

# 基于深度相机的特殊表面定位与修复

费点, 陈建林, 刘东生, 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** 由于成像原理限制, 深度相机无法获得非朗伯体高反射面和透明物质的真实深度值, 造成模型缺失甚至重建失败。针对特殊表面的三维重建问题, 提出一种定位特殊表面并优化位姿追踪的方法。通过累计零深度区和时序一致性约束优化对特殊表面进行定位。对深度数据置信度进行评估, 并采用空间、时间两个维度上非平均的位姿计算方法来减小透明物质对位姿计算的影响。实验结果表明, 在自然场景中, 所提方法能够仅通过消费级相机采集的深度数据完成包含特殊表面的三维模型重建与修复, 位姿追踪更加准确, 模型精度更高。

**关键词** 机器视觉; 三维感知; 三维重建; 非朗伯体定位; 相机位姿优化

中图分类号 TN219

文献标志码 A

doi: 10.3788/AOS202040.2115002

## Depth Camera-Based Location and Restoration of Special Surface

Fei Dian, Chen Jianlin, Liu Dongsheng, Zhang Zhijiang\*

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

**Abstract** Owing to imaging principle limitations, depth camera is impossible to obtain the true depth value of a non-Lambertian high reflective surface and transparent substances, resulting in model missing and reconstruction failure. Aiming at the problem of three-dimensional (3D) reconstruction on special surfaces, a method of positioning special surfaces and optimizing position tracking is proposed herein. In the proposed method, positioning for special surfaces is based on the statistics of the zero depth area and timing consistency constraint optimization. The confidence of the depth data is evaluated, and the non-uniform pose calculation method in space and time is used to reduce the influence of transparent materials on pose calculation. The experimental results show that in natural scenes, the proposed method can reconstruct and repair 3D models with special surfaces using only the depth data collected by a consumer camera, which makes pose tracking more accurate and increases model accuracy.

**Key words** machine vision; three-dimensional sensing; three-dimensional reconstruction; non-Lambertian localization; camera pose optimization

**OCIS codes** 150.6910; 150.5670

## 1 引 言

三维重建是计算机视觉领域的经典课题之一。2010年, 微软公司推出 Kinect v1 传感器, 其通过投射并捕捉固定的红外散斑来直接获得图像的深度信息, 后续推出的 Kinect 2 代传感器改用通过计算发射与接收红外脉冲信号的时间差来获得深度信息的方法。此后各类消费级深度相机逐一问世, 将三维重建带入了一个更加应用化、商业化的时代。

基于深度相机数据的三维重建主要分为两个部分, 一部分是根据深度图像序列估计相机位姿, 另一部分是通过融合相机位姿得到三维表面的稠密模

型。其中, 基于图像序列的相机位姿估计是三维重建中的重点。以 KinectFusion 为代表的许多算法都可以通过消费级相机采集的 RGB-D 数据较好地恢复出朗伯表面的稠密三维模型。但主流的深度相机都依赖红外线的漫反射来获取场景深度, 故当工作空间中存在透明物质和高反光表面时, 深度图上会出现集中的深度孔洞和深度畸变, 影响重建质量。特殊表面的存在会对三维重建造成两个负面影响: 大量深度噪声会严重影响相机位姿估计的准确性, 甚至导致相机丢失跟踪, 重建失败; 未经修复的深度图难以还原特殊表面的三维结构, 使模型出现孔洞, 降低模型质量。

收稿日期: 2019-11-25; 修回日期: 2020-06-19; 录用日期: 2020-07-15

基金项目: 国家自然科学基金面上项目(61572307)

\* E-mail: feidian@shu.edu.cn

针对相机位姿求解的问题,最早的 KinectFusion 算法<sup>[1]</sup>采用迭代最近点(ICP)算法估计相机位姿。这种方法基于稀疏的特征点匹配,直接将深度缺失部分排除在有效数据范围之外,受特征点数量制约,在深度图像存在大量零深度区域时,几乎都会跟踪失败。Bylow 等<sup>[2]</sup>提出最小化截断有向距离函数(TSDF)方法,该方法不再基于特征点,而是稠密地最小化有效深度和隐式表面之间的有向距离值,以计算当前相机位姿。为了解决高精度位姿追踪和高质量重建的问题,Slavcheva 等<sup>[3]</sup>提出最小化所有体素网格帧间 TSDF 算法。尽管目前的位姿求解算法都增加了计算的稠密性,提升了位姿估计的成功率和精度,但在存有特殊表面的重建中,大量错误深度值会被无差别地用于计算,故位姿估计准确度很差,导致模型质量恶化。针对特殊表面的深度图修复与重建的问题,Alhwarin 等<sup>[4-5]</sup>在深度相机的基础上增加一个红外线发射设备,该方法根据红外图像视差获得特殊表面的深度估计值,提升深度图像质量,但不关注模型质量。对于透明物体,Yeung 等<sup>[6]</sup>在已知、可控的变化光照下采集玻璃制品的图像序列,通过估计外表面法线来整合表面深度。Wu 等<sup>[7]</sup>搭建了一套完整的光学系统,将玻璃雕塑放置在特定条纹屏幕前,通过变形条纹和轮廓来重建透明物质。对于含有高反射表面的物体,乔玉晶等<sup>[8]</sup>提出一种基于工业相机光照补偿图像的均值漂移算法,该算法结合深度图像,解决了高反光、弱纹理的表面重建问题。文学东等<sup>[9]</sup>提出了一种基于机载点云、地面点云及倾斜多视纹理融合特征的针对建筑物的三维模型重建方法。上述方法对采集设备的要求比较高,受限于特定的实验环境。Yan 等<sup>[10-11]</sup>采用消费级深度相机采集 RGB-D 图像,用最小化有向距离方法求解相机位姿,并提出了一种基于有向距离统计量来定位透明物质区域的算法,但没有考虑增加的噪声对位姿计算的影响。

