

指导教师： 杨涛

提交时间： 2016/3/14

CVPR2015 Paper Translation

No: 01

姓名： 王劲松

学号： 2013302530

班号： 10011303

室内场景的“鲁棒”重建

摘要

目前，我们提出了一个方法来用 RGB-D 视频室内场景重建，它的关键点是场景几何片段与强大的全局最佳优化相结合。其中，几何配准由于传感器噪音而混淆，从而导致片段细节无法在现场进行辨别不同（环境状况）表面。他们这种明显多于正常数值错误的几何路线的优化方式被禁止。实验结果表明，我们所提出的方法基本上可以提高重建场景模型的精确度。

导论

高保真室内场景重建被看作为一个极具有挑战性的问题，许多室内重建系统简化的假设和重建模型都依赖于“用户”之间的交换。其他的系统模式都依赖于大量的硬件设施（激光雷达扫描仪）

消费者深度相机提供了非常强大的系统重建的机会，但它本身并不可以解决相关的挑战。虽然现在可以方便的创建真实世界的三维模型，质量和可靠性却不能足以被实现。不同于一个对象，它可以完全在摄像机的视野内，特别大的景象必须从复杂轨迹获取的图像来进行构建，每个景象都只能表达出环境状况的一小部分，相机沿路径所摄取的图像由于及其近距离而导致明显的“测距漂离”和全球注册的所有“表面对比”的分歧。

场景重建用户的深度摄像头是被全球所认可的。然而，现研发出的系统似乎不能可靠的满足并适用于要求特别严格的应用程序以支持室内场景重建。这是由 Xiao 等的努力才得以重建大量的室内场景。由于场景管道自动重建的不可靠性，作者使用 manual-labeling 建立不同观点之间的对应关系（目前我们所研发出的自动重建方法还是不够可靠的）

在这项研究中，我们提出了一个由 Xiao 等人工获得的全自动现场重建的管道以及明显超过之前所用的自动方法的室内重建精度。通过我们的方法所得出的重建示例如图 1. 我们的管道几何片段，对当地已被注册的场景片段进行两两对比（构建出模型）。我们解决的一个主要弱点是几何配准的低精度。几何配准算法是会因传感器噪音而产生偏离，从而导致几何细节被混淆并无法从当地的几何特点记性地域区分。有一个困难是由合并已注册的低重叠环路闭合几何片段。在实践中，假设的对比对多于正确的对齐对。

我们解决矛盾并确定使用线流程的全球最佳优化方法。线流程作为基础，以图像恢复的背景下引入自动识别的不连续性使其成为一个独立的优化部分。他们密切相关。线流程的优点是，可以优化目标保留最小二乘形式，并且可以被最小二乘解算器进行

优化。我们向大家展示出这种框架在处理双双登记错误非常有效果。我们的自动修复当它们明显多于正常数值的对比对。非常大量的实验结果证明我们的方法可以极大地提高重建的准确性。

我们的研究包含了一些独立的利益贡献。第一，我们提供了现场重建精准度严格评估基础设施，完善了非常有挑战性的相机轨迹和现实情况中噪声模型的 ICL-NULM 数据集。第二，我们进行场景重建过程中表面情况算法的量化评估。我们研究结果可以证明，大家都知道的算法非常不严谨，而且在过去的几年中介绍的算法只由于旧的方法。第三，除了重建场景的测量精度，我们还叙述了在没有地面真实数据对场景重建质量的定量评价实验过程。

连接工作

具有影响力的 KinectFusion 系统可以演示多组重建与消费者深度摄像头，它基于深度图像融合，视觉里程计，和实时的三维重建。最初的 KinectFusion 系统中使用的平原状态网络切被局限于小的空间，但这种局限目前已被消除，并提出了可以提高系统的准确度替代测距算法。这些系统不会检测环路封闭和有限或者密集的工作区，或简单的步行路线。如果没有与其相匹配的循环封闭处理，那么会破坏复杂的相机路径所需的家具室内场景的全面影像。

一些具有集合回路闭合处理的 RGB 三维重建系统已经被开发。他们都通过使用可视化技能，如 SIFI 或 SURF 关键点，或通过大量密集图像配准匹配单个 RGB-D 图像来检测回路闭合。这种方法提供实时性能，但假设在场景相同位置但不同的图像很相似，那么它是很容易的就被错过匹配图像认证。如所示图 2，我们设置在实时性能不同的是不是同一要求。高质量的离线场景重建在许多应用领域是极具价值的。

离线高保真 RGB-D 重建全球优化由之前的 Zhou 所研究。他们的工作依靠的是场外的已合成环路闭合检测模块，这是不完整的因为容易出现故障。我们的工作提出了一个综合的方法来循环闭合检测和基于几何配准对其进行强大优化。

