

文章编号: 1671-0444(2015)02-0227-05

# 基于多 Kinect 的三维人体重建系统

朱江涛, 吴以凡, 周 亮

(杭州电子科技大学 计算机学院, 浙江 杭州 310018)

**摘要:**结合 KinectFusion 技术,设计并实现了一种基于多台 Kinect 的三维人体重建系统. 该系统使用两台 Kinect 分别采集人体上下两部分的点云数据,通过标定得到两台 Kinect 的姿态关系,并基于标定结果,使用迭代最近点(ICP)算法对两部分点云进行配准得到完整的人体三维模型. 试验表明,系统设备成本低、部署简单,能快速并较高精度地获取人体三维点云.

**关键词:**三维人体重建; Kinect; KinectFusion; 标定; 点云配准

**中图分类号:**TN 401

**文献标志码:**A

## 3D Human Body Reconstruction System Based on Multiple Kinects

ZHU Jiang-tao, WU Yi-fan, ZHOU Liang

(College of Computer, Hangzhou Dianzi University, Hangzhou 310018, China)

**Abstract:** A new 3D human body reconstruction system based on multiple Kinects is proposed using the KinectFusion technology. In this system, two Kinects are used to scan the up and down half of human body and acquire the point cloud data respectively, then the relationship between the pose of two Kinects obtained by calibration and iterative closest point(ICP) algorithm are used to complete the registration and build the whole 3D human body model. Experiments prove that this system can obtain whole point cloud of the 3D human body rapidly with a good precision while the devices are cheap and system is easy to install.

**Key words:** 3D human body reconstruction; Kinect; KinectFusion; calibration; point cloud registration

三维人体重建技术在虚拟现实、动漫产业、纺织业、服装设计等领域有着重要应用,也是近年来计算机视觉领域的一个热门研究课题<sup>[1]</sup>. 基于结构光、激光的三维扫描系统目前已经比较成熟,重建数据精度高,但系统操作复杂、设备昂贵,此类人体扫描、重建系统价格一般在 35 000~500 000 美元,不适合普通用户使用<sup>[2]</sup>. 随着微软 Kinect 的推出,其体积小、价格低廉、易使用等特性,使得深度摄像机走向大众,出现了许多基于 Kinect 的三维重建系统<sup>[3-5]</sup>. 由于 Kinect 摄像头获取的原始深度数据比较粗糙<sup>[6]</sup>,系统在兼顾重建精度和速度方面大多有待提高. 文

献<sup>[7]</sup>提出了一种利用 3 台 Kinect 对有轻微形变的人体进行重建的方法,通过局部配准与全局优化的反复迭代来获得最终人体模型. 该方法需先对人体进行建模,会一定程度地影响效率,对于某些应用场景而言,最终人体模型的质量也不够高且有时存在配准失败问题. 文献<sup>[8]</sup>提出的 KinectFusion 技术,通过手持一台 Kinect 并移动扫描,在任意光照条件下,能够实时、快速地创建房间级大小的复杂场景且富含细节的三维模型. 其建立一个全局表面模型,并将 Kinect 深度图像数据不断融入模型,逐步丰富完善模型,重建精度可达毫米级.

收稿日期: 2013-12-09

基金项目: 国家自然科学基金资助项目(61202093);浙江省重点科技创新团队资助项目(2011R50009)

作者简介: 朱江涛(1989—),男,安徽六安人,硕士研究生,研究方向为计算机应用技术. E-mail: izhujiangtao@163.com

吴以凡(联系人),男,讲师, E-mail: yfwu@hdu.edu.cn

本文将提出一种基于多台 Kinect 的三维人体重建系统,使用两台位置固定的 Kinect 和一个匀速转台.系统运行前,先对两台 Kinect 进行标定,获得两台 Kinect 之间的姿态(水平距离、旋转角度)关系;系统运行过程中,人站立于匀速转动的转台,两台 Kinect 通过 KinectFusion 技术分别采集、融合得到人体上下两部分的点云;采集结束后,利用标定得到的参数对两片点云进行预处理,再使用迭代最近点(iterative closest point, ICP)算法进行配准得到完整的人体点云.该系统硬件成本低,操作部署简单.

