

## 基于 Kinect v2 的实时精确三维重建系统\*

李诗锐<sup>1,2</sup>, 李琪<sup>1,2</sup>, 李海洋<sup>1,2</sup>, 侯沛宏<sup>1,2</sup>, 曹伟国<sup>1</sup>, 王向东<sup>3</sup>, 李华<sup>1,2</sup>



<sup>1</sup>(中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

<sup>2</sup>(中国科学院大学, 北京 100190)

<sup>3</sup>(国家体育总局 体育科学研究所, 北京 100061)

通讯作者: 李诗锐, E-mail: lishirui@ncic.ac.cn

**摘要:** 快速、低成本、精确的三维扫描技术一直是计算机视觉领域研究的热点. 首先, 分析了新一代 Kinect v2 (Kinect for Windows v2 sensor) 的技术参数、测量原理. 设计实验测得其深度精度与测量距离呈线性变换关系. 其次, Kinect v2 深度数据含有大量的噪声, 尤其是在物体边缘, 常用的双边滤波器等去噪算法不能很好地去除这些噪声, 对此设计了一种有效的去噪算法, 以提高重建质量. 最后, 实现了一套基于新一代 Kinect v2 的重建系统. 实验结果表明, 该重建系统能够实时、精确地重建物体, 可以广泛应用于低成本快速三维成型.

**关键词:** ToF; Kinect; 三维重建; 点云去噪; 点云配准

**中图法分类号:** TP391

中文引用格式: 李诗锐, 李琪, 李海洋, 侯沛宏, 曹伟国, 王向东, 李华. 基于 Kinect v2 的实时精确三维重建系统. 软件学报, 2016, 27(10): 2519-2529. <http://www.jos.org.cn/1000-9825/5089.htm>

英文引用格式: Li SR, Li Q, Li HY, Hou PH, Cao WG, Wang XD, Li H. Real-Time accurate 3D reconstruction based on Kinect v2. Ruan Jian Xue Bao/Journal of Software, 2016, 27(10): 2519-2529 (in Chinese). <http://www.jos.org.cn/1000-9825/5089.htm>

### Real-Time Accurate 3D Reconstruction Based on Kinect v2

LI Shi-Rui<sup>1,2</sup>, LI Qi<sup>1,2</sup>, LI Hai-Yang<sup>1,2</sup>, HOU Pei-Hong<sup>1,2</sup>, CAO Wei-Guo<sup>1</sup>, WANG Xiang-Dong<sup>3</sup>, LI Hua<sup>1,2</sup>

<sup>1</sup>(Key Laboratory of Intelligent Information Processing (Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100190, China)

<sup>3</sup>(China Institute of Sport Science, General Administration of Sport, Beijing 100061, China)

**Abstract:** Fast, low-cost, and accurate three-dimensional scanning technology is a hot research topic in computer vision. First, this paper introduces the technical parameters and measurement principle of the Kinect v2 (Kinect for windows v2 sensor) and analyzes the advantages and disadvantages of Kinect v2 comparing to the first generation Kinect (Kinect v1). Secondly, considering the characteristics of Kinect v2 depth data noise, it designs a denoising method for depth data, and applies this method to the 3D reconstruction system. Finally, it implements a real-time accurate 3D reconstruction system based on Kinect v2. Experimental results show that the proposed 3D reconstruction system can acquire an accurate 3D object model in real-time.

**Key words:** ToF; Kinect; 3D reconstruction; point cloud denoising; point cloud registration

\* 基金项目: 国家自然科学基金(61227802, 61379082); 国家体育总局体育科学研究所合作项目(20145030); 国家财政部基本科研业务费(基本 14-02)

Foundation item: National Natural Science Foundation of China (61227802, 61379082); Project Funded by China Institute of Sport Science (20145030); Basic Research Funding from Ministry of Finance of the People's Republic of China (Basic 14-02)

收稿时间: 2016-01-21; 修改时间: 2016-03-25; 采用时间: 2016-05-09; jos 在线出版时间: 2016-08-08

CNKI 网络优先出版: 2016-08-09 15:38:17, <http://www.cnki.net/kcms/detail/11.2560.TP.20160809.1538.018.html>

快速、低成本、精确、方便地获取物体三维模型一直以来都是计算机视觉与图形学领域的研究热点.基于图片的被动式的三维重建虽然能够方便、低成本地获取到物体的三维模型,但其特征提取和匹配依赖于光照环境和物体表面的纹理特征,且重建点云大多较为稀疏.Kinect v2 的全称是 Kinect for Windows v2 sensor,是微软第二代 Kinect 外设,采用了与第一代完全不同的 ToF(time-of-flight)测量技术.Kinect v2 和第一代 Kinect v1 都是作为微软游戏机 Xbox 深度传感器硬件的外设,主要设计用于获取、分析用户的动作和语音.由于它能够实时、精确地获取物体的深度和彩色图像信息,Kinect 系列产品也可以作为一种理想的三维重建外设,虽然重建精度不能达到工业制造水平,但是利用它重建出来的三维模型依然可以广泛应用于机器人定位导航、虚拟现实、产品展示、影视游戏等等.且 Kinect 用于三维扫描,具有快速、低成本、精确、方便等优点.

基于 Kinect v1 三维重建已有大量优秀的工作参考文献综述<sup>[1-3]</sup>,Newcombe 提出的 KinectFusion<sup>[4]</sup>算法利用 Kinect v1 能够实时地对任意复杂室内场景在任意光照条件下进行精确重建.该算法在深度图预处理阶段只使用了双边滤波器.Roth<sup>[5]</sup>,Whelan<sup>[6]</sup>,Chen<sup>[7]</sup>将 KinectFusion 算法扩展到适用于更大场景的重建.Henry<sup>[8,9]</sup>同时利用 Kinect v1 的彩色和深度信息实现了室内场景的稠密三维建模,该算法虽能得到带纹理的三维模型,但得到的最后三维模型比较粗糙,不适用于重建少纹理结构的物体,也不能实时重建.刘鑫<sup>[10]</sup>基于 Kinect v1 搭建了一套快速物体重建系统,该系统需要一个能自动控制转速的旋转平台,离线的全局优化操作,需要旋转平台给定摄像机外参数.周瑾<sup>[11]</sup>使用 4 个 Kinect v1 构建了一套快速三维人体重建系统.该系统内置了一个通用人体三维模型,此方法不能用于重建除人体以外的其他物体.Tong<sup>[12]</sup>利用 4 个 Kinect 同步采集人体数据重建人体表面,利用简人体模板进行最后的全局优化.达到了很好的人体重建效果.