登记的几何数据已被广泛的进行研究。在一个表面上的典型登记管道样品和另一表面上的匹配结构来计算转换，它的挑战是详细采样进行匹配比较昂贵。在过去十年中，研究人员已经研发出可以用于修复和对应并提出不同类型。尽管这样，对齐不准仍然是常见的，我们的方法是对重建管道稳定的几何配准误差进行全方位的优化处理。

范围扫描的全方位优化是基于假设的成双关系，用 Lu 和 Milos 和机器人。在我们的设置中，所有成对关系是嘈杂和在一定程度上受异常值影响，Huber 和 Hebert 描述了通过在成

对关系曲线图中进行搜索的最一致的避开异常值的算法，类似的方法已被用于重新装配裂隙对象。这种方法假设的场景可以优化成对比对，我们可以覆盖这种不正确的情况。

我们的解决方法是基于线流程处理，这是高性能且有效的最小二乘求解。一个独立的最佳对齐场景，即便他们比不虚伪的比赛异常值多。对机器人进行相关调整，在从运动估计结构，鲁棒性可以适用于适当的惩罚函数或通过识别摄像机姿势之间的成对关系。我们的工作相似相关的，但我们侧重于从各种视频中进行情景再现。我们提出了密集的表面重建，通过对表面排列进行直接优化，使其能与密集通信限制的客观标识异常值的专用方法进行结合。我们的实验可以证明出，该方法明显优于没有进行内置表面密集排列前稳定的优化框架。

概述

片段建设

个别范围图像是不完整的，为了得到部分表面几何片段形状的更可靠情况，我们分别输入 RGB-D 视频成 K 帧段，使用 RGB-D 测距估计摄像机轨迹，和周围范围的图像，从而获得每个部分，这些场景片段集合了一些混乱的范围内数据，并产生可靠的信息。它们在现场比独立的图像有更大的体积，而且测距偏差不显著。片段类似于分图，这是许多机器人的使用，让 $P_i = P\{p\}$ 是顶点集成片段 i 让 R ，是刚

体变换的对齐 p_i 为 P_{i+1} ，通过 RGB-D 测距计算。

几何配准

因为测距漂移，简单地使用变换 $\{R_i\}$ 使可识别碎片产生重建在覆盖场景的重叠部分不连续的片段对齐。出于此原因，我们要查出每对片段的重叠选项。几何配准算法是在每一对使用 (P_i, P_j) 。如果算法对准足够的重叠片段，那么被计算的数据在循环闭合的片段之间建立 i 和 j 具有相关的转换 T_{ij} 。

鲁棒优化

许多成对登记中假定的循环闭合的误报，我们通过优化线流程在环闭合制约中增强一个已经注册的表面致密目标识别这些虚构的环闭合件。仅最小二乘目标共同估计场景的完善配置和每个制约的有效性。这一做法使错误限制在他们大大多于真正的环闭合件时得到修复了。

最终模型

环闭合的最后一组被成功识别后，测距和环路闭合交替使用 ICP 进行细化。姿态图形优化用于获得最后的片段姿态 $\{T_i\}$ 。选择非刚性细化可以用来进一步完善对位。注册片段融合成一个全球性的网络模型。

4. 几何配准

我们以一个目前为止最先进的算法来进行室内场景数据的定量分析。这个分析使我们得以接近想要的结果。扩充 ICL-NUIM 数据集，由此增强了 Handa 等人的合成场景分析和复

杂的摄像机轨迹与噪声模式，该数据集进行了非常详尽的材料补充说明。

设定一个输入范围的视频，第三节所描述的一组片段 $\{P_i\}_w$ ，考虑一对 $\{P_i, P_j\}$ 和 P_i 都是在表面面积变得越来越小的，这组被认定为地面实际状况的回路闭合，如果他们在地面实际情况重叠占面积超过 30% P_i ，在这种情况下，地面实况变换 T_{*ij} 和一组或者点对点对应 K_{*ij} 来进行相关联。

每个计算方法在每组片段都运行 $\{P_i, P_j\}$ ，一个计算转型 T_{ij} 保留了之前拟定的循环闭合，如果 30% $T_{ij}P_i$ 与 P_j 重叠，每个算法提出的环闭合都被用来衡量它的召回率和准确度，这个测量并不足以考虑并提出地面实况环路闭合几何的相交数据，由于算法可能已经很能确定环路闭合 P_i 和 P_j 之间产生一个错误转变，所以，改变 T_{ij} 与地面实际值 T_{*ij} 转换。为了避免在转换空间进行不同自由度的任意选择，我们直接测量的结果 T_{ij} 是在地面实况通信 K_{*ij} 。它带来的这些地面实况通信对齐对改造是否能被接受。特别是 T_{ij} 被认为是纯阳性，如果地面纯对应的 RMSE 低于阈值

$$\frac{1}{|K_{*ij}^*|} \sum_{(p^*, q^*) \in K_{*ij}^*} \|T_{ij}p^* - q^*\|^2 < \tau^2.$$

