

引用格式:张佳预,贾同,李文浩.基于RGB-D图像序列的大场景三维重建方法研究[J].中国传媒大学学报(自然科学版),2023,30(01):53-65.

文章编号:1673-4793(2023)01-0053-18

基于RGB-D图像序列的大场景三维重建方法研究

张佳预,贾同*,李文浩

(东北大学信息科学与工程学院,沈阳110819)

摘要:三维重建技术是计算机视觉领域研究的重点,广泛应用于自动驾驶、逆向工程、文物恢复、观演空间展示等领域。目前,对单物体的重建效果显著,但大场景信息复杂、特征杂乱,现有针对大场景的特征匹配算法和点云配准算法在计算效率和精度上仍存在一定的局限性。因此,本文基于RGB-D图像序列,首先采用了一种结合非极大值抑制的ORB特征提取方法,提出了基于KD树和优先队列相结合的匹配方式,然后构建了融合多元信息的关键帧筛选机制,实现了对局部场景稠密点云的实时生成。其次,提出了一种基于双重阈值约束点云精配准方法,在点云法向量夹角阈值约束的基础上,通过自适应距离阈值约束实现ICP算法中最近邻点对的搜索。最后,在真实大场景中进行实验分析,验证了本文算法的有效性。

关键词:三维重建;特征匹配;点云配准

中图分类号:TP231 文献标识码:A

Research on 3D reconstruction method of large scene based on RGB-D image sequence

ZHANG Jiayu, JIA Tong*, LI Wenhao

(Northeastern University, Shenyang 110819, China)

Abstract: Three-dimensional reconstruction technology is the focus of research in the field of computer vision, which is widely used in automatic driving, reverse engineering, cultural relics restoration, performance space display and other fields. At present, the reconstruction effect of single object is remarkable, but the information of large scene is complex and the features are messy. The existing feature matching algorithms and point cloud registration algorithms for large scene still have some limitations in computational efficiency and accuracy. Therefore, based on RGB-D image sequence, this paper firstly adopts an ORB feature extraction method combined with non-maximum suppression, and proposes a matching method based on KD tree and priority queue, and then constructs a key frame filtering mechanism based on multivariate information, which realizes the real-time generation of dense point clouds in local scenes. Secondly, a point cloud fine-registration method based on double threshold constraint is proposed. Based on the threshold constraint of point cloud normal vector angle, the nearest neighbor pair in ICP algorithm is searched by adaptive distance threshold constraint. Finally, experiments are carried out in real large scenes to verify the effectiveness of the proposed algorithm.

Keywords: 3D reconstruction; feature matching; point cloud registration

基金项目:国家重点研发计划(2018YFB404101);国家重点研发计划(20200197)

作者简介(*为通讯作者):张佳预(1999-),男,硕士研究生,主要从事双目深度感知研究。Email:2170871@stu.neu.edu.cn;贾同(1975-),男,博士,教授,主要从事计算机视觉、模式识别、机器学习等方面的研究。Email:jiatong@ise.neu.edu.cn

1 引言

三维重建^[1]作为连接现实世界与虚拟世界的桥梁,已广泛应用于自动驾驶^[2],逆向工程^[3],文物恢复^[4],观演空间展示^[5]等领域。目前对单物体的三维重建已经取得良好效果^[6],但大场景存在空间尺度大、信息复杂、遮挡严重等因素,因此实现大场景三维重建对设备性能、重建方法等都是重大考验。

近年来,深度传感器飞速发展,涌现出一大批操作简单、价格相对低廉的深度相机,如 Kinect^[7]、RealSense^[8]和 Astra 等。与利用传统的几何方法或者 3D 激光扫描技术进行三维重建相比,利用基于深度传感器获得的 RGB-D 图像序列进行三维重建精度可靠,成本低廉,操作简单,满足效率和精度方面的需求。

基于图像序列的重建方法有直接法和间接法两类,直接法是像素级别的重建方法,直接利用彩色图像进行匹配。间接法又叫特征点法,是基于像素块级别的重建方法,其效率是直接法的数十倍到百倍。与直接法不同的是,特征点法多出了特征检测和描述子计算两个步骤。目前有很多特征点检测方法以及描述子表述方法,不同的待重建场景以及不同的重建任务需求可以根据不同的特征点特性进行选取。

特征点法需要进行特征匹配,进而生成稠密点云。常用的图像特征有 SIFT^[9]、SURF^[10]和 ORB^[11]。特征提取完成后需要进行特征匹配,暴力匹配(Brute Force, BF)是最经典的方法,但是当数据量大时,这种方法的计算效率非常低。目前,常用的匹配方法都是基于最邻近开源库(Fast Library for Approximate Nearest Neighbors, FLANN)^[12]进行的,包括基于 KD 树的搜索算法、基于 K-Means 的搜索算法和基于层次聚类树的搜索算法等。特征匹配完成后结合相机位姿数据便可获得稠密点云数据。

由于大场景环境复杂,一次重建显然无法满足要求,需要多次重建后进行点云配准融合。点云配准一般分为粗配准和精配准两个步骤。常用的粗配准算法有 4PCS 算法^[13]、Super-4PCS 算法^[14]和 SAC-IA 算法^[15]。精配准算法主要有 PNP 算法^[16]和 ICP 算法^[17]等。

本文利用 RGB-D 图像序列进行大场景三维重建,首先改进了特征提取和匹配方法,并进行相机位姿跟踪与优化,生成高质量稠密点云;然后提出一种基于双重阈值约束的点云精配准算法。最终,使用本文方法获得真实大场景高效率、高精度的三维重建结果。

2 基于 RGB-D 信息的稠密点云生成

本文利用 Kinect V1 深度相机提供的 RGB-D(彩色和深度)信息,进行稠密点云生成。不仅能够提升重建效率,而且为后续点云配准提供高质量点云数据。

2.1 图像特征点提取

图像特征包括有颜色、纹理以及边缘轮廓等。稳定的特征点具有四个特点:一是可区分性,即特征点的像素信息与周围像素有较大差别;二是可区别性,即不同的特征点之间具有不同的特征信息;三是高效性,即计算特征点的时间要远小于计算所有像素点的时间;四是抗干扰性,当像素点受到不同程度的信息干扰,特征点仍能够保持原有信息。

2.1.1 常用图像特征分析

满足上述特点的特征点可以作为图像特征点。图像的特征点种类繁多,最常使用的特征点有 SIFT 特征、SURF 特征以及 ORB 特征。SIFT 特征具有尺度不变性,同时对于噪声以及仿射变换等也有较高的鲁棒性,但特征提取过程相对耗时;SURF 是对 SIFT 特征的进一步改进,改进包括特征提取的过程以及特征描述的方式,缩短了特征点的提取时间;ORB 特征是一种兼具快速性和稳定性的特征点,由于 ORB 特征点具有高鲁棒性、高实时性,被广泛应用于多数较为经典的三维重建算法中。

针对上述特征点,本文进行了相关实验。取 200 帧 640×480 图像序列进行了测试,通过参数调节使得三种特征点的数量趋于相等。分别统计了每帧检测耗时,并计算了每个特征点消耗的时间,如表 1 所示:

表 1 三种特征点对比

特征种类	平均特征点数量(个)	每帧检测时间(s)	每点耗时(ms)
SIFT	621	1.74	2.8
SURF	593	0.50	0.85
ORB	588	0.02	0.04

