

单目视频图像序列三维重建方法

沙晓鹏, 曹加奇, 李文静, 秦 晔

(东北大学秦皇岛分校 控制工程学院, 河北 秦皇岛 066004)

摘 要: 针对单目场景下的三维重建, 由于运动模糊和小基线图像的存在导致图像模糊、特征点匹配低、重建精度差等问题, 提出一种单目视觉系统的三维重建方法. 首先, 根据图像的边缘信息利用小波变换检测原始数据中的模糊图像并剔除, 通过几何鲁棒性准则筛选出宽基线图像, 得到用于三维重建的清晰图像; 其次, 提出了基于区域划分的错误匹配特征点剔除算法, 剔除重复匹配和错误匹配特征点; 最后, 提出3种不同区域增长方式获取更多特征点进行三维点计算. 结果表明, 提出的方法有效地去除了错误匹配特征点, 提高了特征点匹配的准确率, 获得了更多的点云数量, 提高了重建模型的精度和完整性.

关键词: 三维重建; 关键帧筛选; 特征提取; 稀疏稠密化; 区域增长

中图分类号: TP 391.4 文献标志码: A 文章编号: 1005-3026(2024)12-1680-08

Three-Dimensional Reconstruction Method of Monocular Video Image Sequences

SHA Xiao-peng, CAO Jia-qi, LI Wen-jing, QIN Ye

(School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China. Corresponding author: SHA Xiao-peng, E-mail: shaxiaopeng@neuq.edu.cn)

Abstract: For 3D reconstruction in uncalibrated monocular scene, due to the existence of motion blur and small baseline image, there are still problems such as blurred image, low feature matching, and reconstruction accuracy. A 3D reconstruction method of monocular vision system is proposed. Firstly, the fuzzy image in the original data is detected and removed by wavelet transform algorithm according to the edge information of the image, and the wide baseline image is screened by geometric robustness criterion to obtain clear images for 3D reconstruction. Then, an algorithm for eliminating mismatched feature points based on region division is proposed to eliminate repeated matching and mismatched feature points. Finally, three different regional growth methods are proposed to obtain more feature points for three-dimensional point calculation. The results show that the proposed method can effectively remove mismatched feature points, improve the accuracy of feature point matching, obtain more point clouds, and improve the integrity and accuracy of the reconstructed model.

Key words: 3D reconstruction; key frame screening; feature extraction; sparse and dense; regional growth

三维重建技术广泛应用于增强现实、考古、生物医学等领域. 随着手机、摄像机等拍摄设备的快速普及, 大量高分辨率图像和视频数据可被用来进行重建物体的三维模型. 然而, 对于这些海量图像数据, 由于运动模糊等原因, 会导致图像模糊、检测出的特征点较少、有较多的误匹配特征点等问题, 因此重建的精度较低^[1-2].

相邻帧之间视角变化太小会导致基线过短, 计算三维点时深度不确定性较大. 因此, 筛选出视频序列中关键帧是为了从大量的图像数据中筛选出基线大、清晰度高、特征点匹配度较高的图像, 从而减小误匹配在三维点计算时的不准确性. 现有关键帧筛选方法主要基于特征匹配比率、几何鲁棒性准则 (geometrical robust information

收稿日期: 2023-07-06

基金项目: 河北省自然科学基金资助项目(F2021501021).

作者简介: 沙晓鹏(1983-), 女, 河北邢台人, 东北大学秦皇岛分校副研究员.

criterion, GRIC) 和结构相似准则 (structural similarity, SSIM). Fang 等^[3]提出一种基于深度补全校正和位姿融合的自动选择关键帧方法, 利用边缘梯度变化剔除模糊图像比值和结构相似性准则筛选出关键帧. Zhang 等^[4]基于多特征融合的四元数傅立叶变换, 构造了一种利用融合特征映射提取关键帧的自适应关键帧滤波准则, 实现了关键帧的准确提取.

