iBoW-LCD 利用动态岛屿的概念,这是一种简单但有效的机制,用于将时间上接近的类似图像分组,从而减少了通常与贝叶斯框架相关的计算时间。该方法已经使用多个在不同环境条

件下获取的室内和室外公共数据集进行了验证，取得高精度，并超越了其他最先进的解决方案。SLAM 方法使用 iBoW－LCD，使得本文系统可以动态适应各种环境，尽可能地避免用于训练离线词袋的数据与真实运行环境之间的差异导致的误差。

2　实验验证与分析

为了评估 RLMV－SLAM 的性能，设计的实验分为：①在弱光照图像增强数据集上使用本文提出的图像增强网络与其他主流图像增强方法的实时性对比。②在公开的数据集和实际的弱光照环境中使用 RLMV－SLAM 与其他主流 SLAM 算法进行定位精度和跟踪时长的实验，并对实验结果进行对比。

2.1　硬件平台

使用如图 3 所示基于 ROS 操作系统的移动机器人 TurtleBot2 作为实验平台，在 TurtleBot2 上架设 RGB 相机 Intel RealSense D455。使用一台 Laptop PC 对 TurtleBot2 进行控制，PC 搭载 Intel i5－10200H 处理器，主频为 2.40 GHz 和 NVIDIA GTX 1060 GPU，运行 Ubuntu 20.04 操作系统。其中视觉传感器使用的是 Intel RealSense D455 相机，该相机操作范围为 0.6～6 m，分辨率设为 848 像素×480 像素，频率为 30 Hz。PC 与相机之间采用 USB 3.1 连接。TurtleBot2 平移速度最高为 70 cm/s，最大旋转速度为 180(°)/s。

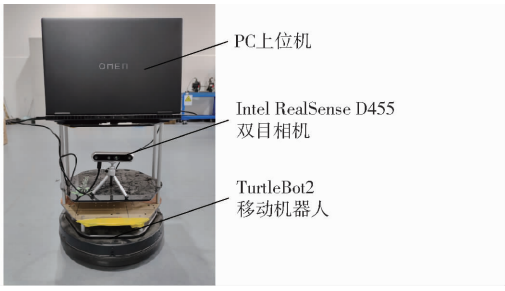


图 3　TurtleBot2 移动机器人

Fig. 3　TurtleBot2 mobile robot

2.2　图像增强实验

在视觉同步定位和建图任务中，相机以固定的帧率实时采集环境中的图像，算法必须能够以相同或近似的帧率对图像进行分析和处理，从相邻图像中估计出位置和姿态变化。理论上图像采集的帧率越高，相邻图像位置和姿态变化越小，无论是对于基于光流还是特征点的视觉 SLAM 算法其准确性和鲁棒性就会越高。因此，对于一个视觉 SLAM 系统来说，实时性不仅体现在算法能否正常运行，也体现在跟踪和定位准确性和鲁棒性上。选取目前主流的图像增强方法与本文方法在 LOL－V1[29] 数据集上对图像增强实时性进行了测试，结果如表 1 所示，本文图像增强模块在实时性方面明显优于其余图像增强方法。只有本文提出的方法能够达到 SLAM 系统对输入图像进行实时处理的要求。

表 1　图像增强方法实时性对比

Tab. 1　Comparison of real-time performance of image enhancement methods

方法	参数量	处理时间/s	帧率/(f·s^{-1})
KIND＋＋	8.28×10^{7}	4.80	0.20
Bread	－	6.10	0.16
EnlightenGAN	8.63×10^{7}	6.50	0.15
IAT	9.00×10^{5}	0.16	6.25
本文方法	2.00×10^{5}	0.04	25.00

对于非学习的图像增强方法如文献[22]中所使用的直方图均衡化和伽马矫正，虽具有较高的实时性，但这类方法只是随机改变图像直方图分布，无法去噪，且可能导致色彩失真。从 TUM 数据集中的 freiburg2_pioneer_slam2 视频序列中截取一帧图像如图 4 所示，并分别使用伽马矫正、直方图均衡化以及本文方法进行图像增强后提取 ORB 特征点，结果显示，使用本文图像增强方法提取到的特征点明显更多。