我们使用了相当宽松的门槛 $\tau = 0.2$ 在所有试验里。

表 1 列出平均传回和对增强 ICL-NUIM 数据集不同的计算方法的准确度。OpenCV 是由 Drost 等研究出的被注册表面算法。全部的查找表中估算计算

加速能力，4PCS 是 Aiger 等人的算法。

【2】和超级 4PCS 是近期 Mellado 等人的算法。我们与超级 4PCS 的创造者合作，来确定他们的方法是最佳参数值。PPF 积分是我们结合 Drost 等的点对方式自定义的表达。【16】与二次抽样是积分不定量【42，20】。PCL 是一个点云库用来实现 Rusu 等人的算法【52，3】。PCL 润色是 Rusu 算法的改良，现在我们来用材料进行补充说明。

| | OpenCV | 4PCS | Super 4PCS | PPF Integral | PCL | PCL modified |
|---------------|--------|------|------------|--------------|------|--------------|
| Recall (%) | 5.3 | 20.0 | 17.8 | 32.5 | 44.9 | 59.2 |
| Precision (%) | 1.6 | 8.9 | 10.4 | 7.1 | 14.0 | 19.6 |
| Runtime (sec) | 10 | 380 | 62 | 83 | 3 | 8 |

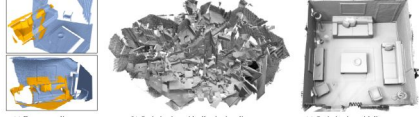
图表 1. 性能几何配准算法。AVER 年龄运行时间调整两个段是用 Intel Core i7CPU 锁定在 3.5GHz 单个流程进行测量。

我们惊讶的发现，Rusu 等人的算法竟然比包括我们最近使用的算法还要好。基于这个实验，我们的管道用这成对几何配准的 PCL 来改进算法。

就像结果表 1 显示，即便是性能最高的几何配准算法的精确度都在 20% 以下，我们认为这主要是因为表面积和形状鉴别有限且被控制在分辨率和现场进行采样范围内，并且由于噪声和失去原本真相所损坏。这个被允许合理召回，但却被由此影响了精度。如图 3，我们单独的考虑一些假阳性对齐是非常有根据的。因此，我们并没有处于试图开发可以被大量召回且近乎完美的登记的成对表面中，在第 5 节我们会向大家展示出，这些特点可以通过对场景进行全局分析来实现。

5. 鲁棒优化

在第四节我们分析出表面两两匹配来确定环路闭合是否误报。我们现在展示出全局优化在被召回过程中几乎没有被减少，以此来到达近乎完美的环闭合精度。



考虑顶点的姿势图 $\{P_i\}$ 和边 $\{R_i\} \cup \{T_{ij}\}$ 。我们的目的是计算一组姿势 $T = \{T_i\}$ 在全局坐标系的片段。表示为以下形式

$$E(T) = \sum_i f(T_i, T_{i+1}, R_i) + \sum_{i,j} f(T_i, T_j, T_{ij}). \quad (1)$$

我们面临的挑战是，大多数 T_{ij} 变化是不正确的，会损坏优化配置，如图 3。因此我们再添加一行过程 $L = \{l_{ij}\}$ 在假设的循环闭合中。变量 l_{ij} 射程于 $[0, 1]$ 和对应有效循环闭合的模型，将 L 和 T 同时优化：

$$E(T, L) = \sum_i f(T_i, T_{i+1}, R_i) + \sum_{i,j} l_{ij} f(T_i, T_j, T_{ij}) + \mu \sum_{i,j} \Psi(l_{ij}). \quad (2)$$

现有术语 $\Psi(l_{ij})$ 表示，拟定环闭合是一个真实事件， $\Psi(l_{ij}) = (pl_{ij} - 1)^2$ ，直接看这个术语目的在最大限度地选择环闭合数量 ($l_{ij} \rightarrow 1$)。然而，如果限制一个扭曲的配置会导致在对准方面不成比例地增加，所以这个禁用 ($l_{ij} \rightarrow 0$)。一个需长期进行调整 $f(T_i, T_j, X)$ 测量姿势之间的不一致性 T_i 和 T_j 以及

相对 X ，我们把它的功能定为对表面进行致密排列。让 K_{ij} 处于该组对应 $X P_i$ 和 P_j 是距离内 $E=0.05m$ ，定义 $f\{T_i, T_j, X\}$ 在对应点之间的平方距离内。

$$f(T_i, T_j, X) = \sum_{(p,q) \in K_{ij}} \|T_i p - T_j q\|^2 \quad (3)$$

$$\approx \sum_{(p,q) \in K_{ij}} \|T_i p - T_j X p\|^2 \quad (4)$$

$$= \sum_{(p,q) \in K_{ij}} \|X^{-1} T_j^{-1} T_i p - p\|^2. \quad (5)$$

线(4)使用对应对，这是保证结构的接近： $(p, q) \in K_{ij} \Rightarrow \|X p - q\| < \epsilon$ 。

使用标准的局部参数来表示 $X^{-1} T_j^{-1} T_i$ as a 6-vector $\xi = (\omega, t) = (\alpha, \beta, \gamma, a, b, c)$ ，收集一个旋转分量 ω 和另一个分量 t ，在这时进行计算，当 $T_j^{-1} T_i \approx X$ 时

$$X^{-1} T_j^{-1} T_i \approx \begin{pmatrix} 1 & -\gamma & \beta & a \\ \gamma & 1 & -\alpha & b \\ -\beta & \alpha & 1 & c \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (6)$$

