计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(7):299−304 [doi: 10.15888/j.cnki.csa.009172]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

特征点维度的静态场景三维重建运动目标剔除①

杨永刚[1], 宋　泽[1], 李思萌[2], 申郑茂[1]

[1](中国民航大学, 天津 300300)
[2](中国城市发展规划设计咨询有限公司, 北京 100120)
通信作者: 李思萌, E-mail: 1330344629@qq.com

摘　要: 基于语义分割的图像掩膜方法常用来解决静态场景三维重建任务中运动物体的干扰问题, 然而利用掩膜成功剔除运动物体的同时会产生少量无效特征点. 针对此问题, 提出一种在特征点维度的运动目标剔除方法, 利用卷积神经网络获取运动目标信息, 并构建特征点过滤模块, 使用运动目标信息过滤更新特征点列表, 实现运动目标的完全剔除. 通过采用地面图像和航拍图像两种数据集以及 DeepLabV3、YOLOv4 两种图像处理算法对所提方法进行验证, 结果表明特征点维度的三维重建运动目标剔除方法可以完全剔除运动目标, 不产生额外的无效特征点, 且相较于图像掩膜方法平均缩短 13.36% 的点云生成时间, 减小 9.93% 的重投影误差.

关键词: 三维重建; 运动目标剔除; 特征点; 目标检测; 语义分割

引用格式: 杨永刚,宋泽,李思萌,申郑茂.特征点维度的静态场景三维重建运动目标剔除.计算机系统应用,2023,32(7):299–304. http://www.c-s-a.org.cn/1003-3254/9172.html

Moving Object Elimination for 3D Reconstruction of Static Scenes in Feature Point Dimension

YANG Yong-Gang[1], SONG Ze[1], LI Si-Meng[2], SHEN Zheng-Mao[1]

[1](Civil Aviation University of China, Tianjin 300300, China)
[2](China Urban Development Planning & Design Consulting Co. Ltd., Beijing 100120, China)

Abstract: The image masking method based on semantic segmentation is often used to solve the interference problem of moving objects in three-dimensional (3D) reconstruction tasks of static scenes. However, a small number of invalid feature points will be produced when the mask is used to eliminate moving objects. To solve this problem, a method for eliminating moving objects in the dimension of feature points is proposed. The convolutional neural network is used to obtain the moving target information, and the feature point filtering module is constructed. Then, the moving target information is used to filter and update the feature point list for the complete elimination of the moving target. The ground image dataset and aerial image dataset and the processing algorithms of DeepLabV3 and YOLOv4 are used to verify the proposed method. The results show that the moving object elimination method in 3D reconstruction in the feature point dimension can completely eliminate the moving object without generating additional invalid feature points. Compared with the image masking method, the proposed method shortens the point cloud generation time by 13.36% and reduces the reprojection error by 9.93% on average.

Key words: 3D reconstruction; moving object elimination; feature point; object detection; semantic segmentation

SFM (structrue from motion) 三维重建[1] 是一项由多幅不同视角图像恢复相机位置及目标物体三维结构的计算机视觉任务. 航拍三维重建[2] 在智慧城市、城市规划、国土资源管理等需要大比例尺地形图的项目

① 基金项目: 国家自然科学基金 (62173332)
收稿时间: 2023-01-01; 修改时间: 2023-02-13; 采用时间: 2023-02-23; csa 在线出版时间: 2023-05-12
CNKI 网络首发时间: 2023-05-16

中具有广阔的应用场景[3]. 然而, 航拍图像视野宽阔、信息丰富, 较容易获取静态环境中的运动物体, 例如行人、车辆. 运动物体对 SFM 算法中相机位姿计算的精度具有干扰, 进而影响到三维重建模型的精度. 因此, 三维重建任务中运动物体的剔除对提升重建精度有着重要的作用.