关键帧筛选后, 图像特征的准确检测及图像之间特征点的精确匹配, 对于从图像中准确地恢复出物体的三维结构信息至关重要^[5]. 在图像匹配中, 经常会出现一些误匹配的特征点, 这些误匹配点可能是由图像噪声、图像变形或者其他因素引起的, 这将导致重建精度降低或者直接失败. 常用匹配算法有暴力 (brute-force) 匹配、K-最近邻匹配等. 随着深度学习与卷积神经网络理论不断发展, 越来越多的图像特征点检测算法诞生, 如 MatchNet, D2-Net 等. 基于深度学习的特征点检测算法在检测效果以及描述子匹配度等方面优于传统算法, 但在检测速度、数据集的依赖度和算法泛用性等方面依然存在问题. RANSAC (random sample consensus)^[6]即随机采样一致性算法, 是一种鲁棒性较强的特征点匹配算法, 常用于解决匹配过程中存在误匹配或外点的情况和匹配点剔除、鲁棒拟合等问题.

在三维信息重建中, 利用运动相机在不同角度和方向获取的图像通过特征点的准确匹配后恢复出场景的三维结构信息. 运动恢复结构 (structure from motion, SFM) 是典型的三维重建算法^[7]. 但通过 SFM 完成稀疏三维点云重建后展示效果相对较差, 为了更好地展示场景和物体的三维结构信息, 可以结合多视角立体 (multi-view stereo, MVS) 方法来实现对场景和物体的稠密三维点云重建. 基于 MVS 的方法包括深度图融合、空间 Patch 点云扩张、体素等稠密重建方式. Ito 等^[8]将归一化互相关与双边权值和视点之间的几何一致性相结合, 提高了目标边界和纹理差区域深度和法线图的估计精度, 通过加权中值滤波器和基于一致性的多视图几何滤波器, 从重建的三维点云中去除异常值. Li 等^[9]提出了从未标定的小运动视频片段中重建场景三维模型的方法, 利用图像中的关键点和线段特征来缓解小基线引起的三维点计算不确定性问题, 改进了 PatchMatch 立体匹配模块, 提高了稠密三维点云重建的精度.

本文针对模糊图像导致的特征点提取与匹配困难、图像特征点错误匹配去除和稀疏三维点云数目较少等问题, 提出了单目视觉系统的三维重建方法. 首先, 融合哈尔小波变换算法和 GRIC 算法对图像的边缘信息检测进行关键帧筛选, 剔除模糊图像; 其次, 提出基于区域划分的误匹配特征点去除算法剔除 SIFT (scale-invariant feature transform) 特征点匹配后存在的重复匹配特征点及错误匹配特征点; 最后, 提出改进基于区域特征点增长方法获得点云数量, 得到更多的三维点云信息, 提高重建精度和完整度.

1 方法简介

本文提出的单目视觉系统的三维重建方法流程如图 1 所示, 由图像预处理 (模糊图像检测和关键帧)、特征点检测与匹配以及三维点云重建组成.

1.1 图像预处理

图像清晰度与图像间的差异决定着三维重建的精度. 相机在拍摄时, 每秒可生成 30~60 帧图像, 为了更好地从视频序列中筛选出清晰度高、具有关键信息的图像, 本文提出模糊图像和关键帧筛选算法, 如图 2 所示.

1) 模糊图像剔除. 本文采用基于哈尔小波变换 (Haar wavelet transform, HWT) 的模糊图像检测方法, 剔除输入图像序列中模糊图像. 给定同一场景下的 N 幅图像 $\{I_i\}$ 对图像进行水平、垂直、对角等变换, 构建边缘图像, 如式 (1) 所示:

$$E_i(x, y) = \sqrt{\omega_{H_1}^2(x, y) + \omega_{V_1}^2(x, y) + \omega_{D_1}^2(x, y)}. \quad (1)$$

其中, $\omega_{H_1}^2(x, y)$, $\omega_{V_1}^2(x, y)$, $\omega_{D_1}^2(x, y)$ 分别是在图像上进行的水平、垂直、对角变换操作.

Dirac-Structure 和 Astep-Structure 是边缘图像结构^[10], 模糊时 2 种结构类型的边界点会消失, 由此可以判断图像是否模糊. 计算 2 种结构与所有类型边的比值, 得到模糊度量因子, 如式 (2) 所示:

$$\left. \begin{aligned} \text{perDA} &= \frac{N_{\text{DA}}}{N_{\text{edge}}} > B_1, \\ \text{perAS} &= \frac{N_{\text{AS}}}{N_{\text{edge}}} > B_1. \end{aligned} \right\} \quad (2)$$

其中: N_{edge} 是边界点总数目; N_{DA} 和 N_{AS} 分别是 Dirac-Structure 和 Astep-Structure 结构下边界点数目; B_1 为接近并大于零的模糊度量因子. 若满足式 (2) 所示条件, 即 N_{DA} 和 N_{AS} 都不为 0, 则认为当前图像为清晰图像, 否则为模糊图像.