这样

$$X^{-1} T_j^{-1} T_i p \approx p + \omega \times p + t.$$

方程(5)可以局部近似为

其中

$$f(T_i, T_j, X) \approx \sum_{(p,q) \in K_{ij}} \|\omega \times p + t\|^2 = \sum_{(p,q) \in K_{ij}} \|[-P]_{\times} I \xi\|^2, \quad (7)$$

$[P]$ 是交叉产物斜对矩阵形式 P ，和 I 是 3×3 单位矩阵。使得 $G_p = [- [P] * [I]]$

$$\begin{aligned}
f(\mathbf{T}_i, \mathbf{T}_j, \mathbf{X}) &\approx \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{K}_{ij}} \|\mathbf{G}_p \xi\|^2 \\
&= \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{K}_{ij}} \xi^\top \mathbf{G}_p^\top \mathbf{G}_p \xi \\
&= \xi^\top \left(\sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{K}_{ij}} \mathbf{G}_p^\top \mathbf{G}_p \right) \xi. \quad (8)
\end{aligned}$$

设定 G_p 恒定, $f(\mathbf{T}_i, \mathbf{T}_j, \mathbf{X})$ 可以按二次型近似 $\xi > \Lambda \xi$ 只需要对其计算一次

$$\Lambda = \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{K}_{ij}} \mathbf{G}_p^\top \mathbf{G}_p \quad (9)$$

该数据使现有术语和对准强度平衡。我们设定的对准方面（如上述推导）， u 被定义为与对应急的平均基数 K_{ij} 成正比，记作 $\kappa: \mu = \tau \cdot 2 \cdot \kappa$ 。它是在所使用的距离阈值，并具有相同的基础含义。直接的讲，当错误 $f(\mathbf{T}_i, \mathbf{T}_j, \mathbf{C})$ 超过 u ，它就超过了相应的对应项。

目标 (2) 被 g_{20} 优化并且被 $I_{ij} < 0.25$ 修建。剩余的环封闭被描述用作构造最终模型。

该方法是非常有效的。表 2 总结了扩充 ICL-NUIM 数据集所呈现的制剂的效果。优化增加了由五个因素设置循环闭合的平均精确度，从低于 20% 到 97.7%。平均召回仅下降 1.4%。

| | Before pruning | | After pruning | |
|---------------|----------------|---------------|---------------|---------------|
| | Recall (%) | Precision (%) | Recall (%) | Precision (%) |
| Living room 1 | 61.2 | 27.2 | 57.6 | 95.1 |
| Living room 2 | 49.7 | 17.0 | 49.7 | 97.4 |
| Office 1 | 64.4 | 19.2 | 63.3 | 98.3 |
| Office 2 | 61.5 | 14.9 | 60.7 | 100.0 |
| Average | 59.2 | 19.6 | 57.8 | 97.7 |

表 2 是鲁棒优化的效果。环闭合的平均准确率优化在 19.6% 设定为 97.7%

基本构想 (2) 为线流程的应用和在之前已经使用了的姿势图形优化。我们的工作通过将表面取向到不同角度。为了评估这一提法的影响，我们测量了 Sunderhauf 和 Protzel 的基本切换约束方法实现的循环闭合精度，这是 Lee 等人最期望的算法。并且我们的构想，在增强 ICL-UNIM 数据的集中结果，用现有的方法高环路闭合集的精确度，但这重建的改善并不能令我们满意。如图 6.2，我们的构想达到了近乎完美的精确度并且明显的提高了重建精度。

| | Original | SC [57] | EM [40] | Ours |
|---------------|----------|---------|---------|--------------|
| Living room 1 | 27.2 | 54.6 | 39.6 | 95.1 |
| Living room 2 | 17.0 | 23.5 | 20.5 | 97.4 |
| Office 1 | 19.2 | 39.6 | 33.7 | 98.3 |
| Office 2 | 14.9 | 25.2 | 19.7 | 100.0 |
| Average | 19.6 | 35.7 | 28.4 | 97.7 |