早期, 相关学者利用帧差法[4]、光流法[5] 获取运动物体信息, 优化三维重建效果. 然而, 帧差法只适用于位置固定的相机, 光流法只适用于序列图像, 存在一定的局限性. 基于卷积神经网络的图像处理算法对特定类别物体具有高精度的识别与定位能力, 解决了帧差法与光流法在一些场景中的局限性问题. GitHub 开源的 Colmap 三维重建算法[6] 具备语义分割图像接口, 将预处理的掩膜混合图像作为三维重建初始图像, 成功剔除指定类别物体. 何继峰[7] 使用 Mask-RCNN[8] 语义分割算法获取多视角图像的掩膜, 剔除室内三维重建任务的运动物体. 汪雷[9] 将光流法与语义分割相结合, 解决了煤矿巷道场景内运动物体的干扰问题, 提高了探测机器人定位与建图系统的精度与鲁棒性. 文献 [6,7,9] 使用的图像掩膜方法, 在图像维度修改运动物体的像素值, 以达到剔除运动物体的目的. 然而, 图像掩膜方法没有考虑混合掩膜图像中像素值的改变对特征点提取产生的影响.

区别于掩膜方法在图像维度的处理, 本文从特征点维度出发, 构建过滤模块, 利用完整特征点列表和卷积神经网络所提供的运动物体信息, 更新特征点列表, 剔除三维重建中的运动物体. 除此之外, 常用的掩膜方法采用语义分割算法获取掩膜, 本文在语义分割方法的基础上, 尝试将目标检测算法运用到三维重建运动物体剔除中, 具有一定的参考意义. 实验结果表明, 特征点维度的运动物体剔除方法可以完全剔除三维重建中的运动目标, 且不会额外生成无效特征点, 同时, 缩短了点云生成时间, 减小了点云重投影误差.

1 图像掩膜方法

SFM 三维重建的典型步骤为: 输入图像、提取特征点、特征点匹配、求解基础矩阵 F 及本质矩阵 E、求解相机旋转矩阵 R 及平移矩阵 T、空间三角化求解、增量式三维重建、生成稀疏点云.

SFM 三维重建中, N 幅多视角图像中相对应的二维点序列被转换为真实世界的三维点. 图像掩膜方法从原始图像中切除运动物体, 致使特征点检测器提取不到包含运动物体信息的二维特征点, 从而无法生成代表运动物体的三维点, 实现剔除运动物体的目标.

如图 1, 图 2, 图像掩膜方法结合语义分割与 SFM, 剔除运动物体. 首先, RGB 初始图像被语义分割算法处理, 生成相同尺寸的 mask 掩膜, 掩膜的每一个像素值都代表一个类别. 其次, 初始图像被转化成通道为 1 的 gray 灰度图. 然后, gray 灰度图与 mask 结合, 灰度图中 mask 指定类别所对应位置的像素值被调整为 0, 形成剔除运动物体之后的灰度图. 最后, 提取此残缺灰度图的 SIFT 特征点[10], 输出特征点列表 *kpts* 及 *decs*. 此特征点列表不包含运动物体信息, 图像掩膜方法以此实现三维重建中运动物体的剔除.

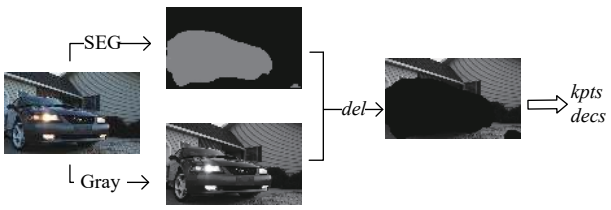


图 1 图像掩膜方法

2 特征点维度的运动目标剔除方法

为解决图像掩膜方法产生少量无效特征点的问题, 本文构建过滤模块, 在特征点维度完成三维重建任务中运动目标的剔除.

2.1 方法原理

特征点维度的运动目标剔除方法如图 3. 其中, Convs 表示卷积神经网络; *DelInfo* 表示运动物体的类别及位置信息; *KPTS*、*DECS* 表示初始完整图像的特征点列表及描述符列表; F 表示特征点过滤模块; *kpts*、*decs* 分别表示过滤更新后的特征点列表及描述符列表.