## 1 三维人体重建系统概述

### 1.1 硬件平台

Kinect 默认模式的可视范围是  $0.8 \sim 3.5$  m, Kinect 与待扫描物体间的推荐距离为  $1.5 \sim 2.0$  m, 可视角度为垂直方向  $43^\circ$ 、水平方向  $57^\circ$ , 垂直方向上下倾斜角度为  $\pm 27^\circ$ [9]. Kinect 的可视空间如图 1 所示.

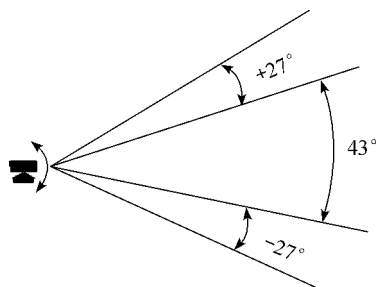


图 1 Kinect 可视空间

Fig. 1 Kinect's field of view

综合考虑采集数据的精度和测量范围,系统的平台搭建如图 2 所示.

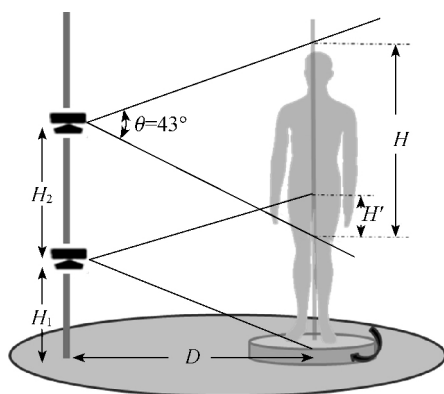


图 2 硬件平台

Fig. 2 Hardware platform

待重建人体站立于匀速转台上,与转台中心距离为  $D$  处垂直树立一固定杆.在固定杆离地高度

$H_1$  处固定第一个 Kinect,在固定杆上第一个 Kinect 上方距离为  $H_2$  处固定第二个 Kinect,两台 Kinect 均面向人体.

Kinect 与待扫描物体间的距离越近,测量越精确,为达到最佳测量效果,本文设定 Kinect 与待重建人体之间的距离为  $1.5$  m, 即  $D=1.5$  m. 设在距离为  $D$  处两台 Kinect 分别能扫描的人体高度为  $H$ , 已知 Kinect 垂直方向可视角度  $\theta=43^\circ$ , 则由几何关系可得:

$$H = 2 \times \tan(\theta/2) \times D \quad (1)$$

则计算可得  $H=1.18$  m.

为确保第一台 Kinect 能完全扫描到人体的双脚,其离地最大高度为  $H/2=0.59$  m, 考虑待测人体的初始站位可能偏离中心靠外,故取  $H_1=0.5$  m 使得第一台 Kinect 的中心离转台中心更近,以确保双脚的精度.设待测人体身高为  $h$ , 则要求两台 Kinect 的扫描高度不小于  $h$ , 即

$$H_1 + H_2 + H/2 \geq h \quad (2)$$

$h$  越大,则要求相应的  $H_2$  越大.若  $H_2$  越大,两台 Kinect 的重叠部分高度  $H'$  则越小,  $H'$  是后续上下两部分点云配准融合的关键,重叠越多(即  $H'$  越大)则配准融合越精确.权衡可测人体身高范围和后续点云融合精度,取可测人体身高  $h=2.0$  m, 相应的  $H_2$  取  $0.9$  m. 重叠部分  $H'$  为

$$H' = 2 \times (H/2) - H_2 \quad (3)$$

代入  $H$  和  $H_2$  计算可得  $H'=0.28$  m.

在系统运行过程中,转台顺时针匀速转动, Kinect 持续扫描人体,不断融合并逐步丰富完善细节,最终得到人体模型.