由上表可知,在检测速度上,ORB 特征是 SIFT 特征的近 70 倍,是 SURF 特征近 20 倍。同时 ORB 特征具有旋转和尺度不变性,有利于后续特征匹配,因此本文选取 ORB 特征点作为后续特征提取点。

2.1.2 基于非极大值抑制的 ORB 特征点提取

(1) ORB 特征提取

ORB 特征点提取步骤包括 FAST 角点检测以及

BRIEF描述子计算两部分,FAST角点检测原理是检测像素信息之间的差异,计算过程如下。

以图像中某像素点P为圆心,构建半径 $r=3$ 的圆形区域,圆周上共16个点,如图1所示。按照顺时针方向将像素点P的灰度值 I_p 与邻域内16个点灰度值做差后和阈值 δ 比较,比较结果分为三类:

$$|x| = \begin{cases} I_p - I_{pi} < \delta \\ I_p - I_{pi} = \delta \\ I_p - I_{pi} > \delta \end{cases} \quad (1)$$

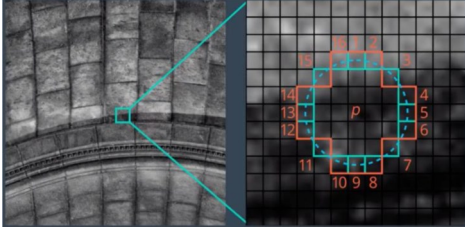


图1 FAST角点检测示意图

若同一类别中有8个像素点连通,则选择像素点P作为FAST角点。根据上式计算,FAST角点本身不包含方向信息。为了能够为FAST角点赋予方向信息,本文引入灰度质心法,具体计算过程如下:

根据FAST角点所确定的圆周范围内的所有像素点的灰度值,确定图像块的零阶距、一阶距:

$$m_{\alpha\beta} = \sum_{x,y \in B} x^\alpha y^\beta I(x,y) \quad (2)$$

式中, $\alpha, \beta \in \{0,1\}$,利用零阶距、一阶距可以确定质心为:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (3)$$

根据质心C点的相对坐标可以求出角点P的方向角 θ :

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}}\right) \quad (4)$$

BRIEF是一种将FAST角点进行二进制向量化的描述算子,该描述算子一般为多维度向量且只包含0和1。0和1是通过FAST角点周围的像素点对 (I_p, I_q) 比得到, I_p 大于 I_q 则为1,反之则为0。若选取了N对像素点对 (I_p, I_q) ,则最终的描述算子维度为N维。BRIEF算子通常取256维。

(2)结合非极大值抑制

由于图像中的不同区域呈现的纹理特征不同,特征点分布会存在较大差异。对于特征较为丰富的区域,会出现特征点堆积现象。这些特征点相似度高,相互形成干扰,同时多余的特征点会增加计算量,降

低匹配效率。为了在不影响特征提取质量的前提下,降低特征点堆积程度。本文在特征点检测过程引入非极大值抑制算法^[18]。

非极大值抑制本质在于搜索局部二维空间的极大值,舍弃其它非极值元素。受非极大值抑制算法思想启发,本文对传统ORB特征提取过程做了相应改进。总体思想为:将目标图像等分为N个图像块,在每块空间中选择响应值最大的特征点作为该区域的特征点,舍弃掉其他特征点。

2.2 特征点对匹配与筛选

特征点提取之后,需要进行特征点匹配。本文采用基于描述子距离与颜色信息的特征点匹配机制。针对匹配实时性问题,采用KD树和优先队列相结合的方法进行对应点搜索;针对匹配准确性问题,采用基于RANSAC方法进行特征点对的进一步筛选。

2.2.1 特征点匹配机制

(1)描述子距离

描述子能够充分表达特征点周围的像素信息,通过计算描述子之间的距离就可有效表征特征点之间的相似程度。在上一节中介绍了ORB特征点的BRIEF描述子,本文通过计算两个BRIEF描述子之间欧氏距离来表示两个ORB特征点的相似程度,具体计算如公式(5)所示。

$$D(P,Q) = \| P - Q \| = \sqrt{\sum_{i=1}^{256} (p_i - q_i)^2} \quad (5)$$

其中,P和Q表示待匹配图像的特征点。 p_i 和 q_i 表示特征点对应的256维描述向量,计算结果D表示特征点之间的相似程度。

(2)RGB信息一致性

颜色信息是基本且直观的图像信息。在灯光稳定、无太多额外光线干扰的场景下,颜色信息在一定时间内具有较高的可靠性。本文将颜色信息作为特征点筛选的重要约束,对特征点周围的像素块计算各自RGB颜色分量的均值,比较待匹配点之间的颜色分量差值并根据阈值筛选。颜色分量的计算如公式(6)所示。

$$I_{RGB} = \frac{1}{9} \sum_{i=1}^9 I_{r,g,b}^i \quad (6)$$

其中计算结果 I_{RGB} 表示各自通道的计算均值,9表示以特征点为中心的3×3像素块的像素数, $I_{r,g,b}$ 表示各个像素点对应的RGB颜色分量。

(3)特征点匹配流程

本文以 BRIEF 描述子距离和 RGB 信息一致性为匹配标准,建立特征点匹配机制,整体流程如图 2 所示。

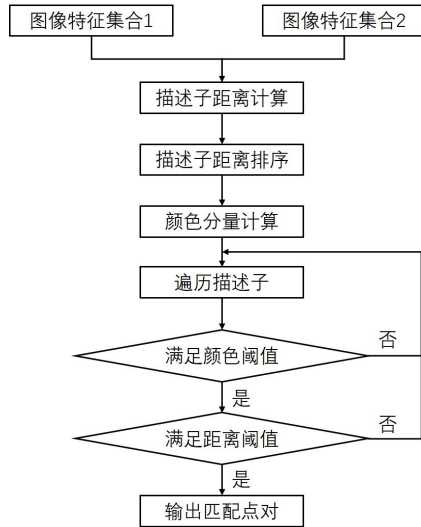


图 2 特征点匹配流程

2.2.2 基于KD树和优先队列的匹配算法改进

在特征点匹配的过程中,一个非常关键的步骤是对应点查找,但查找步骤比较耗时,经典暴力查找的时间复杂度为 $O(n^2)$ 远无法达到实时性要求。为了提升匹配效率,本文联合 KD 树^{[19][20]}和优先队列进行特征点查找。

(1) KD 树

KD 树是一种二叉树数据结构。该结构能够对整个数据空间进行多维度划分,同时可按照不同维度的数据进行索引搜索,是一种高效数据结构。图 3 是 KD 树的空间结构表示,图中的红色、蓝色、黄色以及绿色线分别表示了不同维度的数据划分。而不同维度的划分是根据维度之间的方差大小决定,方差越大表示该维度的向量越具有良好的区分性,是优先划分的维度。因此维度的划分顺序遵循方差由大到小的顺序。基于 KD 树的空间搜索机制包括了两个主要过程,一个是建树过程,另一个是查找过程。