不同于第一代 Kinect 采用的结构光深度测量技术,Kinect v2 采用了 ToF 深度测量技术.Kinect v2 提供了更高的数据获取速度,达到了每秒 60 帧,拥有更高的深度数据保真度,更大的视角,其彩色相机也更为出色.实验结果显示其深度精度随距离呈线性函数变化.而 Kinect v1 采用的结构光技术深度精度随距离呈二次多项式函数变化.实验结果显示,在更大深度有效范围内,Kinect v2 的精度优于 Kinect v1.基于此,研究使用基于 Kinect v2 进行物体三维重建有重要的意义.

本文搭建了一套基于 Kinect v2 的三维重建系统.首先详细分析新一代 Kinect v2 的技术参数、测量原理、深度数据精度以及相对于第一代产品的优缺点;其次,针对于 Kinect v2 的深度数据噪声的特点,设计一套深度图去噪声流程,并将此去噪声方法应用于三维重建系统中,以提高重建精度;最后,实现一套基于新一代 Kinect v2 三维重建系统.实验结果表明,本文中的系统能够实时、精确、方便地对物体进行三维重建.

## 1 Kinect v2 介绍

Kinect v2 是微软第二代 Kinect,搭载于微软新一代游戏机 Xbox one 上,用于感知 3D 图像和音频,Kinect v2 的 ToF 处理器芯片是 TSMC 0.13 1P5M<sup>[13,14]</sup>.Kinect v2,如图 1 所示,包含一个彩色相机、深度相机、红外光发射器.通过 Kinect v2 可以获得场景的深度、彩色、红外图像信息.



Fig.1 Hardware and raw data of Kinect v2

图 1 Kinect v2 硬件结构图和原始数据

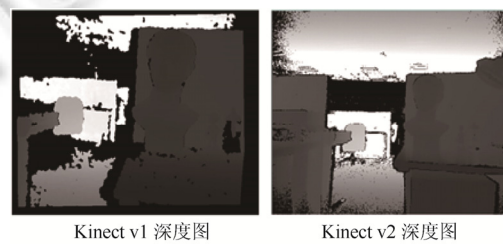


Fig.2 Depth images of two generation Kinect

图 2 两代 Kinect 深度图比较

## 1.1 Kinect v2技术参数

表 1 是两代 Kinect 的技术参数对比.相比于 Kinect v1,Kinect v2 拥有更出色的彩色相机,彩色相机分辨率为 1920×1080,高于第一代的 1280×960,且色彩的保真度更高.Kinect v2 视角更大,其水平视角为 70°,垂直视角为 60°;Kinect v2 有更高的数据传输速度,最大支持深度每秒 60 帧的深度数据获取.Kinect v2 有 3 倍于 Kinect v1 的深度保真度,使得 Kinect v2 能够更容易获取物体更多的细节,如图 2 所示,可以看出,Kinect v2 深度图更精细.Kinect v2 深度数据还有运动模糊小、高动态范围等优点.但是,Kinect v2 的深度分辨率为 512×424,小于 Kinect v1 的 640×480,且深度数据含有更多的噪声数据,关于 Kinect v2 的噪声在下文中会有详细介绍.

Table 1 Technology specification of Kinect v1 and v2

表 1 Kinect v1 与 Kinect v2 的技术参数对比

项目	Kinect v1	Kinect v2
视角	水平 57.5°垂直 43.5°	水平 70°垂直 60°
深度有效值	0.5~5.0m (近场 0.35m)	0.5~5.0m
彩色相机分辨率	1280×960	1920×1080
深度相机分辨率	640×480	512×424
红外相机分辨率	1280×960	512×424

## 1.2 Kinect v2测量原理

Kinect v2 采用 ToF 测距,利用方波调制相机光源,光源的平均频率为 80MHz.通过相位检测来得到发射光和经过物体反射后的光的相位偏移和衰减,从而计算光从光源到物体表面然后再回到传感器的总的飞行时间,根据光的往返飞行时间进而可以求得物体到传感器的距离.通过以下公式(1)计算深度.

$$2d = \frac{\text{phase}}{2\pi} \cdot \frac{c}{f} \quad (1)$$

$d$  为深度, $\text{phase}$  是调制信号相位偏移, $c$  为光速,光在空中的飞行速度约为  $c=3 \times 10^8 \text{m/s}$ , $f$  为传感器的调制频率.

## 2 Kinect v2 深度图去噪

本文针对目标物体和大规模复杂场景进行三维建模.对于目标物体的重建,我们利用直通滤波器选择出特定视场内的物体进行重建,这样既可以避免不必要物体的干扰,又能去除 Kinect v2 质量较差的视场边缘数据,最后选择最为理想的深度数据进行特定物体三维重建.假设深度图表示为  $(u,v,d)$ ,其中, $u,v$  表示深度图的像素坐标, $d$  表示深度.直通滤波器选择满足条件的深度值进行重建.假设深度图中心点的坐标为  $(u_0,v_0)$ ,深度图的长为  $w$ ,宽为  $h$ ,则我们可以通过式(2)选择深度图像中心周围处的深度值.

$$\begin{cases} u_0 - (w/2) \times r \leq u \leq u_0 + (w/2) \times r \\ v_0 - (h/2) \times r \leq v \leq v_0 + (h/2) \times r \end{cases} \quad (2)$$

其中, $r$  表示保留的比率值, $r=1$  表示保留整个深度值.还可以设置深度范围阈值,假设用  $\max$  表示允许的深度的最大值, $\min$  表示最小值,则通过  $\min < d < \max$  可以选择有效的深度值范围.如图 3(c)所示,即为原始深度图 3(d)经过直通滤波器处理后的结果.

