# 关于三维配准与AR一些个人理解

## 论文查阅与一些个人理解

1. 三维配准(3D registration) 核心是：找到相机与场景(或其中某物体)的相对位姿关系{R,t}，例如，pwp3d[1] 在每一帧，寻找模型在2D 图像中的匹配物，求模型物体在相机坐标系下的姿态，也可以等价地转换为，求相机外参； KinectFusion[2]在每一帧，求解当前帧在全局坐标系(全局模型)下的姿态，即相机外参 (KinectFusion全局坐标系一般定义为: R与相机第0帧旋转量R0一致，t相对相机第0帧t0有一个合适的偏移，以便volume包含了我们要扫描的场景)。 无论在[1]还是[2]的算法流程中Tracking都很关键。
2. 增强现实(AR)的一个关键点是：鲁棒地实时检测相机姿态[8]。虽然过去常见于在RGB视频流上叠加虚拟三维模型，但是其后端技术并不仅限于RGB视频流，Izadi S 等人[3]将KinectFusion应用在增强现实中，就是使用仅基于深度图的3D视频流，虽然最后通过RGB视频进行演示，但是RGB数据仅仅用于渲染场景，而不参与三维配准过程。
3. 由Kinect采集到的Depth map，属于range image 的一种，在确定相机内参情况下，可以在相机坐标系下转化成唯一的3D点云[3]。若把深度图看做2D图像，可以将其转化为灰度图（或热度图），但是不同于RGB图的灰度化，深度图生成的灰度图纹理信息非常少；另外也可以通过计算深度图的法向图，法向图在每个点具有三个值，且其优点是：在世界坐标系下，同一物理点在不同视角下可以保持相对稳定的法向值，将法向图转换为RGB图，比深度图灰度化能更好地呈现几何纹理特性，但与一般RGB图相比还有很大差距。
4. 利用RGB图像序列进行三维重建，大部分是基于图像特征点的稀疏重建，它与基于深度图的三维重建，共性在相机外参求解；区别在于Kinect深度图的每一帧都能提供稠密三维点云，而RGB图像只能提供稀疏特征点，因此每帧需要首先通过将像素或特征点三维化生成稀疏(sparse)或半稠密(semi-dense)点云。对于“给定一个已知3D模型，一帧‘图像’数据，如何进行配准”的问题，我认为“一帧数据”的来源非常重要，RGB图像数据与深度图提供的“能力”完全不同，其形式化的问题描述、优化目标也会完全不同，并非“前者的数学原理是基础”。当然，深度图提供的稠密点云确实使得重建更方便、精确[17]，但是并不是说后者是前者问题的**简化**。
5. 我们要处理的问题确实是SLAM问题，尽管SLAM问题确实是由机器人领域兴起[13]的，但是判定问题是否为SLAM的标准并不是其构建场景的尺寸大小，不是其使用的输入数据是RGB还是depth数据，也不是其应用领域是robotics还是CV。SLAM问题的目标是: 同步进行相机定位(外参求解) 与地图构建，这确实是我们要实现的目标；即便抛开“同步”，例如SFM，仍然是要解决Localization+Mapping问题。 另外例如，pwp3d 要解决的不是SLAM 问题，如其标题所述，他的技术路线是分割与跟踪 (segmentation and tracking)，其与SLAM的共性是跟踪部分。在pwp3d作者后来的论文[19, 20]中，也进行进一步稠密点云重建，其中[20]属于monocular SLAM，而[19]不是SLAM，因为其致力于仅重建前景物体，而非整个场景。
6. ~~PTAM论文[18]提到，在场景中增加已知元素，extensible tracking [8, 9, 10, 11, 12]~~ //TODO
7. 影视项目中，黄倩妮&任光阔实现的论文[14]测试报告[16]中，3.2小节演示了相机姿态求解正确性测试，如图1所示。因为论文[14]仅做基于关键帧的相机跟踪，放置立方体只是用于验证算法运行结果的大致正确性，方便直观观察。但是其无法作为精度验证、可视化的方式。而在[2, 3, 9]等算法中，重建结果可直接反馈相机位姿的求解（定性观察），当然，位姿求解精度的定量评估仍要借助运动捕捉系统提供的groundtruth。



图 1 摘自报告[16]小节3.2，图中立方体仅用作演示相机姿态求解的正确性，并不用于配准，也不涉及重建

## 重建精度测试评估思路

SLAM问题，前端工作是相机外参求解，后端是地图拼接融合。我们目前仅致力于通过提升相机外参求解精度，来提升重建结果。具体的误差评估思路是：

1. 数据源准备: 模拟数据[4], 真实数据[5]；真实数据又分为：自采集数据（立方体+石膏模型），公开数据集TUM-dataset[6]、Teapot by Nguyen[7];
2. 测试方法：

① 定性观察[4, 5]不同算法重建结果的局部瑕疵;

