# 深度图进行3D建模方案

## Changelog

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 变更人 | 变更说明 | 变更时间 |
| V1.0 | 张琛 | 初稿 | 2015/02/18 |

目录

[深度图进行3D建模方案 1](#_Toc412024755)

[Changelog 1](#_Toc412024756)

[1. 背景 1](#_Toc412024757)

[2. 当前相关工作 2](#_Toc412024758)

[3. 经典流程 & 算法步骤 4](#_Toc412024759)

[4. 关键点、难点分析 5](#_Toc412024760)

[5. 难点解决 6](#_Toc412024761)

[6. 参考文献 7](#_Toc412024762)

表 1缩略语查找表

|  |  |
| --- | --- |
| 缩略语 | 全称描述 |
| SLAM | Simultaneous Localization And Mapping |
| PTAM | Parallel Tracking And Mapping |
| AR | Augmented Reality |
| SFM | Structure from Motion |
| MVS | Multi-view Stereo |
| ToF | Time-of-Flight |
| SDF | Signed Distance Function |
| TSDF | Truncated Signed Distance Function |
| ICP | Iterative Closest Point |
| DOF | Degrees-of-freedom |
| IR | Infra-red |

## 背景

3D重建技术在计算机视觉及图形学领域具有重要意义，并随着深度相机（如 Microsoft Kinect，或Asus Xtion）的推广，逐渐获得高度关注。由于深度相机能够实时产生深度图，使用这种消费级深度相机进行实时曲面重建近来成为研究领域的关注重点。实时曲面重建的出现，启发了各种交互式应用场景，如增强现实（AR）中，真实世界的物体能够实时重建、渲染为3D模型；机器人自动导航系统可以迅速构建周围环境；甚至用户能够在3D扫描过程中获得实时反馈。

实时重建技术，需要对大量重叠深度图进行增量式融合，并不断提高精度，以得到单个场景（物体）模型。这一工作在既要求实时性能，有要求大规模、高质量重建时会比较困难。因此目前许多实时技术会借助GPU硬件进行加速，但仍然不得不在重建速度、规模、精度上进行一些取舍。

## 当前相关工作

由深度图进行3D模型重建的过程，主要可以分为三部分：深度图（点云）的获取，相机姿态的估计，以及融合不同视点的深度图序列生成3D模型。深度图的获取，主要有基于立体视觉的，基于结构光，以及3D激光扫描等方法。相机姿态估计与3D场景模型生成是一个“先有鸡还是先有蛋”的问题——相机在世界坐标系中的位置与姿态需要依赖场景的3D模型，而场景3D模型的生成反过来也需要先知道相机姿态。因此，相机轨迹姿态的获取与3D模型的生成需要同时进行。

基于立体视觉的方法[1][2]，通过对目标场景或物体不同视角的图像进行特征点匹配，并估计这些配对特征点的3D位置，以获得场景的稀疏点云作为深度图。近年来相机技术的发展，使得基于结构光或ToF技术的相机能够直接获得稠密深度点云数据，而不必通过两幅RGB图像计算视差获取深度图。微软Kinect设备的出现[3]，使得这类技术成本达到了消费级，通过结构光技术，我们能够以VGA分辨率，实时获取场景的稠密深度图。

通过Kinect这类设备获取的原始深度图，含有较大的噪声以及无效区域，Xiaoyi Jiang等人通过结合彩色图像的边缘信息进行图像分割的技术[11]，以及使用多台Kinect设备[12]进行深度图融合预处理，对原始深度图数据进行边缘修复以及孔洞填充，并能够做到实时产生高质量的深度数据。

论文[4]演示了对小物体的高精度实时扫描，但是[5][6]表明更大场景的重建缺失导致精度或效率的降低。基于高度图的表示法[8]能够进行有效的联通曲面数据的压缩，并且能有效应用在较大场景，但是却难以重建复杂的立体结构。