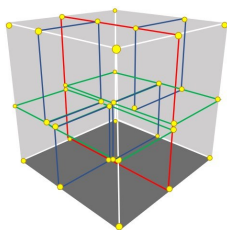


图 3 KD 树空间结构示意图

(2) 基于 KD 树和优先队列结合的匹配算法

KD 树结构处理低维数据 ($k \leq 30$) 时具有快速性的优势,但处理高维数据 ($k \geq 100$) 却有数据维度的灾难。这是因为在寻找查询点 Q 的邻近点过程中,必须要经过回溯步骤。而在回溯的过程,对于高维数据而言,需要回溯并推断的分支数据量庞大,查找效率骤然下降。为了实现特征点的高效匹配,本文设定了最大回溯次数,同时在 KD 树匹配机制的基础上加入优先队列。在优先队列中保存每次遍历得到节点的位置和距离。加入优先队列后,本文可以在有限的回溯次数内,寻找到查询点 Q 的最邻近点。

2.2.3 对应点筛选

特征匹配一般都存在误差,造成对应点误匹配,因此还需要对应点筛选。常用方法有最小二乘和随机采样一致性 (Random Sample Consensus, RANSAC)^[21]方法。最小二乘主要通过构建误差方程,使得误差的平方最小化,从而得到符合样本的模型。但是该方法只适用于数据量较小和小误差的样本。对于数据量较大且存在有大量噪声的样本,该算法效率较低,实时性较差。RANSAC 是通过不断选取数据集中的某些子集,采用多次迭代的方式,从一群包含大量噪声的数据集合中,基于随机性和假设性,计算出符合大部分数据点的数学模型。该方法具有较高的精确度和鲁棒性,广泛应用于数据拟合、矩阵变换等领域。本文采用 RANSAC 方法进行特征点对筛选。

2.3 相机位姿跟踪与优化

Kinect V1 处于工作状态时帧率为 30,直接处理 30 帧图像计算负荷巨大。同时,冗余的图像信息容易使位姿计算与优化过程产生错误。因此,为了能够高效合理的利用摄像头捕获的图像信息,需要选择一些具有代表性的图像,即关键帧。关键帧选取包括相机位姿计算、关键帧筛选以及位姿图优化过程。

2.3.1 相机位姿计算

(1) 相机模型与坐标系变换

在进行相机位姿估计前,需要掌握相机模型与各个坐标系之间的转换关系^{[22][23]}。相机成像本质上是将三维世界的真实场景映射到二维图片的过程,可以用一个几何模型来描述,最具有代表性的就是针孔模型。针孔模型中涉及四大坐标系,分别是世界坐标系、相机坐标系、图像坐标系和像素坐标系。坐标系之间的变换过程也就是相机的成像过程,包括世界坐

标点到相机坐标点的刚体变换、相机坐标点到图像坐标点的投影变换以及图像坐标点到像素坐标点的二次变换。这些变换可以用公式(7)描述。

$$\begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{K} & \vec{0} \\ \vec{0}^T & 1 \end{bmatrix} \frac{1}{Z_c} \begin{bmatrix} \mathbf{R} & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (7)$$

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

式中 (u_c, v_c) 为像素坐标, (X_w, Y_w, Z_w) 为真实场景的世界坐标, \mathbf{K} 为相机内参数, \mathbf{R} 为旋转矩阵, \vec{t} 为平移向量。

(2)位姿计算

相机位姿计算是为了能够求出相机之间的相对运动,常用方法有P3P^[24],ICP和EPNP^[25]。P3P是利用3对3D-2D的对应点进行位姿求解,由于该方法只需要3对特征点,因此求解结果偶然性较大。ICP算法处理的是3D-3D点之间的匹配问题,既可以求已经配对的3D点的变换关系,又可以求未知配对关系的3D点。但ICP算法需要不断重复计算,无法适用于实时场景的位姿计算中。

EPNP方法也是一种利用3D-2D信息计算相机位姿的方法。该方法不仅计算精度较高,且在计算效率上要高于ICP算法,能够满足实际场景的相机位姿计算,因此本文选择EPNP进行相机位姿计算。该方法的几何原理为:空间中任何不共面的四个点(下文称为控制点)总能表示三维空间中的任意一点。表示方法如下:

$$P_i = \sum_{j=1}^4 \alpha_j M_j, \sum_{j=1}^4 \alpha_j = 1 \quad (8)$$

其中 M_j 表示控制点信息, α_j 表示对应于四个控制点的权值信息。

对于同一个三维空间点 P_i ,当相机移动坐标系发生变化时,控制点对应的权值信息不会发生变化,只有控制点空间坐标会随相机移动发生变化。因此对于下一帧三维空间点,仍然满足以下约束:

$$P_i = \sum_{j=1}^4 \alpha_j N_j, \sum_{j=1}^4 \alpha_j = 1 \quad (9)$$

其中 N_j 为变换后坐标系中的控制点。

当相机发生移动时,相机坐标系发生变化,只有对应点的坐标发生了变化。此时选取的三维空间点 P_i 便是由深度图投影得到的三维点。在具体计算过程中,控制点的选取策略较为重要。一般来说,首个

控制点选取整个三维点云的质心,另外三个控制点根据主成分分析^[26]方法,选取点云的三个主要方向。

对于任意一点 P_i ,在已知四个控制点 $M_j(j=1,2,3,4)$ 的三维坐标后,便可求出控制点所对应的权值信息。由于权值信息具有不变性,因此可以求出变换后四个控制点 $N_j(j=1,2,3,4)$ 的三维坐标。根据两组三维坐标 (M_j, N_j) ,最终便可计算出相机坐标系之间的变换关系。

2.3.2 关键帧筛选

为了筛选出具有代表性的图像关键帧,本文构建基于多元信息的关键帧筛选,包括时间信息、空间信息、图像相似度信息和地图质量信息。

(1)时间信息

深度相机正常工作过程中,短时间内相机位姿不会发生较大幅度的变化,此时基于时间间隔的筛选结果相对准确。本文根据时间间隔确定候选关键帧有两个条件:第一,当前帧与上一个关键帧相差MAX=60帧时,确定为候选关键帧;第二,当前帧与上一个关键帧至少相差MIN=30帧,且候选帧数量小于3时,确定为候选关键帧。

(2)空间信息

如果相机发生较快速度的运动时,空间位置信息会发生较快的变化,因此只利用时间间隔很容易造成跟踪丢失。为了能够兼顾相机的剧烈抖动、快速运动等情况,本文构建了基于空间信息的关键帧筛选条件。

根据相机的运动特性,本文利用公式(10)计算相机相对运动的幅值 D ,利用幅值 D 和预先设定的阈值 δ 作比较。大于阈值,则证明当前帧图像相比于上一次关键帧的运动幅值较大,将当前图像确定为候选关键帧;反之则表明运动幅度较小,当前帧图像不作为关键帧。

$$D = \|\Delta \vec{t}\| + \min(2\pi - \|\mathbf{R}\|, \|\mathbf{R}\|) \quad (10)$$

其中相机 \mathbf{R} 和 \vec{t} 表示当前帧和上一关键帧之间的旋转和平移向量, D 表示相对运动幅值。

(3)图像相似度信息

图像相似度可以分为像素级和特征点级两类,像素级方法时间计算成本过高,不适用于实时性的三维点云的稠密重建。因此本文采用特征点级方法,以下是具体计算方法。

对当前关键帧 I_k 和当前帧 I_c 利用本文2.1节方法进行特征提取,记录各自特征点数量 m 和 n ,并按照

2.2节方法进行特征对匹配与筛选,求得匹配点对数 c 。根据公式(11)计算图像 I_k 和 I_c 的相似度 $D(I_k, I_c)$ 。

$$D(I_k, I_c) = \frac{2c}{m+n} \quad (11)$$

利用 $D(I_k, I_c)$ 和预先设定的阈值 δ 进行比较,若 $D(I_k, I_c) > \delta$ 则表明两张图像之间的相似度较高,当前帧图像不具备代表性,无法成为候选关键帧;反之将当前帧作为候选关键帧。