如图 3(c)、图 3(d)所示,Kinect v2 包含较多的噪声数据,尤其是在物体边缘处.类似于 Rusu<sup>[15]</sup>利用物体表面具有平滑性特征,我们可以通过统计物体表面的形状特征,去除引起表面突变的噪声点.不同点在于,Rusu<sup>[15]</sup>只利用单一点云数据的统计特征处理自身噪声.我们在利用历史数据流进行统计物体表面特征后再对当前帧进行去噪声处理.本文基于 KinectFusion 算法加权融合数据流,可以从多个噪声点云加权融合得到低噪声表面.虽然没有从根本上去除噪声误差累计的影响,但是我们可以利用这个光滑表面作为物体表里平滑程度的先验知识,从而修正后面数据流中的噪声,从根本上解决误差累计的问题.假设当前输入为第  $k$  帧数据流,我们统计利用前  $k-1$  帧重建的物体模型的表面光滑程度,如图 3(a)所示.基于时间效率上的考虑,我们不直接在重建表面上进行统计,而是利用光线追踪算法将已经构建好的表面重新投影到  $k-1$  帧相机深度图上,如图 3(b)所示,利用深度图统计此时物体表面的光滑程度.显然, $k-1$  帧融合的数据比当前帧数据更能表示物体表面. $k-1$  帧和  $k$  帧之间只有细微的移动,这样利

用  $k-1$  帧摄像机位置获取的物体表面就能很好地代表  $k$  帧物体表面的光滑特征,从而有效地去除噪声。

以重投影的深度图中每个深度值  $(u_i, v_i, d_i)$  为窗口中心,选取窗口大小为  $winSize \times winSize$  的窗口区域,计算此窗口区域中所有深度  $d$  分布的方差  $\sigma_i$ , 设  $n = winSize \times winSize$ .

$$\sigma_i = \frac{1}{n-1} \sqrt{\sum_{j=0}^n (d_{ij} - \mu)^2} \quad (3)$$

$d_{ij}$  表示以  $d_i$  为中心的窗口中的相邻元素的深度值,  $\mu_i$  为窗口中深度值的平均值. 将所有的方差组成方差集合  $s = \{\sigma_1, \sigma_2, \dots, \sigma_k\}$ , 此时, 每个深度值对应一个  $\sigma_i$ . 计算此方差集合的平局值和方差  $\mu_s, \sigma_s$ . 保留满足公式(6)方差对应的深度值, 其中,  $t$  为一个浮点数阈值. 如图 3(e)所示为滤波之后的效果.

$$\mu_s = \frac{1}{n} \sum_{i=0}^k \sigma_i \quad (4)$$

$$\sigma_s = \frac{1}{n-1} \sqrt{\sum_{i=0}^k (\sigma_i - \mu_s)^2} \quad (5)$$

$$\mu - t \times \sigma < \sigma < \mu + t \times \sigma \quad (6)$$

在经过以上两种方法滤波之后,可能还包含一些孤立的噪声点,如图 3(e)所示.以深度图中每个深度值  $(u_i, v_i, d_i)$  为窗口中心,选取窗口大小为  $winSize \times winSize$  的窗口区域,计数此窗口区域中有效的深度值数  $n$ , 当  $n$  小于某个阈值时,我们认为这个点为孤立的一个点,可以将此孤立点过滤,如图 3(f)所示为我们最后得到的效果.

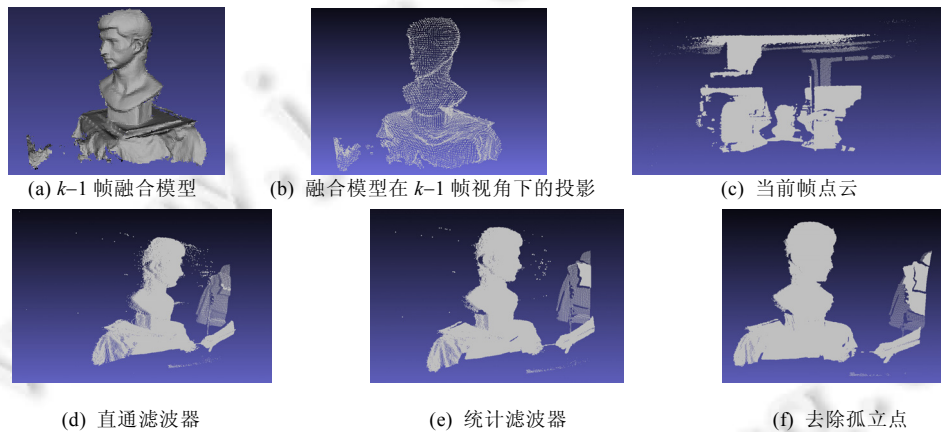


Fig.3

图 3

### 3 Kinect v2 标定

由于只用到了深度图重建物体的三维形状,本系统可以在任意室内光照条件下工作.我们对 Kinect v2 的深度相机进行标定.针对 Kinect v2 的深度相机,采用张正友<sup>[16]</sup>标定法,并利用标定工具<sup>[17]</sup>进行标定.使用 20 个不同角度的深度相机获取的红外图片进行标定.表 2 是标定得到的深度相机内参数.