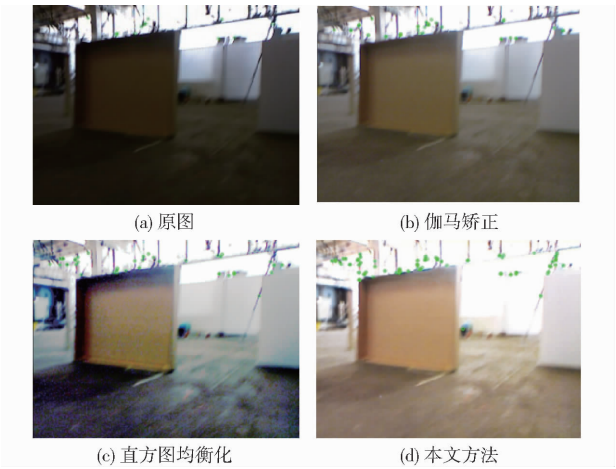


图 4　不同图像增强方法在 TUM 数据集上提取 ORB 特征对比

Fig. 4　Comparison of ORB feature extraction on TUM dataset after using several different image enhancement methods

为了评估增强前后提取的特征点质量和特征匹配的效果，分别使用两帧图像在增强前和增强后进行特征匹配。计算匹配点数量与特征点总数比例、匹配点两个描述子之间的平均欧氏距离和随机采样一致性的内点数量 3 个指标来评估特征点质量。结果如表 2 所示，使用本文方法进行图像增强后 3 个指标均有提高。

为了评估本文所提的图像增强方法对不同导航图像进行颜色、亮度和梯度增强的泛化能力，选

表 2 增强前后特征匹配结果

Tab. 2 Enhanced feature matching results before and after enhancement

评价指标	增强前	增强后
匹配率/%	39	40
平均匹配距离	38.85	35.44
RANSAC 内点数量	75	146

择与训练集没有相同图像的 ExDark[32] 数据集进行验证,ExDark 数据集是用于弱光照图像目标检测训练的数据集,没有包含光照增强参考图像,因此仅进行定性分析,增强结果如图 5 所示。图 5a 为 3 幅低亮度和对比度原始图像,对应增强结果如图 5b 所示。



(a) 原始图像



(b) 增强后图像

图 5 泛化能力测试

Fig. 5 Generalization ability test

2.3 消融实验

实时词袋模型优势在于能够实时生成词袋,无需预先训练。其对系统准确性提升只能在大样本实验中体现且具有偶然性,因此并未进行消融实验。在 TUM[33]、EuRoC MAV[34] 和 ETH3D[35] 3 个数据集的 fr3_long_office_household、EuRoC_MH01、ETH3D_sfm_lab_room_1 共 3 个视频序列上,分别应用地图点补充策略和 Sparse BA 进行位姿估计消融实验。实验测试了这 3 个视频序列上位姿估计的绝对轨迹误差的均方根误差平均值以及内存占用量,测试结果如表 3 所示。虽然 Sparse BA 和地图点补充策略并未能减小绝对轨迹误差,但这 2 项策略都显著降低了 SLAM 系统运行中的内存消耗速度和最大占用内存量。其中,Sparse BA 对内存消耗的降低效果主要取决于闭环矫正次数。另一方面,由于在地图点补充策略中采用更严格的关键帧筛选机制,该策略对内存消耗的降低效果更为显著。对于需要长时间运行的 SLAM 系统来说,降低内存消耗非常有益。

为了评估 RLMV - SLAM 相比其他对光照鲁棒的 SLAM 算法的先进性,将文献[22 - 24]中所用的图像增强方法应用于 ORB - SLAM3,包含有弱光照环境的 ETH3D 数据集中测试绝对轨迹误差均方根误差,结果如表 4 所示,Bread、EnGan 两种图像增强

表 3 Sparse BA 和地图点补充消融实验结果

Tab. 3 Results of ablation experiments for Sparse BA and map point supplementation

