# Kinect Fusion 高精度三维重建 Proposal

## Changelog

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 变更人 | 变更说明 | 变更时间 |
| V1.0 | 张琛 | 初稿 | 2014/10/19 |

## 项目目标&任务

### 项目目标

面向用户对廉价、便利获取三维数字模型日益增长的需求，本项目研究手持 Kinect 对物体的高精度(精度达到 1 mm)重建技术与系统，使得低廉的三维模型获取装置和系统能在家庭用户中推广和普及。重点研究基于深度图插值的小物体重建、累积误差控制等问题，突破 Kinect 深度图低分辨率、噪声明显等局限，研发手持已标定 Kinect 针对图1所示场景的高精度重建平台，在脚型获取、数字鞋楦逆向工程、文物重建方面实现典型应用。



图 1大小已知的高精度方块与待重建小物体

### 任务需求分析

三维重建技术通过物体的二维信息获取物体在空间中的三维信息，对三维物体建立适合计算机表示和处理的数学模型，在计算机中对其进行操作处理并分析相关性质。三维重建技术在数字娱乐、仿真设计、文化遗产保护和工业逆向工程等方面有广阔的应用前景，目前三维重建装置和技术主要有以下三类：

1. 基于激光的高精度三维扫描[1]技术业已成熟，但价格昂贵、操作复杂，难以在普通用户群体中推广；
2. 基于被动型传感器的多视图立体匹配（MVS）技术[2], [3]比较成熟，然而对光照环境、纹理等因素敏感，难以应对稀疏、重复纹理和无明显特征物体；
3. 基于主动型传感器(ToF 、Kinect)三维重建技术[NDI+11, IKH+11, TZL+12, CSC+10,KDS+12]，因其廉价、便携装置，不受光照、纹理等方面影响，在近年广为研究和应用并取得了很大进步，国外出现了 ReconstructMe1、KScan3D2 等基于 Kienct 的商业软件3，但由于 Kinect有效感应距离范围、深度图固有噪声误差等方面局限，上述软件存在小物体重建精度低、环绕拍摄累积误差明显及平面场景重建时易出现明显漂移等方面的不足，限制了其推广使用范围。

因此，解决基于 Kinect 的小物体重建、累积误差控制等关键技术问题，研发实用Kinect重建基础软件平台，实现低成本、便携、高精度的三维重建应用，成为三维重建领域新的研究热点。

### 现有问题描述

Kinect Fusion目前存在如下问题：

1. 配准在旋转到特定视角时存在漂移，如图2a所示；
2. Kinect 分辨率较小，对于小物体，需要将设备靠近采集深度数据，以便获得更多细节。但近距离采集数据容易出现大量无效深度区域，如图2b黑色阴影区域所示；
3. 手持kinect设备进行旋转平移，建模场景时，累计误差可能影响较大（图2c, 2d所示）。

对以上问题，目前解决思路大致如下：

1. 预处理：

从kinect中获取原始深度数据，借鉴雷昊的相机标定工作，得到相机内参，输出我们矫正过的深度数据，与kinect输出的深度图进行对比；

1. 利用彩色图分割实现深度图边界修复，整合刘良建之前的工作，得到修复之后的深度图；
2. 配准：

对深度图数据，ICP配准阶段使用平面、轮廓约束；

在参照物表面，或在场景中附加定标板，使用彩色图像进行定标求解外参 R, T；

1. 使用平面单应关系求解基础矩阵，解决大比例平面点云由单应点对求解基础矩阵易出现配置退化，导致出现图2a所示漂移问题，从而建立起能适应平面场景的三维重建系统。
2. 基于粗略模型对原始低分辨率深度图进行插值，实现对小物体(如图2b所示)的高精度重建。
3. 使用轮廓一致性约束进行全局优化，达到对累积误差(如图2c2d)有效控制。

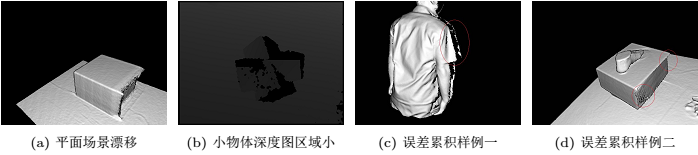


图 2现有Kinect重建技术存在的问题

## 相关工作基础

三维重建技术是图形学、视觉交叉点，通过对实物的二维光度、深度图信息进行分析处理，重建出物体对应三维数字模型供图形学进行交互、编辑、变形和绘制。在产品设计、文物遗产保护、影视游戏、数字娱乐、人机交互及工业逆向工程等方面有着重要的、不可替代的优势，具有广阔的应用前景。

