

基于单类 Kinect 的人体三维重建与优化*

许鸿尧¹, 李重¹, 金小刚², 马利庄³

¹(浙江理工大学 数学科学系, 浙江 杭州 310018)

²(CAD&CG 国家重点实验室(浙江大学), 浙江 杭州 310058)

³(上海交通大学 计算机科学与技术系, 上海 200240)

通讯作者: 李重, E-mail: lizhong@zstu.edu.cn

摘要: 使用单一运动的类 Kinect 深度相机重建和优化静态 3D 全身人体模型. 首先针对类 Kinect 深度相机产生噪声原因提出一种降噪处理方法进行降噪. 结合深度信息和 RGB 信息获取匹配块, 使用高斯混合模型进行局部配准和逐层封闭曲线拟合方法进行全局配准, 结合改进方向距离函数进行合并, 最后使用泊松表面重建方法获取三维模型. 实验结果表明, 该方法能够重建出较高精度的三维人体模型.

关键词: 降噪, 配准, 重建, 类 Kinect, 三维人体模型

中文引用格式: 许鸿尧, 李重, 金小刚, 马利庄. 基于单类 Kinect 的人体三维重建与优化. 软件学报, 2015, 26(Suppl. (2)): 29-41. <http://www.jos.org.cn/1000-9825/15013.htm>

英文引用格式: Xu HY, Li Z, Jin XG, Ma LZ. 3D human body reconstruction and optimization using one kinect-like depth camera. Ruan Jian Xue Bao/Journal of Software, 2015, 26(Suppl. (2)): 29-41 (in Chinese). <http://www.jos.org.cn/1000-9825/15013.htm>

3D Human Body Reconstruction and Optimization Using One Kinect-Like Depth Camera

XU Hong-Yao¹, LI Zhong¹, JIN Xiao-Gang², MA Li-Zhuang³

¹(Department of Mathematical Sciences, Zhejiang Sci-Tech University, Hangzhou 310018)

²(State Key Laboratory of CAD&CG (Zhejiang University), Hangzhou 310058)

³(Department of Computer Science & Engineering, Shanghai Jiao Tong University, Shanghai 200240)

Abstract: This paper presents an approach to reconstruct and optimize static 3D human body models using a single Kinect-like motion depth camera by moving the sensor freely around the human body. First, to reduce the noise in the depth data captured by the Kinect-like depth sensor, an approach is proposed to filter it according to the noise source created from the Kinect-like depth sensor. Then, the search for the corresponding patch is performed by combining with the information of depth and RGB, and the pair frames are aligned with the Gaussian mixture model and the improved signed distance function. For the loop closure problem, a closed curve fitting based method is provided with different layers to globally register all the frames. Finally, the human body surface is constructed by using the Poisson reconstruction. Experimental results show that the presented approach can reconstruct human models with high quality.

Key words: denoise; registration; reconstruction; kinect-like; 3D human body model

三维人体重建应用广泛, 在游戏娱乐、服装工业、体育运动等领域都有涉及. 通常获取模型的设备价格昂贵, 比如激光扫描仪、双目立体相机、基于飞行时间 ToF(time of flight)相机等并且需要专业人员来操作. 近几年推出的类 Kinect 设备如 Kinect、PrimeSense、XtionPRO 等也具备三维扫描的功能. 这类设备最大的优势在于廉价、操作简单, 但是数据噪声大、分辨率低, 直接用来重建三维人体, 效果很不理想. 因此针对类 Kinect 的特点,

* 基金项目: 国家自然科学基金(61272298, 61133009); 浙江省自然科学基金(LY14A010032); 国家教育部留学回国人员科研启动基金([2009]1590); 浙江理工大学 521 人才资助项目

收稿时间: 2014-01-03; 定稿时间: 2014-04-18

研究一种基于类 Kinect 重建三维人体模型的方法非常有意义^[1-3].近年来已有一些学者将类 Kinect 深度相机作为扫描仪应用在三维重建中.比如场景重建^[4-7]、人脸重建^[8]和人体重建^[9-11]等.Newcombe 等人^[4]和 Izadi 等人^[5]的 KinectFusion 方法能实时跟踪和重建室内场景.

Chen 等人^[6]扩展了 Newcombe 和 Izadi 等人提出一种使用内存高效多层 GPU 数据结构的方法用于大规模场景的重建.Roth 等人^[7]使用可移动体积的 KinectFusion 方法在室外大场景的情况下进行重建.这些方法也同样适用于人体重建,但是会出现部分数据缺失和扫描模型闭合处出现不规则形变的现象.Weiss 等人^[9]提出一种人体重建方法,主要通过低分辨率下的图像轮廓和深度数据结合从 SCAPE^[12]模型中估计人体形状,这种方法重建精度不高(脸部,服装等细节不明显).Tong 等人^[10]和周瑾等人^[11]的方法是使用多个 Kinect 扫描站在转台上的静态人体,然后用一个初始模型对扫描数据进行迭代变形来重建人体.这种方法需要一个转台和多个 Kinect 通常受场地限制.本文提出了一种新的操作方便的人体重建方法,可实现在静态人体周围动态移动一台类 Kinect 实时重建三维人体模型.该方法的主要特点有:

(1) 根据类 Kinect 的深度成像和计算原理分析产生噪声和孔洞的原因构造能量方程,通过 Gauss-Newton 算法^[13]求解最小化能量,得到理想的低噪深度图像.点云匹配时,结合深度信息和 RGB 信息,提出匹配块选取的方法.与以往方法不同的是,该方法不用提取特征点,而是首先通过距离和角度查找点云中的匹配块,在无法找到匹配块时再通过图像查找匹配块,这种方法可以减少计算量.

(2) 使用高斯混合模型(Gaussian mixture model,简称 GMM)^[14,15]局部配准,曲线插值全局配准和方向距离函数(signed distance function,简称 SDF)^[4,16]点云合并结合,与直接使用 GMM 的方法相比,这种方法利用匹配块对作为输入来减少计算量,相对 ICP^[17]而言这种方法对带噪声的数据有更好的鲁棒性,当两帧之间变化幅度较大情况下也能进行有效配准.同时通过曲线拟合的方法完成全局匹配并用 SDF^[4,16]完成点云合并能更好地进行多帧配准.

(3) 改进 Newcombe 等人^[4]和 Curless 等人^[16]方法中的方向距离函数.原来方法中将权重 w 取值为常数 1, 本文使用 1 减去归一化后的旋转和平移误差,即用相邻两帧匹配点的加权平均法向的夹角余弦表示旋转误差,用相邻两帧匹配点之间的距离与后一帧中匹配点到原点的距离之间的比值作为平移误差,与原来方法相比能更好的减少点云合并误差.