方法	绝对轨迹误差的均方根误差/m	内存消耗速度/(Mb·s^{-1})	最大内存占用量/GB
RLMV - SLAM(Sparse BA)	0.086 3	16.5	2.18
RLMV - SLAM(非 Sparse BA)	0.081 5	23.6	2.90
RLMV - SLAM(地图点补充)	0.082 2	15.1	1.90
RLMV - SLAM(无地图点补充)	0.078 1	18.4	2.32

表 4 基于不同图像增强方法的绝对轨迹误差

Tab. 4 Absolute trajectory error based on different image enhancement methods

视频序列	ORB - SLAM3			RLMV - SLAM/m
	原始/m	Bread	EnGan 伽马矫正 & 直方图均衡化/m	
einstein_global_light_changes_1	0.21		0.17	0.07
sfm_lab_room_1	1.96		1.54	0.90
sfm house	0.37		0.37	0.14
sfm lab	1.26		1.11	0.65

方法在本文硬件条件下并无法实时运行,证明了本文所提方法在实时性方面的先进性。与原始的和应用了伽马矫正与直方图均衡化的 ORB - SLAM3 相比,RLMV - SLAM 拥有最低的绝对轨迹误差。

2.4 位姿估计精度实验

为了检验 RLMV - SLAM 的有效性,选取 TUM[33]、EuRoC MAV[34] 和 ETH3D[35] 3 个数据集上具有代表性的视频序列对 RLMV - SLAM 和 ORB - SLAM3 的单目模式进行位姿估计进度测试。其中 ETH3D 数据集的 sfm_lab_room_1 和 einstein_global_light_changes_1 序列包含照明不足的环境和急剧的光照变化,用于测试 SLAM 在低光照和复杂光照变化环境中的性能。测试结果如表 5、6 所示,在所有视频序列上,RLMV - SLAM 绝对轨迹误差和相对轨迹误差都优于 ORB - SLAM3 单目模式,其中,在 ETH3D 数据集上 RLMV - SLAM 表现出显著提升。

图 6 为 SLAM 算法在 TUM、EuRoC 和 ETH3D 3 个数据集上的 4 个视频序列上的估计轨迹和真实轨迹。从图 6 可以看出,在表示常规场景的 TUM 和 EuRoC 数据集上,RLMV - SLAM 位姿轨迹效果略优于 ORB - SLAM3,但是,在低光照环境的 ETH3D 数据集上,ORB - SLAM3 无法有效估计出运动轨迹,而 RLMV - SLAM 能够有效估计完整运动轨迹。

图 7 为 SLAM 算法在 TUM 数据集的 fr2_slam3 视频序列上估计位姿分别在 x、y、z 3 个方向和在偏

表 5　绝对轨迹误差对比

Tab. 5　Comparison of absolute trajectory error　m

视频序列	绝对轨迹误差	RLMV－SLAM	ORB－SLAM3
fr1_room	最大值	0.193 0	0.237 4
	平均值	0.070 2	0.122 9
	标准差	0.079 7	0.131 2
fr2_desk	最大值	0.027 1	0.031 0
	平均值	0.009 2	0.011 0
	标准差	0.010 2	0.012 0
fr2_slam3	最大值	0.305 2	4.692 7
	平均值	0.085 9	0.342 4
	标准差	0.106 7	0.832 7
fr3_long_office_household	最大值	0.036 8	0.108 8
	平均值	0.008 3	0.009 0
	标准差	0.008 9	0.010 0
EuRoC_MH01	最大值	0.120 6	0.106 1
	平均值	0.029 0	0.030 3
	标准差	0.035 5	0.037 9
EuRoC_V101	最大值	0.157 2	0.175 3
	平均值	0.081 6	0.085 9
	标准差	0.087 3	0.090 5
ETH3D_sfm_lab_room_1	最大值	0.128 7	0.342 6
	平均值	0.068 2	0.213 9
	标准差	0.078 5	0.223 6
ETH3D_einstein_global_light_changes_1	最大值	1.752 8	4.833 5
	平均值	0.898 0	1.962 2
	标准差	0.951 0	2.016 0

