

基于深度相机的大场景三维重建

刘东生, 陈建林, 费点, 张之江

引用本文:

刘东生, 陈建林, 费点, 等. 基于深度相机的大场景三维重建[J]. *光学精密工程*, 2020, 28(1): 234–243.

LIU Dong-sheng, CHEN Jian-lin, FEI Dian, et al. Three-dimensional reconstruction of large-scale scene based on depth camera[J]. *Optics and Precision Engineering*, 2020, 28(1): 234–243.

在线阅读 View online: <https://doi.org/10.3788/OPE.20202801.0234>

您可能感兴趣的其他文章

Articles you may be interested in

移动高光对三维场景重建的影响与消除

Influence and elimination of moving highlight on three-dimensional scene reconstruction

光学精密工程. 2017, 25(7): 1900–1908 <https://doi.org/10.3788/OPE.20172507.1900>

融合通用形变模型信息的面部三维重建

Face reconstruction fused with generic morphable model

光学精密工程. 2018, 26(1): 184–192 <https://doi.org/10.3788/OPE.20182601.0184>

基于关系图的无人机影像三维重建

On diagram-based three-dimensional reconstruction of UAV image

光学精密工程. 2016, 24(6): 1501–1509 <https://doi.org/10.3788/OPE.20162406.1501>

用于弱纹理场景三维重建的机器人视觉系统

Robot vision system for 3D reconstruction in low texture environment

光学精密工程. 2015, 23(2): 540–549 <https://doi.org/10.3788/OPE.20152302.0540>

采用混合回环检测与闭环优化的TSDF地图创建

TSDF map building based on hybrid loopback detection and closed-loop optimization

光学精密工程. 2018, 26(6): 1497–1506 <https://doi.org/10.3788/OPE.20182606.1497>

文章编号 1004-924X(2020)01-0234-10

基于深度相机的大场景三维重建

刘东生, 陈建林, 费 点, 张之江*

(上海大学 特种光纤与光接入网重点实验室, 上海 200444)

摘要:针对大场景三维重建中,由位姿估计的累积误差而导致的相机漂移和重建模型质量低的问题,提出了减少累积误差的方法。首先,基于由最新 K 对深度和彩色图像融合模型,最小化输入 RGB-D 图像的几何误差和亮度误差来跟踪相机。然后,若相机位置与当前子网格的中心点距离大于给定阈值时,则将子网格平移体素单元整数倍的距离,基于新建的子网格继续跟踪相机并重建局部场景模型。最后,在子网格间以迭代步长式的方法寻找对应表面点,以对应点间的欧氏距离与亮度误差为约束,优化全局相机轨迹。基于数据集的实验结果表明,相机位姿估计精度比主流方法提升 14.1%,全局轨迹优化精度提升 8%。对于自采数据,本文设计的系统可减少位姿估计中的累积误差、重建高质量的场景模型。

关键词:计算机视觉;深度相机;累积误差;三维重建

中图分类号: TN249 **文献标识码:** A **doi:** 10.3788/OPE.20202801.0234

Three-dimensional reconstruction of large-scale scene based on depth camera

LIU Dong-sheng, CHEN Jian-lin, FEI Dian, ZHANG Zhi-jiang*

(Key Laboratory of Specialty Fiber Optics and Optical Access Networks,
Shanghai University, Shanghai 200444, China)

* Corresponding author, E-mail: zjzhang@shu.edu.cn

Abstract: Aiming at the problem of camera drift and low-quality reconstruction caused by accumulated errors in pose estimation for large-scale reconstruction, we proposed a method to reduce accumulated errors. First, based on the model fused by the latest K depth and color images, the geometric and photometric error of the input RGB-D image were minimized to track the camera. Then, if the distance between the camera position and the subvolume center was more than a given threshold, the subvolume was shifted by multiples of voxel size, the camera was continuously tracked and the local scene model was reconstructed based on the newly created subvolume. Finally, the corresponding surface points were searched by the iterative step method between subvolumes, and the global camera trajectory was optimized by Euclidean distance and photometric error between the correspondence. The experimental results based on the dataset show that the camera pose estimation accuracy is improved by 14.1% than the current method, and the global trajectory optimization accuracy is improved

收稿日期:2019-07-01;修订日期:2019-08-27.

基金项目:国家自然科学基金面上项目(No. 61572307)

by 8%. The system designed in this paper can also reduce the accumulated errors in pose estimation and reconstruct a high-quality scene model for self-collected data.

Key words: computer vision; depth camera; accumulated error; three-dimensional reconstruction

1 引言

大型场景的三维重建是图形学和计算机视觉领域的一个研究热点,被广泛应用于室内设计、机器人导航和增强现实。在表面重建系统中,需要同时精确地获取相机的运动轨迹和场景的三维模型。深度相机因其能以视频帧速直接获取物体的位置信息且价格低廉,而非常适合用来研究三维重建。

KinectFusion^[1]系统开创了使用手持式深度传感器进行实时稠密重建的新领域。它采用点到平面的迭代最近点算法(Iterative Closest Point, ICP^[2])将输入的深度图与当前模型配准来跟踪相机,实时性能由 GPU 的并行计算来实现。然而 KinectFusion 只能重建固定于空间中的小场景,并且相机位姿估计不精确。随后,研究者使用移动网格的策略来扩展 KinectFusion 算法,使其能重建大型场景^[3-5]。但对于大型场景重建,长的相机扫描序列会引入越来越大的位姿累积误差,测量的表面信息被不精确地融合到模型中,会进一步恶化位姿估计的结果,最终导致相机漂移,重建的模型质量差,甚至是重建失败。