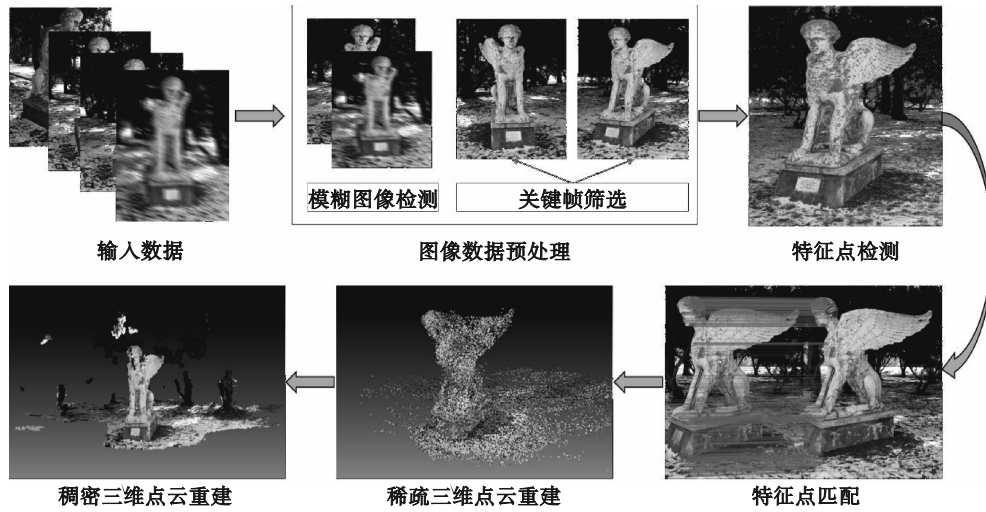


图 1 单目视觉系统三维重建流程图

Fig. 1 Flowchart of 3D reconstruction for monocular vision system

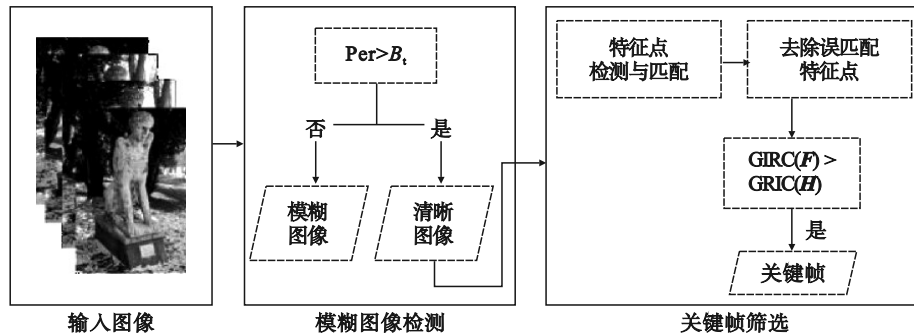


图 2 模糊图像和关键帧图像筛选算法

Fig. 2 Blurred image and key frame image screening algorithm

2) 关键帧筛选. 为了筛选出输入图像中的关键帧, 需要计算出每一张图像的基础矩阵 F 和单应性矩阵 H . F 反映空间一点 P 的像素点在不同视角摄像机下图像坐标系中的表示之间的关系, 根据 RANSAC 算法可计算出每 2 张匹配图像对应的基础矩阵 F 描述了空间中同一平面上的三维点在 2 张图像中的对应关系, 根据 F 与 H 之间的几何关系可以计算出 H 值. 然后使用 GRIC 准则来筛选关键帧, GRIC 数学表达式如式 (3), 式 (4) 所示:

$$GRIC = \sum \rho(e_i^2) + \lambda_1 d \ln r + \lambda_2 \ln(r\lambda_1), \quad (3)$$

$$\rho(e_i^2) = \min\left(\frac{e_i^2}{\sigma^2}, 2(r-d)\right). \quad (4)$$

其中: r 表示数据的维度; λ_1 表示图像中总的特征点数; λ_2 表示选择模型的自由度; d 表示选择模型的尺度; e_i 表示匹配残差; σ^2 表示方差. 参数 d, λ_2 的选择如式 (5) 所示.

$$d = \begin{cases} 2 & GRIC(H), \\ 3 & GRIC(F). \end{cases} \quad (5)$$

$$\lambda_2 = \begin{cases} 8 & GRIC(H), \\ 7 & GRIC(F). \end{cases}$$

当 $d=2, \lambda_2=7$ 时可计算出 $GRIC(H)$ 的值, 当 $d=3, \lambda_2=8$ 时可计算出 $GRIC(F)$ 的值^[1]. 给定一幅图像, 计算矩阵模型 H 和 F 的 GRIC 得分. 若 $GRIC(H) < GRIC(F)$, 则将该图像排除为备选关键帧, 再根据 GRIC 准则筛选出输入图像中的关键帧图像.