Newcombe等人提出的KinectFusion 算法[7]，是第一个高精度实时三维重建的系统，它将场景表示为体素模型，通过截断带符号距离函数（TSDF）表示，结合迭代最近点（ICP）算法，使用GPU进行加速计算，能够以空前的精度水平对场景进行实时三维重建。不过由于使用TSDF体素模型，这一算法也存在一些局限性，包括：刚性的表面模型无法进行模型形变，系统无法工作在无界的大场景中，以及在3D几何结构不明显的场景中易重建失败等。场景规模限制的问题，本质上是因为显存空间不足，可以通过层级空间分割法[9]，或更为通用的，采取体模型跟随相机滑动，对场景进行部分页交换的策略[10]来解决。呙维等人通过多次进行小模型重建，并根据彩色图像特征进行多模型拼接的方法[13]，实现了大型室内场景的三维重建。

## 经典流程 & 算法步骤



图 1系统整体工作流程图

图1演示了系统的整体工作流程，以下是算法步骤描述：

**1. 数据预处理**

Kinect设备输出的RGB以及深度数据分别由两组位置不同的传感器得到，一个是普通的RGB相机，另一个是一对红外线发射/接收感应器。由于RGB相机与红外线接收器位置不完全重合，因此得到的彩色图与深度图存在一定的视差。如果三维重建的过程需要同时依赖彩色图特征信息，需要预先对两个摄像头进行标定，获取二者之间的相对坐标转换关系关系，并将深度图转换到彩色图的视角下，进行对齐。

为减小噪声，对每一帧原始深度图 ，使用双边滤波，进行保留边缘的去噪，得到 。同时，拟采用结合彩色图像的边缘信息进行图像分割的技术[11]，以及使用多台Kinect设备[12]进行深度图融合预处理，对原始深度图数据进行边缘修复以及孔洞填充，产生高质量的实时深度数据。

关键帧的6DOF计算，我们采用点到面ICP（Point-to-plane ICP）算法，进行迭代求解，需要用到第k帧的表面顶点图，以及法向图。同时，为节约计算资源，采用三层图像金字塔方法，需要对每一帧预先求出。

u 𝑡保留边缘的，gned distance)

**2. 相机6DOF姿态估计**

我们所创建的三维模型，是以系统启动时的初始相机坐标系为世界坐标系的。对于k时刻，需要计算相机在世界坐标系下的6DOF姿态 。我们使用ICP算法，利用相邻两关键帧k, k-1帧中的所有三维点信息，求解二者之间的相对变换关系 ，于是得到第k帧相机的姿态：

(1)

(2)

这使不同视点获取的RGB-D图像位于同一坐标系下，为下一步数据融合做准备。

我们使用齐次矩阵 表示 刚性变换：

 (3)

**3. 体素数据融合**

由多个不同视角的深度图，融合得到场景的完整表面，其中心思想是使用一个连续的隐式函数 来表示每一个采样点（体素）。函数D(x)中每一个体素X，包含权重和距离信息(weighted signed distance)。从体素X沿视线观察方向F与物体表面相交，通过计算有向距离函数（signed distance function, SDF）得到体素的值：正值表示沿射线F方向，体素X位于重建表面之前，负值表示位于重建表面之后，而 则表示重建出的物体表面。

新的深度数据不断生成，体素模型需要不断更新。对于同一场景，由于不同视角下获得的深度数据误差不同，因此融合时需要对其加权平均，以避免三维模型出现褶皱。假设深度图序列为 ，对应权重为 ，则

(4)

**4. 光线投射与表面预测**

由步骤3得到的，融合第1,…,k帧后的全局TSDF表示记为 ，其中，，是3D体模型中的某全局坐标点。TSDF每个位置存储两部分：当前TSDF值，以及权重，

(5)

对于每个像素，其对应的光线射线，从像素对应的最小深度开始，直到过零值（对可见表面，为+ve到-ve），即到达曲面表面为止，返回此处的XYZ坐标，即为该像素对应的三维坐标。

将得到的带入步骤2，作为配准目标，继续求解新数据帧第k帧对应的相机姿态，直到不再有新数据输入为止。最终输出为三维场景模型。

## 关键点、难点分析

**场景规模限制**

原始KinectFusion算法对重建的三维场景采用体素模型表示，内存/显存利用率较低，因此对场景规模有限制。当前算法在中等大小（）的室内空间运行良好，但是对于较大建筑内部整体空间的重建，仍然存在较大挑战。

**鲁棒性**

相机姿态的求解，使用了ICP算法，对输入深度图的几何结构要求较高。如果场景中存在较大的平面区域，深度图几何结构不明显，会导致6DOF中有三个自由度没有约束，可能导致三维重建失败。