② 定量评估又可分为：

1. 在模拟数据上，求重建结果（cloud或mesh）与groundtruth的差值[4]（一般热度图演示）。模拟数据可以排除深度图噪声引入的误差。[4]提到CloudCompare, <http://www.danielgm.net/cc/> 工具，//目前尚未调研 2016-4-27 20:22:41
2. 在模拟数据上，将相机外参(R, t)与真值进行对比。
3. 在真实数据TUM-dataset上，对立方体橱柜放在地面的数据组（cabinet）进行评估，我们可以根据其数据中橱柜的尺寸手动生成一个虚拟立方体，运行算法，观察相机**轨迹**误差[5]；当然这里也可以与[5]对比重建结果局部优劣，但是只能定性对比。
4. 在Nguyen[7]采集的Teapot数据集中，恰好也有个立方体，与cabinet数据处理方式类似，我们手动对其建模，然后运行我们的算法。
5. 在自采集的立方体+石膏模型真实数据上，运行算法得到重建结果，与之前用激光扫描仪扫描结果配准、做差（考虑使用 cloud compare）。

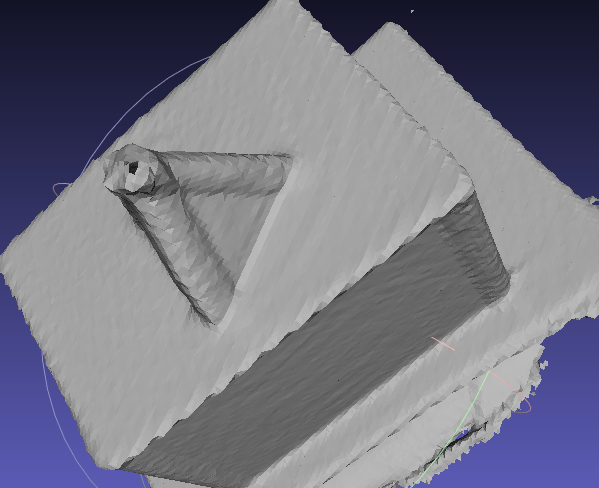
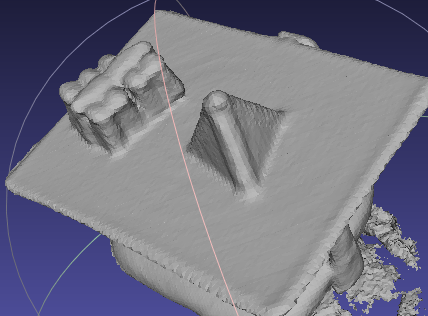
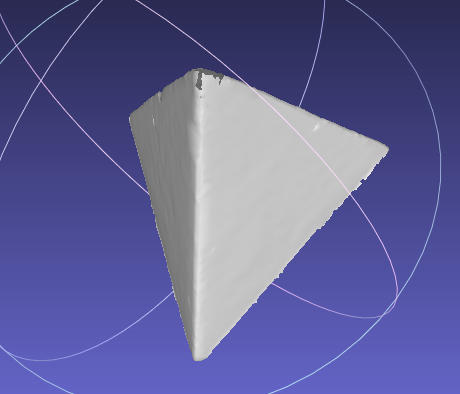
## 一些个人困惑

1. 为什么期望3D**特征退化**的立方体能够提升配准精度？

[5]的出发点是，在即便3D特征退化的场景（大平面，立方体，etc.）中，也能得到可接受的重建结果，其依赖的是整个三维场景的轮廓匹配，核心贡献是阻止了显著漂移的发生；[1]的出发点是，在2D彩色RGB图像上通过模型轮廓与图像轮廓配准优化模型姿态，其算法性能依赖模型在彩色图像上的可辨识度。

他们都需要依赖某种“特征”，如果我们希望使用彩色信息，则场景（立方体基座）应该是彩色纹理丰富的，以便特征描述子SIFT, SURF, ORB等[???]能正常工作，同时对标论文应该是借助RGB数据的RGB-D SLAM[15]，而不是纯基于深度图的SLAM (e.g. KinectFusion[2])。我们目前使用的是纯白色立方体，因为之前讨论定下的目标是不借助RGB的重建。

尽管我们期望高精度的立方体能够提供“帮助”，但是从3D几何特征角度看，它本身是“**特征稀疏**”的，对于使用深度数据的KinectFusion，一个不规则的几何物体更能“帮助”提升配准精度。图1演示的是为了重建正四面体石膏模型，我们在场景中放置立方体，与放置某不规则几何物体的重建结果对比，后者的重建结果优于前者。

(a) (b) (c)

图 2 (a)为立方体做辅助标记物的重建结果; (b)为不规则几何体做标记物的结果; (c)为激光扫描仪重建结果。图(a)在四面体顶部没有闭合，重建缺陷较明显；(b)与(c)相比，原本较锐利的棱边出现明显的“钝化”