为了解决上述问题,Fioraio 等人^[6]使用最新 K 张深度图像构建局部子网格,以帧到模型的方式执行相机跟踪。尽管该方法可极大地减少累积误差,但当场景缺少几何结构特征时,仅利用深度信息进行重建会使相机丢失跟踪。Fu 等人^[6]在跟踪相机的同时检测场景闭环,以闭环为约束来矫正相机轨迹。但检测的闭环存在错误且有部分正确的闭环并未被检测。Choi 等人的离线方法^[7-8]采用基于线流程(Line processes)的鲁棒全局位姿优化来消除检测错误的闭环以提高重建质量,但需耗费大量时间。在 Dai 等人^[9]的 Bundle-Fusion 算法中,新输入的 RGB-D 图像需要和之前所有分块的关键帧匹配,以高斯牛顿法的求解方式极小化位姿对齐误差来优化全局轨迹,可得到精确的相机位姿,但需要两张显卡执行复杂的局部和全局优化。

针对大场景重建中,由位姿估计的累积误差而导致的相机漂移、重建模型质量低的问题,本文设计了一种基于深度相机的场景重建方法。在系统中保持一个 K 长的滑动窗口,基于由最新 K 帧融合的彩色模型跟踪相机,以此来减少局部累积误差。在全局优化中,迭代步长式地在子网格间搜索表面对应点,并以对应点的点到平面距离误差和亮度误差为约束,优化全局相机轨迹。最后通过实验证明了本文方法的有效性。

2 三维重建流程

基于深度相机的静态场景重建方法都有相似的系统框架,如图 1(a)所示。

首先,对输入的 RGB-D 图像执行深度图像预处理以减少噪声和错误的深度值。其次,取决于不同的方法,从预处理过的深度图像中获取不同类型的深度信息。随后,可利用帧到帧、帧到模型的方式计算当前图像到全局坐标系的最优变换。最后,将当前深度图的所有顶点转换到全局坐标系并融合到模型中。

为了重建大型场景,本文方法的流程如图 1(b)所示。随着深度相机在静态场景中扫描,根据时间戳对齐一对深度和彩色图像。在算法启动时,将第一个子网格固定于相机正前方,子网格内的体素由第一对深度和彩色图像初始化,用于后续的相机跟踪。

在每个子网格内,结合彩色和深度测量执行帧到模型的相机跟踪。考虑到短时间内的累积误差小,且物体的颜色亮度可能在不同视角下有些许变化,本文提出使用彩色模型的去融合方案来减少累积误差。核心思想是当第 N 对图像融合到模型时,将 N-K 对图像从模型中去融合,以此来确保输入的图像始终基于最新的 K 对帧估计位姿。

当相机相对于当前子网格中心点的位置超过给定阈值时,为了高效性,将子网格平移体素单元倍数的距离。移动到子网格外面的体素会丢失,而剩余的体素会从旧的子网格复制到新的子网格以继续执行帧到模型的相机跟踪。

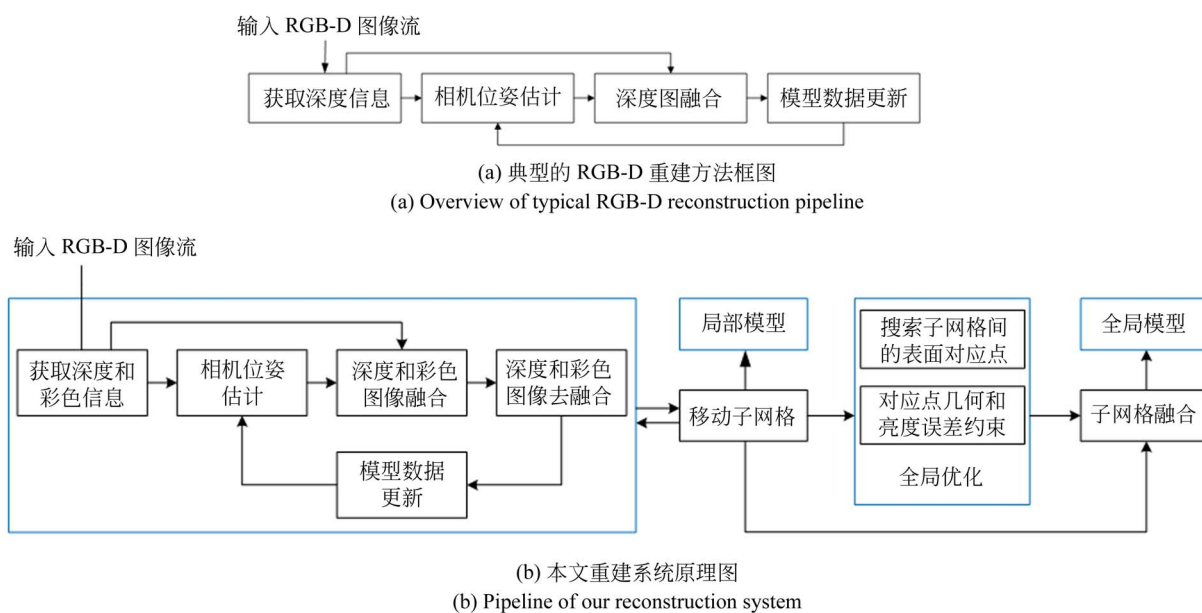


图 1 三维重建流程

Fig. 1 Pipeline of 3D reconstruction system

在全局优化中,本文提出了迭代步长式的方法在子网格间搜索对应表面点,并基于对应点的点到平面距离约束和亮度约束迭代地执行全局优化,目标函数以 sparseLM 的方式求解。相机轨迹优化后,将子网格全局配准,融合到全局网格中^[10]。随后,采用 marching cubes^[11]算法提取三角片以便可视化。

