

# 基于 Kinect 的人体三维重建方法<sup>①</sup>

李翔, 李泽丰, 李治江

(武汉大学 印刷与包装系, 武汉 430072)

通讯作者: 李治江, E-mail: [lizhijiang@whu.edu.cn](mailto:lizhijiang@whu.edu.cn)

**摘要:** 通过 Kinect 体感仪, 实现人体三维重建. 使用 Kinect 体感仪, 扫描获取人体三维数据, 利用深度数据转换算法实现二维顶点的三维化, 再通过红外相机姿态跟踪算法进行顶点集配准, 求解出相机每次的相对位移与转动角度, 实现相机姿态跟踪, 并将每次拍摄到的点集转换到同一全局坐标系下, 使用晶格化显示集成算法将点云集成到提前划分好精度及尺寸的体素晶格中, 最后利用投影映射算法获得可视化的人体三维立体模型. 使用 Kinect 体感仪及三脚架等辅助设备方便快捷地获取人体三维重建结果, 并通过 3D 打印技术对模型进行输出. 该研究实现了人体三维重建中人体扫描、处理、重建、输出全流程.

**关键词:** Kinect; 人体三维重建; 深度数据转换; 红外相机姿态跟踪; 晶格化显示集成; 投影映射; 3D 打印

引用格式: 李翔, 李泽丰, 李治江. 基于 Kinect 的人体三维重建方法. 计算机系统应用, 2018, 27(6): 178-183. <http://www.c-s-a.org.cn/1003-3254/6449.html>

## 3D Reconstruction of Human Body Based on Kinect

LI Xiang, LI Ze-Feng, LI Zhi-Jiang

(School of Printing and Packaging, Wuhan University, Wuhan 430072, China)

**Abstract:** The purpose of the study is to realize 3D reconstruction of human body based on Kinect. We use Kinect to scan and obtain the 3D human body data first. Depth data transformation algorithm converts 2D depth data to the global coordinate system. ICP registration algorithm achieves infrared camera gesture tracking to obtain the states of infrared camera. TSDF algorithm integrates the converted vertex set into the pre-divided voxel lattice. Projection mapping algorithm is used for visualization of the reconstruct results. Using Kinect and auxiliary equipment tripod, we can quickly and easily achieve the results of human body 3D reconstruction and output the model with 3D print. This research realizes the whole process of human body 3D reconstruction including scanning, processing, reconstruction, and 3D model output.

**Key words:** Kinect; human body 3D reconstruction; depth data transformation; ICP; TSDF; projection mapping; 3D print

## 1 引言

随着数字传播技术的发展, 人们越来越不能满足于二维图像带来的视觉效果, 三维立体感更被人们所推崇. 利用三维重建可实现对场景、物体及人体等的立体复制. 其中, 人体三维重建一直有着实际应用价值, 并曾经被普遍应用于各个领域: 在考古领域, 可重建兵马俑、木乃伊等文物<sup>[1]</sup>的三维模型; 在医学领域, 人体

三维重建可用于疾病诊断<sup>[2]</sup>、模拟手术等; 在通讯领域, 人体三维重建可用于基于三维重建与全息投影通讯. 结合 VR/AR 技术, 人体三维重建能够有更普遍的运用, 如虚拟试衣、门禁系统等.

目前, 三维重建的方法主要有: 基于计算机视觉的双视/多视光学成像方法<sup>[3]</sup>利用三维建模软件的方法及基于扫描设备的三维重建. 其中, 基于扫描设备的重建

① 收稿时间: 2017-11-22; 修改时间: 2017-12-15; 采用时间: 2017-12-27; csa 在线出版时间: 2018-05-28

方法可直接获取深度信息、纹理信息,且操作简便,重建效果更真实,因此更适用于人体三维重建.基于红外点云的 Kinect 体感器设备以其价格低廉、操作简单、应用潜力大等优点得到广泛关注,并且可满足一般的精度要求.

