指导教师：杨涛　　提交时间：2016/3/20

CVPR2015 Paper Translation

No: 01

姓名：苏悦

学号：2013302591

班号：10011306

轮廓线索的深度相机跟踪

Qian-Yi Zhou Vladlen Koltun
Intel Labs

摘要

我们提供了一种方法，通过一系列的图像实时跟踪相机的姿态。现有的算法在光滑的表面容易漂移，以致于破坏几何排列。我们表示出从噪音和完整的深度输入提取出的有用的轮廓线索。用这些线索来建立相应的约束，从而提取出关于场景几何和约束姿态估计的信息。尽管输入中存在歧义，但是提供的轮廓约束可靠地提高了跟踪的精确度。基准序列和其他的具有挑战性的样本的结果证明了轮廓线索对于实时相机姿态估计的效用。

1. 说明

对于动物来说，跟踪自运动是视觉感知的主要功能[7, 25]。在计算机视觉中，视觉测程的相应问题强调了一系列的应用并得到了广泛的研究[5, 18, 11]。我们的工作关注在计算机视觉系统中利用越来越多的深度相机。我们的目标是提高深度相机跟踪的精确度，尤其是在目前导致测程法漂移的具有挑战性的场景中。

具有影响力的 KinectFusion 系统[16]展示了实时深度相机跟踪和密集场景重建，通过将深度图像变为场景的立体体现。我们的工作通过整合遮挡的轮廓在优化目标扩展了这些想法，结果显示出显式处理轮廓可以显著地提高跟踪精确度。一种不同的 KinectFusion 方法的扩展是由 Bylow 等人提出的[2]，他们导出了一个原则性的优化算法但是并不跟踪遮挡的轮廓。我们的实验证明轮廓跟踪有显著的效果。

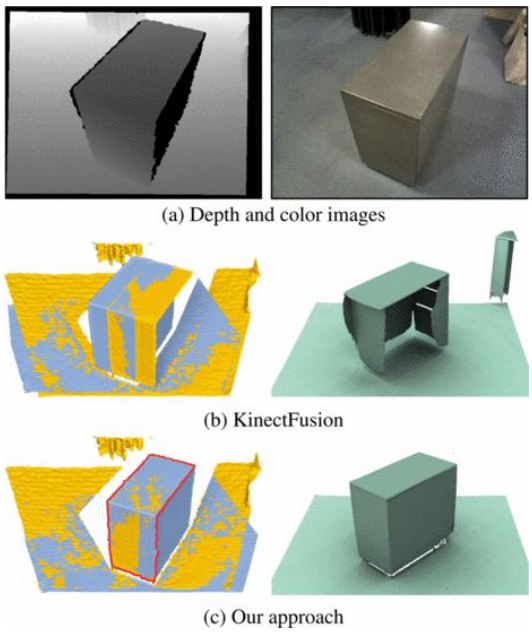

图 1 (a)深度彩色图像来源于一组输入序列 (b)在平滑的平面上，表面配准精确度下降，导致跟踪漂移或者重建失败。(c)我们的方法建立能够稳定实时相机跟踪的轮廓约束(红色)。这个彩色图图片(a,右)是不使用任一方法的原图像。

许多测量法系统使用深度和彩色图像[1, 9, 26]。我们的研究目的在于不依赖一组彩色图像流使跟踪准确度最大化。我们的理由是一些深度相机不伴有彩色摄像机。另外即使有彩色摄像机，它的视角是不同的并且快门可

能不能完美地与深度相机同步。最后，我们的目标是即使在最小照明下系统的功能。

我们的方法是基于跟踪遮挡的轮廓和使用轮廓线索约束配准。这解决了一个常见的几何配准方法的错误模式，这种方法基于迭代就近点算法(ICP)和它的变形[19]，即光滑表面中的不稳定[6]。此问题如图 1 所示，显示了一张用深度相机拍摄的内阁。在一些方面，这个内阁可以看做是大平面表面的集合，导致几何对准滑移和相机跟踪漂移。这个行为在实际中很容易被观察到，并且在基准测程法序列中也是很明显的[22]。我们的解决方法轮廓约束集成入配准目标，从而稳定在具有挑战性的情况下的相机跟踪。

遮挡的轮廓线索在早期计算机视觉中被认为是一个主要的信息来源[14, 12, 3, 8]。他们现在被使用在最先进的多视点立体系统，通知已给定校准相机参数的模型重建[23, 21]。我们的工作利用操作高帧速率深度图像流的实时跟踪系统中的轮廓线索。

