

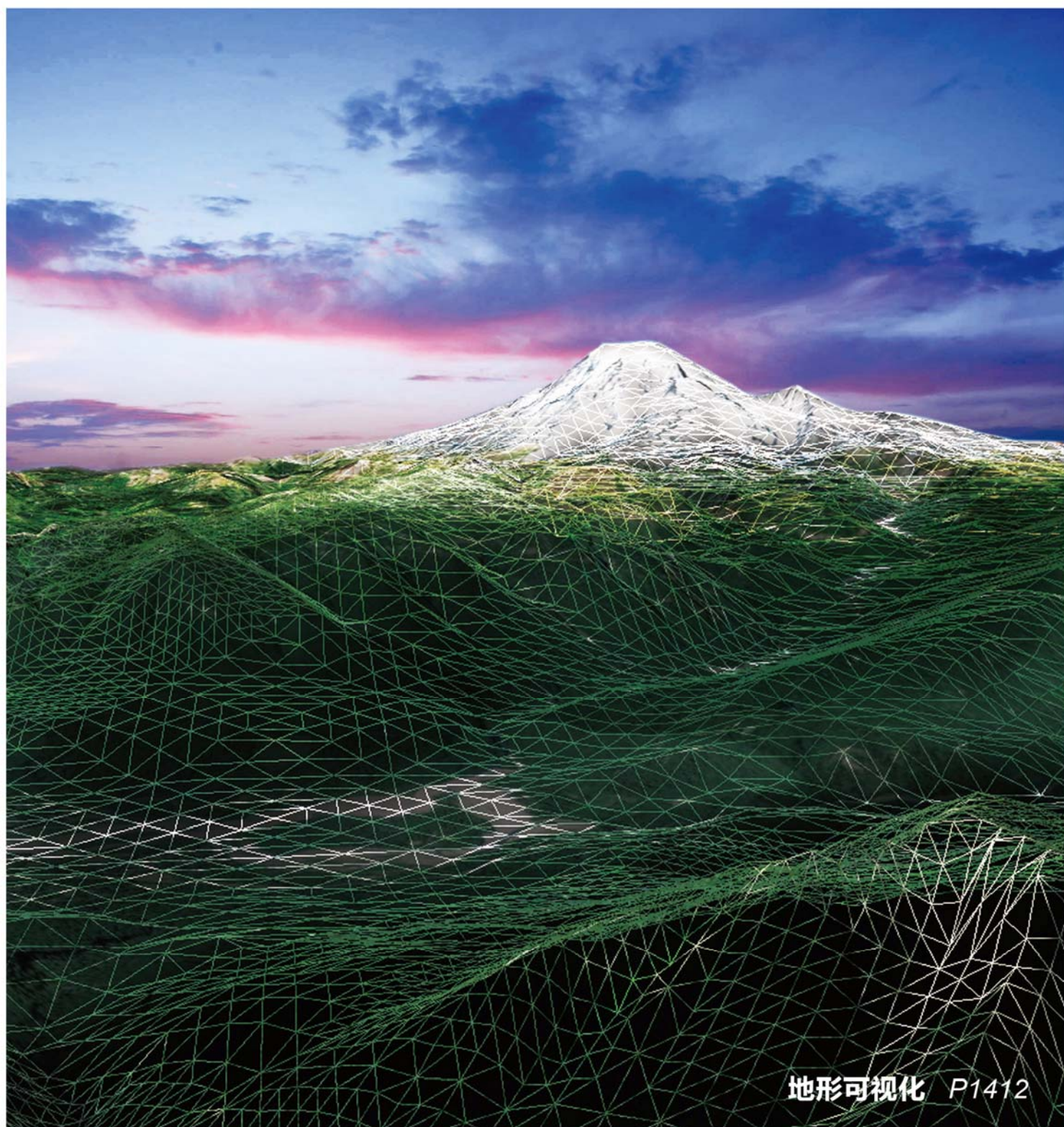


主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
10
VOL.20

ISSN1006-8961
CN11-3758/TB



地形可视化 P1412



第20卷第10期（总第234期）

2015年10月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学会
北京应用物理与计算数学研究所

主 编 李小文
编辑出版 《中国图象图形学报》编辑出版委员会
邮政信箱 北京9718信箱
邮 编 100101
电子信箱 jig@radi.ac.cn
电 话 010-64807995
网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen
Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics
P.O.Box 9718, Beijing, P.R.China
Zip code 100101
E-mail jig@radi.ac.cn
Telephone 010-64807995
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

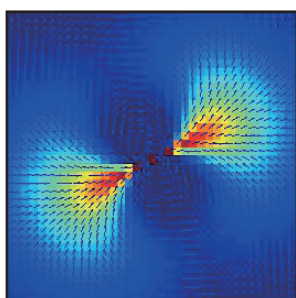
CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ
国外发行代号 M1406
国内邮发代号 82-831
国内定价 50.00元



JPEG图像双重压缩偏移量估计的篡改区域自动检测定位(第1304页)



基于表情传输的交互式照片编辑(第1366页)



结合张量投票和Snakes模型的SAR图像道路提取(第1403页)