Kinect<sup>[4]</sup>是微软公司推出的一款集深度信息获取、骨骼识别、语言识别等多种技术于一身的体感控制器.基于 Kinect 的三维重建已有相干学者对数据处理算法、配准算法、重建算法、纹理贴图等算法进行了相关研究工作.刘鑫等人<sup>[5]</sup>提出了一种结合 Kinect 深度传感器与旋转转台的实时三维重建方案.国防科学技术大学的乔思航等<sup>[6]</sup>使用 3 台 Kinect 摄像机建立了人体三维模型.周瑾等人<sup>[7]</sup>使用分布在不同位置的 4 台 Kinect 传感器同步扫描人体,实现人体的静态三维模型的重构.席小霞<sup>[8]</sup>搭建了一套基于 3 台 Kinect 的体感摄像机的三维图像重建系统.谭红春等人<sup>[9]</sup>、范小娜<sup>[10]</sup>、周文<sup>[11]</sup>、刘洋<sup>[12]</sup>、王欢等<sup>[13]</sup>,从不同方面进行了相关研究,刘雷杰<sup>[14]</sup>提出了一种结合骨架信息进行人物的动态三维模型重建.

以上相关研究中,手持单台 Kinect 采集过程中,由于抖动等原因引入采集误差.固定单台 Kinect,由于固定视场角使得使用场景有所限制.而同时固定多台 Kinect 不仅增加成本,多设备坐标系的融合也使得数据处理过程更加复杂,同时也会引入新的噪声.因此,如何在单台 Kinect 低成本条件下,方便快捷地完成高精度的人体三维数据采集,进而实现人体三维重建中人体扫描、处理、重建、输出,成为本文的研究重点.

## 2 Kinect Fusion 算法流程

本文三维重建方法使用 Kinect V2,并使用其 SDK 中 Kinect Fusion 进行程序开发,实现人体的三维重建.其算法流程为:采集人体单个方向的深度数据,利用深度数据进行配准融合,得到单帧表面重建模型,再经过投影映射得到可视化的重建结果.

1) 深度数据的转换目的是将二维顶点三维化.即把从 Kinect 中获取的原始深度帧数据转换成以米为单位的浮点数据,利用摄像头的坐标信息,将这些浮点数据转换为和 Kinect 相机朝向统一的顶点集,即点云数据.

依据二代 Kinect 红外相机内在的标定矩阵  $K$ ,将特定深度信息值投影变换为以相机为坐标原点空间的

三维顶点  $V(u)$ :

$$v_i(u) = D_i(u)K^{-1}[u, 1] \quad (1)$$

其中  $u$  为已知图像点,  $D(u)$  为深度值.该点法向量表示为  $n_i(u)$ ,再通过相机姿态变换平移矩阵  $T_i$  与旋转矩阵  $R_i$ ,将三维顶点转换到全局坐标,其公式如下:

$$V_i^g(u) = T_i V_i(u) \quad (2)$$

$$n_i^g(u) = R_i n_i(u) \quad (3)$$

2) 姿态跟踪即计算传感器位置和朝向的姿态信息.利用配准算法在摄像头移动时一直获取其姿态,就可以得到当前设备相对于起始帧时的相对姿势.配准利用的是经典 ICP 算法<sup>[15]</sup>,计算得到相邻两帧间的相对位移与旋转,据此能够将相机原点坐标系转换到世界坐标系.具体流程见算法 1.

算法 1. 姿态跟踪算法

1. for 每个图像像素点  $u \in$  深度数据  $D_i$
2. if  $D_i(u) > 0$  then
3.  $V_{i-1} \leftarrow T_{i-1}^{-1} V_{i-1}^g$
4.  $p \leftarrow$  透视投影定点  $V_{i-1}$
5. if  $p \in$  定点图  $V_{i-1}$  then
6.  $V \leftarrow T_i V_i(p)$
7.  $n \leftarrow R_i N_i(p)$
8. if  $\|V - V_{i-1}^g\| < \text{距离阈值}$  and  $Abs(n^*) < \text{法向量阈值}$  then
9. 找到对应点

3) 利用晶格化显示将上述步骤获取的深度数据集成融合.事先预定一个固定三维体分辨率,将其映射到特定的三维物理空间维度.这个体被均匀细分成三维网格体素,即体素晶格.利用 TDFS 算法<sup>[16]</sup>,将全局顶点集成到体素晶格中,得到初步重建结果.随着传感器的移动,深度数据将连续不断的被填充进体素晶格中.具体流程见算法 2.