**我自己的回答**：好的方面是，立方体具有一个明显的几何特征，邻面正交。假如我们能通过这个特性，在每一帧上准确定位立方体，即可以更精确的求解相机外参。如果要利用此特性，必然是使用几何特征，而不是RGB特征。另外，我认为关键障碍是：Kinect V1产生的深度图在物体边缘的噪声太大，因此对立方体“邻面正交”所能提供的定位能力信心不是很足。

1. 假如我们的规划仍然是基于①深度视频流，②KinectFusion，那么我们在KinectFusion基础上最期望解决的问题（贡献）是什么？

**我自己的回答**：观察图1(b)，这是我们目前使用KinectFusion得到的最好重建结果，但是它的瑕疵是：正四面体棱边（以及顶点）明显“钝化”，不再锐利。这个缺陷的误差来源，我们**没有验证过**，即：是否棱边钝化源于相机外参求解的累计误差？

个人计划近期首先测试验证上述误差来源，具体步骤如下：

① 误差源于体素模型的“压缩”导致失真？验证：使用包含尖锐棱边的模拟深度数据+准确相机外参，跳过配准步骤，直接进行TSDF融合，观察结果。若确实棱边钝化，说明假设正确；否则，说明误差与volume数据结构无关。

② 误差源于相机外参求解不准确？验证：使用模拟数据，配准，融合，若与①相比，出现钝化，则印证我们先前的方案目标：需要提升相机外参求解精度。

③ 误差源于Kinect V1产生的深度图的边缘误差太大？验证：使用真实数据，如先前的立方体，在立方体顶面固定放置棋盘格定标板，旋转棋盘格360°，每隔10°静态采样一次，总共采集约30~40帧RGB-D数据；使用棋盘格进行相机外定标，并用此定标结果进行融合。此方法基于假设：“棋盘格外定标法不会在某些视角产生累计漂移误差”。若结果仍然在物体棱边产生钝化，说明是深度数据噪声本身导致的，与外参求解无关；否则，说明确实在外参方面有提升空间。

## 参考文献

[1] Prisacariu V A, Reid I D. PWP3D: Real-time segmentation and tracking of 3D objects[J]. International journal of computer vision, 2012, 98(3): 335-354.

[2] 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.

[3] Izadi S, Kim D, Hilliges O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera[C]//Proceedings of the 24th annual ACM symposium on User interface software and technology. ACM, 2011: 559-568.

[4] Keller M, Lefloch D, Lambers M, et al. Real-time 3D reconstruction in dynamic scenes using point-based fusion[C]//3D Vision-3DV 2013, 2013 International Conference on. IEEE, 2013: 1-8.

[5] Zhou Q Y, Koltun V. Depth camera tracking with contour cues[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 632-638.

[6] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]//Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, 2012: 573-580.

[7] Nguyen C V, Izadi S, Lovell D. Modeling kinect sensor noise for improved 3d reconstruction and tracking[C]//3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on. IEEE, 2012: 524-530.

[8] Bleser G, Wuest H, Strieker D. Online camera pose estimation in partially known and dynamic scenes[C]//Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on. IEEE, 2006: 56-65.

[9] Davison A J, Mayol W W, Murray D W. Real-time localization and mapping with wearable active vision[C]//Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on. IEEE, 2003: 18-27.

[10] Jiang B, Neumann U. Extendible tracking by line auto-calibration[C]//Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on. IEEE, 2001: 97-103.

[11] Park J, You S, Neumann U. Natural feature tracking for extendible robust augmented realities[C]//Proc. Int. Workshop on Augmented Reality. 1998.

[12] Subbarao R, Meer P, Genc Y. A balanced approach to 3d tracking from image streams[C]//Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2005: 70-78.

[13] Davison A J. Real-time simultaneous localisation and mapping with a single camera[C]//Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003: 1403-1410.

[14] Dong Z, Zhang G, Jia J, et al. Efficient keyframe-based real-time camera tracking[J]. Computer Vision and Image Understanding, 2014, 118: 97-110.

[15] Endres F, Hess J, Engelhard N, et al. An evaluation of the RGB-D SLAM system[C]//Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012: 1691-1696.

[16] 任光阔, 黄倩妮, 《影视项目摄像机反求功能测试报告》.

[17] Wasenmüller O, Meyer M, Stricker D. CoRBS: Comprehensive RGB-D benchmark for SLAM using Kinect v2[C]//Winter Conference on Applications of Computer Vision (WACV). 2016.

[18] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on. IEEE, 2007: 225-234.

[19] Prisacariu V A, Kahler O, Murray D W, et al. Real-time 3d tracking and reconstruction on mobile phones[J]. Visualization and Computer Graphics, IEEE Transactions on, 2015, 21(5): 557-570.

[20] Dame A, Prisacariu V, Ren C, et al. Dense reconstruction using 3d object shape priors[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1288-1295.