(4) 地图质量信息

地图质量指的是当前帧重建的点云地图对更新三维地图的贡献程度。贡献程度低表明当前帧对应的图像内容和上一个关键帧重复率越高;贡献程度越高表明当前帧对应的图像内容是最新的场景信息,有助于不断扩大更新三维场景地图。

根据当前帧和上一个关键帧的匹配关系,计算出当前帧图像在世界坐标系中的位姿数据。根据位姿矩阵,将当前帧的ORB特征点投影到世界坐标系中。假设集合 $I = \{P_1, P_2, \dots, P_n\}$ 表示当前帧经过投影变换后在场景地图中的坐标集合。集合 $I' = \{P'_1, P'_2, \dots, P'_n\}$ 表示集合 I 对其每个元素 P_i 经过KNN

搜索得到的最邻近点集合,即集合 I 和集合 I' 中各个元素相互对应。根据公式(12)计算投影点 P_i 与原始点 P'_i 之间的欧氏距离,并判断当前帧的投影点是否与原始地图点有关联。

$$d(P_i, P'_i) = \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2} \quad (12)$$

比较欧氏距离 $d(P_i, P'_i)$ 和距离阈值 δ (实验中取0.2)的大小,如果 $d(P_i, P'_i) < \delta$ 则表明当前三维特征点和已有地图点表示了相同的空间点,因此该特征点不是新地图点;反之记录所有满足 $d(P_i, P'_i) > \delta$ 的特征点个数 m ,根据特征点个数 m 和当前帧所有点的个数 n 之间的比值结果判断当前帧是否为关键帧。

2.3.3 位姿图优化

随着相机不断工作,相机光心轨迹逐渐变长,地图中点云数量逐渐增多,因此相机位姿的累计误差以及全局累计误差会随之变大。为了能够最大限度降低长时间运行造成的位姿累计误差,本文采用优化的方式降低累计误差。

目前常用的优化模型共有两种,分别是图4(a)所示的图优化模型以及图4(b)所示的位姿图优化模型。

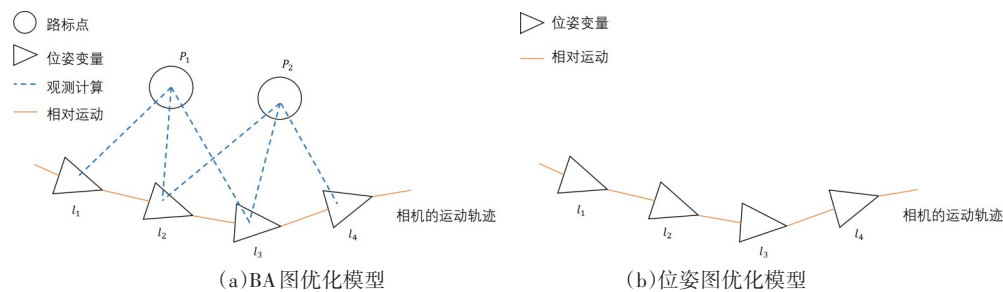


图4 优化模型示意图

在BA图优化模型^[27]中,每次增加一帧图像,就会与多个路标节点建立关系,而路标节点便是稀疏点云地图中的特征点。由于特征点数量巨大,每次增加一帧图像需要进行大量计算,尽管BA优化过程中的模型矩阵 \mathbf{H} 具有稀疏结构,可以显式的由图优化方法表示^{[28][29]},对于整体的优化具有加速作用,但是在大规模构建地图过程中,计算过程仍然难以满足实时性需求。

本文采用位姿图优化模型,其中位姿节点表示待优化的变量,相连的边表示误差项。为了最大限度地降低计算量,位姿图优化模型只优化了位姿节点,边表示的误差项是相邻帧之间的运动估计误差。设位姿节点为 $\mathbf{T}_1, \dots, \mathbf{T}_n$,假设估计了 \mathbf{T}_i 和 \mathbf{T}_j 之间的误差项 \mathbf{T}_{ij} ,则基于李群和李代数^[23]的误差项数学模型为:

$$\mathbf{T}_{ij} = \mathbf{T}_i^{-1} \mathbf{T}_j$$

根据李群和李代数,误差项的最小二乘表达 \mathbf{e}_{ij} 为:

$$\mathbf{e}_{ij} = \ln(\mathbf{T}_{ij}^{-1} \mathbf{T}_i^{-1} \mathbf{T}_j) = \ln(\exp(-\xi_{ij}) \exp(-\xi_i) \exp(\xi_j))$$

式中 $\exp(\cdot)$ 为李群和李代数中的指数映射, ξ_i 和 ξ_j 为待优化变量。

总优化目标函数为:

$$\min_{\xi} = \frac{1}{2} \sum_{ij \in \varepsilon} \mathbf{e}_{ij}^T \Sigma_{ij}^{-1} \mathbf{e}_{ij}$$

式中 ε 表示所有边的集合, Σ_{ij} 为 \mathbf{e}_{ij} 的协方差矩阵。

2.4 稠密点云生成

2.4.1 稀疏点云初始化

对RGB-D图像序列进行2.3节中的关键帧筛选及其优化步骤,随相机运动拍摄可获得一系列关键帧

图像,将所得关键帧依次编号为1,2,...,N。记录第一帧关键帧的相机位姿,并恢复特征点对应的深度值,同时剔除深度值为0的特征点。将彩色图与深度图对应的特征点信息投影到三维空间中,生成初始稀疏点云地图。

2.4.2 稠密点云生成

根据RGB-D图像序列提供的像素信息 (u_c, v_c) 和深度信息 d 生成稠密三维点云。根据本文2.3.1节中相机模型和坐标系转换关系,结合公式(7)可以推导出相机坐标和像素坐标的变换关系,即生成单个关键帧的稠密点云,如公式(16)所示。

$$\begin{cases} X_c = \frac{(u_c - c_x) \cdot Z_c}{f_x} \\ Y_c = \frac{(v_c - c_y) \cdot Z_c}{f_y} \\ Z_c = d \end{cases} \quad (16)$$

三维地图初始化过程中,将第一帧关键帧的相机坐标系作为世界坐标系。因此,从世界坐标到相机坐标的变换矩阵 \mathbf{R} 和 \vec{t} ,则是关键帧中第一帧的相机坐标系与后续帧相机坐标之间的变换关系,相机坐标之间的变换以及优化过程可以从2.3.1节中位姿计算得出。由此可以得到多个关键帧点云融合的稠密点云。

3 点云配准

由于大场景空间信息复杂,单次扫描难以实现对整个场景的完整重建,所以需要进行多次扫描,得到局部稠密点云之后再行点云配准。本文首先联合多种点云滤波模型对初始点云数据进行精简;然后采用Super-4PCS方法进行点云粗配准,为点云精配准提供一个距离较近的初始位置;最后利用基于双重阈值约束的ICP算法进行精配准,得到完整的三维点云模型。