算法 2. 晶格化深度数据集成融合

1. for 对于每一个处于  $x, y$  体切面的体素晶格  $g$
2. while 从前向后操作
3.  $V^g \leftarrow$  将  $g$  从网格转换到全局坐标系
4.  $V \leftarrow T_i^{-1} V^g$
5.  $P \leftarrow$  透视投影定点  $V$
6. if  $V$  在相机视图的投影范围中 then
7.  $sdf_i \leftarrow \|t_i - V_g\| - D_i(p)$
8. if  $(sdf_i > 0)$  then
9.  $tsdf_i \leftarrow \min(1, sdf_i / \text{max truncation})$
10. else
11.  $tsdf_i \leftarrow \max(-1, sdf_i / \text{min truncation})$
12.  $W_i \leftarrow \min(\text{max\_weight}, W_{i-1} + 1)$

13.  $tsdf^{avg} \leftarrow \frac{tsdf_{i-1}W_{i-1} + tsdf_iW_i}{W_{i-1} + W_i}$
14. 保存  $W_i$  和  $tsdf^{avg}$  到体素  $g$  中.

算法 2 中,  $tsdf_i$  是该相机坐标系点  $V_g$  到本次相机原点  $t_i$  的距离与本次观测深度  $D_i(p)$  的差值. 用 **max truncation** 表示选取的截断范围, 若是差值为正, 表示本次测量的面在晶格的前面,  $tsdf_i$  赋值  $[0, 1]$  之间, 值越接近 0 则越靠近观测面; 如果差值为负, 本次测量的面在该晶格的后面,  $tsdf_i$  赋值  $[-1, 0]$  之间, 越靠近观测面的地方值越接近 0. 选取本次计算值的  $tsdf_i$  的权值  $W_i$ , 加权平均求出  $tsdf^{avg}$ , 将  $W_i$  和  $tsdf^{avg}$  存储在对应的晶格, 进行下个晶格的扫描操作.

经过以上处理, 立方体晶格中存储的  $tsdf$  值表示其位置. 在重建模型外是负值, 模型内部是正值, 模型表面是 0 值的形式. 随着传感器的移动, 深度数据将连续不断的被填充进体素晶格中.

4) 利用基于 GPU 的辐射线方式, 实现和生成在网格体内用于显示追踪的隐式表面. 以并行的方式, 每个 GPU 线程沿着一条辐射线, 呈现一个输出图像的像素点. 给定一个有起始位置和方向的射线, 每个 GPU 线程沿着射线遍历每一个像素点, 并通过检测零交点的位置来提取隐含的表面. 最后, 从传感器视点对新建立立方体进行光线投射, 便能够得到可视化的结果.

通过上述步骤, Kinect Fusion 可实时重建出人体三维模型, 与传统的重建方法相比, 精度与效率都较高, 极大提高了三维重建的便利性, 为三维重建创造了良好的用户体验.

### 3 重建场景流程及去噪优化

#### 3.1 重建场景流程

实验场景流程如图 1 所示: 使用三脚架固定 Kinect, 采集对象站立在可旋转台面上. 通过旋转台面, 对人体进行 360 度采集重建; 本文重点重建区域为人体胸部以上部分, 因对人体而言, 该部分辨识度最高. 为获得更好的精度与重建结果, Kinect 与三脚架的距离保持在 0.5-1.0 米之间, 同时可通过三脚架高度与角度调节, 控制 Kinect 采集视角.