近年来出现了一些结合深度学习的深度修复算法。Zhang 等<sup>[12]</sup>训练了基于 VGG-16 的对称编解码全卷积网络,输入单张彩色图像,输出局部表面法向量和闭合轮廓,再结合深度图的全局信息获得零深度区域的深度值。Qiu 等<sup>[13]</sup>在文献[12]的网络中结合由激光雷达获得的稀疏深度数据,完成对室外场景深度图像的孔洞修复。Sajjan 等<sup>[14]</sup>对透明物质特性进行改进,增加了错误深度消除操作和边缘区域处理操作,并将算法应用于机器人对透明物质的抓取。上述算法都将关注点放在对单张深度图

像的修复,用修复后的图像完成重建,忽略了图像序列的时空关系,且修复效果受数据集影响。

为了给特殊表面的定位和修复问题提供一种消费级的解决方案,本文在文献[10]的基础上提出改进,采用 Kinect v1 相机采集深度图像,通过调整深度数据置信度来优化位姿,提高位姿追踪准确度。所提方法定位并修复了包含透明物质和高反射表面的特殊表面,提高了重建质量。

本文工作在于:

1) 提出一种基于几何信息和前后帧一致性约束的深度数据置信度评估方法。在工作空间中存在透明物质的情况下,深度图中每个数据可信度的差别很大。评估分为二维图像和三维体素两个角度。在二维层面上,可信度最差的为深度边缘和深度孤立点,可在单帧深度图上根据几何关系检测;在三维层面上,前后帧数据应受一致性约束,由有向距离函数(SDF)的变化量衡量。

2) 提出基于置信度评估结果的权重调整方案,减小低置信度深度值对位姿计算的影响,使位姿信息主要在高置信度区域获得,优化相机位姿。同时认为,最近的帧序列从时序上来说可信度最高,在位姿估计过程中逐帧去除较远帧的影响,保证相机位姿追踪的成功率。

3) 提出一种基于零深度区域投影视锥体和空间一致性约束的特殊表面定位方法。将视锥体求解问题转换为基于概率的二分类问题,以帧融合前后的 TSDF 变化量作为一致性度量,结合置信度信息,完成零深度区域修复,从而形成完整的稠密三维模型。

## 2 系统框图

系统整体框图如图 1 所示,分为深度数据处理、位姿追踪与优化、表面重建与修复 3 个阶段。本系统的输入为包含透明物质的深度图像序列。

深度数据处理:本阶段主要包括深度图预处理和深度值置信度判断。根据各种噪声类型的特点,对原始深度图进行降噪处理,根据深度分布,分割出感兴趣的深度区间(depth ROI)作为工作空间。提取深度图像中的边缘,并将其作为下一阶段调整置信度的依据。

位姿追踪与优化:本阶段将深度数据转换为 TSDF 的隐式表达方式,采用非平衡置信度的 TSDF 方法计算第  $N-1$  帧到第  $N$  帧的相机位姿变换,包括旋转量  $\mathbf{R}$  和平移量  $\mathbf{T}$ 。同时,根据 SDF

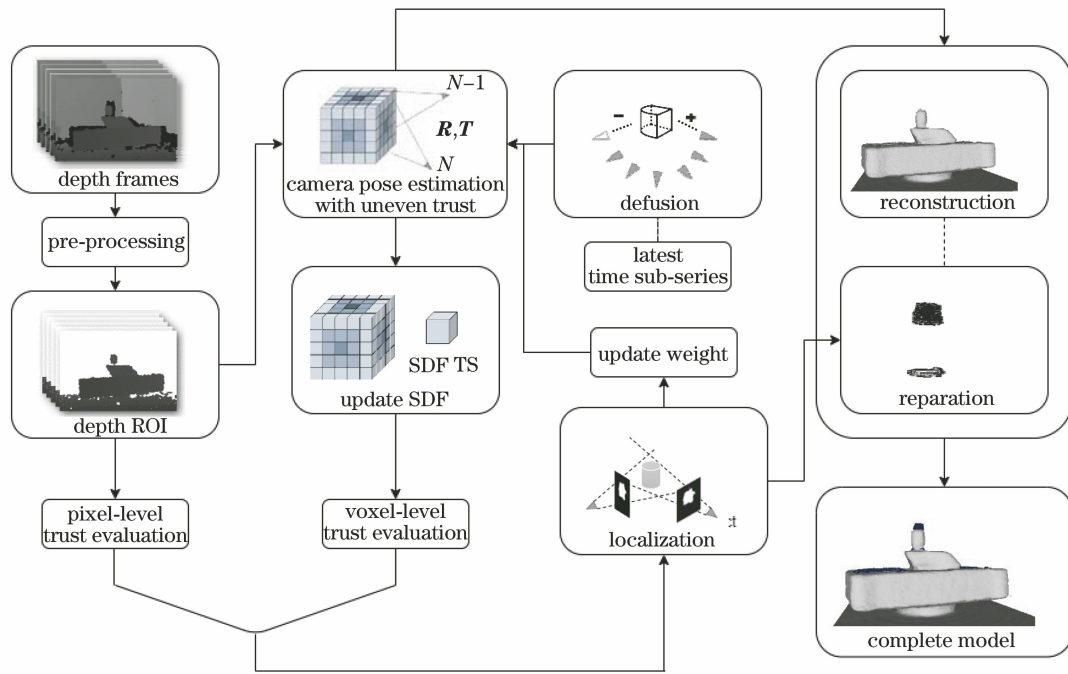


图 1 系统示意图

Fig. 1 Diagram of system

的残差(TS)评估当前体素网格的置信度,优化追踪。在整个位姿追踪的过程中,仅以最近的  $L$  帧作为参考帧,对超出时间序列范围的深度数据进行去融合处理,以减小累计误差。

表面重建与修复:本阶段的目标是产生完整的稠密三维模型。对优化后的位姿和有效深度值进行融合获得朗伯表面的模型。根据多视角下零深度区域视锥并结合前后帧一致性约束,获得非朗伯表面的模型,修复原模型中的孔洞,从而重建出完整的三维结构。