Merrell 等人 [15]曾经使用实时表面重建中的可见性约束，但是这种相机通过其他方法被假定为小范围的。Wang 等人[24]使用轮廓线索用于宽基线范围扫描校准，但是他们的构想没有被设计在实时跟踪中。我们的方法共同优化投影对应和强大的轮廓约束在一个高性能的实时帧结构中。

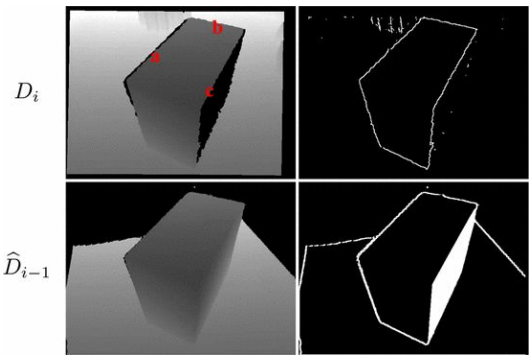

图 2 顶部：原始深度图像 D_i (左)和被探测道德封闭轮廓(右)。丢失的数据可能导致轮廓不准确。底部：深度图像 $\hat{D}_{i-1}$ 从构造出的体积表示(左)中合成。强大的法线估计和结果产生的轮廓对应候选(右)。尽管输入存在歧义，D_i (右上)中被检测到的轮廓在候选组(右下)中有合适的对应关系。

在挑战性的输入序列的实验结果证明我们的方法显著地提高了深度相机的跟踪准确度。

2. 方法

由于相切性，封闭轮廓提供了强有力的几何约束。对于沿着轮廓发生器的所有点，法线是垂直于视图[3]。这是我们的方法所实现的功能。我们根据深度图像梯度恢复出沿着轮廓发生器的法线。使用这些被恢复的信息，我们把轮廓约束到一个表面配准框架。联合优化目标集成界面对应条款和轮廓限制。这稳定了配准并显著降低了在具有挑战性的场景的漂移。

我们建立了 KinectFusion 框架工程[16]。特别是，我们保留截短的符号距离函数 F 作为场景的体积表示。每一个深度图像 D_i 被定义为 F 来估计相机的姿态 T_i，然后求积分并更新 F 的值。之前定义， D_i 通过双边滤波器

变得平滑并背投影到传感器的坐标系中作为一组点 V_i。为了从 F 中建立 3D 点和法线，每个像素的投影被用来合成一副代理深度图像 $\hat{D}i\text{-}1$ 在姿态 T_{i-1} 下。法线 $\hat{N}_{i-1}$ 投影后为一组点 $\hat{V}_{i-1}$，变形到全球坐标系最为 $\hat{V}_{i-1}^{g}$ 和 $\hat{N}_{i-1}^{g}$。定义的目标用来找到变形 T_i 和序列 T_iV_i 和 $\hat{V}_{i-1}^{g}$。

2.1. 轮廓探测

我们开始检测深度图像 D_i 的封闭轮廓。为了准备图像，我们沿着水平线扫描和用父亲相邻深度值来填充缺失的间隔，修复图像缺失的深度信息的区域。这个封闭区域的填充是由在红外投影仪和装有结构光深度传感器的相机之间的间隔导致的[13]。在修复的深度图像 D'_i 中，深度不连续的像素被认为是轮廓生成的。这些轮廓产生器用 C_i 表示：

$$C_i = \{s \in D_i : \exists t \in N_S^8, s.t. D_i'(s) - D_i'(t) > \delta\} \quad (1)$$

N_s^{δ} 是 s 在 D'_i 中的 8 个邻点，δ 是一个深度不连续门槛，基于典型传感器噪声幅值设置为 0.05 米[10]。

注意，在掠射角观察的表面可以从深度图像完全消失，由于投影的图案和菲涅尔效应的导致的失真。(图 2，左上，c 点)。这样的情况显然可以通过我们的方法解决。消失的区域被填充，内部边界被自动标记为轮廓发生器，如图 2 所示。检测到的轮廓可能没有与真实的轮廓对齐，但是这样的不一致通过后续的处理阶段得到解决。

2.2. 轮廓对应

下一个处理过程是建立 D_i 中检测到的轮廓发生器 C_i 和合成深度图像中的点 $\hat{D}i\text{-}1$ 之间的对应关系，这代表着场景模型。这些对应关系都在 ICP 过程的每个步骤中计算并用于配制轮廓约束条件。这些轮廓条件增加了重新计算得到的姿态下的配准目标和约束。