表 3 表面取向建模的影响，从左到右依次为，优化使用基本切换约束【1】，优化利用【期望最大化 1】，并优化我们使用的方法（手段）（在环路闭合精确处理的情况下）

6. 估价

6.1 数据集

增强 ICL-NUIM 数据集

我们第一个数据集的合成环境是基于 Handa 等提供的。地面实际情况几何表面形状的可用性使我们可以对重建精度进行准确测量。这个数据集包括室内环境两个，一个是客厅，一个是办公室。我们已经采用多种方式使它们增强完整的场景重建管道评估数据集。我们已经和这些扩展数据集的预

期用途作者确认。我们实验的补充材料中已经详尽的描述了客厅 1, 客厅 2, 办公室 1, 办公室 2。对于重建的目的模型从头到尾都使用手持成像。

SUN3D 数据集

我们的第二个数据集是室内场景[由 SUN3D 数据库提供]。原始数据由 Xiao 等人提供。包括调整为自动场景重建的离线系统，我们用这两个进行比较。还有许多人工辅助进行重建，利用交叉式界面，让使用者在输入视频时建立与其对应的模型。Xiao 等人提供了八种模型重建和相对应的人工辅助。我们可以说，人工重建是一个极为有益的参考。

运行时间

我们的补充材料中对管道中所有步骤的运行时间进行了描述。

6.2 合成场景

对我们使用 Handa 等提出的错误的 ICL-NUIM 场景表面重建精度进行评价，特别是平均值和重建的表面与地面实际状况的差距，对于每个序列我们评估四个重建的管道：

我们将由 Xiao 等人提供的 Kintinuous, DVO SLAM 自动调整管道以及我们的实验过程与结果作为参考，我们还要测量地面真轨迹并将其输入深度图像获得准确的模型；这些深度图并不完美，因为模拟的传感器受到了噪声影响。表 4 的平均距离在辅助材料中，我们提出的方法明显的优于其它管道建设。两个相对的影响因以 2.7 倍减小到平均中值误差。在这里我

们要注意，这是重建模型精度的直接评估。

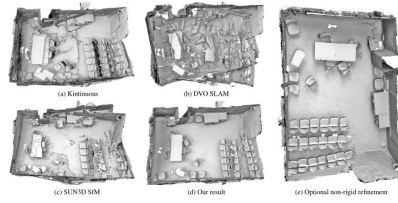


图 4，重建 SUN3D 数据集 mit-32-d507 景象。(a) 由 Kintinuous 重建产生。(b) 由 DVO SLAM 重建产生。(c) 是由 Xiao 等人提供的离线 RGB-D 运动结构产生的管道改良。(d) 是我们的想法。(e) 是使我们选择的非刚性细化 SLAC 产生的结果。

| | Kintinuous | DVO SLAM | SUN3D SfM | Ours | GT trajectory |
|---------------|------------|----------|-----------|-------------|---------------|
| Living room 1 | 0.22 | 0.21 | 0.09 | 0.04 | 0.04 |
| Living room 2 | 0.14 | 0.06 | 0.07 | 0.07 | 0.04 |
| Office 1 | 0.13 | 0.11 | 0.13 | 0.03 | 0.03 |
| Office 2 | 0.13 | 0.10 | 0.09 | 0.04 | 0.03 |
| Average | 0.16 | 0.12 | 0.10 | 0.05 | 0.04 |

表 4 是 ICL-NUIM 序列重建精度，也就是说每个重构模式与地面实况的距离（单位米），我

们的方法是可以降低因素平均误差，是相对于目前发现最接近的一种方法的另一种方法。

为了完整性，我们还使用 Handa 等人描述的 RMSE 度量估计相机轨迹精度对其进行测量，测量的结果在补充材料中。用我们的方法估计轨迹非常准确，相对于另一种方法 RMSE 平均减少了 2.2。我们要注意，轨迹精度只是重建精度的间接测量：在所报告的表面精度测量中，表 4 可以提供更多信息。我们还对最终的重建精度管道的不同部分的影响控制进行了评价，具体的说，我们取代提出的几何环路闭合检

测管道与 Kerl 等人的国家中最先进的基于图像的管道。(对于这些实验,对图像的循环闭合分别设置最大限度的提高精度,每一帧都要进行循环闭合审核)。我们的数据都要进行稳健优化(这些算法在第 5 节)。管道产量最高的重建精度结果列于表 5 中。

6.3 真实世界场景

实验步骤

对真实世界的场景进行定量评价是具有挑战性的,因为没有地面实际情况模型,因此我们开发并研究一个说得过去的评估标准。将配对进行广泛的比较,收集并输入所有重建模型。实验使用亚马逊的 MechanicalTurk 进行成对比较。实验方案描述会在补充材料中进行详尽说明。用 BRE 来将收集的两两数据进行计算数值评分,BRE 分数范围为 $[-1, 1]$,越高越好。

| | SC [57] | | EM [40] | | Ours |
|---------------|---------|-----------|---------|-----------|-------------|
| | [34] | geometric | [34] | geometric | |
| Living room 1 | 0.25 | 0.32 | 0.46 | 0.66 | 0.04 |
| Living room 2 | 0.26 | 0.40 | 0.26 | 0.65 | 0.07 |
| Office 1 | 0.11 | 0.36 | 0.22 | 0.56 | 0.03 |
| Office 2 | 0.52 | 0.27 | 0.56 | 0.48 | 0.04 |
| Average | 0.28 | 0.34 | 0.35 | 0.59 | 0.05 |