### 3 方 法

#### 3.1 深度数据处理

以 Kinect 系列为代表的消费级相机受到设备精度的限制,拍摄的深度图像本身存在大量的噪声,

如由成像原理形成的阴影噪声、随着距离增加产生的误差噪声、物体边缘处表现出的畸变噪声等。当重建目标中包含特殊表面时,深度孔洞和畸变深度所占的比例急剧上升,噪声对图像质量的影响更加显著,这给位姿追踪和表面重建带来巨大挑战,所以必须对粗糙的深度图像进行处理。

图 2(a)为预处理前的深度图,首先采用一次双边滤波法去除原始图像中由一些随机性因素引起的椒盐噪声和孤立深度点,再根据其他噪声的图像特点分别处理。阴影噪声是由物体遮挡了投向其背后区域的结构光引起的,导致物体的右侧出现零深度值区域。像素宽度可根据结构光的成像原理获得<sup>[15]</sup>,再对判定为阴影噪声的区域以背景深度进行填充,如图 2(b)所示。

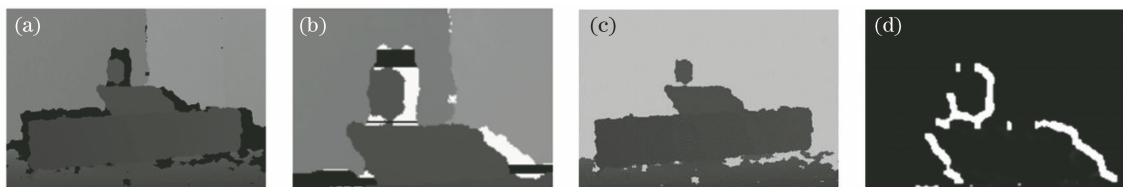


图 2 深度图预处理。(a)原深度图;(b)阴影噪声区;(c)分割 ROI;(d)置信度掩模

Fig. 2 Depth image preprocessing. (a) Raw depth image; (b) shadow noise area;

(c) ROI segmentation; (d) reliability mask

为了去除超出测量范围的深度值,并减少冗余计算,选择深度 ROI。图 3 为去除阴影噪声后的深度直方图,其中  $N_{\text{pixel}}$  为该深度区间内的像素点数

量。待重建的目标为前景物体,该区域深度值连续且较为集中,又明显小于远处的大量背景点。利用这一特点,以前、背景之间的谷值  $d$  作为 ROI 分割

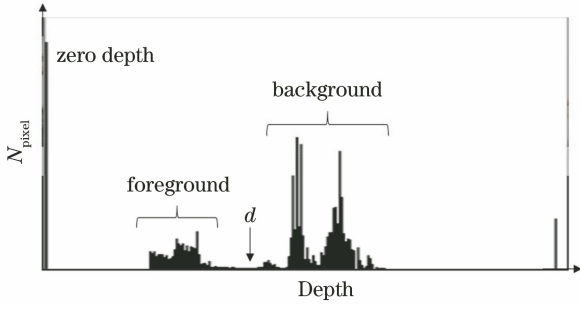


图 3 深度直方图

Fig. 3 Depth histogram

点。为了保证前景完整,根据实验物体设置裕量  $\alpha$  (实验取 100 mm),进入 ROI 的背景点在融合过程中会被逐渐消除。深度 ROI 的分割结果如图 2(c) 所示。

前景物体包含大量边缘噪声,这是因为物体边

缘深度梯度大,相机受到分辨率限制,一个像素点尺度内包含了多个不同深度值。边缘噪声难以在单张图像上恢复,所以,在深度图预处理阶段仅通过 Sobel 算子提取深度图边缘,以掩模的形式指示该区域,并将其作为置信度调整的依据,如图 2(d) 所示。

### 3.2 相机位姿估计

#### 3.2.1 相机位姿初始化

将第一帧图像的相机坐标系定为世界坐标系,后续帧的相机坐标到世界坐标的变换即是相机位姿。以李代数的形式,可以用六个自由度来表示坐标系的变换:

$$\xi = (t_1, t_2, t_3, \omega_1, \omega_2, \omega_3), \quad (1)$$

式中:  $t$  为三个自由度的相机平移;  $\omega$  为三个自由度的相机旋转。其对应的李群可表示为

$$S_E(3) = \left\{ \mathbf{T}_{g,k} = \begin{bmatrix} \mathbf{R}_{g,k} & \mathbf{t}_{g,k} \\ \mathbf{0}^T & 1 \end{bmatrix} \in \mathbf{R}^{4 \times 4} \mid \mathbf{R}_{g,k} \in \mathbf{S}_O(3), \mathbf{t}_{g,k} \in \mathbf{R}^3 \right\}, \quad (2)$$

式中:  $\mathbf{R}_{g,k}$  为相机坐标系到世界坐标系的旋转量,为  $3 \times 3$  的矩阵;  $\mathbf{t}_{g,k}$  为相机坐标系到世界坐标系的平移量,为 3 维向量。记空间点为  $\mathbf{P} = (P_x, P_y, P_z)^T \in \mathbf{R}^3$ , 在世界坐标系中的三维坐标  $\mathbf{P}_g$  可以由相机变换获得:

$$\mathbf{P}_g = \mathbf{T}_{g,k} \mathbf{P}_k. \quad (3)$$

且当前相机坐标系和二维图像之间存在投影关系,即

$$\boldsymbol{\pi}(\mathbf{P}) = \mathbf{p} = (x, y)^T \in \mathbf{R}^2, \quad (4)$$

$$D(\mathbf{p}) = P_z, \quad (5)$$

式中:  $\boldsymbol{\pi}(\cdot)$  为取投影点;  $D(\cdot)$  为取深度值。TSDF 是三维信息的一种隐式表达方式,即当前体素到三维表面的截断有向距离,

$$d = \varphi[\mathbf{T}_{g,k} D(\mathbf{p}) K^{-1} \mathbf{p}], \quad (6)$$

式中:  $\varphi$  为空间点到三维表面对应点  $\mathbf{p}$  的转换公式;  $K$  为相机内参。对于深度图像上的有效深度值,其对应空间点的 SDF 值应为 0。