3 场景重建的算法

3.1 相机模型及符号表示

对于一个三维空间点 $P(x, y, z) \in \mathbb{R}^3$, 可以使用针孔相机模型获取其对应于二维图像平面的像素点 $X(u, v) \in \mathbb{R}^2$

$$(u, v)^T = \pi(x, y, z) = \left[\frac{u - c_x}{f_x} z, \frac{v - c_y}{f_y} z, z \right]^T, \quad (1)$$

其中 f_x, f_y 和 c_x, c_y 分别是焦距和光轴中心点坐标, π 为投影方程。

本文重建方法使用单个深度传感器, 给定 RGB-D 图像流 $\{D_k\}_{k=1}^M$, 第 k 对深度和彩色图像可表示为:

$$I_d^k: \mathbb{R}^2 \rightarrow \mathbb{R}^+ \quad I_c^k: \mathbb{R}^2 \rightarrow [0, 1]^3. \quad (2)$$

欧氏群 $SE(3)$ 中的变换矩阵 $\mathbf{T}_{i,k}$ 将第 k 帧图像的相机坐标系映射到局部子网格 $\{V_s\}_{s=1}^W$ 参考

坐标系中:

$$\mathbf{T}_{i,k} = \begin{bmatrix} \mathbf{R}_{i,k} & \mathbf{t}_{i,k} \\ \mathbf{0}^T & 1 \end{bmatrix} \in SE(3), \quad (3)$$

其中: 三维正交矩阵 $\mathbf{R}_{i,k} \in \mathbb{R}^{3 \times 3}$ 表示旋转分量, $\mathbf{t}_{i,k} \in \mathbb{R}^{3 \times 1}$ 表示平移分量。

3.2 鲁棒的相机跟踪及数据融合

本文采用文献[12]的方法估计相机位姿。如图 2 所示, 使用加权截断有向距离函数 (TSDF^[10]) 表示三维空间:

$$\psi: [\psi_{rgb}, \psi_d] \rightarrow [\mathbb{R}^3, \mathbb{R}], \quad (4)$$

其中 ψ_{rgb} 与 ψ_d 分别表示重建表面的色彩和几何信息。

给定由前 N 对深度和彩色图像融合的表面 ψ^{f-n} , 和由第 $N+1$ 对深度和彩色图像的反投影得到的表面 ψ^{s-n+1} , 其几何位置和颜色亮度应尽可能一致。误差函数可定义为:

$$E(\mathbf{T}_{i,N+1}) = \sum_{p \in \psi^{s-n+1}} (\phi_d^{f-n}(\mathbf{T}_{i,N+1} \hat{P})^2 + \theta \phi^{f-n}(\mathbf{T}_{i,N+1} \hat{P})^2), \quad (5)$$

其中: 第 1 项表示几何误差, 第 2 项表示亮度误差, 并由参数 θ 加权, \hat{P} 为 P 的齐次形式。其中:

$$\phi^{f-n}(\mathbf{T}_{i,N+1} \hat{P}) = \|\Omega(\phi_{rgb}^{f-n}(\mathbf{T}_{i,N+1} \hat{P}) - I_c^{N+1}(\pi(P)))\|, \quad (6)$$

点 P 定义于第 $N+1$ 帧的相机坐标系, Ω 为一个

三维对角矩阵,将 RGB 格式的彩色图转换为灰度图^[13]:

$$\mathbf{\Omega} = \text{diag}(\sqrt{0.299}, \sqrt{0.587}, \sqrt{0.114}). \quad (7)$$

误差函数(5)可使用 Gauss-Newton 算法迭代求解直至收敛。

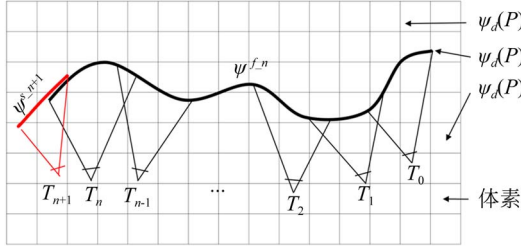


图 2 表面 ψ^{n+1} 与 $\psi^{f,n}$ 的几何位置和颜色亮度应尽可能一致

Fig. 2 Geometric position and color intensities of ψ^{n+1} and $\psi^{f,n}$ should be as consistent as possible

本文在每个体素 v 中分配 6 个值,其中 Φ 是从 v 到最近表面的 TSDF, R, G, B 是三通道颜色值, W_d 和 W_c 分别为 TSDF 和颜色值的确信度。在估计完第 $N+1$ 对图像的位姿后,可使用加权平均法^[10]更新 TSDF:

$$\Phi^{N+1} = \frac{\Phi^N W_d^N + \psi_d^{N+1} \omega_d^{N+1}}{W_d^N + \omega_d^{N+1}}. \quad (8)$$

类似地,以相同的方式更新颜色信息:

$$\Gamma^{N+1} = \frac{\Gamma^N W_c^N + \gamma^{N+1} \omega_c^{N+1}}{W_c^N + \omega_c^{N+1}} \mid \Gamma \in (R, G, B), \quad (9)$$

$$\gamma \in \{r, g, b\}.$$

3.3 基于深度和彩色信息的去融合

针对位姿估计存在累积误差的问题,本文在文献[5]的基础上,结合深度和彩色信息对模型去融合。

考虑到短时间内累积误差可忽略不计,本文基于由最新的 K 帧深度和彩色图像融合的模型跟踪相机。在当前帧融合到模型后,将 $N-K$ 帧的信息从模型中去融合,在程序中该过程对应于一个 K 长的 FIFO 队列,如图 3 所示,直到第 $N-1+C$ 帧时生成一个新的子体积。去融合过程可表示为:

$$\Phi^{N'} = \frac{\Phi^N W_d^N - \psi_d^{N-K} \omega_d^{N-K}}{W_d^N - \omega_d^{N-K}}, \quad (10)$$

$$W_d^{N'} = W_d^N - \omega_d^{N-K}. \quad (11)$$

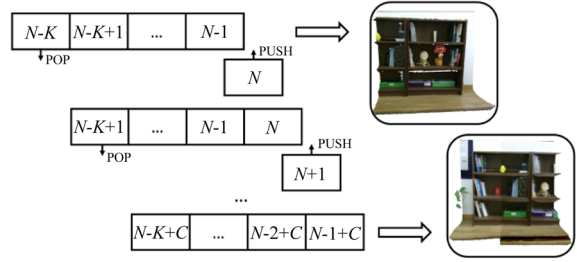


图 3 子网格的融合和去融合

Fig. 3 Fusion and de-fusion of subvolumes

对于彩色测量的累积误差,以同样的方式:

$$\Gamma^{N'} = \frac{\Gamma^N W_c^N - \gamma^{N-K} \omega_c^{N-K}}{W_c^N - \omega_c^{N-K}} \mid \Gamma \in \{R, G, B\}, \quad (12)$$

$$\gamma \in \{r, g, b\}.$$

若小于 K 帧图像被跟踪,则不对子网格去融合。一旦超过 K 帧被处理且满足子网格的移动条件,则将当前子网格保存在内存中。

3.4 子网格的移动策略

当相机的位置距离子网格的中心点超过给定阈值时,需要移动子网格来重建大型场景。

如图 4 所示(彩图见期刊电子版),相机从路标 L_1 移动到路标 L_2 。橙色线表示相机光轴,相机在 L_1 处采集到第 1 帧图像,其位姿为 $\mathbf{I}_{4 \times 4}$,并新建子网格 V_1 ,中心点为 O_1 ,如绿色实线框所示。在任意时刻 t ,第 i 帧图像和当前子网格 V_j 的绝对位姿可分别表示为 $\mathbf{T}_{g,i}$ 和 $\mathbf{P}_{g,j}$, $\mathbf{T}_{j,i}$ 为第 i 帧到局部子网格 V_j 的变换:

$$\mathbf{T}_{j,i} = \mathbf{P}_{g,j}^{-1} \mathbf{T}_{g,i} = \begin{bmatrix} \mathbf{R}_{j,i} & \mathbf{t}_{j,i} \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (13)$$

在移动子网格时,如果同时考虑平移和旋转分量,则当下列条件满足时:

$$\| \mathbf{t}_{ji} \| > Th_{old,t} \text{ 或 } a_{ji} = \|\text{rodrigues}^{-1}(\mathbf{R}_{j,i})\| > Th_{old,R}, \quad (14)$$

V_j 被移动到 V_{j+1} ,如图 4 中的蓝色虚线框所示。其中 $Th_{old,t}$ 和 $Th_{old,R}$ 分别为距离和角度阈值, $rodrigues$ 为罗德里格斯变换。此时,对于新建立子网格 V_{j+1} ,有:

$$\mathbf{T}_{j+1,i} = \mathbf{I}_{4 \times 4}, \mathbf{P}_{j+1} = \mathbf{T}_{j,i} \mathbf{T}_{j+1,i}^{-1} = \mathbf{T}_{j,i}, \quad (15)$$

V_j 中的体素需要被三线性插值到 V_{j+1} 中,会耗费大量时间。同时,由于体素在三线性插值后通常位于非整数坐标处,因此插值方法会影响重建的准确性。

本文采用图 4 所示的子网格移动方法,其中

黑色实线箭头表示 O_1 与 L_2 处光轴的距离, 交光轴于点 C 。当该距离超过给定阈值 d_{th_1} 时, 移动 V_1 :

$$d_1 > d_{th_1}. \quad (16)$$

考虑到三线插值非常耗时且会引入不精确的 tsdf, 本文将 V_1 平移体素单元倍数的距离。同时, 为了使平移到新建的子网格内用于相机跟踪的有效体素数量尽可能的多, 点 O_1 被平移到蓝色虚线框的中心点 O_2 , 生成一个新的子网格, 如图 4 红色虚线框所示。此时, L_2 处满足子网格移动条件的相机视角在新的子网格内具有很大的覆盖范围, 如图 4 黄色箭头所示。因此, 随后用于相机跟踪的实体素数量众多, 相机跟踪会更加精确鲁棒。 V_1 被平移整数体素单元的距离, 最终生成的子网格如红色实线框所示, 此时子网格之间的位姿有以下关系:

$$\mathbf{P}_{j,j+1} = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \text{vs} \left[\frac{\|t_{j,i}\|}{\text{vs}} \right] \\ \mathbf{0}^T & 1 \end{bmatrix}, \mathbf{T}_{j+1,i} = \mathbf{P}_{j,j+1}^{-1} \mathbf{T}_{j,i}, \quad (17)$$

vs 为单个体素的尺寸。

需要注意的是, 当相机沿着固定光轴向前或向后移动超过一定距离时, V_1 同样需要被平移:

$$d_2 > d_{th_{2-l}} \parallel d_2 < d_{th_{2-s}}, \quad (18)$$

其中 d_2 是点 C 到光心 L_2 的距离, $d_{th_{2-l}}$ 和 $d_{th_{2-s}}$ 为两个阈值。当式(17)或式(18)满足时, 平移 V_1 。移动到子网格外的体素将会丢失, 剩余的体素将从旧的子网格复制到新的子网格。