设备移动速度过快，或大幅摇晃震动，也会导致相邻两帧之间深度图差异较大。由于ICP算法的最终迭代结果为局部最优解，此时可能无法正确配准前后两帧，也会导致三维重建失败。

**精确性**

尽管相对设备造价而言，Kinect生成的稠密深度点云质量不错，原始深度数据中存在较大噪声，特别是深度图中存在大量“孔洞”区域。造成这种现象的原因可能是场景中物体表面材质不反射IR光，或者三维结构非常薄，或者表面与IR光入射夹角非常大。同时，与一般彩色相机类似，当设备快速移动时，同样会因为运动模糊导致数据损失。

原始KinectFusion算法仅仅使用深度数据，配准精度完全依赖场景的3D几何结构。如果场景中存在较大的平面区域，会导致6DOF中有三个自由度没有约束，造成较大的配准误差，产生场景漂移的后果，甚至三维重建失败。

**不可自动恢复性**

当由于设备移动速度过快，或场景三维结构过于平整，造成设备轨迹追踪失败时，系统无法从错误状态自动恢复。此时，不得不对系统进行重置，重新进行三维重建，新生成的三维模型与出错之前的三维模型难以拼接为一个整体。

## 难点解决

1. 对于原始深度数据噪声大、精度不高的问题，可以采用[11][12]等方法，对原始深度图进行修复。

2. 对于设备大幅晃动导致配准失败的问题，可以采用与IMU单元相结合的方式。IMU单元对设备突然晃动较敏感。当设备慢速平稳移动时，使用ICP进行相邻关键帧（或关键帧-模型）配准，求解相机姿态；当设备产生迅速大幅晃动时，使用IMU产生的加速度、角加速度数据，对设备姿态进行粗略估计，然后进一步用ICP算法进行精细求解。

3. 对于原始KinectFusion算法的场景规模限制，有一些算法致力于解决此问题。例如[14]提出一种空间哈希算法，对TSDF体素表示进行压缩，使内存占用节省为原始算法的10%，以便适用于更大的场景，同时获得更高的实时帧率。

## 参考文献

[1] Pollefeys M, Van Gool L, Vergauwen M, et al. Visual modeling with a hand-held camera[J]. International Journal of Computer Vision, 2004, 59(3): 207-232.

[2] Newcombe R A, Davison A J. Live dense reconstruction with a single moving camera[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 1498-1505.

[3] http://www.microsoft.com/en-us/kinectforwindows/

[4] Weise T, Wismer T, Leibe B, et al. In-hand scanning with online loop closure[C]//Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on. IEEE, 2009: 1630-1637.

[5] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. The International Journal of Robotics Research, 2012, 31(5): 647-663.

[6] Stuckler J, Behnke S. Integrating depth and color cues for dense multi-resolution scene mapping using rgb-d cameras[C]//Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on. IEEE, 2012: 162-167.

[7] Newcombe R A, Izadi S, Hilliges O, et al. KinectFusion: Real-time dense surface mapping and tracking[C]//Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on. IEEE, 2011: 127-136.

[8] Gallup D, Pollefeys M, Frahm J M. 3d reconstruction using an n-layer heightmap[M]//Pattern Recognition. Springer Berlin Heidelberg, 2010: 1-10.

[9] Zeng M, Zhao F, Zheng J, et al. Octree-based fusion for realtime 3D reconstruction[J]. Graphical Models, 2013, 75(3): 126-136.

[10] Whelan T, Johannsson H, Kaess M, et al. Robust real-time visual odometry for dense RGB-D mapping[C]//Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, 2013: 5724-5731.

[11] Schmeing M, Jiang X. Edge-aware depth image filtering using color segmentation[J]. Pattern Recognition Letters, 2014, 50: 63-71.

[12] Schmeing M, Krauskopf E, Jiang X. Real-time depth fusion using a low-cost depth sensor array[C]//3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2014. IEEE, 2014: 1-4.

[13] 呙维, 胡涛, 朱欣焰. 基于 Kinect 的深度数据融合方法[J]. 计算机应用研究, 2014, 31(1): 285-288.

[14] Nießner M, Zollhöfer M, Izadi S, et al. Real-time 3d reconstruction at scale using voxel hashing[J]. ACM Transactions on Graphics (TOG), 2013, 32(6): 169.