#### 3.2 重建场景的去噪优化

本文重建目标为人体重建, 在深度数据采集过程中, 视场范围内的场景信息对人体影响较大. 主要表现在人体与背景边界处由深度跳跃产生的噪声数据, 如

图 2(a) 所示, 人体边界部分产生明显噪声. 对于该部分噪声, 我们通过设置深度阈值<sup>[17]</sup>对其进行去除. 假设用 **max** 表示深度相机的最大采集距离, **min** 表示最小采集距离, 则可通过下式选择适当的阈值  $d$ :

$$\min \leq d \leq \max \tag{4}$$

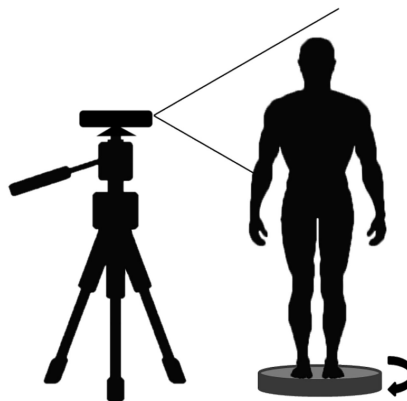


图 1 Kinect 采集实验场景



图 2 去噪优化

通过深度阈值, 可将视场内相较于人体的前景数据与背景数据全部去除, 如图 2(b) 所示, 去除背景之后, 人体边界处相较图 2(a) 由深度跳跃产生的噪声去除明显, 细节如图 2(d) 和 (e) 所示, 噪声去除后的人体边界更加平滑.

除环境背景对三维重建的影响外, 人体深度范围内的其他物体对重建结果也会产生干扰. 如图 2(b) 所示, 与人体在相同深度范围内的物体, 无法通过设置深度阈值去除. 在此, 我们利用直通滤波器去除该部分干扰.

假设深度图像素点表示为  $(x, y, d)$ , 其中  $x, y$  表示深度图像素坐标,  $d$  为深度大小; 深度图的大小为  $(w, h)$ , 中心点坐标为  $(x_0, y_0)$ . 则满足公式 (5) 的深度图区域, 为

重建所需的目标人体区域,即实现满足重建要求的直通滤波器范围。

$$\begin{aligned} x_0 - (w/2) \times r \leq x \leq x_0 + (w/2) \times 2 \\ y_0 - (h/2) \times r \leq y \leq y_0 + (h/2) \times 2 \end{aligned} \quad (5)$$

在该范围内的物体将被保留,在该区域外的物体将被去除,结果如图2(c)所示,成功去除了重建场景内的其他物体噪声干扰。

## 4 实验

### 4.1 实验设备与环境

#### (1) 采集设备: Kinect 二代体感器

Kinect 二代体感器是微软推出的体感周边外设。该体感器不仅能够获取彩色视频信息,还能够通过深度相机获取场景、物体的深度信息,利用这个功能,可实现人体的三维重建,其低廉的价格是其能够普及应用的一大优势。并且其精度也可满足一般消费者需求。

Kinect 二代体感器外观如图3。通体为黑色,长250 mm,宽85 mm,高65 mm,重1.25 kg。相对于一代,Kinect 二代体感器在深度信息采集、视频清晰度、相机视场等方面都有显著提高。



图3 Kinect 二代体感器外观

#### (2) 采集辅助设备: 三脚架

为避免手持体感仪采集数据时,由于抖动等带来的误差,本实验使用三脚架作为辅助设备(见图4)。同时,也更好地控制体感仪与采集对象的距离,提高采集效率,标准化三维数据采集流程。该三脚架可通过中轴升降装置可将 Kinect 采集高度扩展到0.5~1.53m,且手柄可实现360°旋转,大大扩展了 Kinect 设备的采集视场,提高采集范围,使得 Kinect 有了更好的适应性。如下图所示,体感仪固定在三脚架液压云台上。