Table 2 Intrinsic parameters of depth camera

表 2 深度相机内参数

焦距(focal length)	fx	366.66mm
	fy	366.63mm
主点(principal point)	u	261.96
	v	207.61
畸变系数(distortion)	k1	0.082 67
	k2	-0.205 45
	k3	-0.000 46
	k4	0.000 30
	k5	0

## 4 Kinect v2 三维重建

本文在构建基于 Kinect v2 的重建系统时,整体的三维重建算法遵循了 KinectFusion<sup>[4]</sup>的算法框架.在此基础上,本系统的主要贡献是:第一,利用 Kinect v2 SDK 提供的基本的 API 驱动,构建了重建系统的深度数据获取模块.第二,增加了深度数据噪声处理、深度摄像机标定模块.第三,实现了一套完整的低成本实时精确三维扫描解决方案.

### 4.1 KinectFusion算法简介

KinectFusion<sup>[4]</sup>基于 Kinect v1,能够实时地对任意复杂室内场景在变化的光照条件下进行精确重建.整个 KinectFusion 算法框架包含 3 个主要模块.首先,物体表面测量.根据深度相机返回的深度图和深度传感器的内参得到物体表面顶点坐标和法线信息.第二,相机姿态估计.整个姿态估计是一个点到平面(point-to-plane)的 ICP 算法框架<sup>[18]</sup>.采用帧与模型(frame-to-model)的点云选取策略.将当前帧点云与整个融合模型在前一帧位置的光线投影侧面点云进行配准,而不是将当前帧与前一帧点云进行配准.这种策略在一定程度上能够解决闭合环路(loop closure)问题.由于 Kinect 能够提供每秒 30 帧的深度数据,那么相邻两帧之间运动很小,在点与点匹配的选取上,采用了投影数据结合(projective data association)<sup>[19]</sup>算法.第 3 个主要模块是场景表示.整个物理场景利用一个全局三维体<sup>[20]</sup>表示,每个体素单元中包含一个距离值和一个权值.距离值采用截断符号距离函数(TSDF)表示空间某点与物体表面的距离,当 TSDF 值为 0 时表示物体表面,TSDF 值为负表示物体内部,TSDF 值为正表示物体到传感器之间的可见空间.当新一帧的测量数据产生时,根据姿态估计将当前帧的数据 TSDF 值和权值加权融合到全局三维体中,最后利用移动立方体(marching cubes)算法得到最终的三维网格模型.

KinectFusion 算法能够实时、精确地获得物体的三维模型,这在业界引起了广泛的研究与应用.目前,KinectFusion 算法集成到了微软 Kinect 的 SDK 中,但是源代码并没有公开,PCL(Point Cloud Library)<sup>[21]</sup>提供了基于 Kinect v1 的 KinectFusion 算法开源实现,但其目前并不支持 Kinect v2 的数据读取.据我们所知,现在还没有看到基于 Kinect v2 的物体重建开源系统.

### 4.2 Kinect v2物体重建系统

Kinect v2 对平台硬件软件要求比较苛刻,操作系统需 Win 8/Win 8.1,软件开发平台为 Visual Studio 2012 以上,显卡支持 DX11,内置的 USB 3.0,内存最好在 4G 以上,CPU 要求 64 位,最好是 i7 2.5GHz 以上.本文系统构建的环境是,Win 8.1/Visual Studio 2012/GeForce GTX 660/USB3.0/8G/i7-4770k 3.5G.

图 4 是整个重建系统的算法框架.在 PCL<sup>[21]</sup>开源 KinectFusion 算法的基础上修改构建而成,修改了数据获取模块,增加了深度图滤波和摄像机标定模块.PCL 及其 PCL 依赖的开源库(vtk,boost,qull,flann,eigen)都要在 Visual Studio 2012 环境下编译使用.以下是各个步骤的详细介绍.

1) 获取深度数据模块,PCL 在实现 KinectFusion 算法数据获取时采用了 OpenNI,OpenNI 现在不支持 Kinect v2,且目前 OpenNI 网站已经关闭.我们使用 Kinect for Windows SDK 2.0 提供的接口函数来构建数据获取模块.将此模块替换原有的 OpenNI 深度数据获取模块.

2) 深度图滤波模块,主要实现了本文设计的去噪算法.

3) 摄像机标定,采用张正友标定算法对深度相机进行标定,获得摄像机标定参数,用于将深度图转化为点云.

4) 测量表面获取,根据深度数据和标定参数,计算获取表面顶点和顶点的法线信息.

5) 姿态估计,根据对全局模型做光线投射得到的预测表面与新输入的测量表面进行点到平面的 ICP 算法,得到当前摄像机的位置.

6) 融合,将当前帧的数据和摄像机的位置信息结合,从而将当前数据帧融合到全局模型中.

7) 表面预测,根据前一帧摄像机的位置,投影全局的 TSDF 模型得到模型的前一帧位置的一个侧面.

整个算法实时运行,能够达到每秒 25 帧左右,最后对全局模型使用移动立方体算法得到最终的三维网格模型.



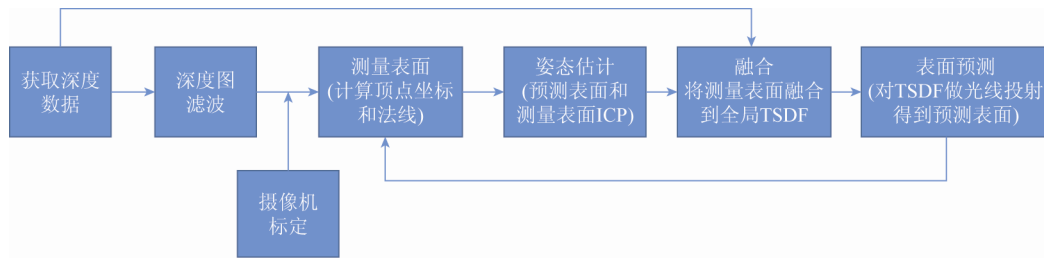


Fig.4 The framework of Kinect v2 reconstruction system

图4 基于 Kinect v2 的重建系统算法框图

## 5 实验结果

### 5.1 Kinect v2深度测量精度分析

为了确定 Kinect v2 的随机误差与距离的关系,我们采用平面拟合<sup>[22,23]</sup>来进行测试.选取办公室墙面作为被采集的平面,将传感器垂直于墙面放置.虽然 Kinect v2 有效空间最小可以在 0.5m 以内,但是这个距离拍摄的墙面含有较大的空洞(即没有深度数据),实验从 0.7m 到 4.9m 向后移动 Kinect v2 深度传感器,间隔 0.6m 获取一次墙面点云,总共采集 8 次,如图 5(a)所示.对于获取到的点云数据,我们选取特定区域采用 Geomagic 进行平面最佳拟合,得到点云到拟合平面距离分布的标准差,也就是残差的分布情况,如图 5(b)所示.