表 6　相对轨迹误差对比

Tab. 6　Comparison of relative trajectory errors　m

视频序列	相对轨迹误差	RLMV－SLAM	ORB－SLAM3
fr1_room	最大值	0.075 0	0.456 2
	平均值	0.007 2	0.008 0
	标准差	0.009 8	0.020 0
fr2_desk	最大值	0.029 2	0.029 1
	平均值	0.002 7	0.002 9
	标准差	0.003 1	0.003 4
fr2_slam3	最大值	0.190 8	4.692 7
	平均值	0.007 0	0.342 4
	标准差	0.146 0	0.832 7
fr3_long_office_household	最大值	0.047 8	0.099 0
	平均值	0.004 5	0.004 0
	标准差	0.005 4	0.006 0
EuRoC_MH01	最大值	0.100 7	0.107 9
	平均值	0.023 0	0.023 1
	标准差	0.028 4	0.028 6
EuRoC_V101	最大值	0.097 4	0.121 7
	平均值	0.022 6	0.022 8
	标准差	0.026 2	0.026 5
ETH3D_sfm_lab_room_1	最大值	0.860 4	2.196 7
	平均值	0.283 6	1.051 2
	标准差	0.358 6	1.271 9
ETH3D_einstein_global_light_changes_1	最大值	1.594 1	3.564 0
	平均值	0.023 4	1.029 3
	标准差	0.047 3	2.128 5

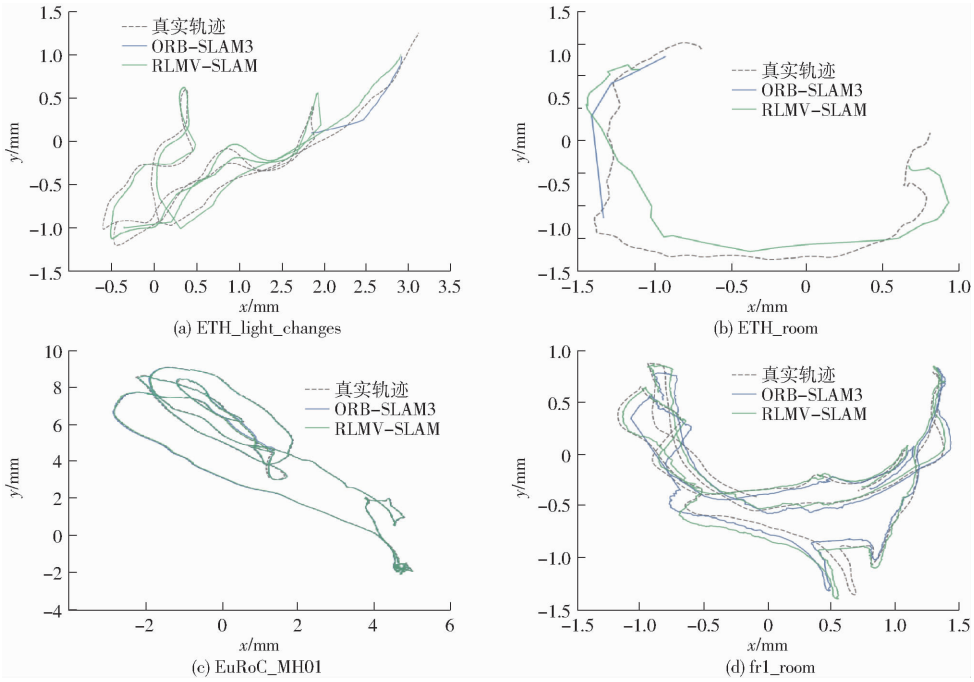


图 6　公开数据集上预测轨迹

Fig. 6　Predictive trajectory on public datasets

航角、俯仰角、滚转角 3 个角度误差曲线。从图 7 可以看出，两种 SLAM 算法在 x、y、z 3 个方向误差曲线基本拟合了真实值，但是 ORB－SLAM3 在一些节点上会出现显著的漂移，如在第 80 秒之后绿色曲线在 3 个方向和 3 个角度都相对真实曲线发生了显著的偏离。