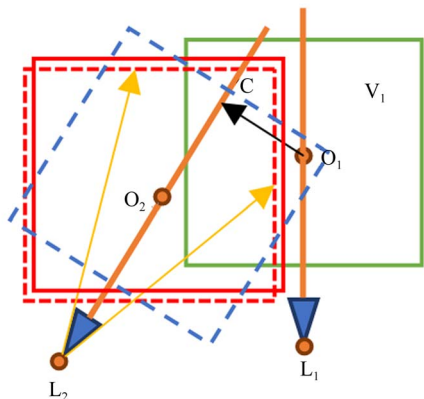


图 4 将子网格从 L_1 平移到 L_2

Fig. 4 Shifting subvolume from L_1 to L_2

3.5 全局优化

虽然相机在每个子网格内是低漂移的, 但由于长的扫描序列, 子网格之间不可避免地存在累积误差。本文基于文献[5]的全局优化框架, 提出了迭代步长式地寻找表面对应点。如图 5(a) 是未优化的子网格, 图 5(b) 是优化后的子网格。

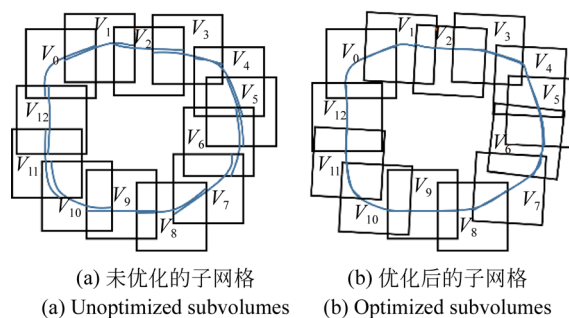


图 5 全局优化的示意图

Fig. 5 A schematic diagram of the global optimization

3.5.1 迭代步长式寻找表面对应点

如图 6, $P = \{p_i\}$ 和 $Q = \{q_i\}$ 是子网格 V_j 和 V_k 的表面点集, p_i 和 q_i 是一对对应点, $T_{g,j}$ 和 $T_{g,k}$ 为子网格 V_j 和 V_k 的全局变换。若 $T_{g,j}$ 和 $T_{g,k}$ 都被精确地估计, 则可建立以下方程:

$$q_i = T_{g,k}^{-1} T_{g,j} p_i, \quad (19)$$

为了表述简洁, 公式中默认存在点的齐次形式与非齐次形式间的转换。

通常, 由于噪声的存在, 式(19)不会相等。为了得到 $T_{g,j}$ 和 $T_{g,k}$ 的精确值, 需要在 V_j 和 V_k 中寻找更加精确的对应点。

对于任意表面点 P , 假设其 tsdf 值为 ψ_d , 可获取归一化的梯度:

$$\mathbf{n} = \hat{\nabla} \psi_d(P) = \frac{\nabla \psi_d(P)}{\|\nabla \psi_d(P)\|}. \quad (20)$$

因 tsdf 表示一点距表面的有向距离, 故 V_k 中的对应点可通过以下方式搜索:

$$q_i = T_{g,k}^{-1} T_{g,j} p_i - \alpha \psi_d(T_{g,k}^{-1} T_{g,j} p_i) \hat{\nabla} \psi_d(T_{g,k}^{-1} T_{g,j} p_i). \quad (21)$$

在式(21)中, V_j 中的表面点 p_i 被转换到 V_k 中, 并沿着转换后的点的法线方向行进 $\alpha \psi_d$ 的距离。理论上仅仅需要行进距离 ψ_d , 但在实验中发现, 由于位姿估计和 tsdf 计算过程中存在误差, 寻找到的对应点不一定是 V_k 的表面点。因此, 本文引入步长参数 α , 基于实验经验, α 由 0.8 增

长到 1.2, 增量为 $\Delta\alpha = 0.05$, 直到找到 tsdf 值足够小的点 q_i 。如果 α 增长到 1.2 之后, q_i 仍然远离表面, 则认为 p_i 没有对应点。

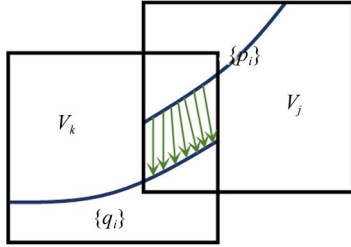


图6 V_j 和 V_k 之间寻找对应点

Fig. 6 Search corresponding points between V_j and V_k

3.5.2 迭代式的全局优化

对于一对对应点 (p_i, q_i) , 其三维空间位置和

$$\Phi'_j = \frac{\sum_{V_i | i \neq j} \{ \Phi^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) W_d^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) + \Phi^i(p) \omega_d^i(p) \}}{\sum_{V_i | i \neq j} \{ W_d^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) + W_d^i(p) \}}. \quad (24)$$

类似地, 对于彩色信息:

$$\Gamma'_j = \frac{\sum_{V_i | i \neq j} \{ \Gamma^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) W_c^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) + \Gamma^i(p) W_c^i(p) \}}{\sum_{V_i | i \neq j} \{ W_c^j(\mathbf{T}_{g,j}^{-1} \mathbf{T}_{g,i} p) + W_c^i(p) \}} \quad | \Gamma \in \{R, G, B\}. \quad (25)$$

随后使用 marching cubes^[11] 算法提取三角片以获得精细的三维场景模型。

4 实验结果及分析

本节使用两类不同的数据分别从定性和定量两个方面验证本文算法的有效性。

4.1 场景重建结果的定性展示

图7是由 Kinect v2 从实验室中采集的 450 张深度和彩色图像重建的模型。图7(a)和图7(b)分别从两个不同的视角展示了重建的结果。可以看出, 尽管场景中的物体繁杂多样, 但最终重建的结果清晰地展示了实验室内部的结构。图7(c)展示了书桌部分的局部细节图, 可以看到书桌、显示器屏幕、纸箱表面都平坦光滑, 甚至纸箱表面上贴的透明胶带都清晰可见。图7(d)展示了书架部分的局部细节图, 垃圾桶、植物、书架上摆放的书本、玩偶都实现了较高水平的还原。