3.1 点云预处理

大场景稠密点云数据量大,常常达到百万级,计算负荷极大,严重影响了配准的效率。因此,本文联

合统计滤波和体素网格滤波对离群点和噪点进行剔除。既不破坏点云特征,又能达到精简点云数量,提高配准效率的目的。

3.1.1 点云离群点剔除

使用相机对目标场景信息进行采集时,不可避免地会采集到目标场景以外的信息,在后续重建的过程中会产生离群点。离群点能够代表的信息量很少,而且会对点云配准过程产生影响,降低配准精确度。因此,本文使用统计滤波的方法对离群点进行剔除。

统计滤波算法^[30]是对点云集合中每个点的邻域进行统计分析的算法,其工作原理如图5所示。首先对点云中各个点与其邻域中最近的 n 个点的平均距离 $d_i (i = 1, 2, \dots, n)$ 进行计算。点云集合中的一点 $P_n(x_n, y_n, z_n)$ 到另一点 $P_m(x_m, y_m, z_m)$ 的距离为:

$$d_i = \sqrt{(x_n - x_m)^2 + (y_n - y_m)^2 + (z_n - z_m)^2} \quad (17)$$

假设所有的 d_i 构成高斯分布,计算任意两点之间距离的平均值 μ 和标准差 σ ,如公式(18)、(19)所示:

$$\mu = \frac{1}{n} \sum_{i=1}^n d_i \quad (18)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \mu)^2} \quad (19)$$

假设标准差倍数为 λ ,当点云中某点与其邻域内 n 个点的平均距离在 $(\mu - \sigma \cdot \lambda, \mu + \sigma \cdot \lambda)$ 范围内时保留该点,反之将其定义为离群点进行剔除操作。

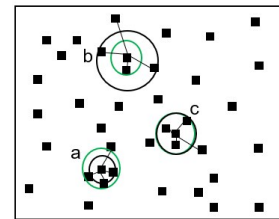


图5 点云统计滤波原理图

3.1.2 点云体素网格滤波

根据上述方法过滤掉离群点之后,继续使用体素网格滤波方法^[31]对点云数量进行精简,体素网格滤波的原理相对简单,如图6所示。

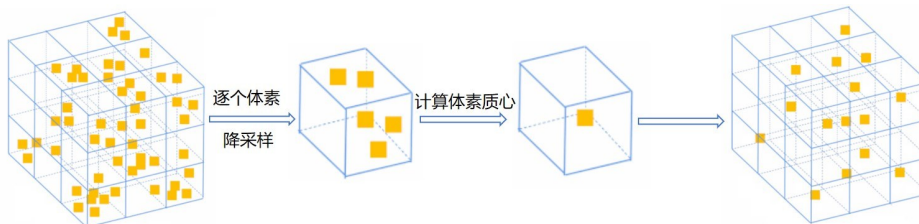


图6 点云体素网格滤波原理图

首先根据点云数据建立一个三维的体素网格,每个网格都可以看做是一个立方体;然后根据体素中的点计算出体素的质心,用质心代替体素中的其它点;最后对所有体素按照上述方式进行处理,就能够获得精简之后的点云数据。假设体素网格的质心用 $P(\bar{x}, \bar{y}, \bar{z})$ 表示,三维体素网格的边长为 L 。质心 P 的表达形式如公式(20)所示:

$$\begin{cases} \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \\ \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \\ \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i \end{cases} \quad (20)$$

3.2 基于 Super-4PCS 的点云粗配准

在对点云进行精配准之前,需要先对两片点云进行粗配准,使两片点云的位置尽可能靠近。Super-4PCS 方法是对常用粗配准方法 4PCS 的改进。

4PCS 算法的基本思想是根据点云刚体变换的仿射不变性进行点云配准,即四个点组成的一个平面无论进行多大角度的旋转,它们之间依旧保持着某种拓扑关系。因此,4PCS 算法与 RANSAC 算法的思想类似,但是在点的选取上增加了四个点需要共面的约束,这在一定程度上比 RANSAC 方法更具有鲁棒性。4PCS 算法的四点共面原理如下图 7 所示。

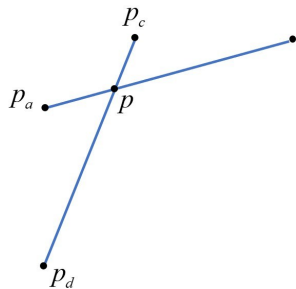


图7 四点共面构成的查询单元

4PCS 算法的主要步骤如下:

1) 在点对的选取上,先在源点云中随机选择 3 个点,由于三个点确定一个平面,在由这三个点确定的平面内选择一个点作为第 4 个点,这 4 个点称为基础对,用 $B = \{p_a, p_b, p_c, p_d\}$ 表示,相交点用 p 表示。

2) 采用公式(21)计算基础对 B 的拓扑信息,用距离参数 r_1 和 r_2 表示。

$$\begin{cases} r_1 = \frac{\|p_a - p\|}{\|p_a - p_b\|} \\ r_2 = \frac{\|p_c - p\|}{\|p_c - p_d\|} \end{cases} \quad (21)$$

3) 在目标点集中寻找与源点云中基础对 B 具有相同拓扑结构的点,采用的计算如公式(20),得到的点如图 8 所示

$$\begin{cases} e_1 = q_1 + r_1(q_2 - q_1) \\ e_2 = q_1 + r_2(q_2 - q_1) \end{cases} \quad (22)$$

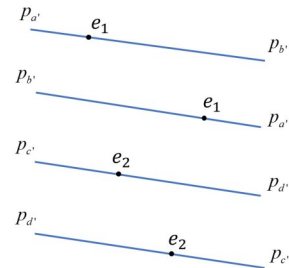


图8 交点类型

4) 在满足拓扑关系的多个点对中,为了确定唯一的变换矩阵,设定一个距离阈值 δ ,对源点集进行变换,将对应点距离小于 δ 的点对作为内点,记录内点数量最多时对应的变换矩阵。

Super-4PCS 方法对 4PCS 方法从两个方面进行了改进。一是改进了寻找对应点对的方式,将任意查找方式改为了圆球形寻点的方式,降低了算法的复杂度,从 $O(n^2)$ 有效降至 $O(n)$ 。二是在计算拓扑结构过程中,增加了两条线段之间的夹角信息,有效地减少了寻找对应点对的次数,极大节省了原始 4PCS 方法运行的时间。

3.3 基于双重阈值约束的点云精配准

ICP 算法是最常用的点云精配准算法之一,可以细分为阈值法和单向临近匹配法^[32]。本文在传统的 ICP 算法的基础上进行了更为深入的研究。针对传统算法在配准效率和配准精度上存在的不足,提出了基于双重阈值约束的点云精配准算法,通过对点云向量夹角设定阈值和设置自适应距离阈值两种方法,对传统算法进行了优化。

3.3.1 传统 ICP 算法存在的不足

传统的 ICP 算法通过反复迭代的方式,计算对应点之间的欧式距离,对最小化误差函数以及旋转矩阵和平移矩阵进行求解,不断优化点云之间的刚性变换关系,如果对应点之间的欧式距离小于设定的阈值,