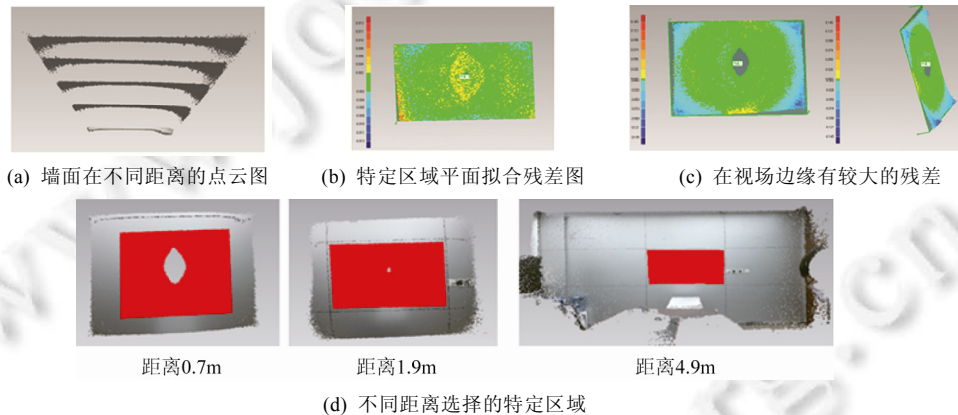


Fig.5

图5

对于特定区域的选取,前两次选择视角中间区域合适大小的区域,后面 6 次选择墙面的同一个区域,如图 5(d)所示.这样选择的原因是,第一,我们选取的墙面区域比较大,这样就使得选取的墙面区域在传感器距离比较远时仍然能够占据视场较大的区域.第二,实验发现,在视场边缘处有较大的残差,如图 5(c)所示,这个残差由于在 ToF 发射的红外光在视场边缘位置反射回来的光振幅较小<sup>[24]</sup>和深度相机的畸变系数所致.选取视场中间部分可去除视场边缘点云的影响,图 7 是我们最后得到的 Kinect v2 标准差与距离的关系.

为了与第一代 Kinect 相对比,我们采用相同的方式得到了 Kinect v1 平面标准差与距离的关系,如图 6 所示.实验结果表明,Kinect v2 标准方差与距离基本呈线性变化,而 Kinect v1 标准方差与距离呈二次多项式变化.这是由设计原理所致,Kinect v1 原理本质上体现为立体视觉原理,其深度分辨率与距离呈二次曲线变化关系.而 Kinect v2 采用的是 ToF 技术,精度依赖于光源频率和信噪比.在更大深度范围内,Kinect v2 的精度大大优于 Kinect v1,这表明,Kinect v2 比第一代更适合远距离大场景重建,如图 8 所示,将两代 Kinect 固定于同一位置,对距离 2m 的人脸模型进行重建,Kinect v2 重建出来的模型明显比 Kinect v1 重建出来的模型更为精细.

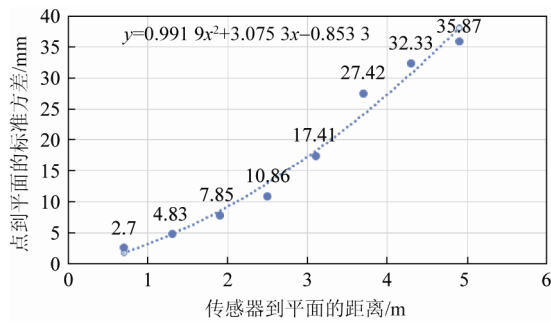


Fig.6 Relationship between standard deviation and deviation for Kinect v1

图 6 Kinect v1 标准差与距离的关系

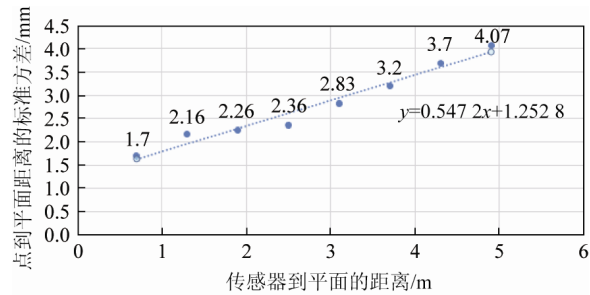


Fig.7 Relationship between standard deviation and distance for Kinect v2

图 7 Kinect v2 标准差与距离的关系

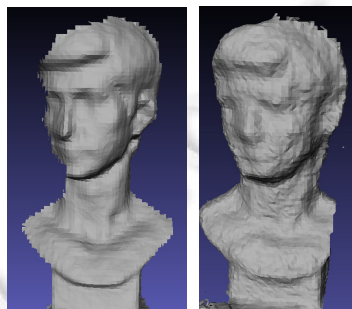


Fig.8 Reconstruct far away object, the left one is reconstructed by Kinect v2 and the right one is reconstructed by Kinect v1, the distance between object and sensor is 2m around

图 8 对远处物体的重建,左图是 Kinect v2,右图是 Kinect v1,物体距离传感器为 2m 左右

## 5.2 物体重建

对特定物体实时高精度重建,我们的这个系统支持两种方式,一种是手持 Kinect 传感器绕着固定物体旋转,另一种是将特定物体置于旋转平台上,将 Kinect 传感器固定,对旋转平台进行旋转.我们选择整个场景表示的三维体物理大小  $1 \times 1 \times 1 \text{m}^3$ ,体素分辨率为  $512 \times 512 \times 512$ ,将物体置于传感器 1m 左右的位置.图 9 所示为实验结果.其结果表明,对于第 1 种方式,能够从更多的视角在不同的深度上扫描被重建的物体,相较于第 2 种方法,重建结果包含更多的细节信息.但是,由于运动的不规则性,鲁棒性不如第 2 种方法,且在方便性上也不如第 2 种方法.Kinect v2 与 Kinect v1 重建的结果相比较,在这种近距离下没有明显的不同.