三维重建技术作为一个从图形学领域兴起以来，至今仍然历久弥新的研究课题。从90 年代出现的激光扫描重建（Stanford 数字化米开朗基罗计划），到本世纪初开始热门的多视图立体匹配技术，一直到近几年来相当热门的基于 ToF 、Kinect 等主动传感器的移动便携重建装置和技术，三维重建技术一直受到广泛关注和研究。本课题主要研究基于主动传感器的重建技术，目前这方面已有大量研究成果，这些成果为本课题的顺利实施奠定了坚实的基础；下面从主要研究技术及相关应用等方面综述研究现状与趋势。

### 基于主动传感器的三维重建技术进展综述

近年来，基于 ToF 、Kinect 等主动传感器的三维重建技术迅猛发展。Izadi 等[NDI+11]提出基于 GPU加速的实时定位与重建技术，并实现动态场景的增强现实[IKH+11]和交互[CIF12]应用；Tong 等[TZL+12]使用多个固定位置的 Kinect 对站在转台上旋转的人体进行重建；刘鑫等[6]等对基于 Kinect 对转台上物体进行快速。Cui 等[CSC+10]提出使用ToF 相机对三维物体进行扫描，通过对深度图进行超分辨率化，实现较高精度的重建方法。应用方面涌现了 ReconstructMe4、KScan3D5等基于 Kienct 的商业软件。

基于主动传感器获取获取深度图进行重建三维的技术一般采用图3所示流程模型，通过内参矩阵直接将传感器采集的深度图投射为三维点对，由点云间 ICP 配准求解出各帧间相对变换矩阵，将各帧深度图或点云转换到同一世界坐标系下，最终通过构造隐式表面或对整体点云进行去冗、网格化重建出三维模型。

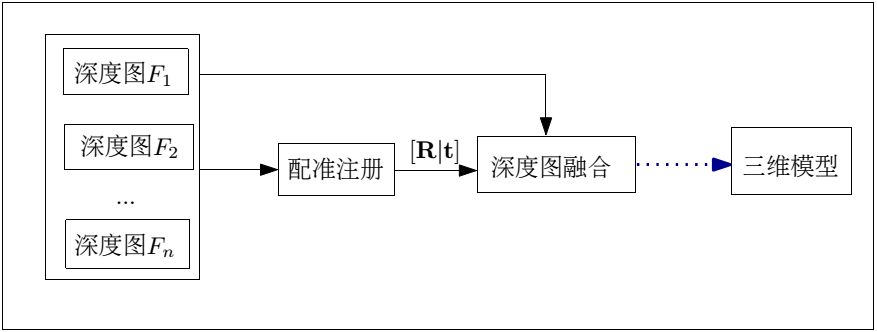


图 3基于深度图的常用重建流程

### KinectFusion及其扩展

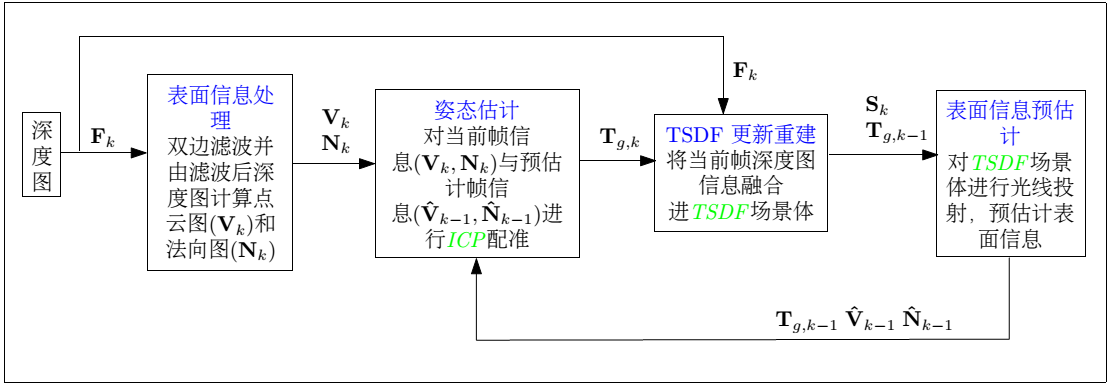


图 4 KinectFusion 流程

Izadi 等[NDI+11]提出如图4所示的 KinectFusion 流程：使用体表示场景，每个体素包含符号距离函数(TSDF)值；姿态估计阶段 ICP 使用点到法平面距离作为度量项，实时采集中相邻帧角度变化较小，ICP 配准可简化为线性最小二乘问题求解；同时各阶段实现 GPU 加速，达到实时重建。由于 KinectFusion 采用体素模型，其场景尺度受限于 GPU 内存空间，且当场景三维几何信息较少时易出现跟踪失败(如图2a所示) 。 Zeng等[ZZZL12]使用八叉树结构提高 KinectFusion GPU 内存效率，可处理场景空间增至 8 倍；Whelan 等[WKF+12]通过在相机移动超过阈值时提取点云和网格模型，利用 RGB 图像特征对各场景体点云和网格拼合，实现大场景重建。 **总言之，现有 KinectFusion 改进工作主要集中在扩展其大尺度重建能力，而改善重建精度方面工作较少。**