则停止迭代,从而实现点云精配准。其步骤为:首先,查找距离源点云 P 中任意一点 p_i 最近的三个点 q_j ($j = 1,2,3$);然后分别计算 p_i 到 q_j ($j = 1,2,3$)之间的距离,如果距离小于阈值,则将 q_j 中与 p_i 距离最近的点作为对应点,否则认为对应点匹配失败;最后将查找到的对应点对保存起来,并记录为 $S_i = \{(p_i, q_j) | i = 1,2, \dots, n\}$ 。

ICP精配准算法虽然步骤简单,配准精度较高,但是也存在一定的局限性。当需要配准的点云数据量较小时,该方法配准速度较快,但是对于本文所研究的大场景点云配准来说,点云的数据量较大,能达到百万级,需要求解的最近邻点集数据量庞大,算法时间复杂度增加。而且ICP算法在反复迭代的过程中会使得两点之间的欧氏距离逐渐缩小,如果固定阈值 ε 过大,会使得算法陷入局部最优;如果设定的阈值 ε 过小,会使得算法在初始阶段就剔除了过多的对应点,导致算法无法正确收敛。因此传统的固定阈值方法可能会造成误匹配问题,有一定局限性。

3.3.2 基于双重阈值约束ICP算法

为了进一步提高点云配准的效率并解决配准过程中造成的误匹配问题,本文提出基于双重阈值的配准机制,分别在法向量夹角和对应点之间的欧氏距离上进行约束,以达到提高配准效率并减少误匹配的目的。

(1) KD树点云近邻点搜索

大场景点云数据量大,使用暴力方法对点云邻点进行查找,效率低下。因此,与本文2.2.2节类似,为了提高点云邻点搜索的速度,提高后续点云配准效率,本文在ICP算法中加入KD树结构。

(2) 基于法向量夹角阈值的点云精简

在三维曲面上,与曲面垂直的向量称之为曲面的法向量,而对于三维空间中的点而言,它们所在的平面为其邻域范围内许多点拟合而成的局部平面,则它们的法向量可以定义为垂直于该局部平面的向量,因此可以根据点的法向量的夹角来判断点云所在曲面的变化幅度^[33],如图9所示,其中对于点 p_i 和 p_j ,其法向量分别为 n_i 和 n_j 。

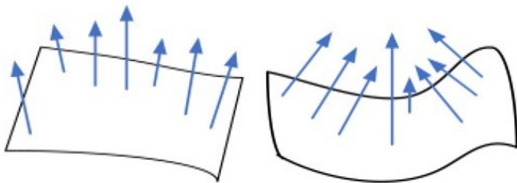


图9 不同局部平面法向量示意图

利用KD树可以搜索到 p 点的邻域点集 $p_i = \{p_1, p_2, p_3, \dots, p_k\}$,然后利用最小二乘法将点 p 和点集 p_i 拟合成一个局部平面,那么通过计算该局部平面的法向量就可以得出点集 p_i 的法向量,即 p 点的无向法向量^[34]。

首先计算邻近点集的均值,即重心 O ,计算如公式(23)所示。

$$O = \frac{1}{k} \sum_{i=1}^k p_i \quad (23)$$

其中, k 为邻域点的个数。

然后使用最小二乘法对邻域点集拟合的局部平面的法向量进行求解。相当于求解公式(24)中函数的最小值。

$$f = \sum_{i=1}^k \|(p_i - \vec{O}) \cdot \vec{n}\| \quad (24)$$

上述最小值的求解也可以转化成对公式(25)中协方差矩阵 E 的最小特征值 λ_j 和对应特征向量 \vec{m}_j 的求解。

$$E = \frac{1}{k} \sum_{i=1}^k (p_i - \vec{O})(p_i - \vec{O})^T \quad (25)$$

$$E \cdot \vec{m}_j = \lambda_j \cdot \vec{m}_j \quad (26)$$

由上述步骤获取到点云的法向量之后,就可以估计出点所在局部平面的变化幅度,进而对特征进行筛选,具体步骤如下:

a) 对邻域点集 p_i 求其法向量 \vec{n} ;

b) 计算邻域点集 p_i 与其邻域所有点的法向量夹角 θ_{ij} ;

c) 计算所有法向量夹角 θ_{ij} 的均值,如公式(27)所示。

$$\bar{\theta} = \frac{1}{k} \sum_{j=1}^k \theta_{ij} \quad (27)$$

设定一个合适的阈值 ε ,如果 $\bar{\theta} \geq \varepsilon$ 则认为该点及其邻域所处曲面变化幅度较大,特征较明显,并将其存入特征点集中;如果 $\bar{\theta} \leq \varepsilon$,则认为该点及其邻域所处曲面变化幅度较小,曲面相对较为平坦,特征表现不明显,并将其删除。

通过上述方法可以对三维空间点云特征进行筛选,根据特征变化程度精简了点云数量,为后续点云配准提供了更有用的信息,为提高点云配准精度打下了基础。

(3) 基于自适应距离阈值的点云配准

针对传统固定阈值ICP算法的局限性,本文提出一种基于自适应距离阈值的约束方法,根据ICP算法

迭代时计算出的所有对应点之间欧式距离的平均值 $d_i (i = 1, 2, \dots, n)$, 求解出一个动态的自适应距离阈值 ε 。主要步骤如下:

a) 首先计算对应点对距离集合 d_i 的平均值 μ_d , 如公式(28)所示:

$$\mu_d = \frac{\sum_{i=1}^n d_i}{n} \quad (28)$$

b) 计算对应点距离集合的标准差 σ_d :

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \mu_d)^2}{n}} \quad (29)$$

c) 那么, 可以得到自适应距离阈值 ε 的计算公式, 如公式(30)所示:

$$\varepsilon = \mu_d + \alpha \sigma_d \quad (30)$$

其中, α 为超参数, 本次实验中 $\alpha = 0.1$ 。

4 实验与结果分析

4.1 实验环境

本文在整个研究过程中使用的硬件和软件系统的配置如表2所示。

表2 本文硬件/软件系统详细配置

设备/软件名称	具体型号/版本号
深度相机	Kinect V1
电脑、内存	NUC8i5BEH、8G
电源	DBK 移动电源 H50
操作系统	Windows 10、Ubuntu 16.04
计算机视觉库	OpenCV 3.4.0

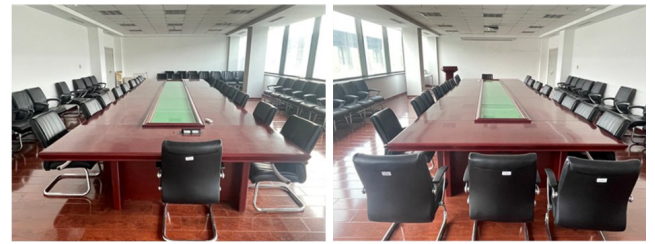
3D点云库	PCL 1.8.0
位姿优化库	G2O
开发环境	Visual Studio2015、VsCode 1.55.2

4.2 真实场景三维重建实验

针对实际大场景进行三维重建, 首先是利用 Kinect V1 拍摄的一系列图像进行特征点匹配; 然后进行位姿估计与优化, 获得局部场景的稠密点云; 最后将局部稠密点云通过本文点云配准算法进行配准, 得到一个完整三维模型。本文选择大型会议室作为实际场景, 并与 RGBDSLAMV2 和传统 ICP 算法进行对比实验, 验证本文方法在局部稠密点云生成以及点云配准方面的有效性。