故在初始化相机位姿时的能量公式为

$$E(\xi) = \sum_{x,y} \varphi(\xi)^2, \quad (7)$$

式中:  $(x, y)$  为深度图像上的像素坐标。

#### 3.2.2 非平衡置信度的位姿追踪

##### a) 空间非平衡置信度

TSDF 数据实际上是通过将捕捉到的深度信息显示在三维空间内得到的。当输入为深度帧序列

时,体素网格中的 TSDF 值表述的是多帧信息的融合。

采用点到点的距离作为(7)式的对应点关系,即

$$d_{\text{point-point}}(\mathbf{p}) = z_i - D(\mathbf{p}). \quad (8)$$

对应 TSDF 为

$$\bar{d}(\mathbf{p}) = \begin{cases} -\mu, & d(\mathbf{p}) < -\mu \\ d(\mathbf{p}), & |d(\mathbf{p})| \leq \mu \\ \mu, & d(\mathbf{p}) > \mu \end{cases} \quad (9)$$

式中:  $\mu$  为截断值。

TSDF 由多视角深度值加权融合获得,即

$$\bar{d}(\mathbf{p}) = \frac{\sum \omega_i(\mathbf{p}) \bar{d}_i(\mathbf{p})}{\Omega(\mathbf{p})}, \quad (10)$$

$$\Omega(\mathbf{p}) = \sum \omega_i(\mathbf{p}). \quad (11)$$

但深度图中的多种噪声和非朗伯表面造成的噪声都会使某些深度数据的可信度降低,故在融合的过程中要进行置信度调整。对深度数据的可信度评价有两个层面:第一,二维深度图像上像素点深度值有可信度的差异;第二,信息融合到三维体素网格中,根据前后帧间的无效深度和不一致性,各个三维网格的 TSDF 值有可信度的差异。

预处理过程中,在去除噪声的同时产生了指示物体边缘的掩模,对应深度图像中可信度最低的区域。对于理想的深度跳变边缘,应设置像素置信度  $\omega_m = \text{mask}(\mathbf{p}) \in \{1, 0\}$ 。然而由于误差问题,深度

图上的边缘也并非物体实际的边缘,但可以确定实际的边缘存在于深度边缘宽为  $\sigma_0$  的邻域范围内。所以将  $\omega_m$  扩展为

$$\omega_m = \frac{1}{\exp(1/\sigma)} (\sigma \leq \sigma_0). \quad (12)$$

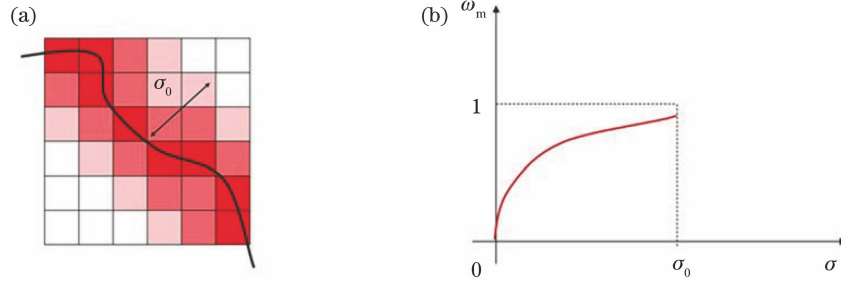


图 4 深度边缘权值设置。(a)  $\sigma_0$  邻域范围;(b)  $\omega_m$  函数曲线

Fig. 4 Weight setting near depth edge. (a)  $\sigma_0$  neighborhood range; (b)  $\omega_m$  function curve

三维体素网格中的深度数据因为特殊表面的存在而有可信度上的差异。特殊表面多数表现为零深度,所以当体素网格的投影落入零深度区域时,当前体素可能属于特殊表面区域也有可能当前帧被特殊表面遮挡,这两种情况下的 TSDF 值显然都是无效值。设置可能存在特殊表面区域的置信度参数  $\omega_{TL} \in \{1,0\}$ 。

特殊表面在少数像素上表现为错误深度值,同一体素的错误深度值会因结构光入射角度的不同,在各个视角下差异较大,即可以通过搜索前后帧不一致的区域来搜索可能存在的特殊表面。文献[10-11]在搜索可能区域时采用 TSDF 方差作为统计量。但是根据(10)式可知,TSDF 实质为前向深度帧序列的加权平均,多次求取均值会使较早的深度帧在计算中有更大的影响。且较早的深度帧在计算位姿时参考序列短,奇异值对 TSDF 的影响显著,

深度边缘的权值设置方式如图 4 所示。极端情况下,位于边缘的像素置信度  $\omega_m$  为 0,即不参与位姿计算。丢失的有效深度信息可以由图像序列中其他视角补全,所以  $\omega_m$  的设置减小了大量深度跳变干扰的同时并不会造成信息的丢失。

所以采用方差作为统计量是不合理的。针对这一问题,采用体素第  $k$  帧融合前后的 TSDF 变化量和全局平均 TSDF 变化量作为评估标准,表达式分别为

$$\eta_{TS} = (\bar{d}_{k-1} - \bar{d}_k)^2, \quad (13)$$

$$\bar{\eta}_{TS} = \frac{1}{N_{ROI}} \sum_{p \in R} \eta_{TS}(p), \quad (14)$$

式中:  $R$  为感兴趣区域;  $N_{ROI}$  为感兴趣区域包含的像素点数。

$\eta_{TS}$  较低的区域表面满足三个条件:存在非噪声区、前后帧共同可视、与相机平面相对平行。所以满足低通滤波条件  $\eta_{TS} < \gamma_l \bar{\eta}_{TS}$  的表面是当前帧追踪相机位姿的理想参考区域,其中  $\gamma_l$  为低通滤波器参数。设置 15 组实验场景,改变滤波器参数,将满足滤波条件的表面反投影到深度图像上,对其和人工评价的理想参考区域进行对比,如图 5 所示,  $\gamma_l \geq$