1.2 特征错误匹配去除算法

SIFT 特征提取算法可以在图像上提取出大量的特征点, 并生成相应的描述子, 但在实际使用时会存在同一图像中的多个不同区域的点可能会匹配到另一张图像的同—区域、同一区域点匹配到多个区域等问题. 为了去除 SIFT 检测出的重复点和错误匹配, 本文提出基于区域划分的错误匹配特征点剔除算法, 算法流程图如图 3 所示.

首先, 对图像 1 和图像 2 使用 SIFT 特征检测器进行特征点提取, 得到特征点的关键点 kpA, kpB , 及对应描述子 $DescA, DescB$ 信息, 根据描述子信息计算欧氏距离, 如式 (6) 所示:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (6)$$

由于初步匹配后的特征点存在较多重复点, 根据数据唯一性去除重复点. 再对去重后的关键点 kpA' 坐标进行区域划分, 划分成多组相邻近的坐标组 $GroupA_{1,2,\dots,N}$, 并得到每组内点 P_{A_i} 对应的索引. 根据索引 $Index_{A_i}$ 对 kpB' 进行分组 $GroupB_{1,2,\dots,N}$, 对 $GroupB_{1,2,\dots,N}$ 分组内的点利用式 (7) 进行相邻判断.

$$|kpB_i - kpB_{i+1,\dots,N}| \leq \max(\text{relTol} * \max(|kpB_i|, |kpB_{i+1,\dots,N}|), \text{absTol}). \quad (7)$$

其中: relTol 为相对容差 (a 和 b 之间允许的最大差值); absTol 为最小绝对容差.

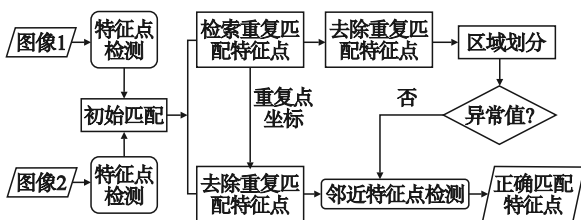


图 3 基于区域划分的错误匹配特征点剔除算法流程图
Fig. 3 Flow chart of error-matching feature point elimination algorithm based on region division

判断对应匹配点 P_{B_i} 是否相邻, 即坐落在同一区域内并利用基于欧氏距离 K-means 的聚类算法判断点之间的相似度, 若坐落在同一区域内且具有较高的相似度则认为该组匹配正确, 若 $GroupA_i$ 中 P_{A_i} 对应点 P_{B_i} 出现异常值, 则根据相邻组且相近点坐标进行判断, 剔除组内错误匹配的点. 依次对每个分组进行遍历, 从而剔除存在的错误匹配点.

1.3 稀疏点云重建算法

为获得更多三维点用于增加三维点云的完整度, 本文提出了基于区域特征点增长的稀疏点云重建算法, 区域增长示意图如图 4 所示.

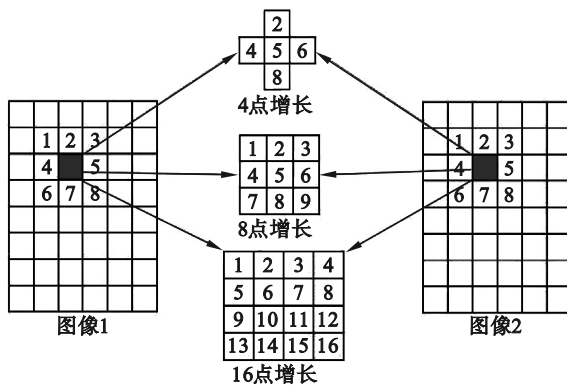


图 4 区域增长示意图

Fig. 4 Schematic diagram of region growing

在本文中主要采用 3 种不同的增长方式获取更多的特征点来进行三维点的计算. 利用光束法平差 (bundle adjustment, BA) 优化后得到的三维点经过反向投影到图像 1 与图像 2 上, 得到对应的投影点, 并利用光度一致性约束判断在 2 幅图像投影点周围的 4 像素点、8 像素点、16 像素点之间的相似性, 筛选出进行二次计算的特征点. 具体如下.