图像处理和编码

基于模式特征的H.264/AVC可逆视频水印

李淑芝, 张翔, 邓小鸿, 吴晓燕1285

二阶总广义变分小波修复模型

许建楼, 郝岩, 张冀1297

JPEG图像双重压缩偏移量估计的篡改区域自动检测定位

赵洁, 郭继昌, 张艳, 张众维1304

图像分析和识别

基于非对称局部梯度编码的人脸表情识别

胡敏, 程轶红, 王晓华, 任福继, 许良凤, 黄晓音1313

采用积分图块间距离检测图像边缘

贾迪, 孟琢, 孙劲光, 李思慧, 赵明远1322

自适应属性加权2维FCM分割算法

侯晓凡, 吴成茂1331

利用层间相关性的岩心CT图像半自动分割方法

徐永进, 滕奇志, 吴晓红, 卿粲波1340

图像理解和计算机视觉

在线特征选取的多示例学习目标跟踪

周志宇, 彭小龙, 吴迪冲, 朱泽飞1346

融合外观和运动特征的在线目标分割

张雷, 李成龙, 汤进, 高思晗1358

基于RGB-D深度相机的室内场景重建

梅峰, 刘京, 李淳芃, 王兆其1366

计算机图形学

带平台伸缩函数的参数曲线变形

张莉, 余慧芳, 檀结庆1374

视景建模中树木纹理图像的随机变形网格方法

施冠羽, 欧阳清1384

基于表情传输的交互式照片编辑

刘锦云, 彭宏京1390

遥感图像处理

结合张量投票和Snakes模型的SAR图像道路提取

符喜优, 张风丽, 王国军, 邵芸1403

GPU Tessellation全球地形可视化方法

李尚林, 郑利平, 张迎凯, 李琳1412

CONTENTS

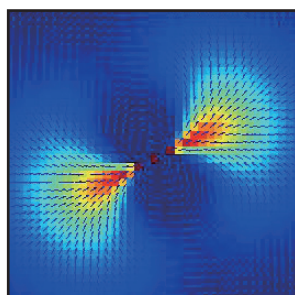
JOURNAL OF IMAGE AND GRAPHICS



Automatic detection and localization of image forgery regions based on offset estimation of double JPEG compression (P1304)



Interactive photo editing based on expression transmission (P1366)



Road extraction from SAR images using tensor voting and Snakes model(P1403)

Image Processing and Coding

- Reversible video watermarking algorithm for H.264/AVC based on mode feature
Li Shuzhi, Zhang Xiang, Deng Xiaohong, Wu Xiaoyan 1285
- Second order total generalized variational model for wavelet inpainting
Xu Jianlou, Hao Yan, Zhang Ji 1297
- Automatic detection and localization of image forgery regions based on offset estimation of double JPEG compression
Zhao Jie, Guo Jichang, Zhang Yan, Zhang Zhongwei 1304

Image Analysis and Recognition

- Facial expression recognition based on asymmetric region local gradient coding
Hu Min, Cheng Yihong, Wang Xiaohua, Ren Fuji, Xu Liangfeng, Huang Xiaoyin 1313
- Edge detection method of block distance combining with summed area table
Jia Di, Meng Lu, Sun Jinguang, Li Sihui, Zhao Mingyuan 1322
- Adaptive weighted two-dimensional histogram FCM segmentation algorithm
Hou Xiaofan, Wu Chengmao 1331
- Semi-segmentation of rock CT images using the correlation of adjacent frames
Xu Yongjin, Teng Qizhi, Wu Xiaohong, Qing Linbo 1340

Image Understanding and Computer Vision

- Object tracking based on multiple instance learning with online feature selection
Zhou Zhiyu, Peng Xiaolong, Wu Dichong, Zhu Zefei 1346
- Online object segmentation via fusing appearance and motion features
Zhang Lei, Li Chenglong, Tang Jin, Gao Sihan 1358
- Improved RGB-D camera based indoor scene reconstruction
Mei Feng, Liu Jing, Li Chunpeng, Wang Zhaoqi 1366

Computer Graphics

- Deformation of parametric curves based on platform extension function
Zhang Li, Yu Huifang, Tan Jieqing 1374
- Method of random transmogrification mesh for tree texture images in scene modeling
Shi Guanyu, Ouyang Qing 1384
- Interactive photo editing based on expression transmission
Liu Jinyun, Peng Hongjin 1390

Remote Sensing Image Processing

- Road extraction from SAR images using tensor voting and Snakes model
Fu Xiyu, Zhang Fengli, Wang Guojun, Shao Yun 1403
- GPU Tessellation method of global terrain visualization
Li Shanglin, Zheng Liping, Zhang Yingkai, Li Lin 1412

中图法分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2015)10-1366-08

论文引用格式: Mei F, Liu J, Li C P, Wang Z Q. Improved RGB-D camera based indoor scene reconstruction[J]. Journal of Image and Graphics, 2015, 20(10): 1366-1373. [梅峰, 刘京, 李淳芃, 王兆其. 基于 RGB-D 深度相机的室内场景重建[J]. 中国图象图形学报, 2015, 20(10): 1366-1373.] [DOI:10.11834/jig.20151010]

基于 RGB-D 深度相机的室内场景重建

梅峰^{1,2}, 刘京^{1,2}, 李淳芃¹, 王兆其¹

1. 移动计算与新型终端北京市重点实验室, 中国科学院计算技术研究所, 北京 100190; 2. 中国科学院大学, 北京 100049

摘要: 目的 重建包含真实纹理的彩色场景 3 维模型是计算机视觉领域重要的研究课题之一, 由于室内场景复杂、采样图像序列长且运动无规则, 现有的 3 维重建算法存在重建尺度受限、局部细节重建效果差的等问题。方法 以 RGBD-SLAM 算法为基础并提出了两方面的改进, 一是将深度图中的平面信息加入帧间配准算法, 提高了帧间配准算法的鲁棒性与精度; 二是在截断符号距离函数(TSDF)体重建过程中, 提出了一种指数权重函数, 相比普通的权重函数能更好地减少相机深度畸变对重建的影响。结果 本文方法在相机姿态估计中带来了比 RGBD-SLAM 方法更好的结果, 平均绝对路径误差减少 1.3 cm, 能取得更好的重建效果。结论 本文方法有效地提高了相机姿态估计精度, 可以应用于室内场景重建中。