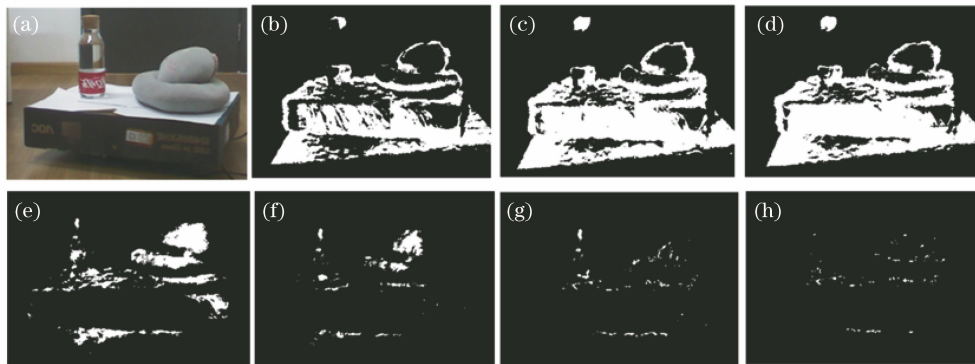


图 5 TSDF 变化量  $\eta_{TS}$ 。(a)彩色图;(b)~(d)  $\eta_{TS} < \gamma_l \bar{\eta}_{TS}$ ,低通滤波器参数  $\gamma_l$  分别取 0.6,0.7,0.8;

(e)~(h)  $\eta_{TS} > \gamma_h \bar{\eta}_{TS}$ ,高通滤波器参数  $\gamma_h$  分别取 1.6,1.7,1.8,1.9

Fig. 5 TSDF variation  $\eta_{TS}$ . (a) Color image; (b)~(d)  $\eta_{TS} < \gamma_l \bar{\eta}_{TS}$ , low pass filter parameter  $\gamma_l = 0.6, 0.7, 0.8$ ;

(e)~(h)  $\eta_{TS} > \gamma_h \bar{\eta}_{TS}$ , high pass filter parameter  $\gamma_h = 1.6, 1.7, 1.8, 1.9$

0.7 时可以完整地得到上述表面,所以选择  $\gamma_1 = 0.7$ 。根据物体表面的连续性,上述体素及其周围体素的置信度相应调整为

$$\omega_{TS} = \frac{\omega_{\max}}{d_j + 1}, \quad (15)$$

式中: $j$  为高置信度体素; $d_j$  为当前体素到高置信度体素  $j$  的距离。同一像素可能受到多个体素影响,故取

$$\omega_{TS} = \max[\omega_{TS}(j)], \quad (16)$$

则(10)式中的  $\omega_i$  可具体表示为

$$\omega_i = \omega_m \omega_{TL} \omega_{TS} \omega_0, \quad (17)$$

式中: $\omega_0$  为基准权重值。

### b) 时间非平衡置信度

从时间序列上来说,位姿关联性随间隔帧数量的增加而减小。与此同时,帧间的误差会一直传递,深度噪声也会在体素网格中持续累积。在对特殊表

面进行重建中,由于噪声量增加,累计误差问题更加严重,如果将  $\{0, 1, \dots, N-1\}$  序列作为第  $N$  帧的参考,极易导致模型质量随融合帧数的增加急剧恶化,甚至追踪失败。但是若只考虑帧到帧的匹配,有效深度过小,位姿会陷入局部最优。所以在理想情况下,每个深度帧  $N$  的位姿应由关系最紧密的  $L$  个近邻帧  $\{N-L, N-L+1, \dots, N-1\}$  获得。

所以设定长度为  $L$  的图像先入先出(FIFO)队列,如图 6 所示。当第  $N$  帧进入队列时,仍然根据置信度加权方法融合 TSDF 数据,同时对处于队列头部的第  $N-L-1$  帧执行融合操作的逆操作,消除其对 TSDF 数据的影响,保持参考帧长度为  $L$ 。这一过程将原本位姿求解过程中的(10)式转换为

$$\bar{d}(\mathbf{p}) = \frac{\sum \omega_i(\mathbf{p}) \bar{d}_i(\mathbf{p})}{\Omega(\mathbf{p})}, \quad (18)$$

$$i = k, k+1, k+2, \dots, k+L.$$

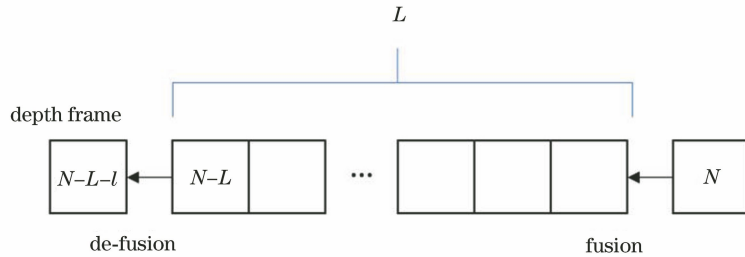


图 6 参考帧 FIFO 队列

Fig. 6 Reference frame FIFO queue

### 3.3 特殊表面修复

由于 Kinect v1 传感器基于红外散斑的发射与接收获取深度,而红外散斑会在特殊表面发生复杂的反射、透射及折射,特殊表面区域在深度图上主要表现为零深度区域,少部分表现为错误深度值。虽然很难从这些深度孔洞中恢复出真实的深度值,但特殊表面存在于相机光心和深度图像零深度区域形成的视锥中<sup>[9-10]</sup>,特殊表面修复问题实际为体素是否属于特殊表面的二分类标签问题。特殊表面引起的错误深度的形成原理复杂、分布随机,可通过和前帧的一致性约束来判断。

基于体素空间的独立性假设,认为工作空间中的每一个体素属于特殊表面的概率是相互独立的,那么透明物质在体素中的分布情况为

$$\Theta = \operatorname{argmax} \left[ \prod_{v \in V_o} P_{\text{obj}}(v) \times \prod_{v \notin V_o} P_{\text{others}}(v) \cdot P(\Theta) \right], \quad (19)$$