图 8 为在 EuRoC_V101 视频序列上的相对位姿

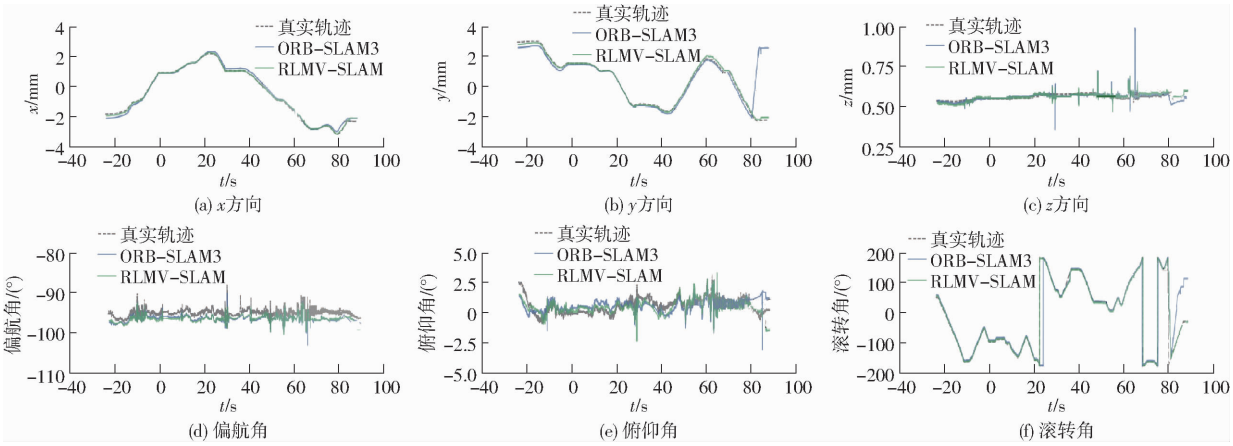


图 7 x,y,z 方向和偏航、俯仰、滚转角误差曲线

Fig. 7 Error curves in x,y,z directions and yaw, pitch, roll angles

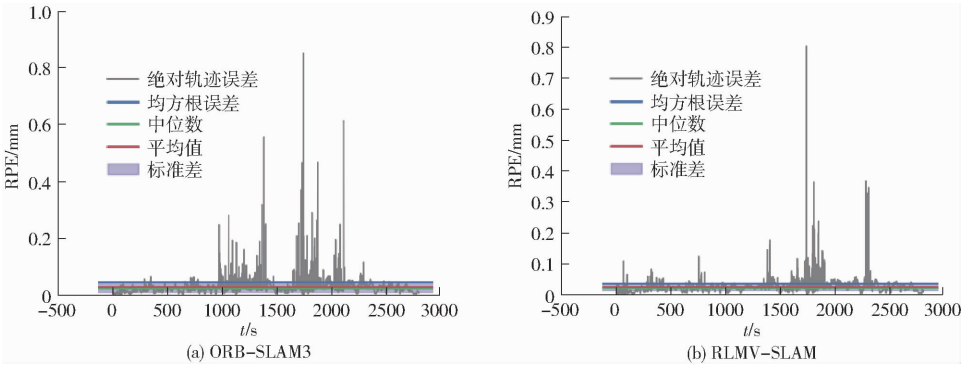


图 8 在 EuRoC_V101 视频序列上的相对误差曲线

Fig. 8 Relative error curves on EuRoC_V101 video sequence

误差曲线。可以看到虽然 RLMV－SLAM 的相对位姿误差的最大值与 ORB－SLAM3 相当,但是误差曲线显然更加平滑,误差波动的幅度更小。图 9 为在 fr3_long_office_household 的绝对轨迹误差曲线。总体来说,2 条曲线都在 0.035 m 的范围内波动,但是 ORB－SLAM3 在 80 s 后出现明显畸变。