#### (3) 开发环境

利用二代 Kinect 体感器进行开发时,它对硬件及操作系统要求较高。本文在 Windows8 操作系统下进行程序开发、数据采集处理。机器配置如下: Intel(R) Core(TM)i5-4570R CPU @2.70 GHz; 内存 (RAM) 为

8.00 GB; 应用程序开发环境为 Visual Studio 2012。



图4 安装三脚架后的 Kinect 二代体感器

#### (4) 输出设备: 3D 打印机

为了更好的观察重建结果,文采用3D打印机将重建结果进行输出。3D打印是一种以数字模型为输入,通过逐层堆叠的方式将模型以实体材料构造打印输出的技术。由于这种技术与印刷中的打印技术相像,均是将输入信息通过另一种材料输出的过程,故而称为3D打印。但它与印刷技术又大有不同,印刷中的打印机是用油墨来重现二维的文字、图形图像信息,而3D打印则是用一定材料以堆叠的形式重现出三维立体模型。



图5 3D 打印机

如图5所示,本文所用打印机型号为创想三维 CR-10,材料为 PLA,使用热熔堆积固化成型技术对模型进行输出,精度可达0.1 mm。

### 4.2 实验结果

利用 Kinect 采集到的深度数据经过转换后, 利用相机姿态统一到同一坐标系下进行晶格化显示集成, 最后经过投影映射为可视化图像, 至此, 三维重建基本实现, 其效果可通过相关软件来查看, 本文使用 3DMax 来查看实验结果, 即三维重建结果。

实验在无室外环境光影响的室内进行, 多组实验后将结果通过 3DMax 可视化. 结果显示人体衣服的褶皱细节丰富生动, 表面光滑平整, 可以区分出衣领等部位; 面部五官中, 鼻子和嘴的重建结果较为稳定, 头发部分无孔洞出现. 整体重建效果较好. 结果如图 6、7 所示.

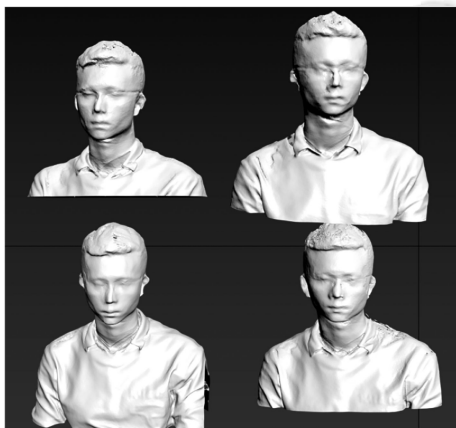


图 6 重建结果正面

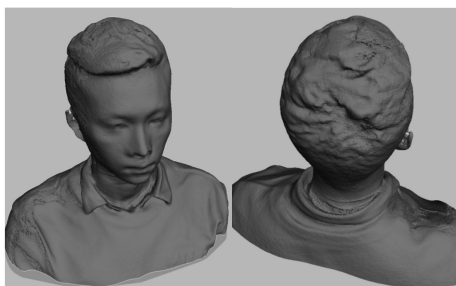


图 7 重建结果侧面与背面

如图 8、图 9, 使用 3D 打印技术对模型输出, 衣领、五官及头发等细节展现较好, 能够满足日常 3D 打印需求.

至此, 本文完成了包括数据采集、数据处理、三维重建、结果输出等的完整人体重建及打印输出流程.



图 8 3D 打印输出结果图



图 9 3D 打印结果细节

### 4.3 实验结果评价

本文对人体三维重建的结果采用客观评价与主观评价两种评价方法, 如表 1、2 所示. 客观评价包括重建设备成本、重建速度、重建精度等指标, 主观评价通过对重建结果视觉感知上的相似性打分, 作为评价指标.