表 5 是管道不同部件的评估。平均距离为地面实况(米):对 ICL-NUIM SE-quences 重建精度,重新把我们强大的优化配方进行切换约束(SC)或期望最大化(EM),结果是重建精度下降。

验证

实验步骤是收集配对比较并计算 ICL-NUIM 序列 BRE 分数。我们评估了由 Kintinuous, DVO SLAM, SUN3DSFM 和我们方法的重建模型。为了使实验

数据更具有信服性,我们还通过整合沿着地面实况轨迹噪声输入 8960 对数据进行成对比较。所得 BRE 分数显示于表 6。BRE 分数的顺序与地面实况距离的平均均值对应。我们要注意,BRE 分数不是线性相关的地面真值,也不是因为距离措施在范围 $[0, \infty]$ 使 BRE 分数在 $[-1, 1]$

| | Kintinuous | DVO SLAM | SUN3D SFM | Ours | GT trajectory |
|---------------|------------|----------|-----------|-------------|---------------|
| Living room 1 | -0.53 | -0.90 | 0.02 | 0.47 | 0.94 |
| Living room 2 | -0.89 | -0.65 | -0.13 | 0.66 | 0.89 |
| Office 1 | -0.71 | -0.41 | -0.15 | 0.09 | 0.98 |
| Office 2 | -0.83 | -0.57 | -0.11 | 0.58 | 0.90 |
| Average | -0.74 | -0.63 | -0.09 | 0.45 | 0.93 |

表 6. 是 ICL-NUIM 序列评估,从两两比较计算 BRE 得分,在亚马逊的 Mechanical 执行计算。BRE 平均得分在辅助材料中对应如图表 4。

实验结果

我们增加到 8 个相同 SUN3D 序列,我们评估了 Kintinuous, DVO SLAM, SUN3DSFM 和我们的方法。作为参考,我们也用由 Xiao 等设置手动辅助重建,收集 17640 进行两两比较。

所得 BRE 分数如表 7。我们所提出的方法优于所有其他自动重建管道。这也是比手动协助重建更高。我们认为这主要是 SUN3D 交换界面的标签,其建立唯一地域对应限制重建模型如图 4。

| | DVO SLAM | Kintinuous | SUN3D SFM | Ours | SUN3D manual |
|------------------|----------|------------|-----------|-------------|--------------|
| hotel_umd | -0.61 | -0.45 | -0.02 | 0.66 | 0.56 |
| harvard_c5 | -0.49 | -0.01 | -0.65 | 0.94 | 0.11 |
| harvard_c6 | -0.97 | 0.05 | -0.01 | 0.96 | -0.15 |
| harvard_c8 | -0.70 | -0.61 | 0.39 | 0.65 | 0.46 |
| mit_32_d507 | -0.78 | -0.28 | -0.02 | 0.74 | 0.36 |
| mit_76_studyroom | -0.52 | -0.47 | 0.35 | 0.50 | 0.19 |
| mit_dorm_next_sj | -0.26 | -0.20 | -0.23 | 0.10 | 0.65 |
| mit_lab_hj | -0.12 | -0.57 | 0.03 | 0.22 | 0.50 |
| Average | -0.56 | -0.32 | -0.02 | 0.60 | 0.33 |

表 7。SUN3D 场景评估,从亚马逊 MechanicalTurk 进行两两比较计算 BRE 分数。此方法优于其它重建。

7. 总结

我们提出的方法需用 RGB-D 视频场景重建，这个主要的想法是线流程过程中将几何配准进行全部优化，因为我们在输入时有些错误是不可避免的，实验结果表明，我们所提出的方法明显地提高了用户级视频制作室内场景模型的保真度。

但我们所提出的算法并不是完美且无懈可击的，首先，如果输入的视频不包含环路闭合，那么会出现模型重建扭曲，但与此同时我们指导操作人员关闭循环会有所帮助。我们认为我们所提出的方法可以如果是实时操作的话那么会有助于获得完整的场景模型。其次，我们的管道没有考虑灾难性原因测距失败，这将会有丢失或片段畸形的可能性。当然这可以通过各个片段的形状来解决模型的不确定性。在这个充满挑战的情况中，惯性数据的整合也将非常有用。

致谢

我们非常感谢与 Andreas Geiger 讨论过程中对我们的帮助，ICL-NUIM 和 SUN3D 数据集的提供者，以及对我们的数据进行测算的 4PCS 的持有者。