1 相关工作

使用类 Kinect 进行三维人体扫描,主要涉及的一些算法操作包括深度图像降噪和修复、点云数据局部配准、全局配准和表面重建等工作.

深度图像中的噪声严重影响重建的精度.为了提高重建精度,降噪是非常重要的过程.在类 Kinect 深度相机推出之前,针对基于飞行时间(ToF)相机提出过降噪的方法^[18],这是一种提高三维深度图分辨率的方法,可通过将多帧带噪声的低分辨率深度图融合为一帧高分辨率的深度图.这种方法可以很好地提高深度图分辨率并修复孔洞,但是 ToF 相机使用的是三角测量的深度计算原理,与光编码(light coding)不同.因此这种方法对于类 Kinect 本身成像和深度计算原理产生的噪声并不适用.在图像降噪上的双边滤波方法也可适用于深度图.这种方法在平滑噪声的同时能很好的保留边缘信息,可是分辨率会降低,具体的细节会丢失.Nguyen 等人^[19]根据类 Kinect 的 Light Coding 原理,在不同的距离上放置一个不同角度的参考平面用类 Kinect 测深度值并记录实际的距离值.通过两者之间的统计关系,提出了一种噪声模型.这种方法可以很好的描述类 Kinect 使用 Light Coding 技术引起的误差分布,但对于太阳光谱,热源等引起的噪声和孔洞并没有很好的处理.

杨琴等人^[20]提出了结合深度信息和颜色信息来提取特征点的方法,分别从两幅 RGB 图像中提取尺度不变特征变换(SIFT)特征,然后找到可能正确的匹配像素对,再以这些匹配像素对作为候选集,结合深度图中顶点的几何信息,使用随机抽样一致(RANSAC)算法找到正确匹配的像素对.虽然这种方法考虑到的 SIFT 特征对图像具有缩放和旋转不变的特性,但是在实际操作中多层的图像金字塔计算需要很大的计算量.

关于两组点云数据之间的局部配准,迭代最近点(ICP)及其各种改进方法^[17]是最常用的配准方法.这种方法

主要通过反复迭代减少点之间的距离和来计算变换矩阵.这种方法对初始状态很敏感,容易产生局部收敛的情况.Myronenko 等人^[15]和 Cui 等人^[14]中提到一种基于概率模型的方法,将其中一组点云看作高斯混合模型的质心,通过最大似然估计使另一组点云与该组点云重合.相对于 ICP,高斯混合模型配准方法对于噪声和孔洞的影响有很强的鲁棒性,并且在点云对相对位置变化较大的情况下有比较好的匹配效果.但是使用最大期望(expectation-maximization,简称 EM)算法迭代计算存在两个问题:(1) 对迭代初始点敏感,不好的初始状态容易导致局部最优解,从而得到的匹配效果不佳.(2) 收敛速度受匹配点对的影响,匹配点对数量越大,收敛速度越慢.导致以上两个问题的关键是匹配点对选取的好坏.

当类 Kinect 绕人体扫描一圈时,由于三维人体是一封闭物体,理想情况下首尾帧之间是重合的,但是在实际情况中由于局部配准的误差影响,所以多帧配准后累积的误差导致最终不重合的现象出现,这个问题被称为环闭合问题(loop closure problem)^[21].针对这个问题,当前比较流行的一些方式是:Weise 等人^[21]提出的方法.该方法定义了一种表面元素的结构,即带方向的一个圆盘,来表示物体表面.通过使用拓扑图结构进行局部刚性变形达到全局配准的效果.Sharp 等人^[22]提出一种新的方法,首先分析旋转和平移误差的分布,然后以每一帧为结点相邻帧之间连接表示成图,分离出图中首末相连的环,对各个环通过最小二乘最小化误差进行全局配准.Gelfand 等人^[23]提出的方法是使用一种积分体积描述子(integral volume descriptor)作为特征,再从特征中找出匹配关系,在查找匹配关系时,使用了距离误差和剪枝算法.

在得到统一的三维点云模型之前,需要将多角度得到的点云进行合并.Newcombe 等人^[4]和 Curless 等人^[14]的方法主要基于方向距离函数.这种方法中有一个权重用来确定沿同一方向不同点云合并后的位置,权重的选取对最后得到统一三维点云模型的精度有影响,Newcombe 等人^[4]和 Curless 等人^[14]的方法将权重设置为常数 1,并未考虑相邻两帧点云之间的误差影响.

三维人体表面重建算法大体可分为两类:显式曲面和隐式曲面两种方法.显式曲面的方法,这类方法通过插值的形式进行网格化,计算量与点云的数量成正比,重建表面会受到点云数据噪声的影响,往往需要后期进行光顺和修补等处理.比如 Delannay 三角剖分^[24]、Voronoi 图^[25]等.隐式曲面的方法,这类方法对拓扑结构复杂和带噪声的三维模型都有很好的鲁棒性.方法的复杂性往往取决于所选用的隐式函数.比如以径向基(RBF)为隐式函数^[26],但是在实际计算时得到的系数矩阵是稠密的且病态的.而近几年提出的泊松表面重建(Poisson surface reconstruction)方法^[27],该方法结合了之前隐式函数方法的优点并解决了基函数解矩阵病态的问题.

2 算法整体流程

本文整个算法流程是:首先使用类 Kinect 获取原始深度图和同步的 RGB 图像,然后对深度图像进行降噪和孔洞修复预处理,将预处理后的深度数据转换为三维点云数据并结合同步的 RGB 图像进行点云匹配,接着将配准后的点云进行方向距离函数合并,最后进行表面重建得到三维模型.算法流程如图 1 所示.