4.2.1 局部稠密重建实验

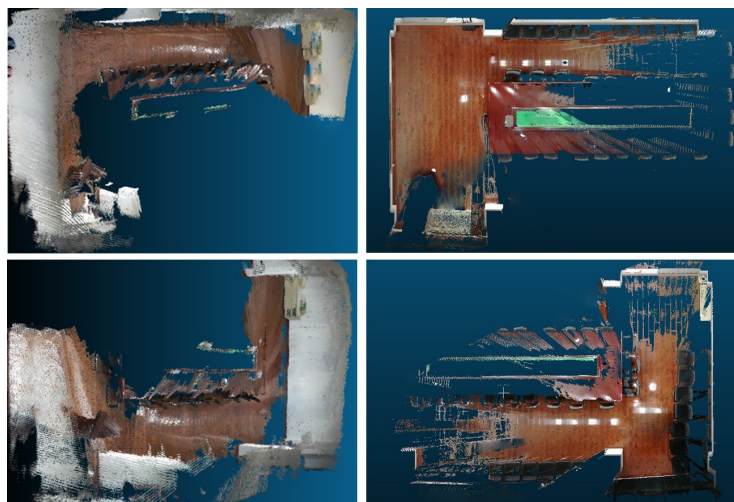
本文对大型会议室场景分两次进行三维重建, 每次采集“L”型区域的图像。会议室的实际场景如图10所示。



(a) 会议室从前到后实际场景 (b) 会议室从后到前实际场景

图10 会议室实际场景

根据本文第2节提出的方法和 RGBDSLAMV2 方法对图10所示的会议室场景进行三维建模, 可以获得两个“L”型区域的稠密点云, 如图11所示。



(a) RGBDSLAMV2算法重建效果

(b) 本文方法重建效果

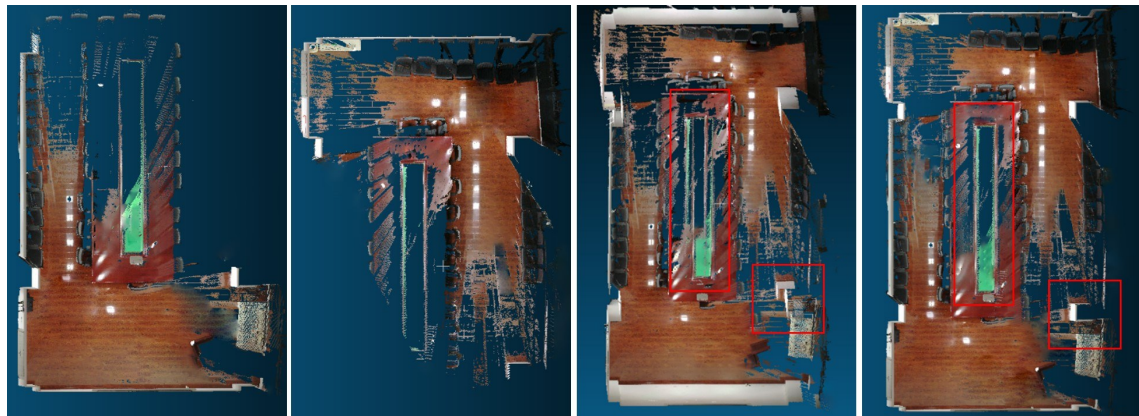
图11 本文方法与 RGBDSLAMV2 方法重建效果对比

可见本文方法和RGBDSLAMV2方法都能够基本保持会议室原有的整体轮廓信息,由于会议室地板和中间桌子材质容易产生反光现象,所以两者重建结果均出现一些孔洞,但是本文方法无论是在配准精度还是在色彩还原度方面优于RGBDSLAMV2

方法。

4.2.2 点云配准实验

将本文方法获得的两片稠密点云,分别使用本文第3节提出方法和传统ICP算法进行点云配准,得到最终配准结果进行实验对比,如图12所示。



(a) (b) (c) (d)

图12 点云配准效果对比:(a)(b)本文方法得到的局部稠密点云;(c)传统ICP算法配准结果;(d)本文方法配准结果)

由(c)(d)中红框部分可以看出,本文算法配准结果基本无偏移现象,而传统ICP配准结果出现了偏移,配准误差较大。

4.2.3 重建精度

为了对本文重建效果进行更加直观的说明,图

13给出了配准之后的场景细节部分,包括会议室主讲台、后门、饮水机、茶几和后排座椅的重建效果,其中第一行为实际拍摄的图片,第二行为RGBDSLAMV2方法的重建结果,第三行为本文方法的重建结果。

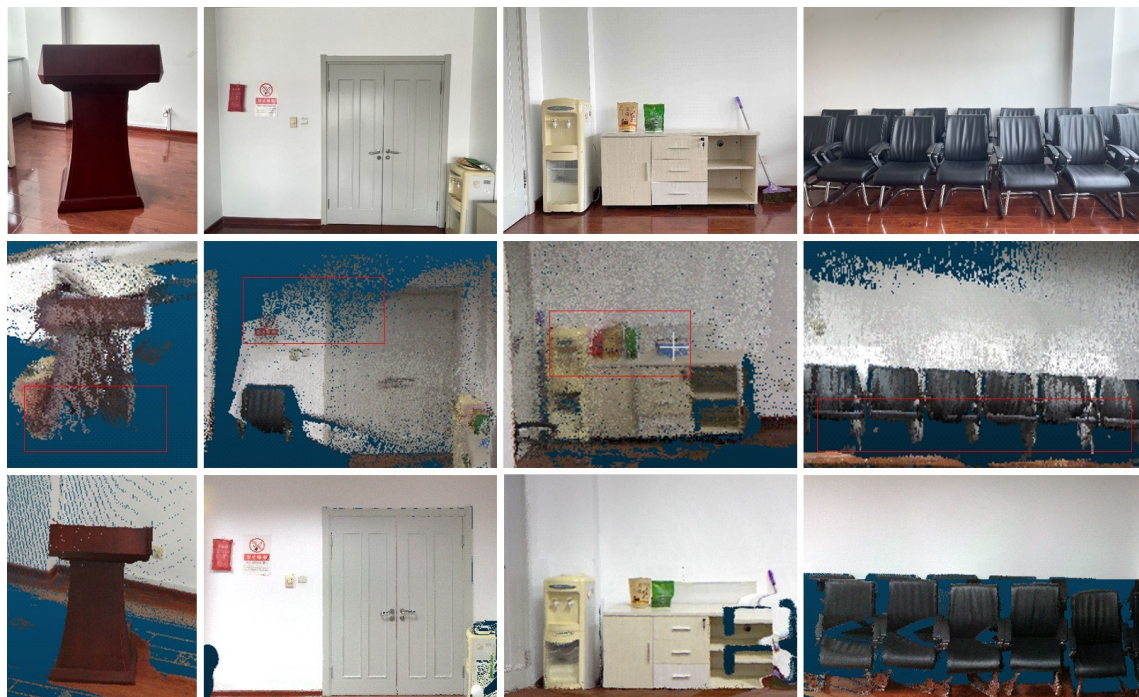


图13 会议室场景细节部分重建结果

由实验结果可以看出,两种方法均能恢复局部场景的基本轮廓信息。但是由于遮挡,光线等因素的影响,RGBDSLAMV2算法的重建结果会出现部分点云模糊、错位以及缺失严重的现象,而本文方法重建出的点云效果更清晰,纹理信息更明显,有效地减少了