此外, 本文还重建了 Augmented_ICL_UNIM^[5] 数据集中的客厅(Living Room2)部分场

景, 并与现有方法进行了比较, 验证了本文提出算法的有效性。

$$e_{i,i} = \| (\mathbf{T}_{g,j} p_i - \mathbf{T}_{g,k} q_i)^\top \mathbf{T}_{g,k} \mathbf{n} \| + \| \psi_{rgb}^V(p_i) - \psi_{rgb}^V(q_i) \|. \quad (22)$$

对于寻找的对应点集, 整体误差函数可定义为如式(23), 可使用 sparseLM 求解:

$$E(\mathbf{T}_g) = \arg \min_{(\mathbf{T}_g, v_0, \dots, \mathbf{T}_g, v_i, \dots, \mathbf{T}_g, v_w)} \sum_{V_j} \sum_{V_k} \sum_{p_i, q_i} \| e_{i,i} \|^2. \quad (23)$$

在本文中, 设置迭代次数 N_{th} , 如果相邻两次迭代的优化误差小于给定阈值或者迭代次数超过 N_{th} 时, 则停止迭代, 输出优化后的相机位姿。

3.5.3 获取全局一致的三维模型

在全局轨迹优化后, 为了得到整体的三维模型, 需将子网格融合到全局 TSDF 中:



(a) 重建模型的全览图 (b) 重建模型的全览图
(a) Overview of reconstruction (b) Overview of reconstruction



(c) 重建模型的细节图 (d) 重建模型的细节图
(c) Details of reconstruction (d) Details of reconstruction

图7 实验室的重建结果

Fig. 7 Reconstruction results of lab

景, 并与现有方法进行了比较, 验证了本文提出算法的有效性。

图8分别展示了4种方法的重建结果。图8(a)为 Bylow 等人^[14] 的重建模型, 与 KinectFusion^[1] 中通过点到平面的 ICP 算法跟踪相机不

同,Bylow 等人直接基于 TSDF 最小化深度测量误差,有效地提高了位姿估计的精确度。但仅通过单一网格持续估计位姿并融合数据,位姿误差不断累积,不准确的位姿导致融合模型出现模糊或者断裂现象,如 8(a)中台灯的灯柱和植物的枝叶所示。图 8(b)为 Fioraio 等人[5]的重建结果,为了减少累积误差,Fioraio 等人沿着相机运动轨迹新建子网格,子网格内局部场景的细节精细度有所提升,模型中台灯的灯柱和植物等细节都基本保留。但其在全局优化中寻找的对应点不够精确,导致子网格之间的相对位姿仍然存在较大误差,重建模型中的墙壁、沙发等处存在明显的

折痕。图 8(c)为 Choi 等人[8]的重建结果,该算法间隔固定帧数(实验中为 50 帧)生成一个新的子网格,但当相机运动速度较快时,单个子网格内覆盖的场景较大,包含的相机运动轨迹较长,重建模型细节处容易出现空洞,如模型中的靠枕和植物所示。图 8(d)为本文提出方法的重建结果,本文结合深度和彩色测量跟踪相机位姿,同时对深度和彩色数据进行融合和去融合操作,估计的位姿更加准确,并且子网格间采用更加精确的步长式方法寻找对应点,极大程度地减轻了折痕现象,同时保证了子网格内局部场景的细节精细度,如图中的靠枕和植物。

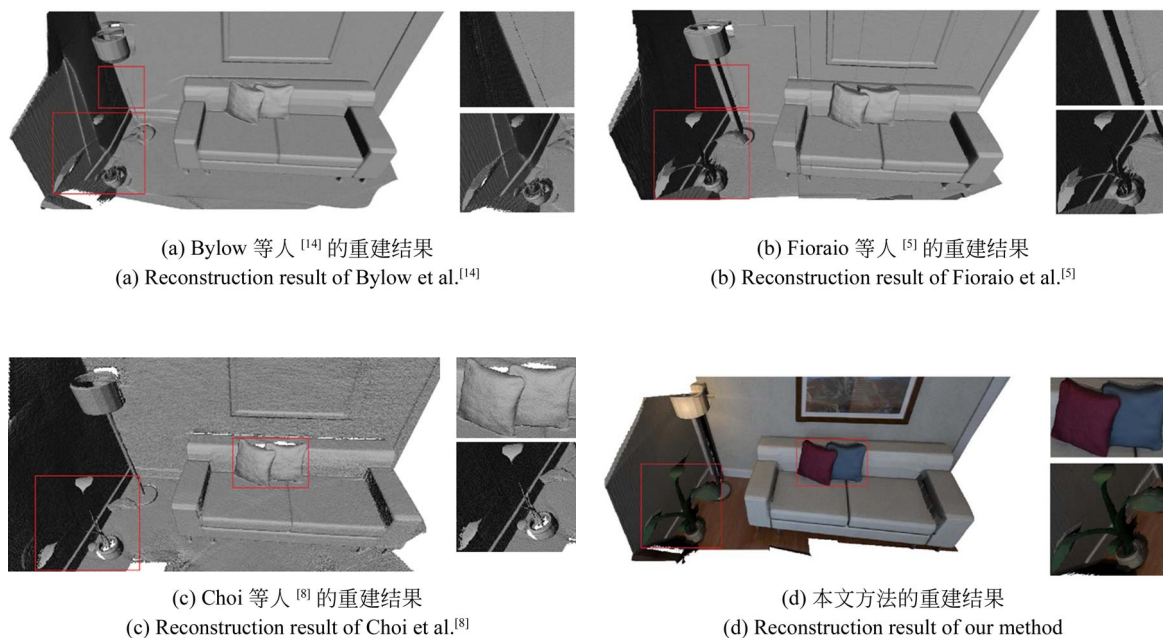


图 8 四种方法的重建结果对比

Fig. 8 Comparison of reconstruction results of four methods

通过自采数据和公开数据集的实验结果分析,本文提出的场景重建方法能够克服较长相机轨迹带来的累积误差问题,重建的模型平滑且完整,同时保留了精致的局部细节,最终生成的三维模型具有更佳的可观性。