### 1.2 软件流程

系统的数据处理流程主要包括标定、采集、配准 3 个部分,如图 3 所示.

系统启动后,首先初始化系统参数,设定 Kinect 深度帧分辨率、帧率,初始化 KinectFusion 相关参数.采集阶段,上下两台 Kinect 同时启动,分别执行相同的数据处理流程:每采集到一帧深度数据, KinectFusion 将其融入全局表面模型,并使用 ICP 算法估计当前 Kinect 位置,采集结束则将其表面模型输出为人体三维点云.配准阶段,先对两片点云分别进行去噪处理,再基于 Kinect 标定所得参数,调整两片点云,对调整后的两片点云用 ICP 算法进行配准,最终得到完整的人体三维模型.其中,标定过程只需在系统硬件平台搭建完成后执行一次.

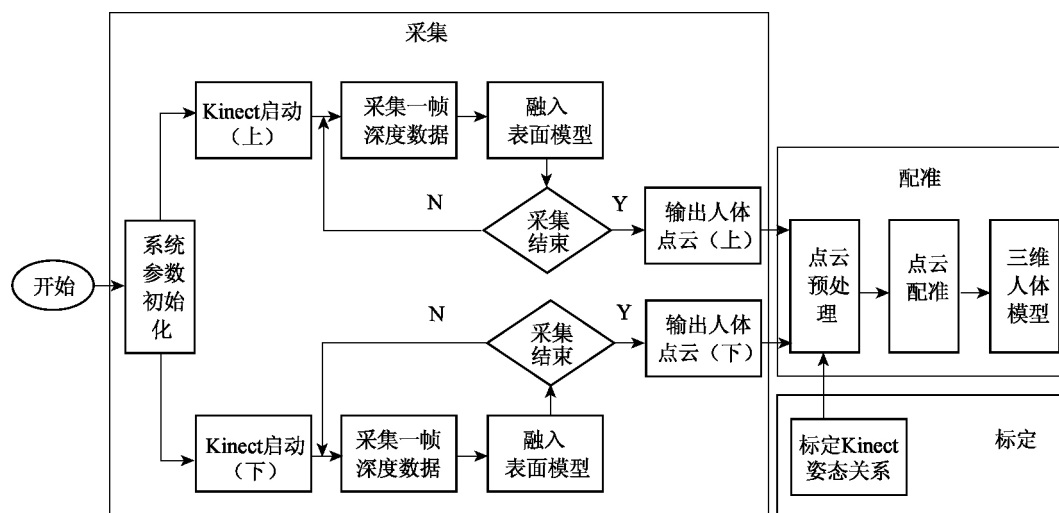


图3 数据处理流程

Fig. 3 Flowchart of data processing

## 2 三维人体点云获取

### 2.1 KinectFusion 技术

KinectFusion 技术通过将 Kinect 实时采集的深度数据帧融入一个全局表面模型,可实时、快速地建立物体的三维模型。全局表面模型是待重建物体所在空间中一个虚拟的立方体(volume),它用来包含物体的三维信息。随着 Kinect 的移动,待重建物体不同视角的深度图像被不断地融入该立方体,三维模型也就不断地精细,最终由该模型得到物体的三维点云。

与已有的一些三维重建技术相比,KinectFusion 有如下一些优势:能够实现实时的相机位置追踪和三维物体重建;由于 KinectFusion 只使用了深度帧数据,因而不受光线影响,重建的精度更高。

### 2.2 点云获取

为了保证深度数据的质量,设定 Kinect 的深度图像帧分辨率为  $640 \text{ 像素} \times 480 \text{ 像素}$ ,帧率为 30 帧/s。初始化 KinectFusion 立方体参数并启动扫描过程。点云获取过程如图 4 所示。对于两台 Kinect