由于 Kinect 等深度相机存在感应范围限制，如 Kincet 只能采集 [0.4m, 5m] 的深度值，其中 [1.2m, 3.5m] 范围内深度值较准，此范围外深度值不可靠；此限制下，待重建物必须置于足够远处，导致小物体重建结果非常粗略。同时，深度图间配准时，由单应点对求解或拟合基础矩阵，误差难以避免，长系列环绕拍摄重建时累积误差明显，重建结果出现如图2所示的明显误差。

## 研究内容

本课题拟重点研究以下内容：

1. 通过内定标、RGBD结合等多种方式对深度图进行校正、修复；
2. 使用多种几何信息约束的深度图注册。采用手持 Kinect 环绕拍摄获取待重建物深度数据，需要恢复出深度图影像的拍摄方位，进而通过融合深度图实现物体的三维重建。拟重点研究：深度图去噪、点云配准技术；综合考虑平面单应和点对单应的基础矩阵求解技术。
3. 使用RGB图像进行相机外参标定。

## 技术路线

### 深度图校正

1. 通过 openni 接口得到原始深度数据 disparity 🡪 使用雷昊的相机内定标程序，获得相机内参 🡪 使用新的内参计算得到处理过的深度图；
2. 整合刘良建先前孔洞填充、物体边界修复的代码，修复深度图无效区域。

### 通过几何信息约束进行外参定标

综合考虑平面夹角、平面距离、单应点对等几何信息的约束方程，进行大比例平面场景深度图配准。主要步骤为：对深度图计算出的点云进行平面拟合，配准阶段先通过平面对关系确定粗略变换矩阵，再通过最小化综合点对误差、平面约束误差(法向夹角和平面距离)的目标函数，对变换矩阵进行求精和细化。

**难点：**

1. 要读懂PCL源码，调试，找到优化代价函数的代码部分；
2. 高精度参照物在深度图中占面积较小，对场景配准结果可能影响较小。

### RGB图像信息约束定标相机外参

在场景中添加棋盘格或椭圆定标板，使用RGB相机定标，计算设备的旋转平移矩阵。当定标板位于相机视窗中，且拍摄角度合适时，采用RGB相机外定标结果；否则，采用几何信息约束求解外参

**难点：**

1. Kinect RGB、深度相机之间的旋转偏移关系未知，RGB相机外定标结果不能直接用于深度相机；

## 项目预期成果

本项目重点研究基于深度图插值的小物体重建、累积误差控制等问题，突破 Kinect 深度图低分辨率、噪声明显等局限，研发手持已标定 Kinect 针对图1所示场景的高精度重建平台，在脚型获取、数字鞋楦逆向工程、文物重建方面实现典型应用。

通过对相机精确的内定标，获得更精确的设备深度数据；通过场景中设置定标板，获得更精确的相机外参，以得到更好的点云配准结果。并与现有三维重建算法/程序进行对比。

## 参考文献

[1] Levoy M, Pulli K, Curless B, et al. The digital Michelangelo project: 3D scanning of large statues[C]//Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000: 131-144.

[2] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge university press, 2003.

[3] Beeler, Thabo, et al. "High-quality single-shot capture of facial geometry." ACM Transactions on Graphics (TOG) 29.4 (2010): 40.

[4] [NDI+11] R.A. Newcombe, A.J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton,D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. Kinectfusion: Realtime dense surface mapping and tracking. In Mixed and Augmented Reality(ISMAR), 2011 10th IEEE International Symposium on, pages 127–136. IEEE,2011.

[5] [KDS+12] Y.M. Kim, J. Dolson, M. Sokolsky, V. Koltun, and S. Thrun. Interactive acquisition of residential floor plans. In Robotics and Automation (ICRA), 2012 IEEE International Conference on, pages 3055–3062. IEEE, 2012.

[6] 刘鑫, 许华荣, and 胡占义. 基于gpu和kinect的快速物体重建. 自动化学报, 38(8), 2012.