2.5 真实环境实验

为了检验算法鲁棒性,选择对于常规 SLAM 方法来说难以有效工作的没有任何主动光源的室内环境,包括寻常起居室、楼梯间和夜间户外场景作为实验区域测试本文的系统的性能。通过同一本地网络中的便携式计算机控制 TurtleBot2 机器人以正常速度移动,并使用该算法实时输出位姿。图 10 为在昏暗起居室中进行实时位姿估计时采集的实验信息。如图 10a 所示,使用原始的 ORB－SLAM3 系统进行位姿估计,特征提取无法在弱光照的视频序列中提取到特征点,因此无法进行位姿估计。然而,如图 10b 所示,使用本文提出的 RLMV－SLAM,有效地增强了原始输入视频序列,使得特征提取器可以有效提取到特征点,成功实现了位姿估计。

表 7 为 3 种 SLAM 系统有效运行的时间占视频

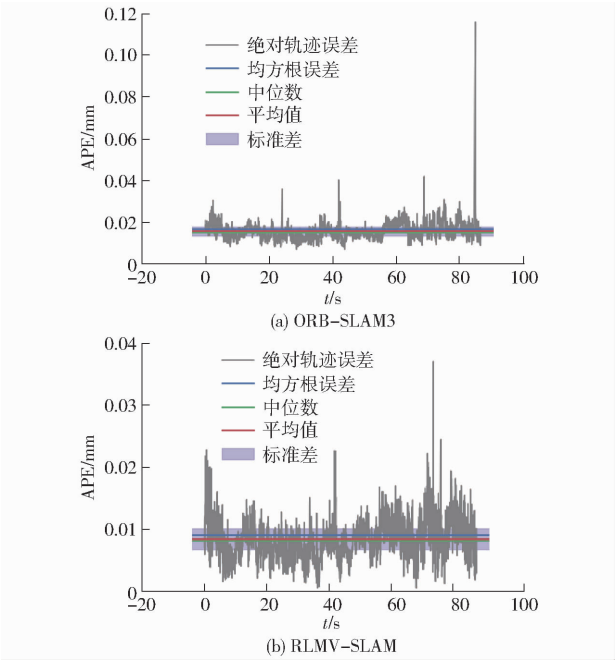
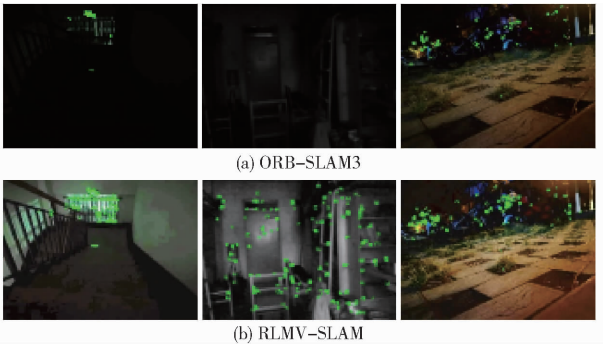


图 9 在 fr3_long_office_household 序列上绝对误差曲线

Fig. 9 Absolute error curves on fr3_long_office_household sequence

序列总时间的百分比。从表 7 可以看出,相比于其他方法,RLMV－SLAM 有效跟踪时间提高 30%,跟踪丢失可能性大幅降低。



(a) ORB-SLAM3

(b) RLMV-SLAM

图 10　RLMV－SLAM 与 ORB－SLAM3 在弱光照环境中的运行情况

Fig. 10　Running conditions of RLMV－SLAM and ORB－SLAM3 in low-light environments

3　结论

(1)设计了一个能够实现亮度和对比度增强，同时消除噪声和矫正色彩误差的轻量化卷积神经网络，将其嵌入到 SLAM 线程的特征提取器之前使其能够实时运行。

表 7　弱光照环境中 3 种 SLAM 系统跟踪跨度

Tab. 7　Tracking spans of three SLAM systems in weak light environments　%

实验场景	RLMV－SLAM	ORB－SLAM3	OV2－SLAM
Dark_dormitory	96.7	52.6	63.1
Dark_office	92.8	55.8	61.3
Dark_outdoor	64.7	29.8	31.5
ETH3D_sfm_lab_room	87.4	11.0	28.1
ETH3D_global_lightchanges	88.5	8.9	15.5

(2)使用一系列优化策略包括地图点补充，配合 Sparse BA 和实时闭环检测方法，在不损失 SLAM 系统稳定性和精度的条件下提升了实时性。

(3)在公开数据集和真实环境中的实验结果表明，本文提出的方法能够在弱光照环境中稳定运行，相比 ORB－SLAM3 和 OV2－SLAM，有效跟踪时长提升 30%。并且，在公开数据集上估计位姿的轨迹误差也有明显降低。

参　考　文　献

[1] LIU J, LI X, LIU Y, CHEN H. RGB－D inertial odometry for a resource-restricted robot in dynamic environments[J]. IEEE Robot. Autom. Lett., 2022,7(4): 9573－9580.