式中: $V_o$  为特殊表面体素的集合; $P_{\text{obj}}$  为体素属于特殊表面的概率; $P_{\text{others}}$  为体素属于非特殊物质或

背景的概率。

$$P_{\text{obj}}(v) = \prod_{i=1}^N \omega_i P \{ I[\boldsymbol{\pi}_i(v)] | v \in V_o \}, \quad (20)$$

式中: $\omega_i$  为权值; $I$  为取深度图像上的像素点。

$$\omega_i = \begin{cases} 1, & n \geq \epsilon \\ e^{-(\epsilon-n)^2}, & n < \epsilon \end{cases}, \quad (21)$$

式中: $n$  为当前体素连续落入零深度区域的帧数, $n$  值大的体素有更大的概率属于特殊表面; $\epsilon$  为帧数截断阈值,可根据实验采集深度图总数调整。尽管预处理过程中已去除了大部分的阴影噪声,但仍然存在其他成像因素造成的零深度噪声区,为了减小噪声干扰,采用非对偶的概率计算方式,即

$$P_{\text{others}}(v) = 1 - \prod_{i=1}^N \{ 1 - I[\boldsymbol{\pi}_i(v)] | v \notin V_o \}. \quad (22)$$

这样的概率计算方式对多个视角下的一致性具有更强烈的约束。当体素连续被投影到零深度区域时,其属于特殊表面的概率逐渐上升,而只要有一次被投影到朗伯表面区域,其概率将急剧下降。同时

可进行弱假设,所有的表面分布情况  $\Theta$  都是等概的,则零深度区域表示为

$$\Theta_{\text{zero}} = \{v \in \Theta \mid P_{\text{obj}} > P_{\text{others}}\}。 \quad (23)$$

透明物质引起的错误深度值通常在真实深度值和背景深度值之间,且空间分布较为随机,所以错误深度值的 SDF 会与前帧融合模型的 SDF 之间差值较大。融合前后的 TSDF 变化量 TS 在位姿追踪的过程中已有计算。将不一致性的度量表示为

$$R_{\text{SDF}} = \sum_{i=1}^N I(\eta_{\text{TS}} > \gamma_h \cdot \bar{\eta}_{\text{TS}})。 \quad (24)$$

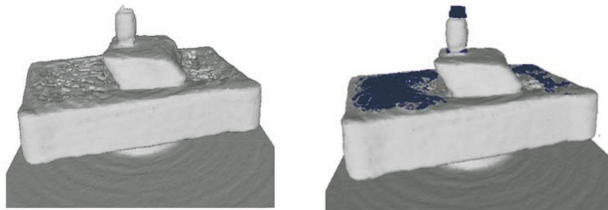
由于当前帧融合前后的 TSDF 一致性和重建表面的梯度变化也有关系,故根据当前帧的平均 TSDF 变化程度设置自适应阈值,高通滤波器阈值系数  $\gamma_h = 1.8$ 。同时将帧数计数作为度量形式时,避免了单帧噪声过大时奇异值对后帧数据的影响。则错误深度区域表示为

$$\Theta_{\text{err}} = \{v \in \Theta \mid R_{\text{SDF}}(v) > \beta \times N\}, \quad (25)$$

式中: $\beta$ 为融合前后 TSDF 变化量过大的帧数占总帧数比例的阈值。

最终透明物质的修复区域为

$$\Theta = \Theta_{\text{zero}} \cap \Theta_{\text{err}} = \{v \mid (P_{\text{obj}} > P_{\text{others}}) \cap$$



(a) (b)

$$[R_{\text{SDF}}(v) > \beta \times N] \}。 \quad (26)$$

## 4 实验结果与分析

### 4.1 特殊表面修复实验

在自然的室内场景下,将包含特殊表面的重建目标放在平稳匀速转动的转台上。转台旋转一周的时间为 30 s。使用一台 Kinect v1 传感器作为采集设备,在 30 s 内采集 600 张不同视角下的深度图作为系统输入。

图 7 为模型修复实验。在图 7(c)场景中,底层纸箱和玻璃瓶下的硬质塑料盖子为高反射表面。如图 2(a)深度图所示,Kinect 相机无法获取表面真实深度,所以修复前的模型上无对应部分。玻璃瓶身部分为透明表面,深度图中瓶身标签上端表现为零深度,标签下端为畸变深度,深度值介于瓶中有色液体深度和真实深度之间。图 7(b)为修复后的模型,图 7(f)为玻璃瓶修复细节。修复后纸箱上表面平滑无孔洞,玻璃瓶缺失和凹陷的部分均被准确定位并得以修补,瓶盖轮廓完整,与瓶身之间的转折清晰。结果表明,所提方法对非朗伯表面的定位与修复是有效的。



(c) (d) (e) (f)

图 7 模型修复实验。(a)修复前模型;(b)修复后模型;(c)实验环境与设置;(d)特殊物体;(e)修复前模型细节;(f)修复后模型细节

Fig. 7 Model reparation experiment. (a) Model before reparation; (b) model after reparation; (c) experimental environment and setting; (d) peculiar object; (e) model detail before reparation; (f) model detail after reparation

### 4.2 位姿追踪对比实验

测量 Kinect 相机到转台中心的距离,根据几何关系获得相机位姿理论值。分别采用 KinectFusion、SDF-tracker 和所提方法追踪相机位姿,用这三种方法的位姿估计结果生成模型,如图 8 所示,并计算在 100,300,600 帧之后的平均相机位

姿误差,如表 1、2 所示。

从图 8 可以看出:KinectFusion 方法因为帧到帧的匹配方式存在累计误差问题,随着帧序列的增加,特殊表面引入的噪声不断累加,导致位姿追踪精度急剧恶化,无法获得目标模型;所提方法和 SDF-tracker 都采用帧到模型的匹配方式,抗噪声能力比