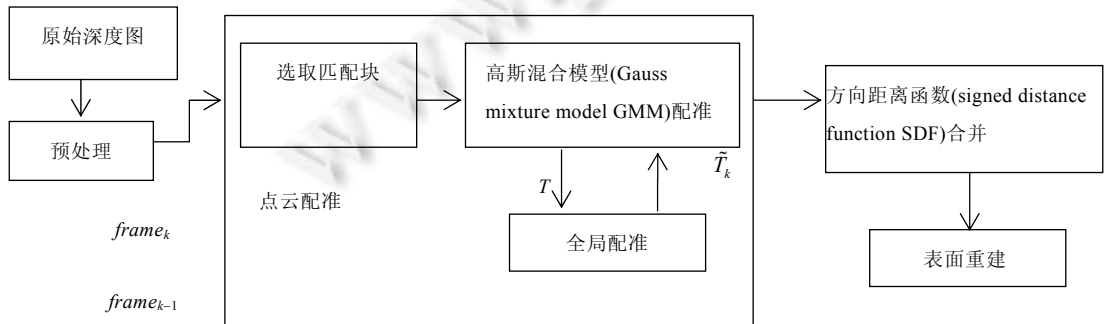


图 1 算法流程图

3 预处理

3.1 基于类Kinect深度数据误差分析

类 Kinect 采用的是一种称为 Light Coding 的技术^[28],即用光源照射空间,根据反射图案进行编码,也属于结构光技术,但与一般的结构光计算有所不同.该技术首先采用了一种称为激光散斑的光源,这是一种激光衍射后随机产生的斑点,会根据距离变换图案.空间中任意散斑图案都是唯一的,可以对整个空间做标记.而之前还有一个参考的图案存储在类 Kinect 的内存中,只要将投射得到的图案与存储的参考图案做一下对比,就可以计算出物体的深度值.Light Coding 技术并没采用三角视差测距,它的精度和光源标定时的参考面密度有关,密度越大越精确.光源标定是指取一个参考面,记录参考面到类 Kinect 的距离,根据不同的距离记录不同参考面上的图案.比如类 Kinect 在 0.8m~4m 范围内,间隔 1cm 取参考面,然后标定后保存 320 个散斑图案.标定间隔越小精度越高.当实际测量物体时,得到 320 个散斑图案与保存的图案比较后,得到一组相关性的图像,在图像中出现峰值的地方,就是物体所在位置,再叠加峰值并进行插值就可得到实际三维物体的深度值.根据类 Kinect 的 Light Coding 技术原理,我们不难知道,太阳光谱和热源会影响类 Kinect 的精度,因为太阳光谱和热源会干扰类 Kinect 红外发射器发射的近红外光源.

3.2 降噪和孔洞修复

从以上分析可知,类 Kinect 深度数据的误差产生主要为两个方面.第一是参考面选取的密度,第二是由于外界太阳光谱和热源的干扰.综合引起类 Kinect 噪声和孔洞的原因,我们从以下 4 个方面减少噪声和孔洞修复:

1. 通过多深度图像融合来修复孔洞.
2. 缩短深度图内相邻点的距离来提高分辨率.
3. 多深度图融合时,引入 Nguyen 等人^[19]的误差分布模型,避免或减少多深度图融合引起的噪声积累现象.
4. 深度图进行光滑处理,减少毛刺现象.

据此,我们提出一种能量模型并通过 Gauss-Newton 算法^[13]求解最小化能量,得到理想深度图 X

$$\min_X (E_{fill} + E_{ppi} + E_{smooth}) \quad (1)$$

其中, E_{ppi} 使相邻点尽可能靠近提高分辨率,可设

$$E_{ppi} = \sum_{k=1} \sum_{j \in N(k)} \|u(k) - u(j)\|^2 \quad (2)$$

$u(k), u(j) \in X$ 表示 X 中相邻点, $N(k)$ 表示与 k 索引对应点的相邻点索引.

针对多帧深度图融合时相邻位置深度值可能出现抖动情况, E_{smooth} 用来对深度图进行光滑处理,减少毛刺现象,可设

$$E_{smooth} = \sum_{k=1} \sum_{j \in N(k)} \|X(u(k)) - X(u(j))\|^2 \quad (3)$$

其中, $X(u(k)), X(u(j))$ 表示 X 中相邻点的深度值. E_{fill} 用来多深度图融合并填补空洞,可设

$$E_{fill} = \sum_{i=1}^n \|W_i \cdot *(D_i - X)\|^2, \quad (4)$$

其中, D_i 表示第 i 帧原始深度图, $*$ 表示相同维度矩阵点乘,

$$W_i(u(k)) = \begin{cases} 0, & D_i(u(j)) - X(u(j)) > Z_{threshold} \\ \exp\left(-\frac{\|u(k) - u(j)\|^2}{2\sigma_L^2} - \sum \frac{\|D_i(u(j)) - X(u(j))\|^2}{2\sigma_z^2}\right), & |D_i(u(j)) - X(u(j))| \leq Z_{threshold} \\ 1, & D_i(u(j)) - X(u(j)) < -Z_{threshold} \end{cases} \quad (5)$$

其中, $j \in \{k\} \cup N(k)$,

$$\sigma_L(\theta, z) = 0.8 + 0.0035 \frac{\theta}{\frac{\pi}{2} - \theta} z \frac{p_x}{f_x} \quad (6)$$

$$\sigma_z(\theta, z) = 0.0012 + 0.0019(z - 4)^2 + \frac{0.0001}{\sqrt{z}} \frac{\theta^2}{\left(\frac{\pi}{2} - \theta\right)^2} \quad (7)$$

p_x 表示深度图中像素的面积, f_x 表示焦距, θ 表示相机坐标 Z 轴方向与物体表面法向之间的夹角. $W_i(u(k))$ 表示第 i 帧深度图 $u(k)$ 位置的权重, 用来避免和减少深度图融合过程中出现的距离和角度产生的误差. 图 2 所示为本文方法处理与双边滤波处理的比较结果.

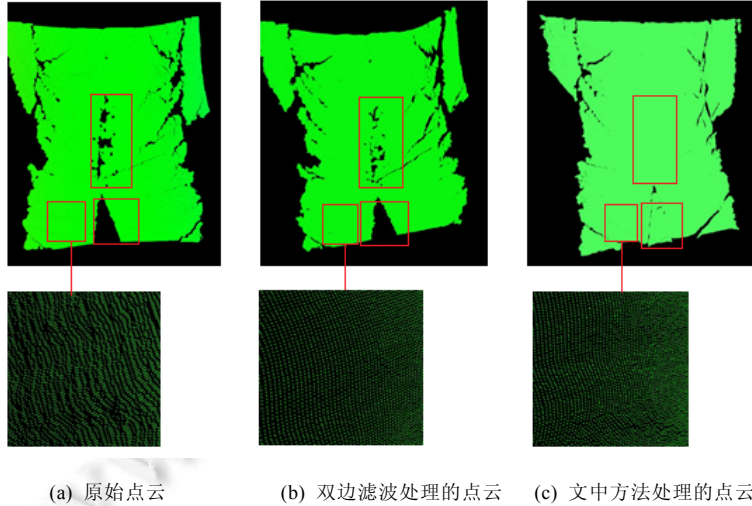


图 2 降噪处理

4 局部配准

这里提到的局部配准主要是指两帧点云数据之间的刚性配准, 即一帧点云中所有点在配准过程中的旋转和平移变换都是相同的. 我们在本文中使用的第 3 节的方法得到的匹配块作为这种算法的匹配点, 构造高斯混合模型, 通过最大期望算法(EM)迭代计算变换矩阵, 然后用变换矩阵变换点云的空间位置进行点云对匹配. 与 Myronenko 等人^[15]和 Cui 等人^[14]提出的基于概率模型的方法以及 ICP 及其各种改进方法^[14]相比, 本文中方法可以较好的解决局部最优解和收敛速度慢的问题.