关键词: RGB-D 深度相机; 同时定位与地图构建; 相机姿态估计; 3 维重建

Improved RGB-D camera based indoor scene reconstruction

Mei Feng^{1,2}, Liu Jing^{1,2}, Li Chunpeng¹, Wang Zhaoqi¹

1. Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** Three-dimensional reconstruction containing texture information is a classical issue in computer vision. Considering the complexity of an indoor scene and the length of sampling image sequence captured from a random moving RGB-D sensor, conventional three-dimensional reconstruction methods suffer from limited scale and perform poor local detail reconstruction effect. **Method** This paper proposes two improvements of the RGBD-SLAM-based three-dimensional reconstruction algorithm to obtain higher quality reconstruction effect. On the one hand, the plane-primitives are incorporated as constraints to enhance robustness and accuracy of the pair-wise registration algorithm. On the other hand, to reduce the influence of RGB-D sensor large distortion, a novel exponential weight function that is motivated by a Gaussian noise model is proposed. **Result** In the experiment, the proposed method yields higher quality results compared with state-of-the-art approaches on the benchmarks dataset of the computer vision group of Stanford. Our method also achieves lower average absolute trajectory error compared with a conventional RGB-D SLAM method. **Conclusion** Experimental results demonstrate that our meth-

收稿日期: 2015-04-03; 修回日期: 2015-06-26

基金项目: 国家自然科学基金项目(61300131); 国家科技部“十一五”科技计划项目(2013BAK03B07); 国家高技术研究发展计划(2013AA013902)

第一作者简介: 梅峰(1990—), 男, 中国科学院大学计算机应用专业硕士研究生, 主要研究方向为虚拟现实、室内场景重建。E-mail: meifeng@ict.ac.cn

通信作者: 刘京, 博士研究生, E-mail: liujing01@ict.ac.cn

Supported by: National Natural Science Foundation of China(61300131); National High Technology Research and Development Program of China(2013AA013902)

od substantially increases the accuracy of camera pose estimation and quality of indoor scene three-dimensional reconstruction.

Key words: RGB-D sensor; simultaneous localization and mapping (SLAM); camera trajectory estimation; 3D reconstruction

0 引言

作为虚拟现实研究的一个重要领域,室内环境的虚拟实现近年来引起了研究者的浓厚兴趣。室内环境的虚拟漫游有着很强的应用背景,比如,虚拟的办公室、博物馆、洞窟等已应用于远程办公、教育、娱乐和旅游等行业。基于激光扫描方法的室内场景3维建模虽然能够获得理想的建模效果,但是系统造价昂贵。RGB-D相机是一种可以同时获得颜色信息和深度信息的传感器。RGB-D相机作为新近涌现的硬件设备,诸如 Microsoft Kinect^[1]、Asus Xtion等,同传统3维扫描设备相比,具有价格低廉、结构小巧、使用方便,且能实时捕获物体表面的深度与彩色信息等优点。RGB-D相机为计算机视觉、计算机图形学等领域的相关研究提供了一种新的手段,在3维重建领域受到了广泛的关注。

由于 RGB-D 相机视角与分辨率的局限性,一帧深度图通常只能重建场景的部分内容,利用 RGB-D 相机进行室内场景重建,一个重要的问题是如何估计相机在扫描过程中的运动姿态,将相机每一帧输入配准到统一坐标系下。同时,重建模型的表示方式也对建模精度与重建效果有着深刻的影响。

2011年8月,微软在 SIGGRAPH 大会上展示了其 KinectFusion^[2]项目成果。其中针对 Kinect 高频输出特点,基于文献[3]提出的体重构技术和点到平面的迭代最近点算法(ICP),采用截断符号距离函数(TSDF)和 GPU 进行并行加速,实现实时精细3维建模,精度可达到毫米级。基于 KinectFusion 提出的方法,Bylow 等人^[4]提出了两种 ICP 优化距离函数,并对比了两种距离函数对相机姿态跟踪结果的影响。Zhou 等人^[5]通过将点云分成片段,利用非刚体配准得到较好的建模效果。Matthias 等人^[6],通过在 Kinect 上加入惯性测量装置,提高了 ICP 算法的鲁棒性,增强了建模效果。基于 KinectFusion 的方法采用 TSDF 体元表示场景模型,建模精度可达到毫米级,但方法中没有利用颜色信息,生成的模型没有纹理信息,同时由于没有环闭合检测与优化,

随着扫描距离增大,重建模型会产生累积误差。

同时定位与地图创建(SLAM)通过运动过程中观测到的地图特征定位自身位置和姿态,再根据自身位置增量式地构建地图,在机器人学中有很长的历史^[7-8]。新型的 RGB-D 相机如 Kinect,能同时提供彩色和深度图像,利用这种设备 SLAM 方法也得到了新的发展^[9-13]。Henry 等人^[9]首次将 RGB-D 数据应用到基于图的 SLAM 方法中,将视觉特征与 ICP 方法相结合,提高了相机姿态配准的精度。Endres 等人^[10]提出了一个类似的系统并在公共数据集^[14]上进行了检验。

以 RGBD-SLAM 算法为基础提出了两方面的改进。针对相机姿态估计问题,通过将深度图中的平面信息引入相机帧间配准算法,提高了帧间配准算法的鲁棒性;在重建模型表示上,使用截断符号距离函数(TSDF)体重建方法,并针对 Kinect 设备深度数据有畸变问题,提出一种基于深度畸变的权重方案,有效提高场景重建精度。

1 算法概述

利用 RGB-D 相机对室内场景进行3维重建。如图1(a)(b)所示,是 Kinect 设备输出的彩色图像和对应的深度图像,图1(c)是利用本文算法融合彩色与深度信息,重建的场景3维模型。

本文算法流程如图2所示,使用基于 RGB-D SLAM 方法解决相机姿态估计问题,在帧间配准算法中,引入平面信息,增加配准算法的鲁棒性与精度;利用估计的相机姿态,将各帧深度图配准到同一坐标系,接着使用基于 TSDF 的体方法,将配准后的深度图融合为高精度的场景模型。具体来说,本文算法包括以下几个阶段:

1)数据预处理阶段。针对 RGB-D 相机的输出的彩色图像提取加速稳健特征(SURF)^[15]特征点;对于深度图像,首先使用双边滤波算法去除噪声,使用文献[2]中的方法,将深度图投影为空间点云,然后采用随机抽样一致性(RANSAC)方法^[16]对空间点云进行平面提取。

2)相机姿态估计阶段。该阶段的目的是估计

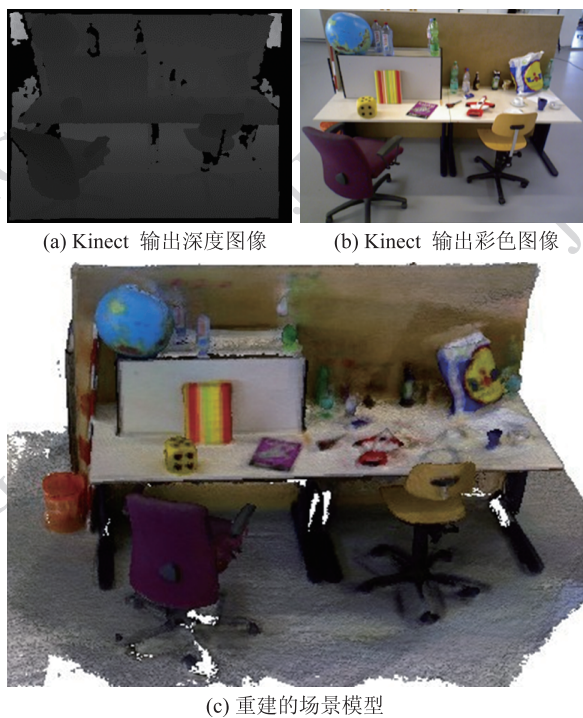


图1 传感器数据与场景重建效果

Fig. 1 RGB-D sensor output data and reconstruction result

相机每一帧在全局坐标系下的位置。相机姿态估计

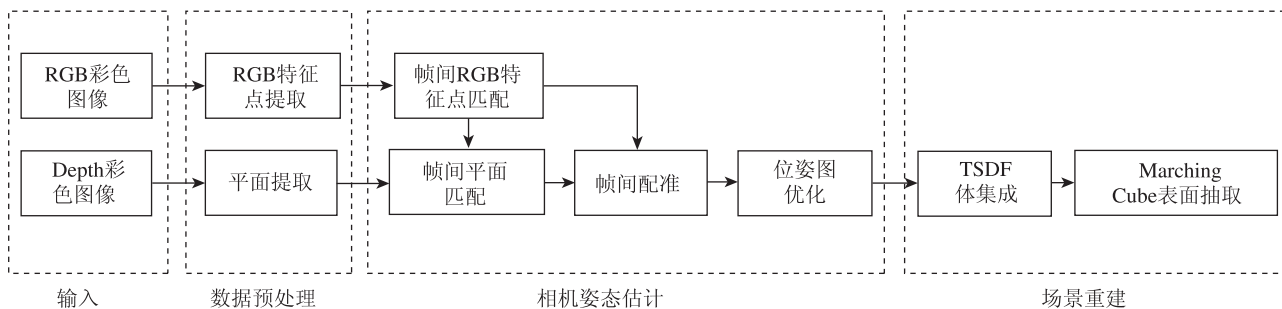


图2 算法流程图

Fig. 2 Overview of algorithm

2 基于平面基元的帧间配准算法

在相机姿态估计中,对于时刻 i, j 的两帧输入,帧间配准算法的目标是计算两帧间旋转平移关系,用 6DOF 的刚体矩阵表示为

$$T_{i,j} = \begin{bmatrix} R_{i,j} & t_{i,j} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (1)$$

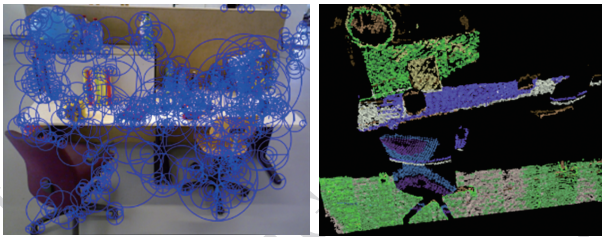
式中, $R_{i,j}$ 为 3 维矩阵,表示旋转量, $t_{i,j} \in \mathbf{R}^3$ 表示平移量。针对两帧输入的彩色图像,使用 FLANN 方法^[19]对 SURF 特征描述子进行特征点匹配。假定深度图像已经与彩色图像对齐,即彩色图像上的点

可以分为位姿图构建与位姿图优化两个部分。在图构建阶段,对于每帧输入,将其与前 20 帧进行配准,若该帧能配准到之前任意一帧,则在位姿图中加入该帧,否则丢弃。实验^[12]显示,当连续帧之间的位移增大时,帧间配准方法的效果随之下降。本文算法将深度图中抽取的平面信息加入帧间配准,由于平面不容易受到测量误差的影响,同时只需要少数的平面匹配对,因此可以应对大的基线距离,提高帧间配准算法的鲁棒性与精度。在图优化阶段,使用 g2o^[17] 框架来求解优化后相机姿态, g2o 框架是一个可扩展的图优化器,可以用来优化以图表示的非线性误差方程,被广泛用于各种 SLAM 与光束平差法 (bundle adjustment) 问题求解当中。

3) 场景重建阶段。求解完相机姿态后,利用 TSDF 体方法对深度图进行融合。如 Curless 等人^[18],本文使用米级分辨率的体素网格来表示 3 维空间,每个网格存储两个值,截断距离与权重。当融合所有输入图像到 3 维体素网格后,使用移动立方体 (marching cubes) 算法进行等值面提取,提取出模型表面。