各自采集的每一帧深度图像,分别做如下处理:第一步,深度图像转化,将原始的深度图像转化为相机坐标系中三维顶点和法线;第二步,相机追踪,使用基于 GPU 实现的 ICP 算法,配准当前帧和上一帧数据,得到当前的 Kinect 相机姿态(位置 and 方向);第三步,立方体融合,已知当前的相机姿态,则可以将当前得到的顶点转化到全局坐标系中,并更新相应小立方体(体素)的数值;第四步,光线投射,对立方体进行光线投射,得到立方体渲染的可视化图像。移动的 Kinect 会从细微变化的不同视角去看一个表面,任何在原始 Kinect 图像中看不到的孔洞都会被填充,立方体中的数据也会被不断改善。考虑到每帧数据处理过程存在着大量的计算,消耗较多的 CPU 和 GPU 资源,非中高端硬件配置的计算机处理 30 帧/s 的帧率会出现不流畅现象,扫描过程中,对帧率作降频处理,保证处理的流畅性。

扫描完成后,由两台 Kinect 各自的立方体数据计算得到人体的上下两部分的网格数据,并将网格数据保存为 PCD(point cloud data)点云格式以便后续的点云预处理和配准。

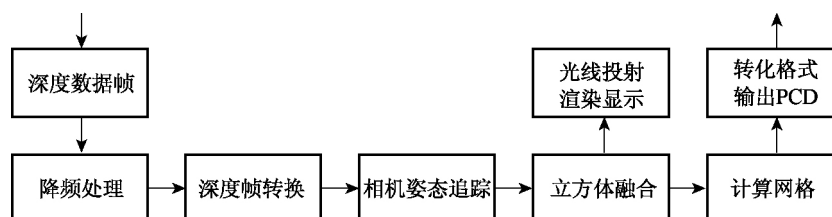


图4 点云获取过程

Fig. 4 Procedures of point cloud acquisition

### 2.3 点云配准

扫描过程最终得到的人体上、下两部分点云(记作  $C_1$  和  $C_2$ )中存在噪声点,需先剔除. 本文使用基于  $K$ -近邻的离群点剔除算法,具体为:对点云中每个点,计算其与  $K$  个近邻点的平均距离. 假设所有点到其  $K$ -近邻的平均距离符合高斯分布(均值为  $M$ , 标准差为  $D$ ),则到  $K$ -近邻的平均距离超过  $M+D$  的点,被视为离群点并移除. 去噪后得点云  $C'_1$  和  $C'_2$ . 由于  $C'_1$  和  $C'_2$  来自两台姿态(位置和方向)不同的 Kinect,其观测视角和坐标系不同,利用 Kinect 姿态关系进行初步调整,对  $C'_1$  进行旋转和平移变换,统一到  $C'_2$  所在坐标系.

两台 Kinect 姿态关系由标定过程得到,采用棋盘格标定方法<sup>[10]</sup>对两台 Kinect 同时标定,如图 5 所示. 其中,  $O_{xyz}$  为摄像机坐标系;  $O_{XYZ}$  为图像坐标系;  $O_{X_w Y_w Z_w}$  为世界坐标系;  $p_1$  和  $p_2$  为空间上一点  $P$  在图像坐标系上的投影.

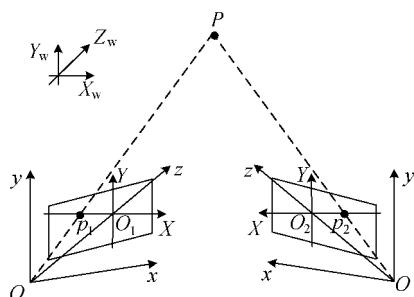


图 5 Kinect 标定

Fig. 5 Kinect calibration

设同一时刻观看空间物体(棋盘格)的同一特征点  $P = (x, y, z, 1)^T$ , 两台 Kinect 各自图像平面坐标系下对应的点分别为  $p_1 = (x_1, y_1, 1)^T$ ,  $p_2 = (x_2, y_2, 1)^T$ .