4.1 匹配块的选取

以往方法^[20,29,30]中提到选取匹配点的方法都是结合了深度信息和 RGB 信息. 通过在 RGB 图上提取特征点, 根据特征点来找到匹配的像素对, 再由匹配的像素对找到对应的深度图中的匹配点. 这种方法相对直接使用深度信息找匹配关系更准确, 但是特征点的提取和特征点之间对应关系的查找都需要很大的计算量. 这种方法理论上基于图像分层, 可以更好地处理图像放缩产生的失真. 但在实际操作中, 两三层以上的图像金字塔计算量非常大而且过于模糊并不适合配准. 本文采用类似杨琴等人^[20]提出的方法, 但是并没有直接使用尺度不变特征变换匹配算法(scale invariant feature transform, 简称 SIFT)算法在图像上选择匹配点, 而是使用两帧点云间的距离和单位法向的角度来判断匹配块. 首先选取两帧点云重叠部分的一个点云块, 然后查找另一帧中最短距离的点云块, 并计算距离和单位法向的夹角, 再根据阈值(这里我们选取夹角为 5° , 距离为 5mm)判断是否为匹配块, 如果计算值小于阈值则为匹配块, 否则不是匹配块. 在无法找到匹配块的情况下, 将点云投射到深度图像中通过深度图像与 RGB 图像的对应关系在 RGB 图像中查找匹配块. 在进行图像查找时, 使用一种新的搜索方向和步长, 与 Barnes 等人^[31]提出的方法比这种方法减少了计算复杂度, 但会降低一定的精度. 因为之后采用的配准算法具有很好的鲁棒性, 此时并不需要很高的精度. 本文这种方法是针对第 4.2 节配准方法提出的, 具体算法步骤如下:

Step 1. 从第 k 帧中取 3 个点云块, 图 3(a) 中用立方体表示. 查找第 $k+1$ 帧点云中与此 3 个点云块距离最近

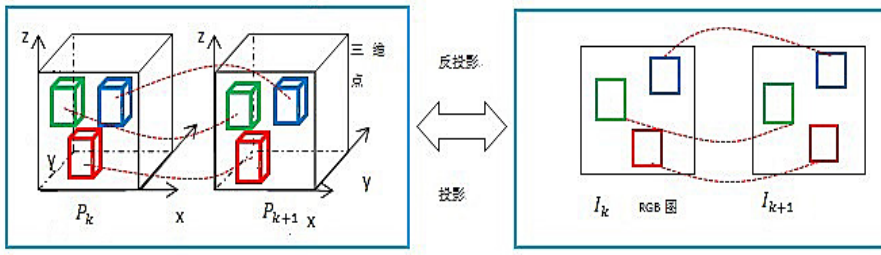
的 3 个点云,然后计算点云块之间的法向角度和距离(用虚线表示).

Step 2. 如果点云块之间的角度和距离都小于阈值则为匹配块,终止步骤.否则将这 6 个点云块分别投影到各自的 RGB 图上.

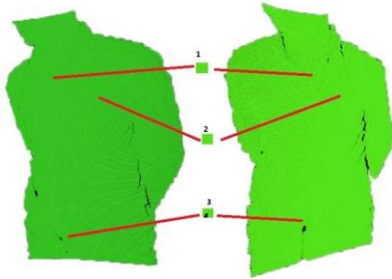
Step 3. 在第 $k+1$ 帧 RGB 图中搜寻 3 个与第 k 帧 RGB 图中近似的 RGB 块.搜索方向主要是以第 1 次投影的 RGB 块为起点,Step 1 中的法向在 RGB 平面上的投影方向为搜索方向.搜索的步长主要是块的边长(这里使用 5×5)与法向夹角的余弦值的乘积.

Step 4. 计算相似度(这里使用 CLab 空间下的欧式距离^[32])如果小于阈值返回 Step 3,否则反投影到点云块返回 Step2.

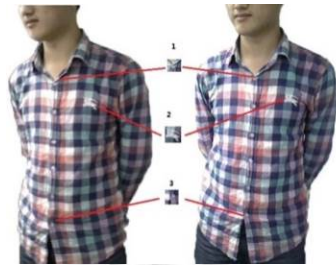
图 3 所示为匹配块选取过程.



(a) 匹配块选取示意图



(b) 点云匹配块



(c) RGB 图匹配块

图 3 匹配块选取过程

4.2 高斯混合模型配准

这里将点云对的配准过程看成高斯混合模型概率密度函数的最大似然估计问题^[14,19].将其中一个点云看作高斯质心,另一点云表示观测样本.在最优化的过程中实现点云的配准. $X_{M \times D}=(x_1, \dots, x_N)^T$ 表示作为高斯质心的点云, $Y_{N \times D}=(y_1, \dots, y_M)^T$ 表示观测样本的点云, D 表示点的维数, N, M 表示点云中点的个数.高斯混合模型的概率密度函数为

$$p(y) = \sum_{m=1}^{M+1} p(m)p(y|m) \tag{8}$$

其中, $p(y|m) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|y-x_m\|^2}{2\sigma^2}\right)$ 表示第 m 个高斯分量的概率密度.而当 $m = M+1$ 时, $p(y|m) = \frac{1}{N}$ 表示噪声和孔洞的均匀分布, $p(m) = \frac{1}{M}$ 表示所有的高斯分量之间混合的概率是相同的.通过加入均匀分布的权重 $w, 0 \leq w \leq 1$, 混合模型又可表示为

$$p(y) = w \frac{1}{N} + (1-w) \sum_{m=1}^{M+1} \frac{1}{M} p(y|m) \tag{9}$$

对数似然函数表示为

$$L(\theta) = \ln p(y) = \sum_{n=1}^N \ln \left\{ \sum_{m=1}^{M+1} p(m) p(y_n | m) \right\} \quad (10)$$

通过期望最大(EM)算法最大化对数似然函数估计参数.根据极值必要条件得

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^N \frac{p(m) p(y_n | m)}{\sum_{m=1}^{M+1} p(m) p(y_n | m)} \left(\frac{\partial p(m) p(y_n | m)}{\partial \theta} \right) = 0.$$

令 $p_{mn} = \frac{p(m) p(y_n | m)}{\sum_{m=1}^{M+1} p(m) p(y_n | m)}$, 则

$$p_{mn} = \frac{\exp\left(-\frac{1}{2\sigma^2} \|x_m - T(y_n)\|^2\right)}{\sum_{k=1}^M \exp\left(-\frac{1}{2\sigma^2} \|x_m - (sRy_k + t)\|^2 + (2\pi\sigma^2)^{D/2} \frac{w}{1-w} \frac{M}{N}\right)},$$

T 表示变换矩阵.通过求解方程得到相应的参数估计.具体算法步骤如下:

输入:点云 $X=(x_1, \dots, x_N)^T$ 和点云 $Y=(y_1, \dots, y_M)^T$.