表 1 实验 I 中位姿追踪误差分析

Table 1 Error analysis of pose tracking in experiment I

Method	Average pose error	Average pose error	Average pose error
	after 100 frames	after 300 frames	after 600 frames
KinectFusion	32.8	38.2	62.6
SDF-tracker	26.1	27.5	33.8
Proposed method	25.7	27.0	29.7

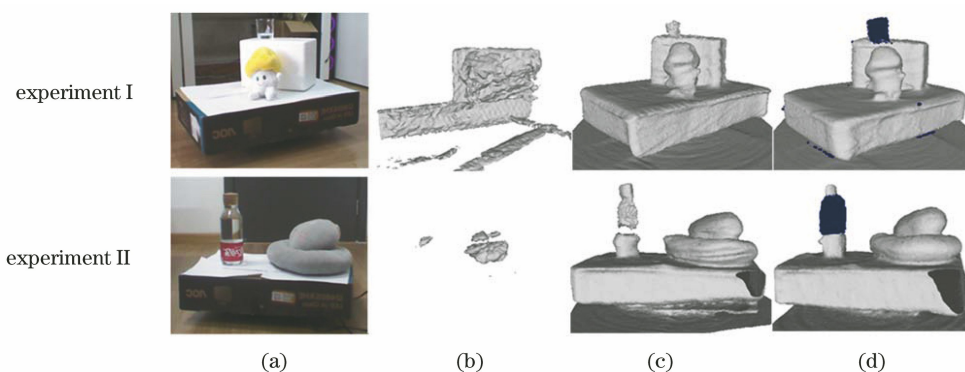


图 8 不同位姿结果对应的模型。(a)实验设置;(b) KinectFusion;(c) SDF-tracker;(d)所提方法

Fig. 8 Models corresponding to different pose results. (a) Experimental setting; (b) KinectFusion; (c) SDF-tracker; (d) proposed method

表 2 实验 II 中位姿追踪误差分析

Table 2 Error analysis of pose tracking in experiment II

mm

Method	Average pose error after 100 frames	Average pose error after 300 frames	Average pose error after 600 frames
KinectFusion	51.0	100.2	811.2
SDF-tracker	28.9	33.4	38.8
Proposed method	30.5	33.2	33.7

帧到帧的方式更强。从表 1、2 可以看出:在初始的较短序列内,参考模型受错误深度值影响不显著,所提方法和 SDF-tracker 追踪结果相似;在完整的时间序列内,所提方法表现出了更强的追踪能力,获得的位姿更加精确。

### 4.3 重建效果对比实验

所提方法在文献[10]算法(后文简称算法 I)的流程中增加了对相机位姿追踪的优化过程,改进了错误深度区域的定位方法。用图 9、10 两组实验数据对比所提方法与算法 I 的重建效果。在自然室内场景中,手持 Kinect v1 相机拍摄包含透明物质在内的目标物体,采集 600 帧不同视角下的深度图。

图 9 为一组简单的几何结构实验,手持相机逆时针旋转。图 9(c)为算法 I 的重建结果,由于没有考虑透明物质对相机位姿追踪的影响,在靠近透明物体的一侧时位姿追踪估计精度较低,模型

上出现错误褶皱,靠近杯体处明显。图 9(d)为所提方法的重建结果,相对图 9(c),表面平滑、规则,杯体部分修补质量较高。图 10 为复杂几何结构实验,增加了重建难度,手持相机顺时针旋转。图 10(c)、(d)为两个视角下算法 I 的结果。受物体遮挡和采集角度的限制,零深度投影仅定位玻璃杯的上半部分,TSDF 方差对玻璃杯下半部分定位失败,这是由位姿问题误差使盒体上侧和瓶盖上侧出现错误褶皱引起的。位姿追踪起始阶段的误差过大,持续累计进入后向序列,难以修正,导致了图 10(b)所示的位姿丢失。书包背带部分严重残缺,透明区修复效果也很差。图 10(e)、(f)为对应视角下所提方法获得的模型,模型整体平滑,书包背带部分完整,零深度区域定位准确,SDF 一致性探测修补了错误深度部分,修复后的玻璃瓶模型形状完整。

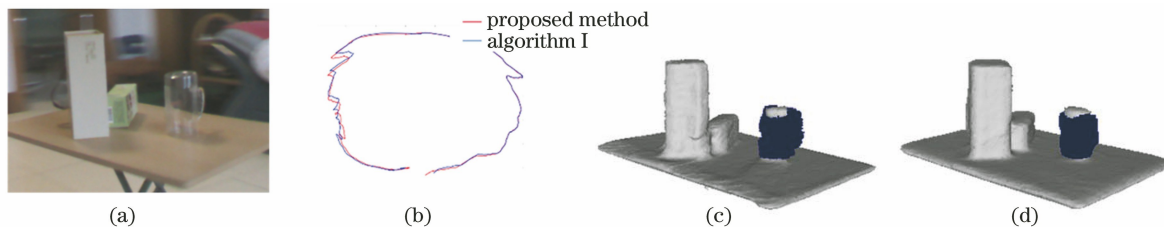


图 9 重建结果对比 1。(a)实验环境与设置;(b)相机位姿;(c)算法 I 结果;(d)所提方法结果

Fig. 9 Comparison of reconstruction results 1. (a) Experimental environment and setting; (b) camera pose; (c) result obtained by algorithm I; (d) result obtained by proposed method