假设共 N 幅 RGB 图像作为三维重建的初始输入图像. 首先, 每幅 RGB 图像被送入经训练的卷积神经网络中, 输出图像中预设类别的运动目标信息 *DelInfo*.

$$DelInfo_i = \mathrm{Convs}(\mathrm{RGB}_i) \tag{1}$$

式 (1) 及后续公式中, $i = 1,2,\cdots,N$, N 为输入图像总数. 使用 DeepLabV3[11] 语义分割及 YOLOv4[12] 目标检测两种卷积神经网络算法, 算法不同时, 输出的

DelInfo 也不同.

$$DelInfo=\begin{cases} PR, & \text{Convs=DeepLab} \\ [a,b,c,d,S]^r, & \text{Convs=YOLO} \end{cases} \tag{2}$$

使用 DeepLabV3 算法时, *DelInfo* 为张量 *PR*. *PR* 的尺寸与初始图像尺寸一致, 每一个元素均同时包含位置与类别信息. 使用 YOLOv4 算法时, *DelInfo* 为尺寸 $(R, 5)$ 的列表, R 为目标检测矩形框数量, $r=1,2,\cdots,R$. 式 (2) 中 a、b、c、d 为矩形框坐标值, S 为目标类别, 每一个五维向量包含一个目标的位置及类别信息.

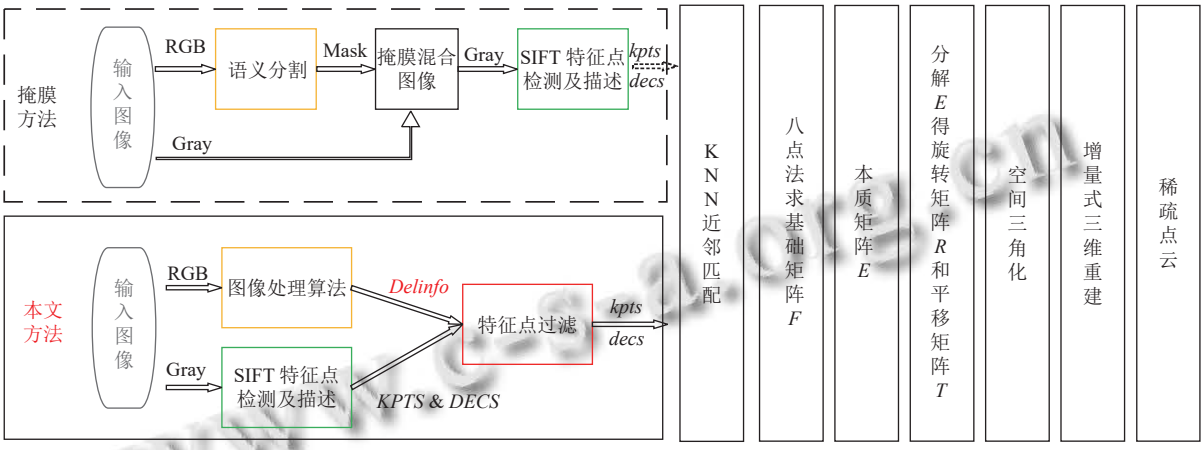


图 2 SFM 三维重建

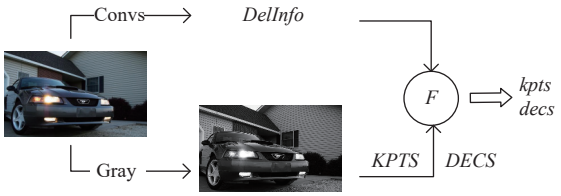


图 3 特征点维度的运动目标剔除方法

然后, 获取每幅灰度图的 SIFT 特征点信息 *KPTS* 及 *DECS*. 由于对完整图像提取特征点, *KPTS* 及 *DECS* 为包含运动物体的完整特征点信息.

$$KPTS_i, DECS_i = \text{SIFT}(gray_i) \tag{3}$$

此时获取的特征点信息分别为:

$$KPTS_i = [k_i^1, k_i^2, \cdots, k_i^j, \cdots, k_i^J] \tag{4}$$

$$DECS_i = [d_i^1, d_i^2, \cdots, d_i^j, \cdots, d_i^J] \tag{5}$$

其中, k_i^j和d_i^j分别表示第 i 幅图像的第 j 个特征点及其描述符. J 为第 i 幅图像的特征点个数.