通过本系统可以方便地对物体进行快速、精确的三维重建,选取了有代表性的常见物体进行重建,包括:具有简单表面的抱枕和花盆,具有复杂表面的玩具老虎,还有具有现代设计感的电风扇,以及应用广泛的人脸和脚模型,如图 12(a)所示为这些物体的彩色图片,图 12(b)是这些物体重建后的结果.我们采用旋转台的方式,手动驱动旋转台,旋转一周所花时间大概在 15~40s 之间.本系统所有算法都采用 GPU 并行加速实现,重建的速度能够达到每秒 25 帧左右,能够做到实时重建.所以,对于特定物体重建的时间也在 15~40s 这个区间内,亦即旋转一周转台所花费的时间,表 3 统计了每个物体的实际物理尺寸、重建模型的点个数和重建时间.

接下来的实验对比了增加去噪声算法后的结果.如图 10(a)、图 10(b)所示,边缘噪声对场景重建结果有很大的影响,在物体边缘部分明显有错误重建的地方.实验结果显示,去噪算法能够有效地去除这种噪声.如图 10(c)、图 10(d)所示,去噪算法能够有效地减少表面噪声.

将本系统的重建结果与微软提供的 Kinect v2 KinectFusion SDK 算法进行比较.利用 Kinect Studio 采集数据,以相同的数据流作为输入,如图 11(a)、图 11(c)所示为微软 KinectFusion SDK 重建结果,图 11(b)、图 11(d)

是本系统的重建结果.由图 11(a)、图 11(b)对比可以看出,在旋转一圈重建整个物体时,微软 SDK 含有较为严重的误差累计.由图 11(c)、图 11(d)可以看出,微软 SDK 不能重建较为精细、复杂的物体,而本系统成功地对此物体进行了重建.

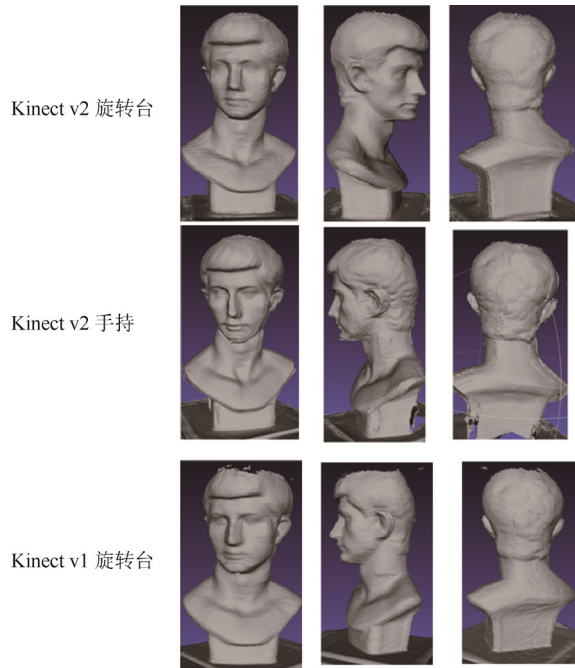


Fig.9 Reconstruct object on rotating platform or by handheld Kinect

图 9 物体置于旋转台与手持 Kinect 重建

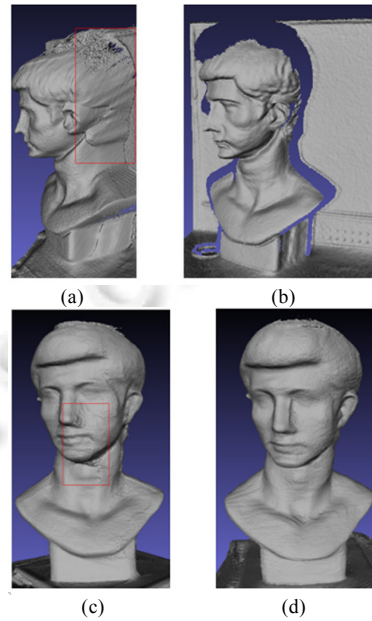


Fig.10 (a)(c) Reconstruction without denoise

(b)(d) Reconstruction with denoise

图 10 去噪声后重建的效果,(a)和(c)是未经去噪处理的重建结果,(b)和(d)是经过去噪处理的结果

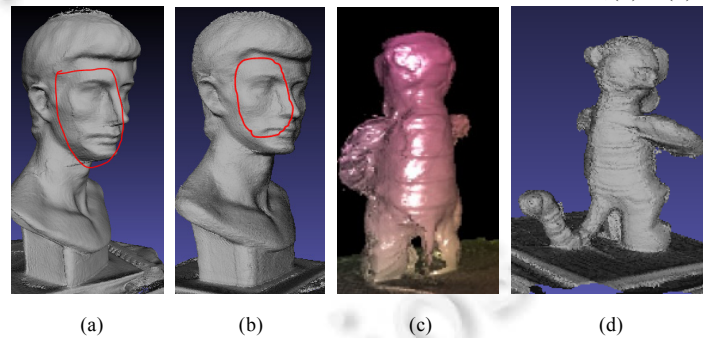


Fig.11 (a)(c) Reconstruction by Microsoft KinectFusion SDK (b)(d) Reconstruction by our method