首先, 将 BA 优化后的稀疏三维点云作为种子点反向投影到 2 幅图像上, 投影点为 P 和 P' . 根据光度一致性约束计算投影点周围 4, 8, 16 像素点灰度值大小所得的归一化互相关 (normalized cross correlation, NCC) 系数值, 如式 (8) 所示, 并与预设值比较.

$$P_{NCC} = \frac{\sum_{i=1}^n (P_i - \bar{P}_i)(P'_i - \bar{P}'_i)}{\sqrt{\sum_{i=1}^n (P_i - \bar{P}_i)^2 \sum_{i=1}^n (P'_i - \bar{P}'_i)^2}}. \quad (8)$$

其中, P_i 为种子点 p_i 周围的 Patch 块. 采用归一化互相关系数作为判断标准. 若 P_{NCC} 大于预设值, 则将 P 和 P' 周围新生成的点列入二次三维点计算范围内, 依次对稀疏三维点遍历, 直到遍历结束. 再对通过区域增长获得的特征点利用三角测量的方式计算出对应三维点坐标, 将获得的三维点通过点云融合加入初始点云中, 构成新的三维点云.

2 实验分析

2.1 模糊图像剔除与关键帧筛选

为验证本文提出的图像预处理方法的有效性, 利用手机从任一角度对图 5 所示重建物体进行视频拍摄, 视频图像序列共 278 帧, 其中模糊图像 66 帧, 清晰图像 212 帧. 实验结果如表 1 所示, 通过对比 Harr+ 对应约束和 Harr+SSIM 这 2 种方法, 本文所提出的方法最终确定关键帧为 175 帧, 可获得更多的关键帧用于三维重建.



图 5 实验目标数据采集

Fig. 5 Experimental target data acquisition

表 1 模糊图像剔除与关键帧筛选
Table 1 Fuzzy image elimination and key frame screening

筛选方法	总图像	模糊图像	清晰图像	非关键帧	关键帧
Harr+对应约束	278	66	212	74	138
Harr+SSIM	278	66	212	53	159
本文方法	278	66	212	37	175

2.2 特征点匹配

为了验证提出的基于区域划分的错误匹配特征点剔除算法的有效性,分别用RANSAC错误匹配算法及本文方法对文献[12-13]提供的开源数据集中的LU sphinx, Golden_statue 图像数据和

本文采集的图像数据进行了匹配实验.在RANSAC置信度为70,迭代次数为100的情况下与本文所提出的误匹配特征点去除算法进行对比,如图6和表2所示.