最后, *DelInfo*、*KPTS* 和 *DECS* 作为特征点过滤模块的输入, 过滤器 F 根据运动物体的位置及类别信息对特征点列表进行更新, 输出不包含运动物体特征点的 *kpts*、*decs*.

$$kpts_i, decs_i = F(PR_i, KPTS_i, DECS_i) \tag{6}$$

$$kpts_i, decs_i = F([a,b,c,d,S]_i^r, KPTS_i, DECS_i) \tag{7}$$

此时获取的特征点信息分别为:

$$kpts_i = [k_i^1, k_i^2, \cdots, k_i^k, \cdots, k_i^K] \tag{8}$$

$$decs_i = [d_i^1, d_i^2, \cdots, d_i^k, \cdots, d_i^K] \tag{9}$$

其中, k_i^k和d_i^k分别表示第 i 幅图像的第 k 个特征点及其描述符. K 为本文方法剔除运动目标后的第 i 幅图像的特征点个数. $K<J$, 且$J-K$即为本文方法剔除掉的特征点个数.

过滤更新后的特征点信息 *kpts*、*decs* 作为 KNN 匹配及后续步骤的输入, 完成三维重建运动目标剔除的任务. 代表运动物体的无效特征点被剔除, KNN 匹配时间减少, RANSAC[13] 内点率提高, 迭代次数减少, 最终点云生成时间相应减少. 运动物体被剔除, 相机姿态估计的干扰减小, 精度更高, 点云的重投影误差随之减小.

2.2 Convs 模块

如式 (1)、式 (2), Convs 模块提供两种方案. 如图 4, 使用语义分割算法时, 运动物体的可视化方案为掩膜图像, 像素点的坐标值及像素值包含位置与类别信息; 使用目标检测算法时, 运动物体的可视化方案为矩形框标记的图像, 矩形框的位置及类别编号包含运动物体的位置和类别信息.

2.3 特征点过滤模块

如式 (6)、式 (7), 构建特征点过滤模块如图 5, 利用运动物体信息 *DelInfo* 过滤更新特征点列表.

(a) 语义分割　　(b) 目标检测

图 4　Convs 输出可视化

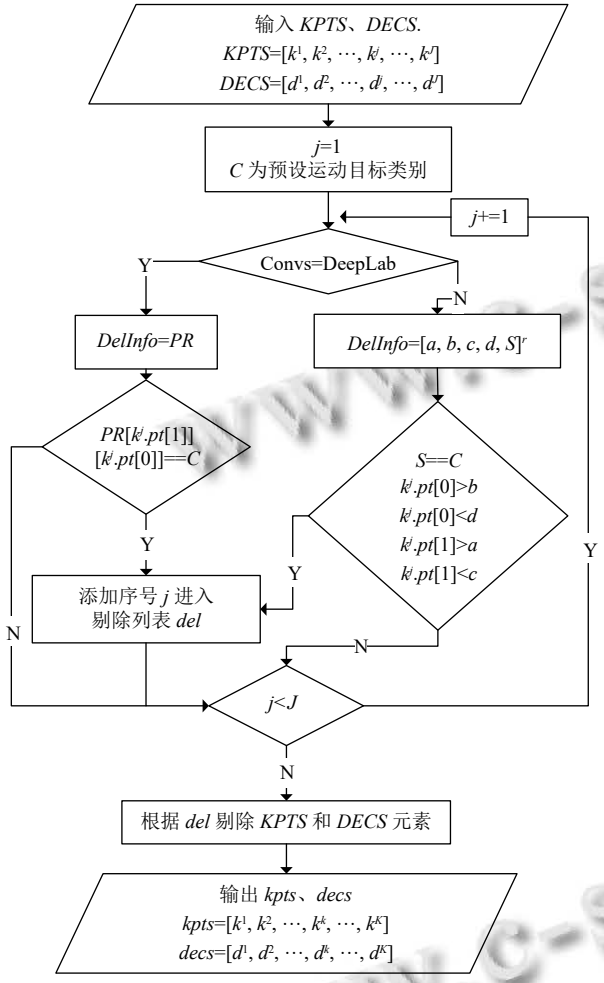


图 5　特征点过滤模块

3　实验分析

3.1　实验准备

实验在相同环境下进行, 具体如表 1.