与深度图像相同位置的点是一一对应的,使用 Newcombe 等人^[20]的方法将特征点反投影到 3 维空间, $p_k \in \mathbf{R}^3, p_k = [x, y, z]^T$ 为 3 维空间中的点。对两帧彩色图像进行特征点匹配后,随机选取 3 对匹配点 (用于计算旋转平移矩阵的最小数量) 估计旋转平移矩阵,将此变换应用到所有匹配的特征点对并计算内点数。在本文中,当匹配点对变换后的欧氏距离小于 3 cm 时,被认为是内点。随后,使用这些内点计算改进的旋转平移矩阵。以上步骤被迭代地进行,拥有最多内点的旋转平移矩阵被最终使用。设 $p_{i,k}$ 与 $p_{j,k}$ 为对应的匹配点对,其中 $k = 1, \dots, M, M$ 为匹配点对的数量。

对于输入的深度图,使用文献[2]中的方法构造顶点图与法向图,然后使用文献[13]中的RANSAC方法对空间点云进行平面抽取。平面被表示为 $\pi = [n_{\pi}^T, d]^T$,其中 $n_{\pi}^T = [n_x, n_y, n_z]^T$ 为平面的法向量, d 为相机坐标系原点到平面的距离。通过寻找平面内的特征点进行平面匹配,当两平面内同时拥有的匹配点对数超过阈值时(本文使用5),则认为这两个平面匹配。当平面没有足够匹配特征点则不认为两平面匹配,平面是作为额外的约束项优化估计的旋转平移矩阵,缺少平面约束项,本文算法依然可以使用特征点匹配对进行优化。得到匹配的平面对 $\pi_{i,t} = [n_{i,t}^T, d_i]^T$ 与 $\pi_{j,t} = [n_{j,t}^T, d_j]^T$,其中 $t = 1, \dots, N, N$ 为平面匹配对的数量。



(a) 彩色图中SURF特征点示意 (b) 点云中的平面基元抽取, 同种颜色的点属于同一平面

图3 特征点提取与平面基元提取

Fig. 3 SURF feature extraction and plane primitive extraction

针对两帧匹配的特征点对于平面对,对应点之间的点到平面距离,与对应平面对间平面法向量的距离之和为

$$E(T_{i,j}) = \sum_k \| (T_{i,j} p_{j,k} - p_{i,k})^T N(p_{i,k}) \|^2 + \sum_t \| T_{i,j} n_{j,t} - n_{i,t} \|^2 \quad (2)$$

通过优化方程式(2)得到最佳的相对位姿变换矩阵 $T_{i,j}$,采用文献[20]中的方法,将优化问题转化为线性最小二乘问题,设

$$x = [\beta, \gamma, \alpha, t_x, t_y, t_z]^T \in \mathbf{R}^6 \quad (3)$$

$$T_{i,j} = \begin{bmatrix} 1 & \alpha & -\gamma & t_x \\ -\alpha & 1 & \beta & t_y \\ \gamma & -\beta & 1 & t_z \end{bmatrix} \quad (4)$$

通过优化式(2)得到最佳的相对位姿,采用线性化方法,将优化问题转化为一个最小二乘优化。通过计算一个线性方程组如式(5)来计算最优 x 。

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} x = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (5)$$

式中

$$A_1 = [[p_{j,k}]_{\times} | I_{3 \times 3}] N(p_{i,k})$$

$$b_1 = N(p_{i,k})^T (p_{i,k} - p_{j,k})$$

$$A_2 = \begin{bmatrix} n_{j,1} \\ \vdots \\ n_{j,N} \end{bmatrix}, b_2 = \begin{bmatrix} n_{i,1} \\ \vdots \\ n_{i,N} \end{bmatrix}$$

上述目标函数可以利用 Levenberg-Marquardt 优化算法求解其最优估计 $\hat{x}$ 。对于每帧输入,将其与前20帧进行配准,若该帧能配准到之前任意一帧,则在位姿图中加入该帧,否则丢弃。构建好位姿图后,本文使用g2o框架进行位姿图优化,得到最终的相机姿态。

3 基于 TSDF 体方法的 3 维重建

利用相机姿态求解的结果,将每一帧深度图像对齐到统一坐标系下,考虑到一帧 640×480 的深度图投影为点云后,大约产生30万个点,每次场景扫描序列多达上千帧,直接使用点云表示场景不仅数据量巨大,存在大量的数据冗余,而且由于深度畸变的影响,存在大量的噪点。本文使用TSDF体方法融合各帧的深度图像。如文献[21],本文使用分辨率为 m 的体素网格来表示3维空间,即空间每一维度都被划分为 m 块,每个网格存储两个值,截断距离与权重,距离函数 F 的定义如文献[20],即

$$F: [0, \dots, m-1]^3, \mathbf{R}^3 \rightarrow \mathbf{R} \quad (6)$$

$$W: [0, \dots, m-1]^3, \mathbf{R}^3 \rightarrow \mathbf{R} \quad (7)$$

F 为网格点处的截断距离, W 表示网格处的权重,截断距离函数是该栅格到物体模型表面的距离,同时使用正负来表示该栅格是在表面被遮挡一侧还是可见一侧,而过零点就是表面上的点,如图4中所示,为2维平面内示意图。