References

- [1] P. Agarwal, G. Grisetti, G. D. Tipaldi, L. Spinello, W. Burgard, and C. Stachniss. Experimental analysis of dynamic covariance scaling for robust map optimization under bad initial estimates. In ICRA, 2014. 3
- [2] D. Aiger, N. J. Mitra, and D. Cohen-Or. 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics*, 27(3), 2008. 3, 4
- [3] A. Aldoma, Z. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze. Point cloud library: Three-dimensional object recognition and 6 DoF pose estimation. *IEEE Robotics and Automation Magazine*, 19(3), 2012. 4
- [4] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1), 1996. 1, 3, 6
- [5] M. Bosse, P. M. Newman, J. J. Leonard, and S. J. Teller. Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework. *International Journal of Robotics Research*, 23(12), 2004. 3
- [6] E. Bylow, J. Sturm, C. Kerl, F. Kahl, and D. Cremers. Real-time camera tracking and 3D reconstruction using signed distance functions. In RSS, 2013. 2
- [7] R. Cabral and Y. Furukawa. Piecewise planar and compact floorplan reconstruction from images. In CVPR, 2014. 1
- [8] A. Chatterjee and V. M. Govindu. Efficient and robust large-scale rotation averaging. In ICCV, 2013. 3
- [9] J. Chen, D. Bautembach, and S. Izadi. Scalable real-time volumetric surface reconstruction. *ACM Transactions on Graphics*, 32(4), 2013. 2
- [10] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós. Mapping large loops with a single hand-held camera. In RSS, 2007. 3
- [11] N. Corso and A. Zakhor. Indoor localization algorithms for an ambulatory human operated 3D mobile mapping system. *Remote Sensing*, 5(12), 2013. 1
- [12] D. J. Crandall, A. Owens, N.

- Snavely, and D. P. Huttenlocher. SfM with MRFs: Discrete-continuous optimization for large-scale structure from motion. *PAMI*, 35(12), 2013. 3
- [13] M. J. Cummins and P. M. Newman. FAB-MAP: probabilistic localization and mapping in the space of appearance. *International Journal of Robotics Research*, 27(6), 2008. 3
- [14] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH*, 1996. 2, 3, 4
- [15] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *ICCV*, 2003. 2
- [16] B. Drost, M. Ulrich, N. Navab, and S. Ilic. Model globally, match locally: Efficient and robust 3D object recognition. In *CVPR*, 2010. 3, 4
- [17] H. Du, P. Henry, X. Ren, M. Cheng, D. B. Goldman, S. M. Seitz, and D. Fox. Interactive 3D modeling of indoor environments with a consumer depth camera. In *UbiComp*, 2011. 1
- [18] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard. 3-D mapping with an RGB-D camera. *IEEE Transactions on Robotics*, 30(1), 2014. 1, 2, 3
- [19] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Reconstructing building interiors from images. In *ICCV*, 2009. 1, 3
- [20] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In *Symposium on Geometry Processing*, 2005. 3, 4
- [21] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *PAMI*, 14(3), 1992. 1
- [22] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6(6), 1984. 1
- [23] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4), 2010. 3, 5
- [24] Y. Guo, F. A. Sohel, M. Bennamoun, M. Lu, and J. Wan. Rotational projection statistics for 3D local surface description

- and object recognition. *IJCV*, 105(1), 2013. 3
- [25] J. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *CIRA*, 1999. 3
- [26] A. Handa, T. Whelan, J. McDonald, and A. J. Davison. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *ICRA*, 2014. 1, 4, 6
- [27] R. I. Hartley, K. Aftab, and J. Trumpf. L1 rotation averaging using the Weiszfeld algorithm. In *CVPR*, 2011. 3
- [28] P. Henry, D. Fox, A. Bhowmik, and R. Mongia. Patch volumes: Segmentation-based consistent mapping with RGB-D cameras. In *3DV*, 2013. 2
- [29] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *International Journal of Robotics Research*, 31(5), 2012. 1, 2
- [30] Q. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann. Reassembling fractured objects by geometric matching. *ACM Transactions on Graphics*, 25(3), 2006. 3
- [31] D. F. Huber, O. T. Carmichael, and M. Hebert. 3-D map reconstruction from range data. In *ICRA*, 2000. 1
- [32] D. F. Huber and M. Hebert. Fully automatic registration of multiple 3D data sets. *Image and Vision Computing*, 21(7), 2003. 3
- [33] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Incremental smoothing and mapping. *IEEE Transactions on Robotics*, 24(6), 2008. 3
- [34] C. Kerl, J. Sturm, and D. Cremers. Dense visual SLAM for RGB-D cameras. In *IROS*, 2013. 6, 7, 8
- [35] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for RGB-D cameras. In *ICRA*, 2013. 3
- [36] K. Khoshelham and S. O. Elberink. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, 12(2), 2012. 5
- [37] Y. M. Kim, J. Dolson, M. Sokolsky, V. Koltun, and S. Thrun. Interactive acquisition of residential floor plans. In *ICRA*, 2012. 1
- [38] G. Klein and D. W. Murray. Parallel tracking and mapping