表 1　实验环境

名称	类型
系统	Windows 10
深度学习框架	TensorFlow
编程语言	Python
CPU	AMD EPYC 7542
GPU	NVIDIA RTX3090
RAM	256 GB

采用地面拍摄图像及低空航拍图像两组数据集进行实验, 如图 6. 地面拍摄数据以房屋为三维重建主体, 设置人、自行车为运动物体. 航拍数据集以天津世纪钟转盘为重建主体, 设置行驶的汽车为运动物体. 分别采用 DeepLabV3 语义分割和 YOLOv4 目标检测作为计算 *DelInfo* 的 Convs. 特征点过滤模块设置两个通道, 可根据 *DelInfo* 的不同做出调整.

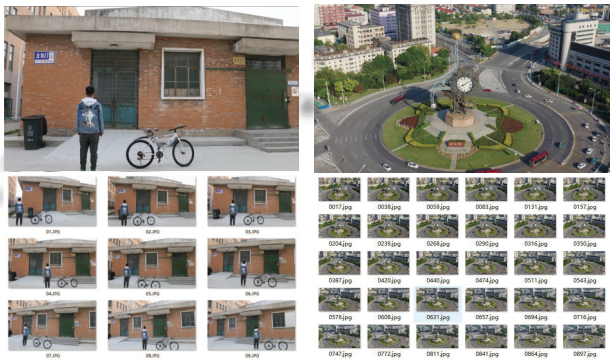


图 6　实验数据集

3.2　特征点数量对比分析

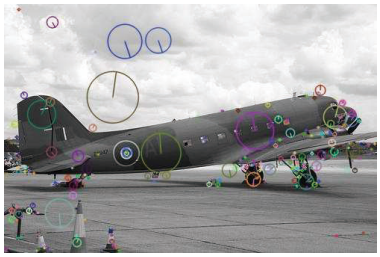
首先采用示例图像进行方法验证, 如图 7、表 2. 图 7 分别为原图 SIFT 特征点可视化, 图像掩膜法剔除运动物体后 SIFT 特征点可视化, 以及本文方法剔除运动物体后的 SIFT 特征点可视化. 由图可得, 掩膜方法修改飞机所在位置像素值为 0, 较多包含飞机信息的特征点被剔除, 然而尾翼、机身、起落架及掩膜边缘位置均产生新的无效特征点. 相比较, 本文方法不修改像素值, 利用特征点过滤模块更新特征点列表, 飞机区域特征点被剔除的同时, 没有生成新的无效特征点. 由表 2 可知, 本文方法剔除 169 个无效特征点, 优于图像掩膜方法的 119 个特征点.

图 8、图 9 为地面图像数据和航拍图像数据的特征点数量对比. 原图特征点数量最多, 掩膜方法及本文方法剔除运动物体后特征点均少于初始图像. 二者中, 本文方法剩余特征点少于掩膜方法, 即本文方法剔除特征点数量大于掩膜方法.

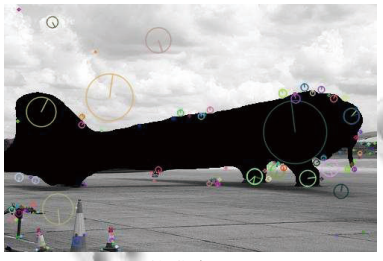
3.3　点云生成时间及重投影误差对比分析

地面图像数据、航拍图像数据以及 5 种重建算法: SFM 方法、Colmap 方法、掩膜剔除的 SFM 方法、本文所提两种方法共 10 组实验的对比如表 3. 航拍数据的稀疏点云如图 10. 设置点云生成时间、点云数量和

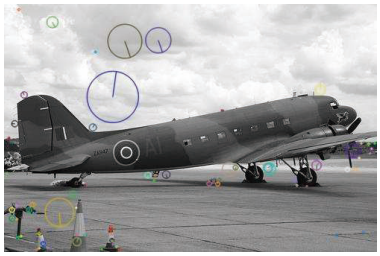
点云重投影误差作为评价指标. 其中, Colmap 利用 GPU 而不是 CPU 进行计算, 仅分析其点云数量及重投影误差.