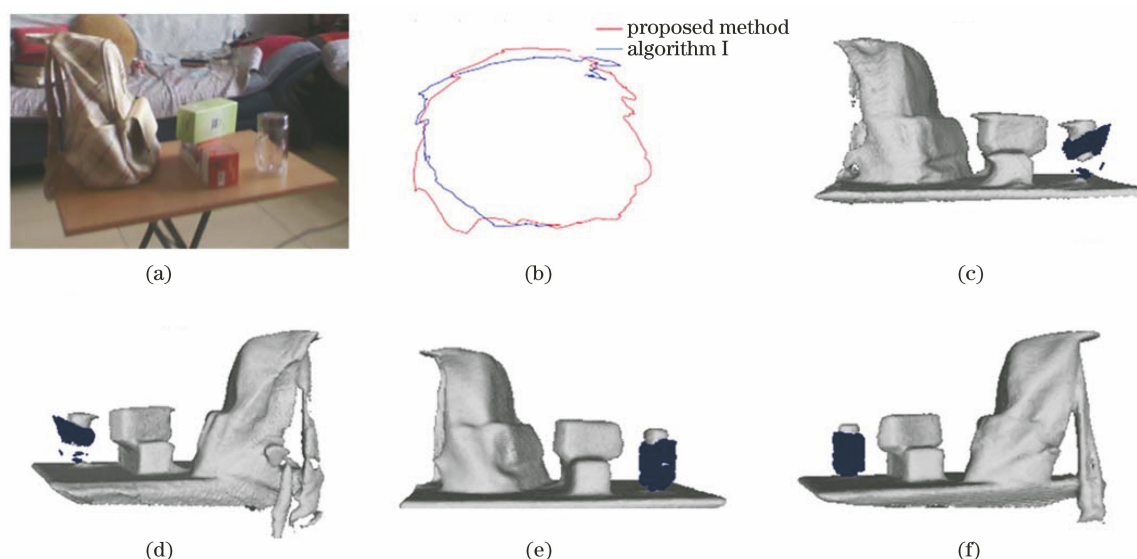


图 10 重建结果对比 2。(a)实验环境与设置;(b)相机位姿;(c)(d)算法 I 结果;(e)(f)所提方法结果

Fig. 10 Comparison of reconstruction results 2. (a) Experimental environment and setting; (b) camera pose; (c)(d) result obtained by algorithm I; (e)(f) result obtained by proposed method

## 5 结 论

针对深度相机的三维重建问题,提出了一种改进的特殊表面定位和修复办法。通过对深度数据和空间网格数据进行置信度评估,减小特殊表面所引入的噪声对位姿计算的影响,并通过去融合方法减小累计误差,提高追踪算法的准确性。实验结果表明,在自然场景下,所提方法有效地优化了特殊表面定位和位姿追踪质量,修复了原模型孔洞的三维结构,改善了模型中的畸变和毛刺。所提方法没有讨论重建的时间效率,今后的工作可以将所设计系统移植到 GPU 上进行并行计算,同时优化算法效率,使其向实时重建的方向发展。

### 参 考 文 献

- [1] Newcombe R A, Izadi S, Hilliges O, et al. KinectFusion: real-time dense surface mapping and tracking [C] // 2011 10th IEEE International Symposium on Mixed and Augmented Reality, October 26-29, 2011, Basel, Switzerland. New York: IEEE, 2011: 127-136.
- [2] Bylow E, Olsson C, Kahl F. Robust online 3D reconstruction combining a depth sensor and sparse feature points [C] // 2016 23rd International Conference on Pattern Recognition (ICPR), December 4-8, 2016, Cancun, Mexico. New York: IEEE, 2016: 3709-3714.
- [3] Slavcheva M, Kehl W, Navab N, et al. SDF-2-SDF registration for real-time 3D reconstruction from RGB-D data [J]. International Journal of Computer Vision, 2018, 126(6): 615-636.
- [4] Alhwarin F, Ferrein A, Scholl I. IR stereo kinect: improving depth images by combining structured light with IR stereo [M] // Pham D N, Park S B. PRICAI 2014: trends in artificial intelligence. Lecture notes in computer science. Lecture notes in computer science. Cham: Springer, 2014, 8862: 409-421.
- [5] Yun Y, Seo D, Kim D. Recognition of transparent objects using 3D depth camera [C] // 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), June 28 -July 1, 2017, Jeju, South Korea. New York: IEEE, 2017: 882-883.
- [6] Yeung S K, Wu T P, Tang C K, et al. Normal estimation of a transparent object using a video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(4): 890-897.
- [7] Wu B J, Zhou Y, Qian Y M, et al. Full 3D reconstruction of transparent objects [J]. ACM Transactions on Graphics, 2018, 37(4): 1-11.
- [8] Qiao Y J, Zhang S Y, Zhao Y H. Surface robust reconstruction method for high light and weak textured objects [J]. Acta Photonica Sinica, 2019, 48(12): 1212002.  
乔玉晶, 张思远, 赵宇航. 高光弱纹理物体表面鲁棒重建方法 [J]. 光子学报, 2019, 48(12): 1212002.
- [9] Wen X D, Chen W M, Xie H, et al. A method for building model reconstruction based on multi-source feature fusion [J]. Geomatics and Information Science

- of Wuhan University, 2019, 44(5): 731-736, 764.
- 文学东, 陈为民, 谢洪, 等. 一种融合多源特征的建筑物三维模型重建方法[J]. 武汉大学学报·信息科学版, 2019, 44(5): 731-736, 764.
- [10] Yan J, Xia Q, Chen J L, et al. Reparation of transparency for high-quality three-dimensional reconstruction using a single RGB-D sensor [J]. *Optical Engineering*, 2018, 57(10): 103101.
- [11] Xia Q, Yan J. Localization of transparent objects using signed distance functions [J]. *Electronic Measurement Technology*, 2018, 41(14): 88-91.
- 夏青, 严军. 基于有向距离函数的透明物体定位[J]. *电子测量技术*, 2018, 41(14): 88-91.
- [12] Zhang Y D, Funkhouser T. Deep depth completion of a single RGB-D image [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 175-185.
- [13] Qiu J X, Cui Z P, Zhang Y D, et al. DeepLiDAR: deep surface normal guided depth prediction for outdoor scene from sparse LiDAR data and single color image [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 3308-3317.
- [14] Sajjan S S, Moore M, Pan M, et al. ClearGrasp: 3D shape estimation of transparent objects for manipulation[EB/OL]. (2019-10-14) [2019-11-24]. <https://arxiv.org/abs/1910.02550>.
- [15] Freedman B, Shpunt A, Machline M, et al. Depth mapping using projected patterns: US8493496 [P/OL]. 2013-07-23 [2019-11-24]. <https://patents.justia.com/patent/8493496>.