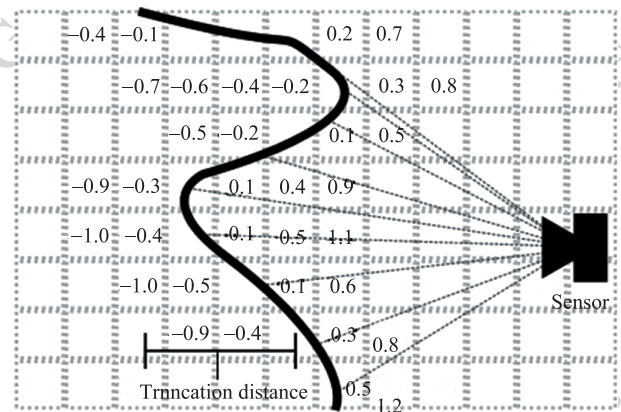


图4 TSDF在2维空间示意图

Fig. 4 TSDF demonstration on 2D

当有新的数据需要加入模型时,进行融合处理

$$F_{i+1} = \frac{W_i F_i + w_{i+1} f_{i+1}}{W_i + w_{i+1}} \quad (8)$$

$$W_{i+1} = W_i + w_{i+1} \quad (9)$$

式中, f_{i+1} 表示输入帧对应网格的截断距离, w_{i+1} 表示对应截断距离的权重。据研究文献[22]显示,深度在 1~2 m 范围内的误差主要集中在 10 mm 上下的范围;在 2~3 m 范围的随机跳变误差主要集中在 10~20 mm 以内,但是零星的大误差开始出现;当距离超过 3 m 时随机跳变误差已经超过 30 mm,开始出现噪点。在融合过程中使用权重函数可以减少深度畸变对建模效果的影响。因此,我们提出了一种新的基于深度噪声模型的指数权重函数,并对比了其他两种不同权重函数对建模效果的影响。3 种权重函数示意图如图 5 所示。

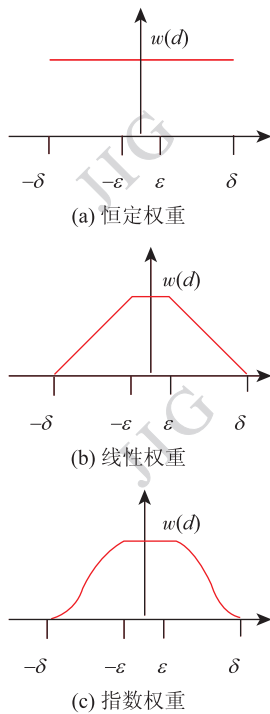


图 5 3 种权重函数图像

Fig. 5 Three different weighting functions

恒定权重函数,对所有的输入给定相同的权重

$$w_{\text{const}}(d) = 1 \quad (10)$$

式中, d 为输入的距离函数值。线性权重函数,由文献[20]提出并用于 Kinect Fusion 中,即

$$w_{\text{lin}}(d) = \begin{cases} 1 & d < \epsilon \\ \frac{\delta - d}{\delta - \epsilon} & \epsilon \leq d \leq \delta \\ 0 & d > \delta \end{cases} \quad (11)$$

式中, δ 为截断距离, ϵ 为穿透距离,在该距离内使用相同权重。与线性权重函数相比,提出一种指数权重函数,它基于深度图像的高斯噪声模型,即

$$w_{\text{exp}}(d) = \begin{cases} 1 & d < \epsilon \\ e^{-\delta(d-\epsilon)^2} & \epsilon \leq d \leq \delta \\ 0 & d > \delta \end{cases} \quad (12)$$

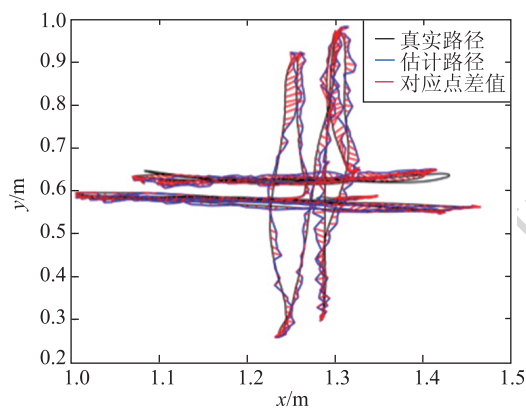
在实验中,这种权重模型将带来更好的建模效果。使用 TSDF 体方法融合每帧深度数据后,再使用移动立方体(marching cubes)算法进行等值面提取,提取出模型表面。在进行深度数据融合的同时,使用文献[23]中的方法,增加一个额外的 3 维网格存储网格单元的 RGB 值,这为等值面提取后的模型提供纹理数据,最终生成带有纹理的场景 3 维模型。

4 实验

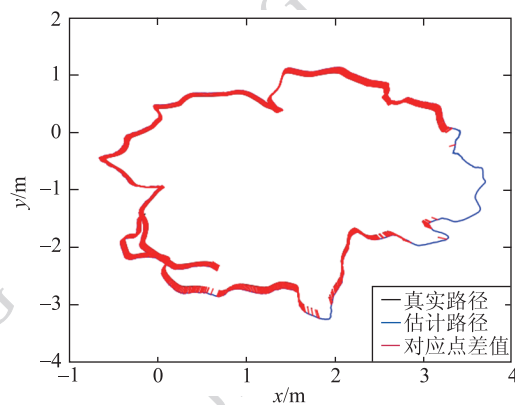
本文使用文献[17]中被广泛使用的数据集来验证帧间配准算法的准确性。该数据集在使用 RGB-D 相机进行场景扫描的同时,使用高精度运动捕捉系统提供了与相机同步的真实相机姿态数据。同时,该数据集还提供了两种度量相机姿态估计误差标准,即相对相机姿态误差(RPE)与绝对路径误差(ATE)。

采用 Handheld SLAM 数据子集,该数据集包含 11 个典型的室内视频序列,其中两个“fr1/xyz”、“fr1/rpy”包含较简单的运动,其他序列包括大规模室内场景与不规则的相机运动。图 6 显示了在不同序列下相机估计姿态与真实值之间的 ATE 误差,图像表示相机姿态在 xy 空间的投影结果,黑色线条表示相机姿态真实值,蓝色线条表示相机姿态估计值,红色线段表示相同时刻真实值与估计值之间的误差。两幅图像相机姿态估计值与真实值之间的 ATE 误差均方根分别为 0.013 m 与 0.052 m,可以看到相机估计姿态与真实姿态非常接近。

同时,本文利用 Freiburg1 (FR1) 数据集将本文的相机姿态估计方法与 RGB-D SLAM 方法进行对比,图 7 显示了本文方法与 RGBD SLAM 方法在“fr3/long_office_household”序列在 ATE 误差度量下的对比结果,其中本文方法的 ATE 均方根误差为 0.038 m 而 RGB-D SLAM 方法为 0.047 m,可以看到,本文方法相机估计姿态更接近真实值。其他序列上的对比结果如表 1,从表中可以看出,在数据集