(a) 原图+SIFT



(b) 掩膜法+SIFT



(c) Ours

图 7 剔除方法对比

表 2 示例图像被剔除特征点的数量对比

原图SIFT点数	剔除点数		剩余点数	
	掩膜方法	本文方法	掩膜方法	本文方法
259	119	**169**	140	**90**

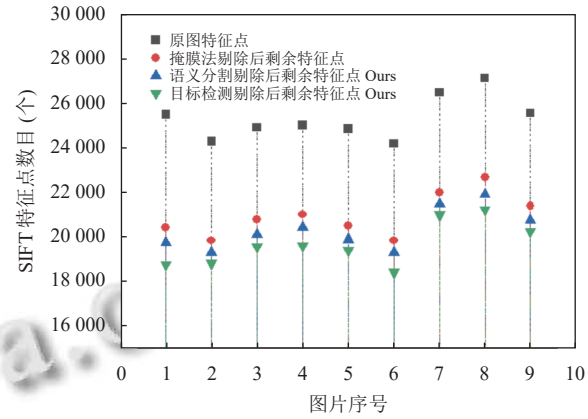


图 8 地面数据特征点数量对比

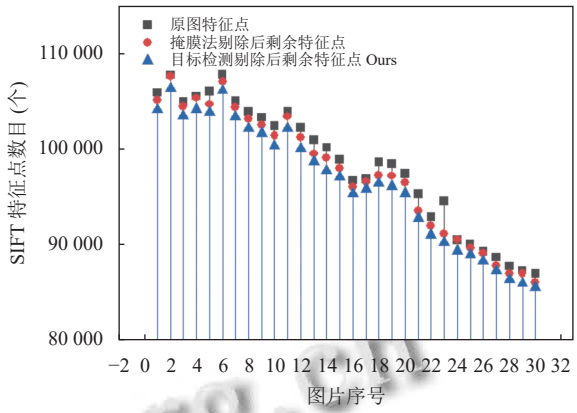


图 9 航拍数据特征点数量对比

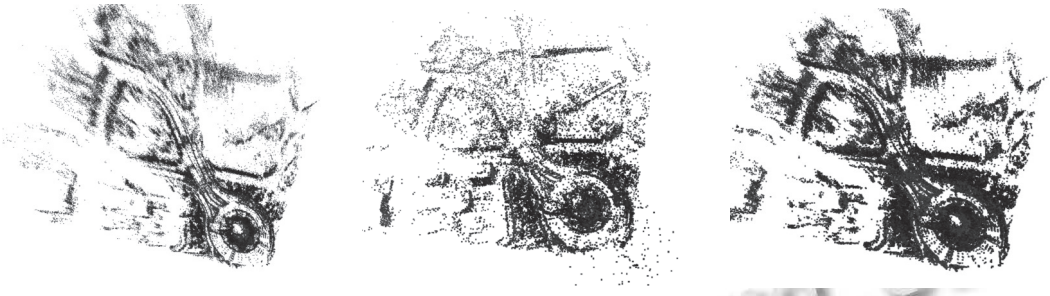
表 3 三维重建运动物体剔除方法对比

方法	地面图像数据					航拍图像数据				
	特征点匹配时间 (s)	最小二乘迭代次数 (次)	点云生成时间 (s)	点云数量 (个)	重投影误差 (像素)	特征点匹配时间 (s)	最小二乘迭代次数 (次)	点云生成时间 (s)	点云数量 (个)	重投影误差 (像素)
SFM	13.504	11	104.547	9332	1.344	851.440	5	6233.917	250884	0.394
Colmap	—	—	—	9454	0.950	—	—	—	28884	0.918
SFM+掩膜法	7.208	6	82.862	9645	1.097	834.027	8	7983.532	253207	0.301
SFM+DeepLab (Ours)	9.441	**3**	64.154	**9732**	**0.927**	822.435	**4**	6085.651	**254865**	**0.288**
SFM+YOLO (Ours)	**6.387**	**3**	**57.924**	9295	1.212	**810.946**	**4**	**5869.728**	254247	0.352

点云生成时间主要受到 KNN 匹配、RANSAC 误匹配剔除和捆绑调整的影响. 本文方法在两组图像数据中的特征点匹配时间分别为 6.387 s 和 810.946 s, 均小于 SFM 方法及掩膜方法花费时间. 由图 8、图 9, 本文方法相比掩膜方法剔除更多无效特征点, RANSAC 局内点比例提高, RANSAC 迭代次数减少. 本文方法的捆绑调整最小二乘迭代次数分别为 3 次、4 次, 小于 SFM 方法的 11 次、5 次和掩膜方法的 6 次、8 次. 特征点匹配和捆绑调整最小二乘迭代过程直接影响稀疏点云生成时间. 本文方法在两组图像中的点云生成时间分别为 57.924 s、5 869.728 s, 小于 SFM 方法的 104.547 s、6 233.917 s, 也小于掩膜方法的 82.862 s、

7983.532 s.