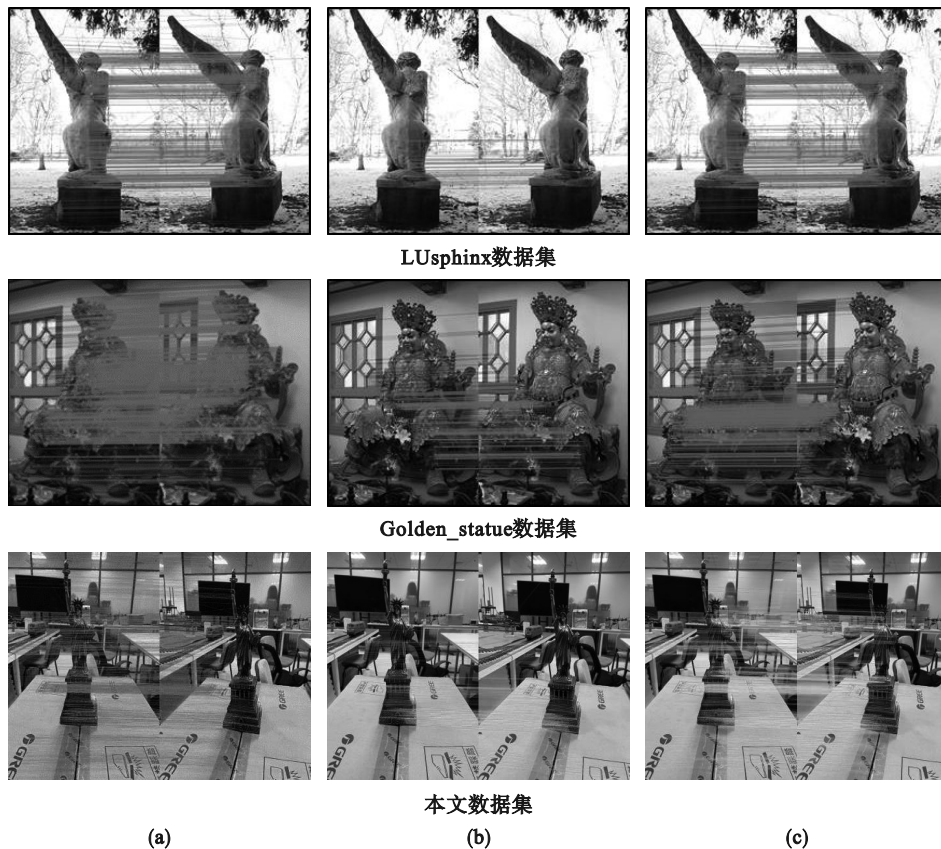


图 6 特征点匹配结果
Fig.6 Feature point matching results
(a)—初始匹配; (b)—RANSAC结果; (c)—本文方法.

表 2 特征匹配对比
Table2 Comparison of feature matching

图像数据	方法	初始匹配/个	去重后/个	正确匹配点/个	正确率/%	时间/s
LU sphinx ^[12]	RANSAC	1 311	1 311	602	46	7
	本文方法	1 311	856	832	63	4
Golden_statue ^[13]	RANSAC	5 386	5 386	2 950	55	23
	本文方法	5 386	5 184	4 281	79	15
本文图像数据	RANSAC	2 133	2 133	405	19	10
	本文方法	2 133	1 800	1 057	50	6

2.3 点云重建

为了验证本文提出的改进稀疏三维点云重建方法的性能和效果, 分别利用运动恢复结构算法 SFM 以及本文方法采用不同的区域增长 (region growing, RG) 半径 (4, 8, 16) 对公开图像数据 LUsphinx (70 张) 和 Golden_statue (18 张) 进

行稀疏三维点云重建实验, 结果如图 7 和图 8 所示. 通过稀疏三维点云重建的效果图和点云数据统计表 3 可以看出, 随着区域增长半径的增加, 三维点云数量逐渐增加, 重建效果也逐渐变好. 这一结果表明, 本文提出的改进方法在稀疏三维点云重建方面具有一定的优势和适用性.

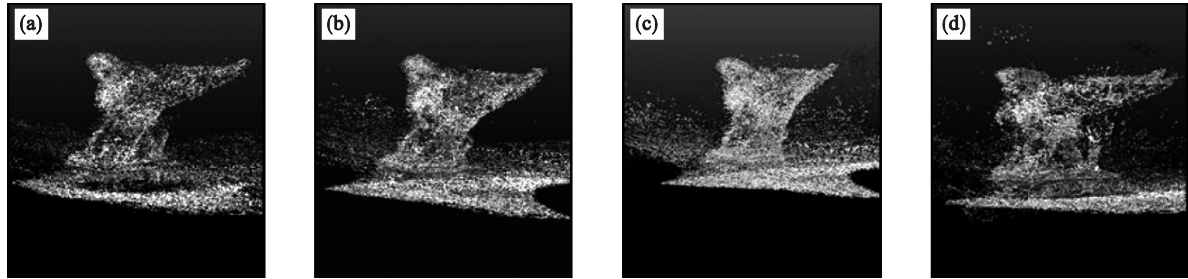


图 7 LUsphinx 稀疏三维点云重建结果

Fig. 7 LUsphinx sparse 3D point cloud reconstruction results

(a)—SFM 方法; (b)—本文方法 RG4; (c)—本文方法 RG8; (d)—本文方法 RG16.