(a) fr1/xyz



(b) fr2/desk

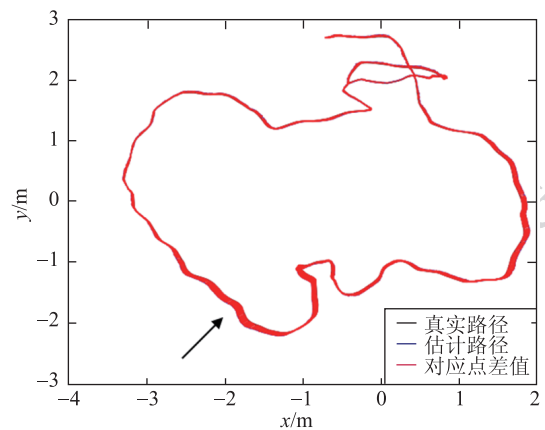
图6 “fr1/xyz”与“fr2/desk”序列绝对路径误差

Fig. 6 Visualization of the absolute trajectory error (ATE) on “fr1/xyz” and “fr2/desk”

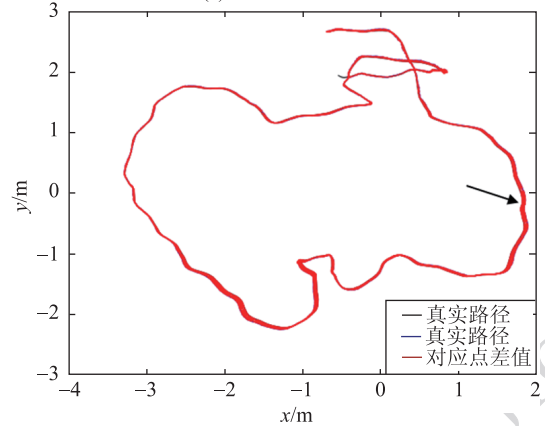
“fr1/xyz”序列中取得了最佳的 0.013 m ATE 误差与 0.020 m RPE 误差。在数据序列“fr1/desk”中,得到的最差 0.043 m ATE 误差与 0.075 m RPE 误差。总体来说,ATE 误差指标与 RPE 误差标准下,本文方法的平均误差较 RGB-D SLAM 小,取得更好的相机姿态估计结果。

利用 Stanford 数据集^[5],对比了本文场景建模效果与文献[5]中的重建效果,如图 8。实验结果显示本文方法能有效地对室内场景进行重建,同时本文方法还提供了有纹理的场景模型。

为了验证不同权重函数模型对建模效果的影响,使用“fr3/long_office_household”序列在真实相机姿态下,利用不同权重函数进行建模,实验结果如图 9 所示。实验结果表明,线性权重函数相比恒定权重函数减少了重建过程中的噪点,但线性权重与常权重函数在凳子腿与桌面部分建模效果差,这些地方距离相机较远,深度畸变较大,采用指数权重函数有效地减少了深度畸变带来的影响。



(a) RGB-D SLAM方法



(b) 本文方法

图7 相对路径误差对比

Fig. 7 Compare with RGB-D SLAM method on “fr3/long office”

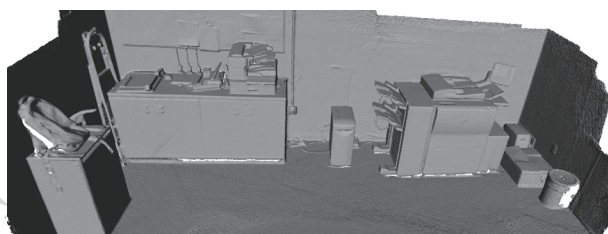
表1 对比 RGB-D SLAM 方法,在不同数据集上 ATE 与 RPE 误差

Table 1 Comparison of our approach with RGB-D SLAM in Absolute Trajectory Error (ATE) and Relative Pose Error (RPE) /m

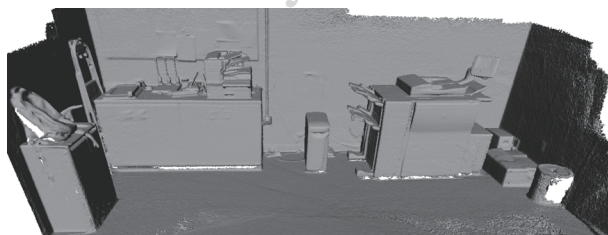
数据集	ATE		RPE	
	本文	RGB-D SLAM	本文	RGB-D SLAM
fr1/360	0.069	0.079	0.110	0.103
fr1/desk2	0.049	0.043	0.090	0.102
fr1/desk	0.043	0.023	0.075	0.049
fr1/plant	0.026	0.091	0.044	0.142
fr1/room	0.069	0.084	0.139	0.219
fr1/rpy	0.027	0.026	0.040	0.042
fr1/teddy	0.039	0.076	0.073	0.138
fr1/xyz	0.013	0.014	0.020	0.021
fr2/desk	0.052	0.095	0.099	0.143
fr2/rpy	0.024	0.019	0.034	0.026
fr2/xyz	0.020	0.026	0.030	0.037
平均	0.039	0.052	0.069	0.093