图 11 (a)和(c)为微软 KinectFusion SDK 重建结果,(b)和(d)是本系统的重建结果

Table 3 Object reconstruction

表 3 重建物体

物体	尺寸(长×宽×高 cm <sup>3</sup> )	重建点个数	重建时间(s)
盆栽	28.5×28.5×37	972 320	20
人脚模型	25×10×12	217 516	18
人脸模型	33×20×50	1 313 543	30
玩具虎	13×13×35	588 880	21
抱枕	43×43×10	1 084 622	19
风扇	30×30×35	837 179	30



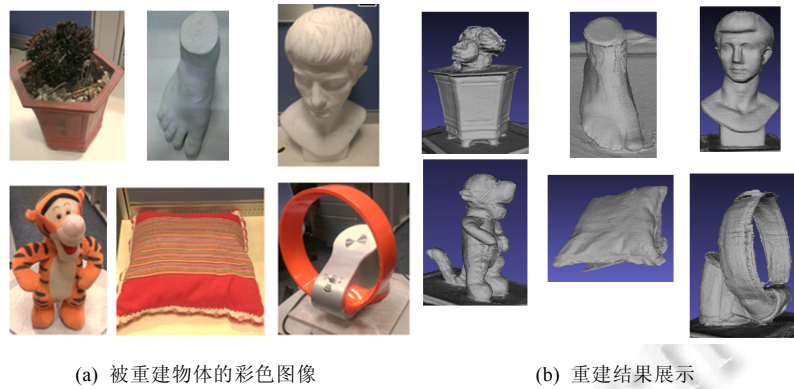


Fig.12

图 12

对于重建好的模型,选取一些测量对象进行测量,如图 13 所示,以花盆为例的测量过程,将测量结果与真实测量值进行比较,统计测量的相对误差见表 4.最后得到的平均相对误差为 1.8%.

最后,我们将本系统用于重建办公桌场景和人体,如图 14 所示.此时,选择场景三维体的物理尺寸为  $3 \times 3 \times 3 \text{m}^3$ ,物体的体素分辨率为  $512 \times 512 \times 512$ .手持 Kinect 进行场景扫描.从实验结果可以看出,Kinect v2 在某些区域有较大的误差,分析发现,Kinect v2 对于电脑屏幕、黑色物体,平面法线与 Kinect v2 视线近垂直时有较大误差.Kinect v2 应用于复杂场景的重建鲁棒性存在问题.我们还将本系统用于重建人体,人坐在旋转椅子上,对旋转椅子进行重建,如图 14 所示.

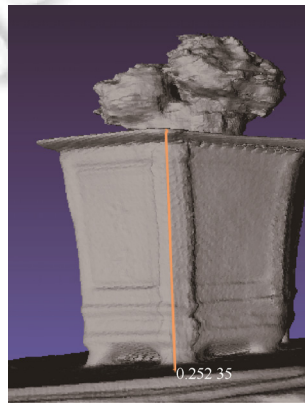


Fig.13 Measure the high of flowerpot using 3D model

图 13 利用三维模型测量花盆的高度

Table 4 Object measurement

表 4 物体测量

编号	待测物体	真实值(mm)	测量值(mm)	绝对误差(mm)	相对误差%(平均值为 1.97)
1	书本厚度	33.0	32.7	0.3	0.91
2	正方体边长	158.0	152.3	5.7	3.61
3	花盆宽度	285.5	276.6	8.9	3.11
4	花盆高度	250.0	252.6	2.6	1.04
5	脚模型长度	250.0	243.9	6.1	2.44
6	脚模型高度	126.8	127.5	0.7	0.55
7	脸模型高度	520.0	528.2	8.2	1.58
8	脸模型宽度	334.2	338.2	4.0	1.19

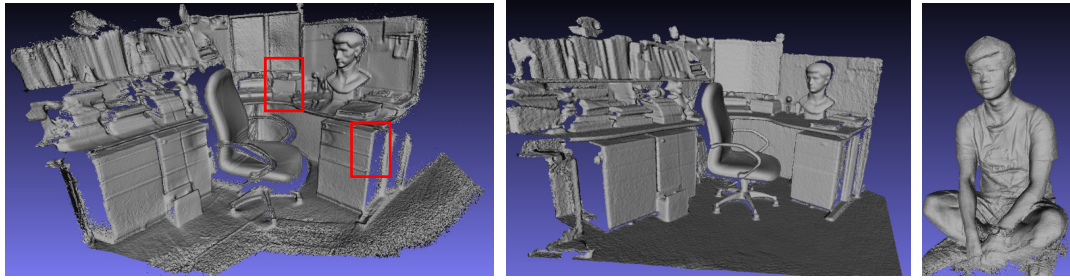


Fig.14 Kinect v2 scene and human body reconstruction

图 14 Kinect v2 场景与人体重建结果

## 6 结 论

通过分析, Kinect v2 深度精度随距离变化呈线性变化,且在较大深度范围内具有较高的精度.本文实现的物体重建系统能够方便地对物体进行实时、精确的重建.针对 Kinect v2 深度图进行设计去噪流程提高了重建的鲁棒性和准确性.我们的系统利用单台 Kinect v2 对物体重建,既支持使用简易旋转平台,又支持手持 Kinect v2 重建物体,且适用于对各种物体重建,包括人体.下一步的工作将重点解决 Kinect v2 在重建复杂大场景时的鲁棒性问题.