表 1 客观评价指标

重建设备成本	单台 Kinect+三脚架
重建速度	45 s
重建精度	512 晶格 (1.9531 mm 精度)

表 2 主观评价指标

	1	2	3	4	5	平均
本文	0	3	9	5	3	3.40
文献[7]	9	6	4	1	0	1.85
文献[11]	3	4	7	4	2	2.90

本文重建方法, 使用单台 Kinect 和三脚架辅助设备完成采集过程, 相对于文献[6-8]中的重建系统中 3-4 台 Kinect 设备, 降低了成本, 且增加了系统稳定性. 由于 Kinect Fusion 采用实时重建, 相比于文献[5]中方

法,重建速度更快.本文在晶格化显示时采用512晶格,在保证重建速度的基础上,也有着很高的重建精度.

在主观评价中,邀请20位同学,对以上4个方法重建结果进行打分:1分表示很不像,5分表示很像.对比主观评价,本文重建结果在视觉感知上较符合被重建人.

## 5 结语

本文实现了完整的基于Kinect二代体感器的人体扫描、处理、重建、输出方法.介绍了基于深度数据的实时重建算法,对处理好的深度数据,采用数据转换方法,利用深度相机参数,将平滑后的深度数据转换.再利用ICP算法进行顶点集配准,求解出相机每次的相对位移与转动角度,实现相机姿态跟踪.利用得到的相对位移与转动角度,将每次拍摄到的点集转换到同一全局坐标系下,利用TSDF算法将点云集成到提前划分好精度及尺寸的体素晶格中,然后利用投影映射获得可视化的三维立体.

同时,本文通过对采集场景的去噪优化,提高了本文方法的环境适应性.三脚架的使用,优化了数据采集流程,消除相机抖动误差,保证采集视场,也使得重建过程更加方便快捷.从实验结果来看,重建结果较好,能够满足普通消费者对人体三维重建的基本要求.最后采用3D打印机将重建结果实体化输出.

## 参考文献

- 1 邱兆文,张田文.文物三维重建关键技术.电子学报,2008,36(12):2423-2427.[doi:10.3321/j.issn:0372-2112.2008.12.028]
- 2 吕开阳,叶华茂,李晓光,等.Kinect体感技术在动物外科实验教学中的应用及展望.中国医学教育技术,2012,26(2):171-173.
- 3 蔡钦涛.基于图像的三维重建技术研究[硕士学位论文].杭州:浙江大学,2004.
- 4 Smisek J, Jancosek M, Pajdla T. 3D with Kinect. Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain. 2011. 1154-1160.
- 5 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.
- 6 乔思航,程志全,陈寅,等.基于三个Kinect的个性化人体重建.系统仿真学报,2013,25(10):2408-2411.
- 7 周瑾.基于Kinect深度相机的三维人体重建技术研究[硕士学位论文].杭州:杭州电子科技大学,2013.
- 8 席小霞.基于多台Kinect摄像机的三维重建系统研究[硕士学位论文].太原:中北大学,2016.
- 9 谭红春,耿英保,杜炜.一种高效的人脸三维点云超分辨率融合方法.光学技术,2016,(6):501-505.
- 10 范小娜.基于Kinect三维重构及织物模拟研究[硕士学位论文].杭州:浙江大学,2014.
- 11 周文.基于RGB-D相机的三维人体重建方法研究[硕士学位论文].合肥:中国科学技术大学,2015.
- 12 刘洋.基于RGB-D序列及模板的三维人体建模[硕士学位论文].北京:北京工业大学,2014.
- 13 王欢,汪同庆,李阳.利用Kinect深度信息的三维点云配准方法研究.计算机工程与应用,2016,52(12):153-157.[doi:10.3778/j.issn.1002-8331.1407-0506]
- 14 刘雷杰.基于Kinect人体动态三维重建.电子产品世界,2016,23(9):35-37,41.
- 15 Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling. Quebec, Canada. 2001. 145-152.
- 16 Werner D, Al-Hamadi A, Werner P. Truncated signed distance function: Experiments on voxel size. Campilho A, Kamel M. Image Analysis and Recognition. Cham: Springer International Publishing, 2014. 357-364.
- 17 李诗锐,李琪,李海洋,等.基于Kinect v2的实时精确三维重建系统.软件学报,2016,27(10):2519-2529.