定义 T 作为当前 ICP 迭代中 D_i 的姿态。对于每一个点 $s \in C_i$，我们寻找 $\hat{D}i\text{-}1$ 中的一个点 t 例如在全局坐标帧 s 和 t 之间的距离较小，它们的法向量是被对齐的。距离规范可以表示为：

$$\| TV_i(s) - \hat{V}_{i-1}^{g}(t) \| < \varepsilon. \quad (2)$$

我们设置 ε=0.1 米。执行法线对齐较为困难。由于缺失数据和沿边缘的横向噪音，法线 $N_i(s)$ 的可靠预估很难从 D_i 中获得。局部轮廓几何中的小扰动能引起法线的剧烈扰动。由于我们缺少几何轮廓的准确信息，所以我们无法得到法线的可靠预估。我们的方法是避免同时计算轮廓的法线 $N_i(s)$，仍然执行法线校正标准。我们使用的事实是 $N_i(s)$ 点必须（几乎）垂直于视图的射线。用 $R_i^g(s)$ 来表示视图射线：

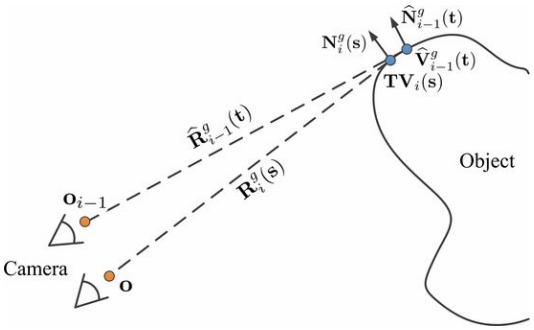

图 3 轮廓对应标准符号

$$R_i^g(s)=\frac{TV_i(s)-o}{\|TV_i(s)-o\|} \quad (3)$$

o 是全局坐标系中相机的原点。让 $N_i^g(s)=TN_i(s)$ 作为全局坐标系中 s 的法线。我们得到：

$$R_i^g(s)^T N_i^g(s)\approx 0 \quad (4)$$

如图 3 所示，我们可以合理地假设 $R_i^g(s)$ 和 t 中的视角射线之间的夹角用 $\hat{R}_{i-1}^g(t)$ 表示，是较小的。因此(4)大约等于：

$$\hat{R}_{i-1}^g(t)^T N_i^g(s)\approx 0 \quad (5)$$

法线的对齐标准可以近似表示为：

$$\hat{R}_{i-1}^g(t)^T \hat{N}_{i-1}^g(t)<\zeta \quad (6)$$

我们用 $\zeta=\cos(75^\circ)$ 和(6)作为和 t 建立一个轮廓对应的必要条件。直观上来看，降低了的标准仍然是在小相近运动情况下轮廓保持近切于视图射线。降低的标准不如 s 和 t 之间法线的对齐标准严格，但是尽管实际中传感器噪音有影响却仍然被使用。

标准(6)的一个显著优点是对于姿态 T 来说它是不变的，因此可以在迭代标配过程之前预先计算。我们因此计

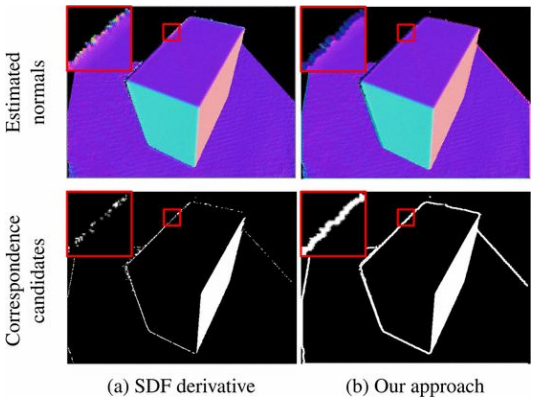

图 4 法线估计。(a)符号距离函数的数值差异沿着边缘不稳定。(b)我们使用一个替代的推导去估计预测深度图像的法线。

算了一系列轮廓对应候选点 $\hat{C}_{i-1}^g$（图 2 底部）并且在整个帧中构建了一个 kd 树结构。在每一次 ICP 迭代中，给定的 T 和 $s\in C_i$ 我们寻找在 $\hat{C}_{i-1}^g$ 中离 $TV_i(s)$ 最近的点并验证方程(2)。

