# Kinect Fusion 算法实现设计文档

## Changelog

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 变更人 | 变更说明 | 变更时间 |
| V1.0 | 张琛, 孙国飞 | 初稿，接口描述待补充 | 2014/10/12 |

目录

[Kinect Fusion 算法实现设计文档 1](#_Toc400958714)

[Changelog 1](#_Toc400958715)

[一． 目标 & 算法概述 1](#_Toc400958716)

[二． 表面测量（surface measurement） 2](#_Toc400958717)

[三． 映射表面重建（mapping as surface reconstruction） 2](#_Toc400958718)

[四． 利用光线追踪进行表面预测（surface prediction from ray casting the TSDF） 3](#_Toc400958719)

[五． 传感器姿态估计（sensor pose estimation） 4](#_Toc400958720)

[六． 参考文献 5](#_Toc400958721)

## 目标 & 算法概述

本文档目的在于给出对论文 [1] 进行重现的设计草稿，使用与其相同的硬件环境，在PC上驱动kinect，完成对场景的实时三维重建（不包括大场景）。

论文 [1] 给出了只利用廉价的深度相机和商业图形硬件在任意光照条件下进行精确的实时的室内场景绘制系统。我们把通过kinect传感器获取的所有深度图像实时地融合为一个全局的隐式曲面模型。当前的相机位置也同时通过利用ICP算法来跟踪当前深度图像和全局模型的相关联系来获得。论文比较利用模型来进行跟踪的方法（本文方法）和利用之前深度图来进行跟踪的方法（frame-to-frame），前者在室内场景中在跟踪和绘制时有着更高的精确度和有限的漂移。论文也给出了其系统在多个方面在数量和质量上的结果，它实现了稠密曲面的实施重建，并且在任何分辨率下都有着很好的细节和鲁棒性。

论文 [1] 算法主要由四部分组成：表面测量（surface measurement），映射表面重建（mapping as surface reconstruction），利用光线追踪进行表面预测（surface prediction from ray casting the TSDF）和传感器姿态估计（sensor pose estimation）。

下图1展示了系统整个工作流程：

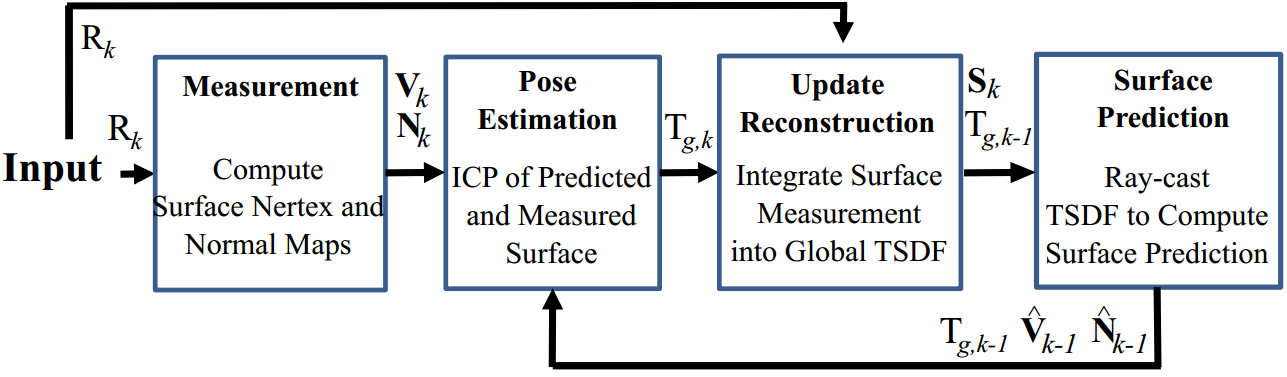


图 1 系统整体工作流程

以下详述算法每一步骤的实现细节。

## 表面测量（surface measurement）



图 2 表面测量流程图

如图1所示， 对于每一帧RGBD数据，表面测量由原始深度图，经过降噪处理，计算出三维点集 ， 法向量集。作为后续配准步骤的输入。

对像素点 u， 其法向计算公式为：



其中，，即向量归一化。

需要注意，为了提高配准效率，此处对提供一个3级图像金字塔接口；由大到小，图像分辨率逐级折半。

## 映射表面重建（mapping as surface reconstruction）



图 3 表面重建流程图

需要重建的三维物体中的每一个点p（对应离散化之后的每一个体素voxel），都可以根据当前（第k次测量）相机位置和当前相机获得的深度图计算出两个值TSDF和权重：

同样p也对应有两个全局的TSDF和权重：

是点p到距离自己最近的表面的截断距离（给定一个阈值μ，如果点p到距离自己最近的表面的归一化距离λ|r|<μ，就取r，否则取μ（或者-μ））。是这一点的权重值（本文中权重是利用该点表面法向和投影方向的夹角的余弦值来得到）。

得到当前相机位置下的局部TSDF和权重之后，结合之前的全局TSDF和权重，可以更新出当前相机位置下的全局TSDF和权重。对，可以设一个阈值，从而保证它不会太大。

## 利用光线追踪进行表面预测（surface prediction from ray casting the TSDF）



图 4 表面预测流程图

如图3所示，表面预测输出结果为 , ，作为下一步传感器姿态估计的输入。

假定一个当前相机位置，根据已有的体素模型，可以计算出模型在该相机位置下的深度图，从而可以计算出此深度图对应的三维点图,对应的法向图用另一种方法计算，即用在表面附近的点（=0或接近0）的TSDF值得梯度值来表示。在对离散化之后的体素进行三线性插值，可以得到更加精确的值（即，离散化之后得到的都是整数值，通过插值，可以得到准确的小数值）。

## 传感器姿态估计（sensor pose estimation）



图 5 传感器姿态估计流程图

对一个未知的相机位置，可以得到它拍摄的物体的深度图，从而可以得到相应的三维点图和法向图。假设该相机的位置为，那么由三可以得到之前一帧预测的，和。所要做的就是优化目标函数

其中，u和是两幅深度图上的对应点，其对应关系通过投影关联算法得到（projective associate algorithm原文中的引用4）。上述目标函数是个非线性的，通过线性化来求解（线性化的方法主要是认为旋转角度很小，从而有cosθ=1，sinθ=θ）。在求解时，通过不同分辨率的N和V的迭代计算，从而减少计算时间，提高效率。

求解目标函数的方法主要采用迭代的算法和矩阵的乔里斯基分解（Cholesky decomposition）来做。对于当前的相机位置，可以通过上一帧的结果左乘一个新的变换得到（这个变换是一个很微小的变换，所以可以用线性化来近似）。

其中（α，β，γ分别是绕x，y，z轴旋转的角度，，，是平移向量）。可以用一个向量来表示。那么原来要优化的目标函数E就变成了如下形式：

只要求出的最小二乘解即可。通过左右同时左乘，就变为（这个方程组是肯定有解的）。对上述方程组求解时用迭代的方法计算，X的初值全部设为0，通过对不同分辨率下的N和V分别迭代一定次数（4,5,和10次），就可以得到要求的X，从而得到相机位置的参数。

## 人员分工

孙国飞：

表面测量，表面重建

张琛：

表面预测，相机姿态估计，模块整合

## 参考文献

[1] Newcombe R A, Davison A J, Izadi S, 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.