由棋盘格标定结果可知  $p_1$  和  $p_2$  到  $P$  点的三维变换分别为

$$p_1 = A_1 \begin{bmatrix} R_1 & T_1 \\ 0 & I \end{bmatrix} P \quad (4)$$

$$p_2 = A_2 \begin{bmatrix} R_2 & T_2 \\ 0 & I \end{bmatrix} P \quad (5)$$

其中:  $A_1$  和  $A_2$  分别为 Kinect 的  $3 \times 3$  内参矩阵;  $R_1$  和  $R_2$  为  $3 \times 3$  旋转矩阵;  $T_1$  和  $T_2$  为  $3 \times 1$  平移矩阵.

联立式(4)和(5)可得  $p_1$  到  $p_2$  的三维变换矩阵(即参数矩阵)  $[R|T]$ :

$$[R|T] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (6)$$

其中:  $R$  和  $T$  分别为旋转矩阵和平移矩阵.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (7)$$

调整算法如下:

(1) 调整开始, 创建点云  $C'_1 = \emptyset$ ;

(2) 遍历  $C'_1$ : 对任意顶点  $P(x, y, z) \in C'_1$  作  $[R|T]$  变换, 得到变换后的点  $P'(x', y', z')$ , 有

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = [R|T] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (8)$$

然后将  $P'$  加入到  $C'_1$  中;

(3) 输出  $C'_1$ , 调整结束.

调整后, 点云  $C'_1$  和  $C'_2$  坐标系相同. 由于 Kinect 标定存在一定的误差, 经过调整后的两片点云配准效果还不理想, 需要再通过 ICP 算法完成两片点云的进一步精确配准.

ICP 算法是经典的点云配准算法, 通过最小二乘法迭代计算最优坐标变换(旋转矩阵和平移矢量), 以最小化误差函数<sup>[11]</sup>. ICP 算法简便直观, 配准精度较好, 但算法的运行速度以及向全局最优的收敛性在很大程度上取决于给定的初始变换估计以及迭代过程中对应关系的确立<sup>[12]</sup>, 上述调整过程可以为算法提供较好的初始位置, 能够有效地提高算法效率. 配准过程: 首先设置算法输入数据为  $C'_1$  和  $C'_2$ , 再对点云进行采样, 确定初始对应点集, 去除错误对应点对, 迭代计算, 求解坐标变换矩阵, 拼接两片点云.

### 3 实验结果

系统开发平台为 Visual Studio 2010, Intel® Core™ i3-3220 处理器, NVIDIA GeForce GTX 650 显卡. 图 6 和 7 分别为实际人体和塑料人台的三维重建过程. 其中图(a)和(b)分别为上下 Kinect 采集的人体三维数据; 图(c)为经过去噪和调整后的两片点云同时显示在同一坐标系中的结果; 图(d)为配准后的最终人体三维点云, 且图(c)和(d)为三维点云数据, 无阴影和光照效果.

图 6(a)中人体所穿背心(厚度约为 2 mm)及图 7(a)中模特衣服纹理均清晰可辨, 表明由 KinectFusion 采集获取的三维人体模型具有较高的精度. 由图 6(c)和图 7(c)可知, 经过去噪、调整, 两片点云整体上有了初步的匹配效果, 但在扫描的重叠部分, 仍存在误差. 由图

6(d)和7(d)可知,通过配准可消除人体上下两部分点云在旋转角度、平移方面的误差,配准后的三维人体点云在拼接处曲面光滑,配准效果比较理想.