输出: R 旋转矩阵和 t 平移向量.

Step 1.初始化旋转矩阵 R 为单位矩阵,平移向量为零向量, w 取 0~1 之间的值, $s=1$.根据如下式子计算初始方

$$\text{差, } \sigma^2 = \frac{1}{D_{NM}} \sum_{n=1}^N \sum_{m=1}^M \|x_n - y_m\|^2.$$

Step 2.计算概率矩阵 $P=[p_{mn}]_{M \times N}$.

$$p_{mn} = \frac{\exp\left(-\frac{1}{2\sigma^2} \|x_n - (sRy_m + t)\|^2\right)}{\sum_{k=1}^M \exp\left(-\frac{1}{2\sigma^2} \|x_n - (sRy_k + t)\|^2 + (2\pi\sigma^2)^{D/2} \frac{w}{1-w} \frac{M}{N}\right)} \quad (11)$$

Step 3.计算各个参数估计,各参数表达式可通过对数似然函数极值必要条件得到.

Step 4.计算得到的各参数值代入对数似然估计函数

$$L(\theta) = \ln p(y) = \sum_{n=1}^N \ln \left\{ \sum_{m=1}^{M+1} p(m) p(y_n | m) \right\} \quad (12)$$

并检查参数和对数似然函数的收敛情况.如果不收敛返回 Step 2,否则结束.

4.3 全局配准

对人体进行扫描的整个过程会产生很多帧点云,通过两两局部配准后,最终我们希望能让首尾帧重合并且重新调整局部配准中的旋转和平移误差,如图 4(a).但实际情况下达不到理想效果,即 n 帧匹配点序列中的首尾匹配点列出现不重合现象,这又被称为环闭合问题(loop closure problem)^[20],如图 4(b)所示.针对这个问题,论文使用基于曲线插值的方法进行全局配准,如图 5 所示.具体操作如下:

1. 通过对匹配块中的点进行重采样,构造人体模型不同高度下(假设取 h 个不同高度)的 n 帧匹配点序列.

2. 首先对首尾帧匹配点进行修正,取首尾帧匹配点的中点统一为封闭匹配点.使用三次 B 样条插值曲线方法构造封闭拟合曲线,每个点 P_i 的参数 t_i 通过弧长参数化得到,并利用这些参数计算点列在封闭拟合曲线中的位置 P'_i .

3. 对每一帧中不同高度下的匹配点列,我们重新计算每一帧的修正旋转和平移矩阵 R_i 和 T_i ,要求总的偏移量达到最小,可通过最小二乘方法求解.

$$\min \sum_{i=1}^h \|R_i \cdot P_i + T_i - P'_i\|^2 \quad (13)$$

4. 使用修正后的变换矩阵调整每一帧,得到全局配准效果,如图 6 所示.

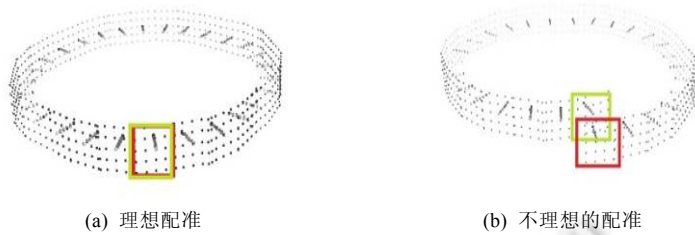


图 4 理想配准与不理想配准

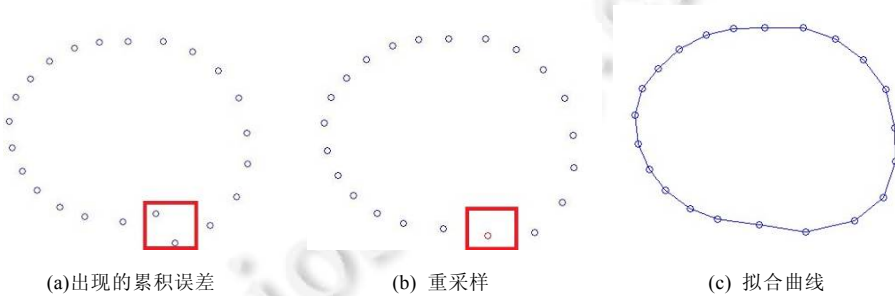


图 5 全局配准

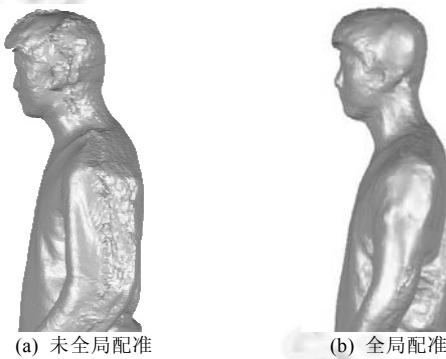


图 6 全局配准效果

5 方向距离函数的合并和表面重建

通过全局配准之后能得到完整人体的点云模型,但是在这个模型中同一个表面可能存在多层点云的情况,因为点云配准可以使多帧点云之间的空间位置非常接近,但不能合并为一帧点云,直接使用这种空间位置接近的多帧点云进行表面重建,会影响重建的效果.这里我们先使用方向距离函数进行点云合并. $\phi(\bar{x})$ 表示一个隐函数,当 $\phi(\bar{x}) \leq 0$, \bar{x} 位于内部区域 Ω ,当 $\phi(\bar{x}) > 0$, \bar{x} 位于外部区域 Ω^+ ,当 $\phi(\bar{x}) = 0$, \bar{x} 位于边界区域 $\partial\Omega$.这样就可以定义一个距离函数表示到边界的距离: $d(\bar{x}) = \min(|\bar{x} - \bar{x}_r|), \forall \bar{x}_r \in \partial\Omega$.从这个距离函数中可以看出当 $d(\bar{x}) = 0$ 时 $\bar{x} \in \partial\Omega$.一个方向距离函数可定义为一个隐函数 $f(\bar{x})$ 对所有的 \bar{x} 满足 $|f(\bar{x})| = d(\bar{x})$,即

$$f(\bar{x}) = \begin{cases} -d(\bar{x}), & \bar{x} \in \Omega^- \\ 0, & \bar{x} \in \partial\Omega \\ d(\bar{x}), & \bar{x} \in \Omega^+ \end{cases} \quad (14)$$

