# 深度图进行3D建模方案

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 |

## 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。

## 关键点、难点分析

场景规模限制

鲁棒性

3D几何结构不明显的场景中易重建失败

相机移动速度导致ICP失败

不可恢复性

精确性

深度图孔洞

累计漂移误差

## 难点解决

## 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.