[2] YE W, LAN X, CHEN S, et al. PVO: panoptic visual odometry[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9579－9589.

[3] CAMPOS C, ELVIRA R, RODRIGUEZ J J G, et al. ORB－SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Trans. Robot, 2021,37(6): 1874－1890.

[4] FERRERA M, EUDES A, MORAS J, et al. OV2－SLAM: a fully online and versatile visual SLAM for real-time applications[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1399－1406.

[5] 刘庆运，杨华阳，刘涛，等. 基于激光雷达与深度相机融合的 SLAM 算法[J]. 农业机械学报，2023，54(11)：29－38.
LIU Qingyun, YANG Huayang, LIU Tao, et al. SLAM algorithm based on fusion of LiDAR and depth camera[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(11): 29－38. (in Chinese)

[6] LI Y, YUNUS R, BRASCH N, et al. RGB－D SLAM with structural regularities[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 11581－11587.

[7] MIN Z, DUNN E. Voldor＋SLAM: for the times when feature-based or direct methods are not good enough[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 13813－13819.

[8] ZHANG X, WANG W, QI X, et al. Point-plane SLAM using supposed planes for indoor environments[J]. Sensors, 2019, 19(17):3795.

[9] GUO C X, ROUMELIOTIS S I. IMU－RGBD camera navigation using point and plane features[C]//2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013: 3164－3171.

[10] WANG R, GERAGHTY D, MATZEN K, et al. VPLNet: deep single view normal estimation with vanishing points and lines[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020:686－695.

[11] ZHANG Y, DI X, ZHANG B, et al. Better than reference in low-light image enhancement: conditional re-enhancement network[J]. IEEE Transactions on Image Processing, 2022,31:759－772.

[12] CHEN C, GENEVA P, PENG Y, et al. Monocular visual-inertial odometry with planar regularities[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023: 6224－6231.

[13] YANG S, SCHERER S. CubeSLAM: monocular 3D object SLAM[J]. IEEE Trans. Robot., 2019,35(4): 925－938.

[14] ROSS P, ENGLISH A, BALL D, et al. A novel method for analysing lighting variance[C]//Australian Conference on Robotics and Automation,2013:8－16.

[15] 薛金林，褚阳阳，宋悦，等. 基于自适应阈值 ORB 特征提取的果园双目稠密地图构建[J]. 农业机械学报，2024，55(6)：42－51，59.
XUE Jinlin, CHU Yangyang, SONG Yue, et al. Construction of dense binocular maps in orchard based on adaptive threshold ORB feature extraction[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(6): 42－51,59. (in Chinese)

[16] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018: 224－236.

(下转第 403 页)