Newcombe 等人^[4]和 Curless B 等人^[15]提出的方法将方向距离函数离散化为 $m \times m \times m$ 的空间体素网格(m 通常取 512),整个空间在一个立方体中.每个体素网格 v 包含两组数据 $\{d, w\}$, d 表示空间网格 v 到表面网格的距离, w 表示同一方向上多个网格距离合并的权重.当 $d < 0$ 时, v 位于表面的内部;当 $d > 0$ 时, v 位于表面的外部;只有

当 $d=0$ 时, v 位于表面上. Newcombe 等人^[4]方法中使用了截取带宽,即只计算 $-T < d < T$ (通常取 0.03m) 范围内的体素网格,其余网格内的数据设为常数.方向距离函数又可以表示为

$$f(\vec{x}) = \begin{cases} -d(\vec{x}), \vec{x} \in \Omega^- \\ 0, \vec{x} \in \partial\Omega \\ d(\vec{x}), \vec{x} \in \Omega^+ \end{cases} \quad (15)$$

$$d(\vec{x}) = \begin{cases} -T, d < -T \\ d(\vec{x}), |d(\vec{x})| \leq T \\ T, d > T \end{cases} \quad (16)$$

另外, Newcombe 等人提出的方法中将每个体素网格中的权重设为常量 1. 而实际合并操作是首先用前一帧的点云计算体素网格中的 d 和 w , 然后加入后一帧点云通过

$$D_{k+1} = \frac{D_k + w_{k+1}d_{k+1}}{w_k + w_{k+1}} \quad (17)$$

$$w_k = w_{k+1} = 1 \quad (18)$$

重新计算体素网格中的距离. SDF 合并通常会产生累积误差, 影响最后的重建效果, 而 SDF 合并产生误差主要包括两方面: 一是由于扫描设备本身的局限性, 获取的点云数据带有误差. 这个问题在前一节中已考虑到了. 二是由于点云匹配计算变换矩阵的误差. 针对这种情况, 本文改进了原来的 SDF 方法^[4,15]中的权重取值方法

$$D_{k+1} = \frac{D_k + w_{k+1}d_{k+1}}{w_k + w_{k+1}} \quad (19)$$

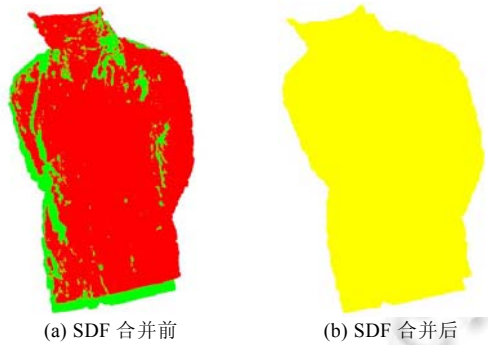
$$w_k = 1 - \text{err}_{R_k} - \text{err}_{I_k} \quad (20)$$

$$\text{err}_{R_k} = \frac{1}{k} \sum_{i=1}^k \cos(n_i \cdot n_{i+1}) \quad (21)$$

$$\text{err}_{I_k} = \frac{1}{k} \sum_{i=1}^k \|P_i - P_{i+1}\| / \sqrt{\|P_{i+1}\|^2} \quad (22)$$

其中, n_{i+1} 表示第 $i+1$ 帧点云中选取 3 块点云的加权平均法向, n_i 表示第 i 帧点云中选取 3 块点云的加权平均法向, 两者的夹角余弦表示旋转误差. 对于平移误差使用第 i 帧和第 $i+1$ 帧中匹配点之间的距离与第 $i+1$ 帧中匹配块到原点距离之间的比值. 当变换矩阵估计误差为 0 时, $w_k=1$, 所以原来的 SDF 方法^[4,15]中的权重取值方法是这里的一种特殊情况. 图 7 是合并的效果图.

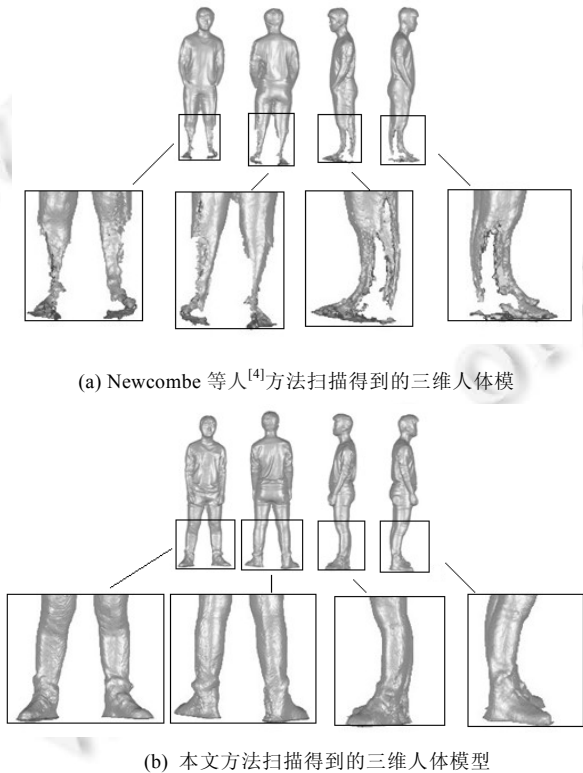
点云合并之后, 需要通过表面重建得到网格模型. 但通常这些点云是不规则的, 从这些点云中直接进行表面重建是一个病态问题, 即没有唯一解. 一个理想的重建过程应该是能处理复杂的拓扑和几何, 同时对于不规则带噪声的点云数据也能重建出一个近似的表面并能保证表面的光滑. 一类方法是显式曲面的方法, 这类方法通过插值的形式进行网格化, 重建表面精确, 但计算量与点云的数量成正比, 重建表面会受到点云数据噪声的影响, 往往需要后期进行光顺和修补等处理. 如 Delannay 三角剖分^[23], Voronoi 图^[24]等. 另一类是隐式曲面的方法, 这类方法可以有效地重建带噪声的点云. 比如以径向基(RBF)为隐式函数的方法^[25], 但是在实际操作时得到的系数矩阵是稠密的且病态的. 本文采用了泊松表面重建方法^[26]. 该方法将表面重建问题表示为泊松方程的解, 通过寻找一个指示函数, 使其最佳近似一个带噪声非均匀的点集, 其主要步骤如下: (1) 将模型表面采样的有向点集转换为模型指示函数梯度样本点; (2) 计算指示函数, 使用隐函数方法进行表面重建; (3) 通过指示函数提取对应等值面, 获得三维重建表面; (4) 采用八叉树技术, 将细节点所在面的等值线段投影到粗节点所在面上, 避免产生裂纹, 优化重建表面. 该方法运行速度快, 构网精度高, 结合了之前隐式函数方法的优点并可以解决基函数解矩阵病态的问题.