4.2 绝对轨迹误差的定量分析

表 1 所示为四种方法对数据集 Augmented ICL-NUIM Living Room2 部分场景评估的绝对轨迹误差(ATE),其中包括均方根误差(RMSE)、平均值(Mean)和标准差(STD),单位均为米。从表中可看出,相较于现有的几种方法,本文方法估

表 1 四种不同方法估计的相机绝对轨迹误差

Tab. 1 Absolute trajectory error evaluated by four different methods (m)

方法	Bylow ^[14]	Fioraio ^[5]	Choi ^[8]	本文
RMSE	0.228 4	0.145 9	0.131 1	0.112 6
Mean	0.202 6	0.127 5	0.096 2	0.093 9
STD	0.105 4	0.080 7	0.072 3	0.061 9

计的绝对轨迹误差最低,均方根误差 RMSE 与 Choi 等人提出的算法相比低 14.1%,说明本文提出的算法在位姿估计准确率上得到了很大提升。

4.3 重建时间的定量分析

表 2 所示为三种方法对自采的 450 张实验室场景数据集评估的重建时间。需要说明的是,本文基于 Fioraio^[5]等人的方法实现整体模型配准,融合效率已在文献[5]中说明。与文献[5]不同的是,此处的对比实验基于串行 CPU,硬件环境为 i5-7300HQ、2.5 GHz 英特尔 CPU。

相比 Bylow^[14]等人直接配准单张深度图像, Fioraio^[5]等人需对深度图像去融合,且需新建子网格,后续还需对子网格融合配准。而本文在文献[5]的基础上加上了彩色图像的配准和去融合,且在子网格间迭代式地寻找对应点,使用迭代式地方式优化全局误差函数,因此耗费时间比文献[14],文献[5]都长。

表 2 三种不同方法的重建时间对比

Tab. 2 Reconstruction time for four different methods

方法	Bylow ^[14]	Fioraio ^[5]	本文
总体重建耗时/s	99	170	190
平均每帧耗时/ms	221	378	423

后续使用 GPU 实现代码加速可极大地减少重建时间,理论上可达到 Fioraio^[5]等人的重建效率。

4.4 实验中的参数分析

除了上述定性和定量的实验,本节还研究了算法中相关参数对重建效果的影响,包括位姿跟踪中用于调节彩色误差在总误差中所占比例的权重参数 θ 和迭代步长参数 α 。实验均基于 Augmented ICL-NUIM Living Room2 数据集。

表 3 不同权重 θ 对绝对轨迹误差的影响

Tab. 3 Influence of different weight θ on ATE (m)

θ	0	1	2	4	6	8	10
RMSE	0.120	0.118	0.116	0.114	0.113	0.113	0.115
占比	0	0.19	0.31	0.45	0.54	0.60	0.65
θ	12	14	16	18	20	50	100
RMSE	0.118	0.121	0.123	0.124	0.127	0.132	0.138
占比	0.69	0.71	0.74	0.75	0.77	0.88	0.92

表 3 所示为不同 θ 值对相机运动轨迹误差的影响。当 $\theta=0$,此时彩色误差占比为 0,位姿跟踪

中仅使用了深度数据;当 $\theta=100$,此时彩色误差占比为 92%,位姿跟踪中彩色信息起决定性的作用。当 $\theta=8$ 时,彩色误差和深度误差共同作用于位姿跟踪, RMSE 最小。

此外,对于全局优化,本文还设置了三组对比实验,分别命名为 A:未优化的绝对轨迹 RMSE。B:迭代步长 $\alpha=1.0$,优化后的绝对轨迹 RMSE。C:本文迭代步长式寻找对应点,优化后的绝对轨迹 RMSE。表 4 展示了实验结果。从表中可看出,迭代步长式寻找对应点方法比直接寻找对应点方法轨迹优化结果提升 8%。

表 4 迭代步长 α 对绝对轨迹误差的影响

Tab. 4 Influence of iterative step α on ATE (m)

实验编号	A	B	C
RMSE	0.161 7	0.122 3	0.112 6

由上述实验结果可知,本文提出的三维重建方法可减少位姿估计中的累积误差,得到的相机轨迹更加精确。

5 结 论

针对大场景重建中,相机位姿估计的累积误差导致相机漂移、重建模型质量低的问题,提出了基于深度相机的高质量三维场景重建方案。针对在缺少几何结构特征的场景中,仅使用深度信息跟踪相机易于失败的问题,结合深度和彩色测量以帧到模型的方式跟踪相机;针对位姿估计中存在累积误差,提出结合深度和彩色信息的去融合方案;在全局轨迹优化中提出迭代步长式地寻找表面对应点,并以对应点的欧氏距离和亮度信息为约束,优化全局相机位姿。

本文在传统的三维重建框架基础上提出了彩色信息去融合方案,并在全局优化中引入迭代步长,提高了相机轨迹的精确度。与基于深度相机的重建方法对比,本文提出的重建方法相机轨迹精度提升 14.1%,基于实际数据实验,本文方法能实现精确的相机跟踪,重建高质量的场景模型。满足实际应用需求。本文的算法基于串行 CPU 实现,重建时间相对较长。同时本文使用归一化体素网格表示三维场景,内存消耗大,因此限制了