## References:

- [1] Berger K, Meister S, Nair R, Kondermann D. A state of the art report on Kinect sensor setups in computer vision. In: Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications. Berlin, Heidelberg: Springer-Verlag, 2013. 257–272. [doi: 10.1007/978-3-642-44964-2\_12]
- [2] Han J, Shao L, Xu D, Shotton J. Enhanced computer vision with microsoft Kinect sensor: A review. IEEE Trans. on Cybernetics, 2013,43(5):1318–1334. [doi: 10.1109/TCYB.2013.2265378]
- [3] Li SR, Tao KL, Wang SY, Li HY, Cao WG, Li H. 3D reconstruction by Kinect sensor a brief review. Computer Aided Drafting, Design and Manufacturing, 2014,24(1):1–11.
- [4] Newcombe RA, Davison AJ, Izadi S, Molyneaux D. KinectFusion: Real-Time dense surface mapping and tracking. In: Proc. of the 10th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR). IEEE, 2011. 127–136. [doi: 10.1109/ISMAR.2011.6092378]
- [5] Roth H, Vona M. Moving volume KinectFusion. In: Proc. of the BMVC. 2012. 1–11. <http://bmv2012.surrey.ac.uk/>
- [6] Whelan T, Kaess M, Fallon M, Johannsson H, Leonard J. Kintinuous: Spatially extended KinectFusion. 2012. [https://www.researchgate.net/publication/279843674\\_Kintinuous\\_Spatially\\_Extended\\_KinectFusion](https://www.researchgate.net/publication/279843674_Kintinuous_Spatially_Extended_KinectFusion)
- [7] Chen J, Bautembach D, Izadi S. Scalable real-time volumetric surface reconstruction. ACM Trans. on Graphics (TOG), 2013,32(4): 113. [doi: 10.1145/2461912.2461940]
- [8] Henry P, Krainin M, Herbst E, Ren X, Fox D. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In: Proc. of the 12th Int'l Symp. on Experimental Robotics ISER. 2010. [doi: 10.1007/978-3-642-28572-1\_33]
- [9] Henry P, Krainin M, Herbst E, Ren X, Fox D. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. The Int'l Journal of Robotics Research, 2012,31(5):647–663. [doi: 10.1177/0278364911434148]
- [10] Liu X, Xu HR, Hu ZY. GPU based fast 3D-object modeling with Kinect. Acta Automatica Sinica, 2012,38(8):1288–1297 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2012.01288]
- [11] Zhou J, Pan JJ, Tong J, Liu LG, Pan ZG. Fast 3D human body reconstruction from Kinects. Journal of Computer-Aided Design & Computer Graphics, 2013,25(6):873–879 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-9775.2013.06.015]
- [12] Tong J, Zhou J, Liu L, Pan Z, Yan H. Scanning 3D full human bodies using Kinects. IEEE Trans. on Visualization and Computer Graphics, 2012,18(4):643–650. [doi: 10.1109/TVCG.2012.56]
- [13] Payne A, Daniel A, Mehta A, Thompson B, Bamji CS. 7.6 A 512×424 CMOS 3D time-of-flight image sensor with multi-frequency photo-demodulation up to 130MHz and 2GS/s ADC. In: Proc. of the 2014 IEEE Int'l Solid-State Circuits Conf. Digest of Technical Papers (ISSCC). IEEE, 2014. 134–135. [doi: 10.1109/ISSCC.2014.6757370]

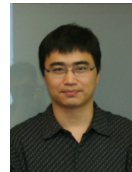
- [14] Sell J, O'Connor P. The Xbox one system on a chip and Kinect sensor. *Micro*, 2014,34(2):44–53. [doi: 10.1109/MM.2014.9]
- [15] Rusu RB, Marton ZC, Blodow N, Dolha M, Beetz M. Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 2008,56(11):927–941. [doi: 10.1016/j.robot.2008.08.005]
- [16] Zhang Z. A flexible new technique for camera calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(11):1330–1334. [doi: 10.1109/34.888718]
- [17] Bouguet JY. Camera calibration toolbox. 2013. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/2013](http://www.vision.caltech.edu/bouguetj/calib_doc/2013)
- [18] Chen Y, Medioni G. Object modelling by registration of multiple range images. *Image and Vision Computing*, 1992,10(3):145–155. [doi: 10.1016/0262-8856(92)90066-C]
- [19] Blais G, Levine MD. Registering multiview range data to create 3D computer objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1995,17(8):820–824. [doi: 10.1109/34.400574]
- [20] Curless B, Levoy M. A volumetric method for building complex models from range images. In: *Proc. of the 23rd Annual Conf. on Computer Graphics and Interactive Techniques*. ACM, 1996. [doi: 10.1145/237170.237269]
- [21] PCL. Point Cloud Library. 2014. <http://www.pointclouds.org/2014>
- [22] Khoshelham K. Accuracy analysis of Kinect depth data. In: *Proc. of the ISPRS Workshop Laser Scanning*, Vol. XXXVIII-5/W12. 2011. 133–138. [doi: 10.5194/isprsarchives-XXXVIII-5-W12-133-2011]
- [23] Smisek J, Jancosek M, Pajdla T. 3D with Kinect. In: *Consumer Depth Cameras for Computer Vision*. London: Springer-Verlag, 2013. 3–25. [doi: 10.1007/978-1-4471-4640-7\_1]
- [24] Foix S, Alenya G, Torras C. Lock-In time-of-flight (ToF) cameras: A survey. *Sensors Journal*, 2011,11(9):1917–1926. [doi: 10.1109/JSEN.2010.2101060]

#### 附中文参考文献:

- [10] 刘鑫,许华荣,胡占义.基于 GPU 和 Kinect 的快速物体重建.自动化学报,2012,38(8):1288–1297 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2012.01288]
- [11] 周瑾,潘建江,童晶,刘利刚,潘志庚.使用 Kinect 快速重建三维人体.计算机辅助设计与图形学学报,2013,25(6):873–879 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-9775.2013.06.015]



李诗锐(1987—),男,重庆人,博士,主要研究领域为计算机图形学,计算机视觉.



曹伟国(1976—),男,博士,助理研究员,CCF会员,主要研究领域为计算机图形学,数字图像处理,模式识别,形状分析.



李琪(1992—),男,硕士,主要研究领域为计算机图形学,计算机视觉.



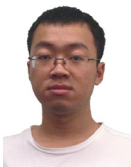
王向东(1973—),男,博士,研究员,博士生导师,主要研究领域为运动生物力学.



李海洋(1986—),男,博士,主要研究领域为计算机图形学,数字图像处理.



李华(1957—),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为计算机图形学.



侯沛宏(1992—),男,硕士,主要研究领域为计算机图形学,计算机视觉.