(a) SDF 合并前 (b) SDF 合并后
图 7 SDF 合并效果图

6 实验结果分析

文中算法在内核 Intel i7,内存 8G,显卡 GT630M 的笔记本上运行.我们对一些人體模型进行三维重建,并将本文重建效果与 Newcombe 等人^[4]的方法进行比较,如图 8 所示.可以发现,Newcombe 等人的方法在重建过程中有时会出现如图 8(a)所示,腿部缺失的现象,而图 8(b)中应用本文方法得到的结果比较完整.



(a) Newcombe 等人^[4]方法扫描得到的三维人体模

(b) 本文方法扫描得到的三维人体模型

图 8 本文方法与 Newcombe 等人^[4]方法的结果比较

我们还使用本文方法对三维人体的不同部位如头部、脚部使用 Kinect 进行三维扫描和重建,实验效果如图 9 和图 10 所示.结果发现,本文方法得到的重建效果能很好的保持局部细节(如脚趾信息和头发特征等).图 11 是我们对另外一些人體模型三维重建的正面和背面效果,结果发现与真实人体较为准确.表 1 是图 11 中不同三维人體模型扫描后的网格模型数据信息和扫描时间,从表 1 中可以看出随着顶点数和三角片数的增加,本文方法的时间超过 Newcombe 方法.扫描过程中,人为操作 Kinect 不当可能会导致扫描时间延长,但超出 5%左右的时间

变化还是可以接受的.图 12 是我们对重建后的三维人体模型经过纹理映射的人体重建效果图,重建效果更为真实.另外,我们还可对重建人体模型通过骨架驱动的皮肤方法进行变形^[33],可以产生一组运动序列,应用于人体运动和动画,如图 13 所示.

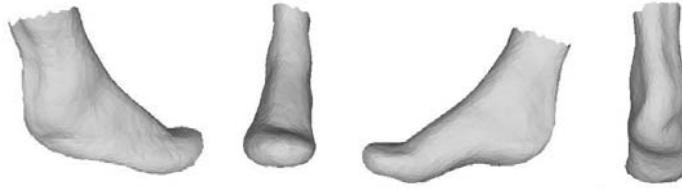


图 9 脚部模型

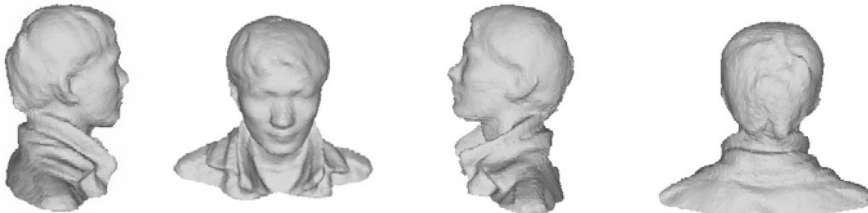


图 10 头部模型

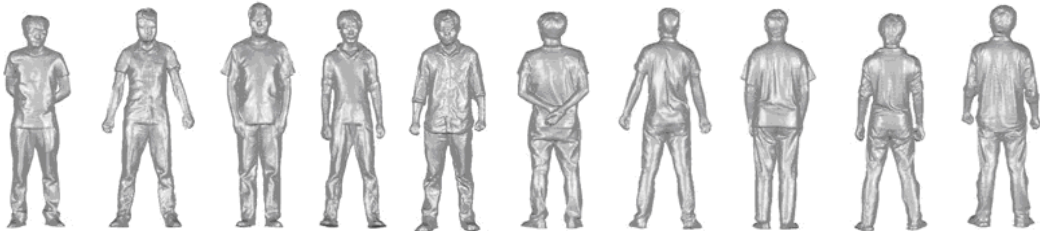


图 11 扫描得到的不同三维人体模型

表 1 图 11 中不同人体模型三维重建的数据信息

模型	重建网格模型顶点数	重建网格模型三角片数	Newcombe 等人 ^[4] 方法重建时间(s)	本文方法重建时间(s)
人体模型 1	89 389	176 084	450	430
人体模型 2	137 651	273 570	460	480
人体模型 3	102 391	204 782	453	460
人体模型 4	95 598	191 196	446	450
人体模型 5	95 356	190 708	440	458

7 结论及未来工作

本文提出了一种通过在静态人体周围动态移动一台类 Kinect 实时重建三维人体模型的方法.该方法首先对原始深度数据进行降噪和孔洞修复的预处理,然后结合深度信息和 RGB 信息选取匹配块并使用高斯混合模型进行局部匹配和曲线拟合方法进行全局配准,对匹配后的点云数据进行方向距离函数合并,最后使用泊松表面重建,得到人体重建模型.该模型可用于游戏娱乐、服装工业、运动分析等领域,由于扫描设备成本低,具有很高的应用价值.

文中工作也存在一些不足.在实际扫描过程中需要人体固定姿态,这对被扫描者来说不是很不方便,人体姿态在一定时间内内容会存在一定变化误差,如何修正人体变化扰动误差及进一步提高扫描速度将是我们今后改进的主要方向.