under plastic film[J]. Water Saving Irrigation, 2011(3): 29 – 32,38. (in Chinese)
[33] 王小兵.膜下高频滴灌棉花耗水量与灌溉制度研究[D].石河子：石河子大学,2009.
WANG Xiaobing. Experimental research on water consumption and irrigation schedule of the cotton under high-frequency drip irrigation mulched with plastic films[D]. Shihezi: Shihezi University, 2009. (in Chinese)
[34] ZHANG J, WANG Z. Study on consumption characteristics of cotton under drip irrigation with film in North Xinjiang[J]. Procedia Engineering, 2012, 28: 413 – 418.
[35] GUZMÁN-CRUZ R, CASTAÑEDA-MIRANDA R, GARCÍA-ESCALANTE J J, et al. Calibration of a greenhouse climate model using evolutionary algorithms[J]. Biosystems Engineering, 2009, 104(1): 135 – 142.
[36] HERRERO J M, BLASCO X, MARTÍNEZ M, et al. Robust identification of non-linear greenhouse model using evolutionary algorithms[J]. Control Engineering Practice, 2008, 16(5): 515 – 530.
[37] 王军,关红杰,李久生.棉花膜下滴灌二维土壤水与作物生长耦合模拟模型率定与验证[J].灌溉排水学报,2014,33(增刊 1):343 – 347.
WANG Jun, GUAN Hongjie, LI Jiusheng. Calibration and verification of two-dimensional coupling simulation model of soil water and crop growth under cotton film drip irrigation[J]. Journal of Irrigation and Drainage, 2014,33(Supp. 1): 343 – 347. (in Chinese)
[38] 苏李君,郭媛,陶汪海,等.基于群智能优化算法的土壤水动力参数反演[J].农业机械学报,2023,54(5):324 – 334.
SU Lijun, GUO Yuan, TAO Wanghai, et al. Inversion of soil hydrodynamic parameters with richards equation based on intelligent optimization algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(5):324 – 334. (in Chinese)

(上接第 391 页)

[17] SARLIN P E, DETONE D, MALISIEWICZ T, et al. Superglue: learning feature matching with graph neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4938 – 4947.
[18] HAGHIGHI Y, KUMAR S, THIRAN J P, et al. Neural implicit dense semantic slam[J]. arXiv preprint arXiv:2304.14560, 2023.
[19] LI H, GU X, YUAN W, et al. Dense RGB SLAM with neural implicit maps[J]. arXiv preprint arXiv:2301.08930, 2023.
[20] KALINOV I, PETROVSKY A, ILIN V, et al. Warevision: CNN barcode detection-based uav trajectory optimization for autonomous warehouse stocktaking[J]. IEEE Robotics and Automation Letters, 2020, 5(4): 6647 – 6653.
[21] PETROVSKY A, KALINOV I, KARPYSHEV P, et al. Customer behavior analytics using an autonomous robotics-based system [C]//2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2020: 327 – 332.
[22] HAO L, LI H, ZHANG Q, et al. LMVI – SLAM: robust low-light monocular visual-inertial simultaneous localization and mapping[C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019: 272 – 277.
[23] SAVINYKH A, KURENKOV M, KRUZHKOV E, et al. Darkslam: GAN-assisted visual slam for reliable operation in low-light conditions[C]//2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring). IEEE, 2022: 1 – 6.
[24] SINGH S P, MAZOTTI B, MAYILVAHANAN S, et al. Twilight SLAM: a comparative study of low-light visual SLAM pipelines[J]. arXiv preprint arXiv:2304.11310, 2023.
[25] HUANG J, LIU S. Robust simultaneous localization and mapping in low-light environment[J]. Computer Animation and Virtual Worlds, 2019, 30(4): 1895.
[26] LIU X, GAO Z, CHENG H, et al. Learning-based low light image enhancement for visual odometry[C]//2020 IEEE 16th International Conference on Control & Automation (ICCA). IEEE, 2020: 1143 – 1148.
[27] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the kitti dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231 – 1237.
[28] GUO C, LI C, GUO J, et al. Zero-reference deep curve estimation for low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1780 – 1789.
[29] WEI C, WANG W, YANG W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.
[30] ZHANG Y, DI X, WU J, et al. A fast and lightweight network for low-light image enhancement[J]. arXiv preprint arXiv: 2304.02978, 2023.
[31] GARCIA-FIDALGO E, ORTIZ A. iBoW-LCD: an appearance-based loop-closure detection approach using incremental bags of binary words[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3051 – 3057.
[32] LOH Y P, CHAN C S. Getting to know low-light images with the exclusively dark dataset[J]. Computer Vision and Image Understanding, 2019, 178: 30 – 42.
[33] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB – D SLAM systems[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012: 573 – 580.
[34] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets[J]. The International Journal of Robotics Research, 2016, 35(10): 1157 – 1163.
[35] SCHOPS T, SCHONBERGER J L, GALLIANI S, et al. A multi-view stereo benchmark with high-resolution images and multi-camera videos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3260 – 3269.