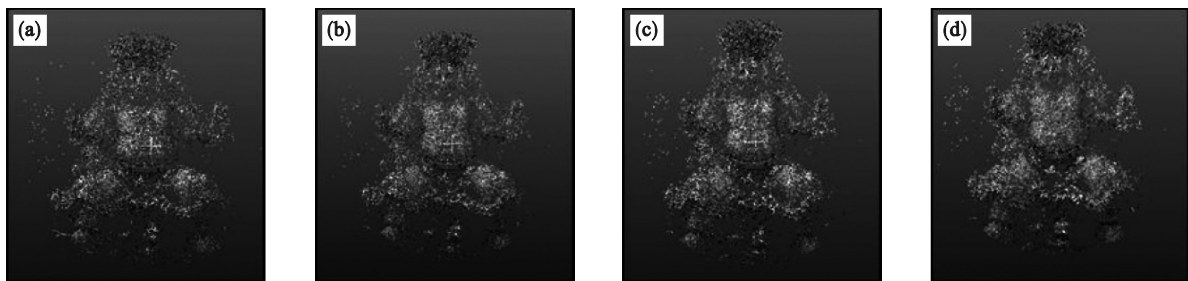


图 8 Golden_statue 稀疏三维点云重建结果

Fig. 8 Golden_statue sparse 3D point cloud reconstruction results

(a)—SFM 方法; (b)—本文方法 RG4; (c)—本文方法 RG8; (d)—本文方法 RG16.

表 3 稀疏三维点云重建数据

Table 3 Sparse 3D point cloud reconstruction data

方法	LUsphinx 稀疏点云	Golden_statue 稀疏点云
文献[12-13]	32 668	39 950
SFM 方法	67 488	15 347
本文方法 (RG4)	114 254	27 712
本文方法 (RG8)	113 381	27 483
本文方法 (RG16)	137 854	32 886

2.4 稠密三维点云重建

为了更好地展示场景和物体的三维结构信息, 本文主要基于空间 Patch 点云扩展的方式, 使用 PMVS 方法对本文方法输出的稀疏三维点云进行稠密重建. 结果如图 9 所示. 通过计算得到稠密三维模型的各个顶点之间的均方距离 RMSE 和 Hausdorff 距离 2 种评价指标, 数据如表 4 所示.

可以看出 2 种方法的模型重建精度相近, 但本文提出的方法在某些细节展示方面具有一定优势, 点云的精度和覆盖率有所提高, 如图 9 中方框区域所示. 但本文算法重建时间还有待进一步优化和提升. 为了更好地表示物体的三维模型, 本文对稠密后的三维点云进行了网格表面重建, 实验结果如图 10 所示. 稠密三维点云在完成

表面重建后的结果更加精细和真实.这是因为稠密三维点云具有更高的采样密度,而本文重建结

果更加清晰,表面形状更加精准,且不存在明显的空洞和不连续的表面,如图 10 中方框区域所示.

表 4 稠密三维点云重建数据
Table 4 Dense 3D point cloud reconstruction data

参数指标数据	方法	点云/个	增加点云数/个	重建时间/s	RMSE	Hausdorff距离
LU sphinx ^[12]	PMVS ^[14]	352 899		303	0.108 399	0.108 399
	本文方法	382 743	29 844	645		
Golden_statue ^[13]	PMVS ^[14]	104 557		83	0.048 637	0.016 452
	本文方法	146 975	42 418	179		

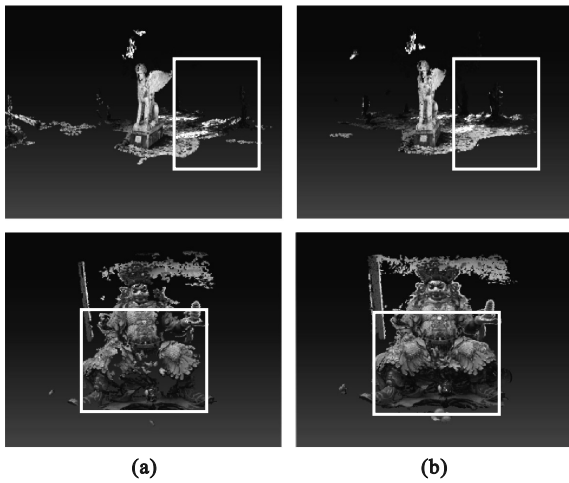


图 9 稠密三维点云重建结果

Fig. 9 Dense 3D point cloud reconstruction results
(a)—SFM 和 PMVS; (b)—本文方法.