图 12 纹理映射后的三维人体模型

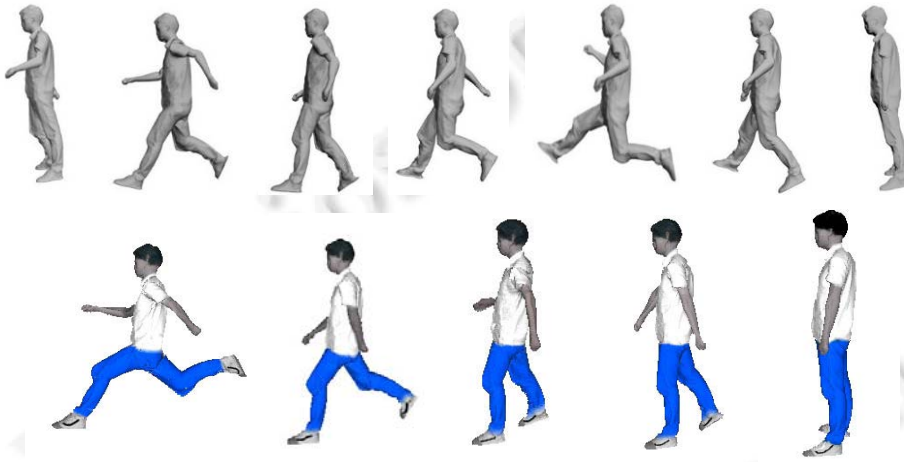


图 13 人体运动序列

References:

- [1] 3D scanner. http://en.wikipedia.org/wiki/3D_scanner
- [2] Microsoft Kinect. <http://www.microsoft.com/zh-cn/kinectforwindows>
- [3] PrimeSense. <http://en.wikipedia.org/wiki/PrimeSense>
- [4] Newcombe R, Davison AJ, Izadi S, Kohli P, Hilliges O, Shotton J, Molyneaux D, Hodges S, Kim D, Fitzgibbon A. Real-Time dense surface mapping and tracking. In: Proc. of the IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR). 2011. 127–136.
- [5] Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, Shotton S, Hodges J, Freeman D, Davison A, Fitzgibbon A. KinectFusion: Real-Time 3D reconstruction and interaction using a moving depth camera. In: Proc. of the 24th Annual ACM Symp. on User Interface Software and Technology. 2011. 559–568.
- [6] Chen J, Bautembach D, Izadi S. Scalable real-time volumetric surface reconstruction. ACM Trans. on Graphics (TOG), 2013,32(4): 113.
- [7] Roth H, Vona M. Moving Volume KinectFusion. British Machine Vision Conf. 2012. 1–11.
- [8] Weise T, Bouaziz S, Li H, Pauly M. Realtime performance-based facial animation. ACM Trans. on Graphics, 2011,30(4): 77.
- [9] Weiss A, Hirschberg D, Black MJ. Home 3D body scans from noisy image and range data. IEEE Int'l Conf. on Computer Vision. 2011. 1951–1958.
- [10] Tong J, Zhou J, Liu LG, Pan ZG, Yan H. Scanning 3d full human bodies using Kinects. IEEE Trans. on Visualization and Computer Graphics, 2012,18(4):643–650.
- [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)
- [12] Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J. SCAPE: Shape completion and animation of people. ACM Trans. on Graphics, 2005,24(3):408–416.
- [13] Nocedal J, Wright S. Numerical Optimization. New York: Springer-Verlag, 1999.
- [14] Cui Y, Schuon S, Thrun S, Stricker D, Theobalt C. Algorithms for 3D shape scanning with a depth camera. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(5):1039–1050.

- [15] Myronenko A, Song XB. Point set registration: Coherent point drift. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(12):2262–2275.
- [16] 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. 1996. 303–312.
- [17] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In: Proc. of the 3rd Int'l Conf. on 3-D Digital Imaging and Modeling. 2001. 1–145.
- [18] Schuon S, Theobalt C, Davis C, Thrun S. LidarBoost: Depth superresolution for ToF 3D shape scanning. IEEE Conf. on Computer Vision and Pattern Recognition. 2009. 343–350.
- [19] Nguyen C, Izadi S, Lovell D. Modeling Kinect sensor noise for improved 3D reconstruction and tracking. In: Proc. of the 2nd Int'l Conf. on 3-D Imaging, Modeling. IEEE, 2012. 524–530.
- [20] Yang S, Qi Y, Shen XK, Zhao QP. Rapid and automatic method for 3D scanned data registration. Ruan Jian Xue Bao/Journal of Software, 2010,21(6):1438–1450(in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3543.htm> [doi: 10.3724/SP.J.1001.2010.03543]
- [21] Weise T, Wismer T, Leibe B, Van G. In-Hand scanning with online loop closure. In: Proc. of the 12th IEEE Int'l Conf. on Computer Vision Workshops (ICCV Workshops). 2009. 1630–1637.
- [22] Sharp G, Lee S, Wehe D. Multiview registration of 3D scenes by minimizing error between coordinate frames. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004, 1037–1050.
- [23] Gelfand N, Mitra N, Guibas L, Pottmann H. Robust global registration. In: Proc. of the Euro Graphics Symp. on Geometry Processing. 2005,2(3):5.
- [24] Kolluri R, Shewchuk J, O'Brien J. Spectral surface reconstruction from noisy point clouds. In: Proc. of the 2004 Euro Graphics/ACM SIGGRAPH Symp. on Geometry Processing. 2004. 11–21.
- [25] Amenta N, Choi S, Kolluri R. The power crust, unions of balls, and the medial axis transform. Computational Geometry, 2001, 19(2):127–153.
- [26] Turk G, O'Brien J. Modelling with implicit surfaces that interpolate. ACM Trans. on Graphics, 2002,21(4):855–873.
- [27] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. In: Proc. of the 4th Euro Graphics Symp. on Geometry Processing. 2006. 61–70.
- [28] Seo JK, Sharp GC, Lee SW. Range data registration using photometric features. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2005. 1140–1145.
- [29] Roth G. Registering two overlapping range images. In: Proc. of the 2nd Int'l Conf. on 3-D Digital Imaging and Modeling. 1999. 191–200.
- [30] Spektor E, Mor Z, Rais D. Integrated processor for 3D mapping. 2013.
- [31] Barnes C, Shechtman E, Finkelstein A, Goldman DB. PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans. on Graphics, 2009,28(3):24.
- [32] Simakov D, Caspi Y, Shechtman E, Irani. Summarizing visual data using bidirectional similarity. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2008. 1–8.
- [33] Baran I, Popović J. Automatic rigging and animation of 3D characters. ACM Trans. on Graphics (TOG), 2007,26(3):72.

附中文参考文献:

- [11] 周瑾,潘建江,童晶,刘利刚.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.
- [20] 杨琴,齐越,沈旭昆,赵沁平.一种快速的三维扫描数据自动配准方法.软件学报,2010,21(6):1438–1450. <http://www.jos.org.cn/1000-9825/3543.htm> [doi: 10.3724/SP.J.1001.2010.03543]



许鸿尧(1986—),男,浙江绍兴人,硕士,主要研究领域为计算机图形学。



金小刚(1969—),男,博士,教授,博士生导师,主要研究领域为计算机动画。



李重(1975—),男,博士,教授,博士生导师,主要研究领域为数字几何处理,服装 CAD。



马利庄(1963—),男,博士,教授,博士生导师,主要研究领域为计算机图形图像。