本文方法在两组图像中的点云数量分别为 9 732 个、254 865 个, 相比 SFM 方法的 9 332 个、250 884 个、掩膜方法的 9 645 个、253 207 个以及 Colmap 方法的 9 454 个、28 884 个均有提升, 尤其大幅多于 Colmap 基于航拍数据中生成的稀疏点云数目.



(a) SFM (b) Colmap (c) Ours

图 10 航拍数据稀疏点云对比

重投影误差指真实三维空间点在图像平面的投影与计算得到的近似投影点之间的差值. 本文方法在两组图像数据中的重投影误差分别为 0.927、0.288, 小于 SFM 方法的重投影误差 1.344、0.394、掩膜方法的重投影误差 1.097、0.301 以及 Colmap 方法的 0.950、0.918.

4 结语

本文对静态场景三维重建中运动物体的剔除方法开展研究. 为解决图像掩膜方法修改像素值导致的无效特征点引入问题, 构建特征点剔除模块, 根据卷积神经网络提供的运动物体信息, 筛选更新特征点列表, 在特征点维度实现三维重建运动物体的剔除. 通过地面图像、航拍图像两个数据集的实验, 结果表明:

(1) 特征点维度的三维重建运动目标剔除方法剔除运动物体的同时不引入额外的无效特征点, 相比图像掩膜方法平均多剔除 2.32% 的无效特征点.

(2) 特征点维度的三维重建运动目标剔除方法在点云生成时间和点云重投影误差方面, 优于传统 SFM 方法和图像掩膜剔除方法, 相比于图像掩膜方法, 平均节省 13.36% 的点云生成时间, 平均减小重投影误差 9.93%.

参考文献

1 Snavely N, Seitz SM, Szeliski R. Modeling the world from internet photo collections. International Journal of Computer Vision, 2008, 80(2): 189–210. [doi: 10.1007/s11263-007-0107-3]

2 原明超, 仇俊. 无人机倾斜摄影测量在三 维模型测图中的应用. 测绘通报, 2020, (7): 116–119, 142.

3 郑太雄, 黄帅, 李永福, 等. 基于视觉的三维重建关键技术研究综述. 自动化学报, 2020, 46(4): 631–652. [doi: 10.16383/j.aas.2017.c170502]

4 陈国军, 陈燕, 韦鑫. 多视图像序列中运动目标分割优化. 系统仿真学报, 2014, 26(9): 2023–2027. [doi: 10.16182/j.cnki.joss.2014.09.028]

5 李沛燃, 黄文杰, 陶晓斌. 基于包含多个刚体运动目标的单目视频三维重建研究. 计算机与数字工程, 2016, 44(10): 2037–2042. [doi: 10.3969/j.issn.1672-9722.2016.10.037]

6 Schönberger JL, Frahm JM. Structure-from-motion revisited. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 4104–4113.

7 何继锋. 基于图像语义的室内动态场景三维重建算法研究与实现 [硕士学位论文]. 长沙: 湖南大学, 2020.

8 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.

9 汪雷. 煤矿探测机器人图像处理及动态物体去除算法研究 [硕士学位论文]. 徐州: 中国矿业大学, 2020.

10 Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]

11 Chen LC, Zhu YK, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: IEEE, 2018. 833–851.

12 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

13 高莎, 袁希平, 甘淑, 等. 集成 SIFT 算法与检测模型优化的 UAV 影像匹配方法. 光谱学与光谱分析, 2022, 42(5): 1497–1503. [doi: 10.3964/j.issn.1000-0593(2022)05-1497-07]

(校对责编: 孙君艳)