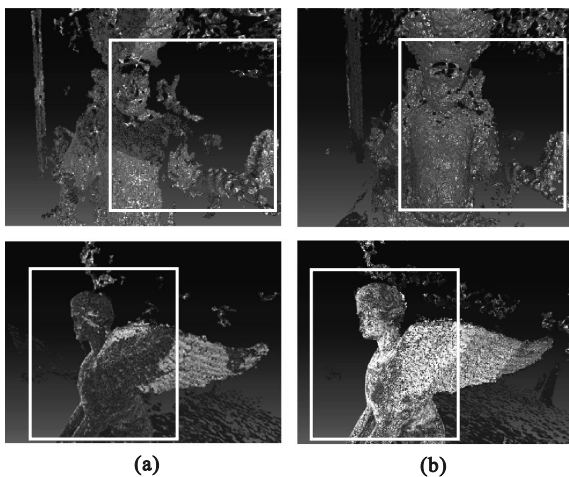


图 10 稠密三维点云表面网格重建

Fig. 10 Dense 3D point cloud surface mesh reconstruction
(a)—SFM 和 PMVS; (b)—本文方法.

3 结 语

本文提出了一种任意场景下视频图像序列的三维重建方法.基于提出的模糊关键帧筛选方法和改进误匹配剔除算法提高了特征点检测和提取的准确率.通过基于区域划分的错误匹配特征点去除算法剔除错误和重复匹配点,提高了图像特征点匹配的准确性.通过改进区域增长算法,获取了更多点云的数量,提高了重建精度和完整度.实验数据表明,所提出的视频图像序列的三维重建方法特征点匹配精度和三维点云结构信息的完整度较好.

参考文献:

- [1] Barrile V, Bernardo E, Bilotta G. An experimental HBIM processing: innovative tool for 3D model reconstruction of morpho-typological phases for the cultural heritage [J]. *Remote Sensing*, 2022, 14: 1288.
- [2] 胡正乙,谭庆昌,孙秋成.基于RGB-D的室内场景实时三维重建算法[J].东北大学学报(自然科学版),2017,38(12):1764-1768.
(Hu Zheng-yi, Tan Qing-chang, Sun Qiu-cheng. RGB-D based indoor scene real-time 3D reconstruction algorithm [J]. *Journal of Northeastern University (Natural Science)*, 2017, 38(12): 1764-1768.)
- [3] Tian F G, Gao Y B, Fang Z J, et al. 3D reconstruction with auto-selected keyframes based on depth completion correction and pose fusion [J]. *Visual Communication Image Represent*, 2021, 79: 103199.
- [4] Zhang Y Z, Zhang J Y, Liu R X, et al. Key frame extraction based on quaternion Fourier transform with multiple features fusion [J]. *Expert Systems with Applications*, 2023, 216: 11119467.
- [5] 何鑫睿,李秀梅,孙军梅,等.基于改进Pix2Vox的单图像三维重建网络[J].计算机辅助设计与图形学学报,2022,34(3):364-372.

- (He Xin-rui, Li Xiu-mei, Sun Jun-mei, et al. Improved Pix2Vox based 3D reconstruction network from single image [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2022, 34(3):364–372.)
- [6] Won J, Park J W, Song M H, et al. Robust vision-based displacement measurement and acceleration estimation using PANSAC and Kalman filter [J]. *Earthquake Engineering and Engineering Vibration*, 2023, 22(2):347–358.
- [7] Nielsen M S, Nikolov I, Kruse E K, et al. Quantifying the influence of surface texture and shape on structure from motion 3D reconstruction [J]. *Sensors*, 2023, 23(1):178–200.
- [8] Ito K, Ito T, Aoki T. PM-MVS: PatchMatch multi-view stereo[J]. *Machine Vision and Applications*, 2023, 34:32–48.
- [9] Li Z X, Zuo W M, Wang Z Q, et al. Robust 3D reconstruction from uncalibrated small motion clips [J]. *The Visual Computer*, 2022, 38(5):1589–1605.
- [10] Tong H H, Li M J, Zhang H J, et al. Blur detection for digital images using wavelet transform [C]//IEEE International Conference on Multimedia & Expo. Taipei, 2004:17–20.
- [11] Torr P H S. Geometric motion segmentation and model selection [J]. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1998, 356:1321–1340.
- [12] Olsson C, Enqvist O. Stable structure from motion for unordered image collections [C]// Proceedings of the Scandinavian Conference on Image Analysis. Berlin: Springer, 2011:524–535.
- [13] Enqvist O, Kahl F, Olsson C. Non-sequential structure from motion [C]//2011 IEEE International Conference on Computer Vision Workshops. Barcelona, 2011:264–271.
- [14] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(8):1362–1376.