外界环境中光线、遮挡等因素的影响。

为了验证本文重建出的三维模型的精确度,进行定量说明,本文将物体实际尺寸与重建尺寸差值的绝对值作为重建误差,选取了重建出的三维模型中的目标物体,并对其重建误差进行统计。



图14 标注距离的会议室局部点云

表4 目标物体的重建误差

特征物体	实际尺寸(cm)	重建尺寸(cm)	重建误差(cm)
主讲台宽度	71.5	70.49	1.01
主讲台高度	96.4	97.44	1.04
后门宽度	155.8	154.70	1.10
后门高度	212.5	211.27	1.23
指示牌宽度	30.0	28.98	1.02

如图14所示通过软件对会议室场景中的三处局部场景点云之间的欧氏距离进行测量,统计得到的数据如下表4所示。表中列举的几类特征物体重建误差在2 cm之内,误差保持在厘米级。

5 结论

本文针对大场景三维重建关键问题,提出基于RGB-D图像序列的三维重建方法。在稠密点云生成阶段,本文首先使用联合非极大值抑制的ORB特征提取方法,其次构建了基于特征点固有信息的匹配机制,提出了基于KD树和优先队列相结合的匹配方式,然后构建了基于多元信息的关键帧筛选机制,并对关键帧位姿进行优化,进而得到高质量稠密点云;在点云配准阶段,提出了一种基于双重阈值约束的点云精配准算法;最后,通过本文方法对真实大场景进行三维重建,验证了本文的有效性。

参考文献(References):

- [1] Kien D T. A review of 3D reconstruction from video sequences[J]. University of Amsterdam ISIS Technical Report Series, 2005.
- [2] Esparza J, Helmle M, Bernd Jahne. Wide base stereo with fisheye optics: a robust approach for 3D reconstruction in driving assistance[C]. German Conference on Pattern Recognition, Springer, Cham, 2014:342-353.
- [3] Buonamici F, Carfagni M, Furferi R, et al. Reverse engineering modeling methods and tools: a survey [J]. Computer-Aided Design and Applications, 2018, 15(3): 443-464.
- [4] Gomes L, Bellon O R P, Silva L. 3D reconstruction methods for digital preservation of cultural heritage: A survey [J]. Pattern Recognition Letters, 2014, 50(1): 3-14.
- [5] Shen S. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes [J]. IEEE Transactions on Image Processing, 2013, 22(5):1901-1914.
- [6] Zeng Z K, Jia T, Ma J C. A Registration Method of Multi-View Point Clouds Using Turntable and Monocular Structured Light [C]//2021 China Automation Congress (CAC). IEEE, 2021: 7144-7149.
- [7] Zhang Z. Microsoft kinect sensor and its effect [J]. IEEE multimedia, 2012, 19(2): 4-10.
- [8] Keselman L, Iselin Woodfill J, Grunnet-Jepsen A, et al. Intel realsense stereoscopic depth cameras [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 1-10.
- [9] Beis J S, Lowe D G. Shape indexing using approximate nearest-neighbour search in highdimensional spaces [C]//Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 1997.
- [10] Bay H, Tuytelaars T, Gool L V. SURF: Speeded up robust features [J]. Proceedings of the 9th European conference

- on Computer Vision -Volume Part I, 2006.
- [11] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C]//IEEE International Conference on Computer Vision, ICCV 2011. IEEE, 2011.
- [12] Muja M, Lowe D G. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration [C]//International Conference on Computer Vision Theory & Application Vis-sapp. 2009.
- [13] Ge X. Automatic markerless registration of point clouds with semantic-keypoint-based 4-points congruent sets [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130:344-357.
- [14] Mellado N, Aiger D, Mitra N J. Super 4pcs fast global point-cloud registration via smart indexing [J]. Computer graphics forum, 2014, 33(5):205-215.
- [15] Rusu R B, Blodow N, Beetz M. Fast point feature histogram (FPFH) for 3D registration [C]. IEEE International Conference on Robotics and Automation, 2009: 3212-3217.
- [16] Lepetit V, Moreno-Noguer F, Fua P. Epnp: an accurate $o(n)$ solution to the pnp problem [J]. International Journal of Computer Vision, 2009, 81(2):155-166.
- [17] Chetverikov D, Svirko D, Stepanov D, et al. The trimmed iterative closest point algorithm [C]. International Conference on Pattern Recognition, IEEE, 2002, 3:545-548.
- [18] Neubeck A, Van Gool L. Efficient non-maximum suppression [C]//18th international conference on pattern recognition (ICPR'06). IEEE, 2006, 3: 850-855.
- [19] 崔明明, 余小庆. 基于新SURFDAISY算法和反向验证随机KD树的图像配准[J]. 航空电子技术, 2018, 49(04): 39-44.
- [20] 岑丽. 基于改进KD树与RANSAC算法的目标识别算法[J]. 电子设计工程, 2017, 25(21):141-144.
- [21] Chum O, Matas J, Kittler J. Locally optimized RANSAC [C]//Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, Proceedings 25. Springer Berlin Heidelberg, 2003: 236-243.
- [22] Hartley R, Zisserman A. Multiple view geometry in computer vision [M]. Cambridge university press, 2003.
- [23] 高翔, 张涛. 视觉SLAM十四讲 [M]. 电子工业出版社, 2018.
- [24] Gao X S, Hou X R, Tang J, et al. Complete solution classification for the perspective-three-point problem [J]. IEEE transactions on pattern analysis and machine intelligence, 2003, 25(8): 930-943.
- [25] Lepetit V, Moreno-Noguer F, Fua P. EP n P: An accurate $O(n)$ solution to the PnP problem [J]. International journal of computer vision, 2009, 81: 155-166.
- [26] Wold S, Esbensen K, Geladi P. Principal component analysis [J]. Chemometrics & Intelligent Laboratory Systems, 1987, 2(1-3):37-52.
- [27] Triggs B. Bundle Adjustment — A Modern Synthesis [C]//International Workshop on Vision Algorithms: Theory & Practice. Springer-Verlag, 1999.
- [28] Kummerle R, Grisetti G, Strasdat H, et al. G2o: A general framework for graph optimization [C]//IEEE International Conference on Robotics & Automation. IEEE, 2011.
- [29] Polok L. Incremental Block Cholesky Factorization for Nonlinear Least Squares in Robotics [J]. IFAC Proceedings Volumes, 2013, 46(10):172-178.
- [30] Zhao Q, Gao X, Li J, et al. Optimization Algorithm for Point Cloud Quality Enhancement Based on Statistical Filtering [J]. Journal of Sensors, 2021, 2021:1-10.
- [31] Xiong B, Jiang W, Li D, et al. Voxel Grid-Based Fast Registration of Terrestrial Point Cloud [J]. Remote Sensing, 2021, 13(10): 1905.
- [32] 李文浩, 贾同, 吕朝辉, 孙小钧, 黄俊文, 张松娜. 大场景三维扫描与建模方法研究 [J]. 中国传媒大学学报(自然科学版), 2022, 29(02):79-88.
- [33] 曾祥磊. 基于几何特征的三维点云配准算法研究 [D]. 山东大学, 2020.
- [34] 蔡文娜, 刘凤连. 一种基于法向量夹角的点云配准方法 [J]. 天津理工大学学报, 2018, 34(3): 17-20.