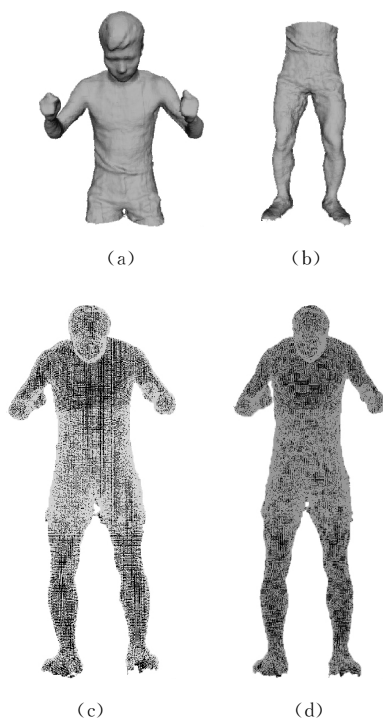


图6 实际人体的三维重建

Fig. 6 3D reconstruction of human body

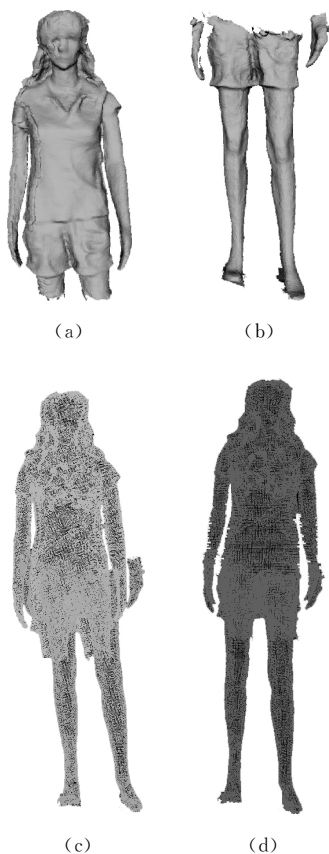


图7 塑料人台三维重建

Fig. 7 3D reconstruction of plastic model

## 4 结 语

由于受成本、操作复杂度等方面的影响,已有的三维人体重建系统应用范围受到限制. 本文介绍了一种基于多台 Kinect 的三维人体重建系统,该系统使用两台 Kinect 同时分别扫描人体上下两部分,扫描结果经由数据处理系统去噪、调整、配准,最终得到完整的人体三维点云. 系统基础设备较少、成本低廉,可在任意光照条件下,快速重建出精度较好的三维人体模型,能满足一般的应用需求. 后续研究将侧重于进一步提高系统的稳定性,解决偶尔出现的 KinectFusion 扫描过程中由相机追踪失败引起的扫描失败问题.

## 参 考 文 献

- [1] 黄敏,杨念,张剑. 基于雅可比算法的着装人体三维重建[J]. 东华大学学报:自然科学版, 2007, 33(5):607-612.
- [2] WEISS A, HIRSHBERG D, BLACK M J. Home 3D body scans from noisy image and range data [C] // IEEE International Conference on Computer Vision. 2011: 1951-1958.
- [3] 宋诗超,禹素萍,许武军. 基于 Kinect 的三维人体扫描、重建及测量技术的研究[J]. 天津工业大学学报, 2012, 31(5):34-41.
- [4] 刘鑫,许华荣,胡占义. 基于 GPU 和 Kinect 的快速物体重建[J]. 自动化学报, 2012, 38(8):1288-1297.
- [5] 周瑾,潘建江,童晶,等. 使用 Kinect 快速重建三维人体[J]. 计算机辅助设计与图形学学报, 2013, 25(6):873-879.
- [6] NEWCOMBE R A, IZADI S, HILLIGES O, et al. KinectFusion: Real-time dense surface mapping and tracking [C] // 10th IEEE International Symposium on Mixed and Augmented Reality. 2011:127-136.
- [7] TONG J, ZHOU J, LIU L G, et al. Scanning 3D full human bodies using Kinects[J]. Visualization and Computer Graphics, 2012, 18(4):643-650.
- [8] IZADI S, KIM D, HILLIGES O, et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera [C] // Symposium on User Interface Software and Technology. 2011.
- [9] Kinect for Windows sensor components and specifications [EB/OL]. [2013-11-21]. <http://msdn.microsoft.com/en-us/library/jj131033.aspx>.
- [10] ZHANG Z Y. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [11] LIU Y H. Automatic registration of overlapping 3D point clouds using closest points[J]. Image and Vision Computing, 2006, 24(7):762-781.
- [12] 朱德海. 点云库 PCL 学习教程[M]. 北京:北京航空航天大学出版社, 2012:302-303.