重建的场景大小。后续工作可将算法移植到 CUDA 平台,用于 GPU 加速。同时使用 Oc-

tree^[19]或 Voxel Hashing^[20]来表示重建空间,以便重建更大场景。

参考文献:

- [1] NEWCOMBE R A, IZADI S, HILLIGES O, *et al.*. Kinect Fusion: Real-Time Dense Surface Mapping and Tracking[C]. *IEEE Computer Society*, 2011:127-136.
- [2] SZYMON RUSINKIEWICZ, MARC LEVOY. Efficient variants of the icp algorithm[J]. *International Conference on 3D Digital Imaging Modeling*, 2001:145-152.
- [3] H ROTH, M VONA. Moving Volume Kinect Fusion[C]. *Proceedings of British Machine Vision Conference*, 2012:1-11.
- [4] WHELAN T, KAESS M, JOHANNSSON H, *et al.*. Real-time large-scale dense RGB-D SLAM with volumetric fusion[J]. *International Journal of Robotics Research*, 2014, 34(4-5):598-626.
- [5] FIORAIO N, TAYLOR J, FITZGIBBON A, *et al.*. Large-scale and drift-free surface reconstruction using online subvolume registration[J]. *IEEE Computer Vision & Pattern Recognition*, 2015: 4475-4483.
- [6] XINGYIN F, FENG Z, QINGXIAO W, *et al.*. Real-time large-scale dense mapping with surfels [J]. *Sensors*, 2018, 18(5):1-19.
- [7] ZHOU Q Y, MILLER S, KOLTUN V. Elastic Fragments for Dense Scene Reconstruction [C]. *IEEE International Conference on Computer Vision*, 2013:473-480.
- [8] CHOI N S, ZHOU Q Y, KOLTUN V. Robust Reconstruction of Indoor Scenes[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015:5556-5565.
- [9] DAI A, MATTHIAS NIEßNER, MICHAEL ZOLLHÖFER, *et al.*. BundleFusion: real-time globally consistent 3D reconstruction using on-the-fly surface re-integration[J]. *ACM Transactions on Graphics*, 2017, 36(4): 3054739.
- [10] CURLESS B, LEVOY M. A volumetric method for building complex models from range images [J]. *In SIGGRAPH International Conference*, 1996:303-312.
- [11] LORENSEN W E, CLINE H E. Marching cubes: a high resolution 3D surface construction algorithm [J]. *ACM Siggraph Computer Graphics*, 1987, 21(4):163-169.
- [12] ERIK BYLOW, CARL OLSSON, FREDRIK KAHL. Robust Camera Tracking by Combining Color and Depth Measurements[C]. *IEEE International Conference on Pattern Recognition*, 2014:4038-4043.
- [13] WHELAN T, JOHANNSSON H, KAESS M, *et al.*. Robust Real-Time Visual Odometry for Dense RGB-D Mapping[C]. *IEEE International Conference on Robotics and Automation*, 2013:5724-5731.
- [14] BYLOW E, JÜRGEN STURM, KERL C, *et al.*. Real-time camera tracking and 3d reconstruction using signed distance functions[J]. *Robotics: Science and Systems*, 2013:35-42.
- [15] 李劲澎, 姜挺, 龚志辉, 等. 稳健李代数旋转平均用于 GPS 辅助无人机影像三维重建[J]. *光学精密工程*, 2017, 25(6): 1607-1618.
- ZHANG J Q, JIANG T, GONG ZH H, *et al.*. GPS-supported 3D reconstruction of UAV images based on robust Lie-algebraic rotation averaging [J]. *Opt. Precision Eng.*, 2017, 25(6): 1607-1618. (in Chinese)
- [16] 薛婷, 阮维鹏, 张少杰. 基于激光扫描的密集三维重建与优化[J]. *光学精密工程*, 2017, 25(3): 555-561.
- XUE T, RUAN W P, ZHANG SH J. 3D reconstruction and optimization of dense bubbly flow based on laser scanning[J]. *Opt. Precision Eng.*, 2017, 25(3): 555-561. (in Chinese)
- [17] 高瞻宇, 顾营迎, 吕耀宇, 等. 融合通用形变模型信息的面部三维重建[J]. *光学精密工程*, 2018, 26(1): 184-192.
- GAO ZH Y, GU Y Y, LÜ Y Y, *et al.*. Face reconstruction fused with generic morphable model [J]. *Opt. Precision Eng.*, 2018, 26(1): 184-192. (in Chinese)
- [18] LIU H, LI C, CHEN G, *et al.*. Robust key-frame-based dense slam with an RGB-D camera [J]. *IEEE Computer Vision and Pattern Recognition*, 2015:1-12.
- [19] FRANK STEINBRÜCKER, KERL C, CREMERS D. Large-Scale Multi-resolution Surface Reconstruction from RGB-D Sequences[C]. *IEEE*

International Conference on Computer Vision,
2013:3264-3271.

[20] NIENER M, ZOLLHFER M, IZADI S, *et al.*.

Real-time 3D reconstruction at scale using voxel
hashing [J]. *ACM Transactions on Graphics*,
2013, 32(6):1-11.

作者简介:



刘东生(1994—),男,湖北十堰人,硕士研究生,2017年于上海大学获得学士学位,主要从事计算机视觉与三维重建方面的研究。E-mail: liudongsheng_10@163.com

通讯作者:



张之江(1968—),男,上海人,博士,教授,博士生导师,1991年于哈尔滨科技大学获得工学学士学位,1996年、1999年于哈尔滨工业大学分别获工学硕士和工学博士学位,主要从事图像处理、视觉检测、数字全息和三维显示等方面的研究。E-mail: zjzhang@shu.edu.cn