(a) 本文方法带纹理重建效果



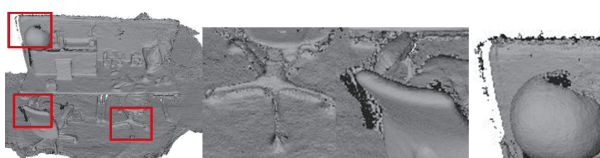
(b) 文献[5]方法重建效果



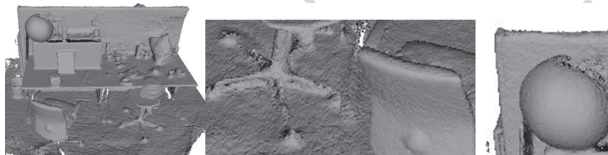
(c) 本文方法重建效果

图8 建模效果对比

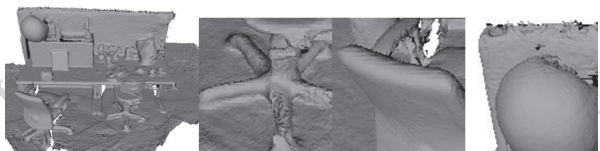
Fig. 8 Reconstruction result comparison with reference[5]



(a) 恒定权重函数



(b) 线性权重函数



(c) 指数权重函数

图9 3种权重函数的重建结果

Fig. 9 Reconstruction on different weighting functions

5 结论

本文提出了一种改进的 RGB-D 相机室内场景重建方法。在室内场景重建过程中,针对相机姿态

估计问题,将深度图中的平面信息加入帧间配准算法中,提高算法的鲁棒性。同时针对 RGB-D 相机的深度数据畸变,提出了一种指数权重函数,减小深度畸变对重建模型的影响。使用 Computer Vision Group 提供的数据集进行了验证。实验结果表明,对于大多数数据集序列,本文方法相机估计姿态与真实值之间的 RME 误差只有 4 cm 左右;同时,对比了 RGB-D SLAM 方法,实验结果显示,本文方法在大多数数据集上取得了比 RGB-D SLAM 方法更好的结果,平均 ATE 误差较 RGB-D SLAM 算法小 1.3 cm。

在未来的工作中,将继续研究相机姿态跟踪中的环比和检测问题;同时,在相机位姿图优化中,增加子图优化与图剪枝等技术。

参考文献 (References)

- [1] Zhang Z. Microsoft kinect sensor and its effect[J]. MultiMedia, IEEE, 2012, 19(2): 4-10.
- [2] Izadi S, Kim D, Hilliges O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera [C]// Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. NY, USA: ACM, 2011:559-568.
- [3] Endres F, Hess J, Engelhard N, et al. An evaluation of the RGB-D SLAM system[J]. IEEE International Conference on Robotics and Automation, 2012, 162(4):1691-1696.
- [4] Bylow E, Sturm J, Kerl C, et al. Real-time camera tracking and 3d reconstruction using signed distance functions[C]//Robotics: Science and Systems (RSS) Conference 2013. Berlin, Germany: Robotics: Science and Systems,2013:309-313.
- [5] Zhou Q, Miller S, Koltun V. Elastic fragments for dense scene reconstruction [C]//2013 IEEE International Conference on Computer Vision. Sydney:IEEE, 2013:473-480.
- [6] Nießner M, Dai A, Fisher M. Combining inertial navigation and icp for real-time 3d surface reconstruction [J]. Eurographics 2014-Short Papers, 2014,10(3): 13-16.
- [7] Thrun S. Robotic mapping: a survey[J]. Exploring Artificial Intelligence in the New Millennium, 2002,5(11): 1-35.
- [8] Nuchter A, Lingemann K, Hertzberg J, et al. 6D SLAM with approximate data association [J]. International Conference on Advanced Robotics. icar., 2005,3(7):242-249.
- [9] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: using depth cameras for dense 3D modeling of indoor environments [C]//Experimental Robotics. Berlin, Heidelberg: Springer, 2014: 477-491.
- [10] Endres F, Hess J, Engelhard N, et al. An evaluation of the

- RGB-D SLAM system[J]. IEEE International Conference on Robotics and Automation, 2012, 162(4):1691-1696.
- [11] Kerl C, Sturm J, Cremers D. Robust odometry estimation for RGB-D cameras [C]//IEEE International Conference Robotics and Automation. Germany, Karlsruhe:IEEE, 2013: 3748-3754.
- [12] Steinbrucker F, Sturm J, Cremers D. Real-time visual odometry from dense RGB-D images [C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Colorado Springs, USA: IEEE, 2011: 719-722.
- [13] Taguchi Y, Jian Y D, Ramalingam S, et al. Point-plane SLAM for hand-held 3D sensors [C]//proceedings of IEEE International Conference Robotics and Automation, Germany, Karlsruhe: IEEE, 2013: 5182-5189.
- [14] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C]//proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington DC IEEE, 2012: 573-580.
- [15] Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features[J]. In ECCV, 2006, 8(3):404-417.
- [16] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Readings in Computer Vision, 1987, 24(6):726-740.
- [17] Kummerle R, Grisetti G, Strasdat H, et al. G2o: a general framework for graph optimization [C]//proceedings of IEEE International Conference on Robotics and Automation. Sydney, Australian: IEEE, 2011:3607-3613.
- [18] Curless B, Levoy M. A volumetric method for building complex models from range images [C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. New York, USA:ACM, 1996: 303-312.
- [19] Muja M, Lowe D G. Fast approximate nearest neighbors with automatic algorithm configuration. [J]. VISAPP International Conference on Computer Vision Theory and Applications, 2009, 84(99):331-340.
- [20] Newcombe R A, Davison A J, Izadi S, et al. KinectFusion: real-time dense surface mapping and tracking [C]//The 10th IEEE International Symposium Mixed and Augmented Reality. Basel, Switzerland:IEEE, 2011: 127-136.
- [21] Baumgart B G. Winged edge polyhedron representation [R]. State of California Stanford Univ. Cadept of Computer Science, 1972.
- [22] Khoshelham, Kelberink S O. Accuracy and resolution of Kinect depth data for indoor mapping applications. [J]. Sensors, 2012, 12(2):1437-1454.
- [23] Whelan T, Johannsson H, Kaess M, et al. Robust real-time visual odometry for dense RGB-D mapping [C]//Proceedings of IEEE International Conference on Robotics and Automation. Germany, Karlsruhe: IEEE, 2013: 5724-5731.