- for small AR workspaces. In ISMAR, 2007. 2
- [39] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. G2o: A general framework for graph optimization. In ICRA, 2011. 4, 6
- [40] G. H. Lee, F. Fraundorfer, and M. Pollefeys. Robust pose-graph loop-closures with expectation-maximization. In IROS, 2013. 3, 6, 7, 8
- [41] F. Lu and E. E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4(4), 1997. 3
- [42] S. Manay, D. Cremers, B. Hong, A. J. Yezzi, and S. Soatto. Integral invariants for shape matching. *PAMI*, 28(10), 2006. 4
- [43] N. Mellado, D. Aiger, and N. J. Mitra. Super4PCS: Fast global pointcloud registration via smart indexing. *Computer Graphics Forum*, 33(5), 2014. 3, 4
- [44] A. S. Mian, M. Bennamoun, and R. A. Owens. Automatic correspondence for 3D modeling: an extensive review. *International Journal of Shape Modeling*, 11(2), 2005. 3
- [45] R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In CVPR, 2010. 2
- [46] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In ISMAR, 2011. 1, 2
- [47] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger. Real-time 3D reconstruction at scale using voxel hashing. *ACM Transactions on Graphics*, 32(6), 2013. 2
- [48] D. Nistér, O. Naroditsky, and J. R. Bergen. Visual odometry. In CVPR, 2004. 2
- [49] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. N. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles. Detailed real-time urban 3D reconstruction from video. *IJCV*, 78(2-3), 2008. 2
- [50] R. Roberts, S. N. Sinha, R. Szeliski, and D. Steedly. Struc-

- ture from motion for scenes with large duplicate structures. In CVPR, 2011. 3
- [51] S. Rusinkiewicz, O. A. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. *ACM Transactions on Graphics*, 21(3), 2002. 2
- [52] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In ICRA, 2009. 3, 4
- [53] A. Sankar and S. M. Seitz. Capturing indoor scenes with smartphones. In UIST, 2012. 1
- [54] S. Se, D. G. Lowe, and J. J. Little. Vision-based global localization and mapping for mobile robots. *IEEE Transactions on Robotics*, 21(3), 2005. 3
- [55] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard. Point feature extraction on 3D range scans taking into account object boundaries. In ICRA, 2011. 3
- [56] F. Steinbrücker, C. Kerl, and D. Cremers. Large-scale multi-resolution surface reconstruction from RGB-D sequences. In ICCV, 2013. 2
- [57] N. Sünderhauf and P. Protzel. Switchable constraints for robust pose graph SLAM. In IROS, 2012. 3, 6, 7, 8
- [58] N. Sünderhauf and P. Protzel. Switchable constraints vs. max-mixture models vs. RRR - a comparison of three approaches to robust pose graph SLAM. In ICRA, 2013. 3, 6
- [59] E. Turner, P. Cheng, and A. Zakhor. Fast, automated, scalable generation of textured 3D models of indoor environments. *IEEE Journal of Selected Topics in Signal Processing*, 9(3), 2015. 1, 3
- [60] F. Wauthier, M. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In ICML, 2013. 8
- [61] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald. Robust real-time visual odometry for dense RGB-D mapping. In ICRA, 2013. 2, 6, 7
- [62] T. Whelan, J. McDonald, M. Kaess, and J. J. Leonard. Deformation-based loop closure for large scale dense RGB-D SLAM. In IROS, 2013. 1, 2
- [63] K. Wilson and N. Snavely. Network principles for SfM: Disambiguating repeated structures with local context. In ICCV, 2013. 3

- [64] J. Xiao and Y. Furukawa. Reconstructing the world's museums. In ECCV, 2012. 1
- [65] J. Xiao, A. Owens, and A. Torralba. SUN3D: A database of big spaces reconstructed using SfM and object labels. In ICCV, 2013. 1, 3, 6, 7
- [66] C. Zach. Robust bundle adjustment revisited. In ECCV, 2014. 3
- [67] C. Zach, A. Irschara, and H. Bischof. What can missing correspondences tell us about 3D structure and motion? In CVPR, 2008. 3
- [68] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In CVPR, 2010. 3
- [69] Q.-Y. Zhou and V. Koltun. Dense scene reconstruction with points of interest. ACM Transactions on Graphics, 32(4), 2013. 1, 2, 3
- [70] Q.-Y. Zhou and V. Koltun. Color map optimization for 3D reconstruction with consumer depth cameras. ACM Transactions on Graphics, 33(4), 2014. 1
- [71] Q.-Y. Zhou and V. Koltun. Simultaneous localization and calibration: Self-calibration of consumer depth cameras. In CVPR, 2014. 3, 4, 7
- [72] Q.-Y. Zhou, S. Miller, and V. Koltun. Elastic fragments for dense scene reconstruction. In ICCV, 2013. 3