2.3 法线估计

为了识别出轮廓对应候选点 $\hat{C}_{i-1}^g$，我们需要估计法线 $\hat{N}_{i-1}^g(t)$ 对 $\hat{D}i\text{-}1$ 中的点 t。表面法线能从符号距离函数 F[16]的数值梯度中预估出来。这个方法在平滑表面效果很好但是沿着边缘不稳定，如图 4 所示。我们开发了一个替代的法线估计过程，用来支持强大的轮廓对应估计。

我们从综合过的深度图像 $\hat{D}i\text{-}1$ 中估计法线。这并不很简单。因为 $\hat{D}i\text{-}1$ 是通过投影变换产生的而且我们寻找一个快速的基于图像的能在全局坐标系中产生表面法线的过程。

我们首先使用 2.1 节中提到的修复过程来填充缺失的深度值。$h(u,v)$ 表示通过图像域中被修复的深度图像 $\hat{D}'i\text{-}1$ 来定义的深度函数。这个函数能替代表示为在相机坐标系中隐式函数 $F(x,y,z)$ 的零水平点集：

$$F(x,y,z)=h(u,v)-z \tag{7}$$

$$x=\frac{1}{f_x}(u-c_x)h(u,v) \tag{8}$$

$$y=\frac{1}{f_y}(v-c_y)h(u,v) \tag{9}$$

(c_x,c_y) 是光学中心，(f_x,f_y) 是焦距。

在 $s=(u,v)$ 中法线被定义为 $F(x,y,z)$ 的梯度。

$$\hat{N}_{i-1}(s)^T=\nabla F(x,y,z)=\left(\frac{\partial F}{\partial x},\frac{\partial F}{\partial y},\frac{\partial F}{\partial z}\right) \tag{10}$$

令 $g(u,v,z)=(x,y,z)$ 作为函数在 $R^3\to R^3$ 中。$(F\circ g)(u,v,z)$ 的梯度可从公式(7)中推导出：

$$\nabla(F\circ g)(u,v,z)=\left(\frac{\partial h}{\partial u},\frac{\partial h}{\partial v},-1\right) \tag{11}$$

$\left(\frac{\partial h}{\partial u},\frac{\partial h}{\partial v}\right)$ 是通过 7×7Sobel 滤波器求得的图像梯度。通过链式法则，

$$\nabla(F\circ g)(u,v,z)=\nabla F(x,y,z)J_g(u,v,z) \tag{12}$$

雅可比矩阵 $J_g(u,v,z)$ 能从方程(8)和(9)中计算得到。我们解决了在线性系统中去估计法向量 $\hat{N}_{i-1}(s)$。这个过程实时在图形硬件上执行，并且生成整幅图像较为可靠地法线估计。关键是，这个包括沿着封闭轮廓的发现，如图 4(b)所示。

2.4 优化

让 $K=\{(s,t)\}$ 对应对组成，对应对从投影数据关联和通过 2.2 节中提到的轮廓对应过程中得到。我们的优化目标是整合轮廓序列和表面序列：

$$E(T)=\sum_{(s,t)\in K}w_{s,t}((TV_i(s)-\hat{V}^g_{i-1}(t))^T\hat{N}^g_{i-1}(t))^2$$

$\omega_{s,t}$ 是确定表面对应和轮廓对应之间关联影响的一个权重。目标在迭代中被优化。我们对表面对应将设 $\omega_{s,t}$ 为 1，对轮廓对应设为 ω_0。设 $\omega_0=0$ 降低目标 E(T)去标定点到平面 ICP，并且让我们的系统等同于 KinectFusion。增加 ω_0 可以增加轮廓约束的权重。我们的方法对 ω_0 的小变化不敏感。在 1 到 16 之间的任何值强制轮廓约束并使跟踪稳定。我们在所有的试验中设 $\omega_0=4$。

3. 结果

我们评估从 TUM RGB-D 基准[22]序列得到的方法。我们关注演示目标 3D 扫描的四个序列。我们的主要比较是不依赖其他信息渠道的两个纯深度相机跟踪方法：KinectFusion [16] (PCL

	KinectFusion [16]	Bylow et al. [2]	Our approach	Kerl et al. [9]	Whelan et al. [26]
fr3/cabinet	0.624	0.020	**0.015**	0.323	0.021
fr3/large_cabinet	0.275	0.109	**0.051**	0.103	0.055
fr3/structure_notexture_far	0.124	0.037	**0.026**	0.047	0.029
fr3/structure_notexture_near	0.252	**0.016**	0.023	0.384	0.023
Average	0.319	0.046	**0.029**	0.214	0.032

表 1 估计相机轨迹到 TUM RGB-D 基准序列的准确性。(RMSE 单位为米)

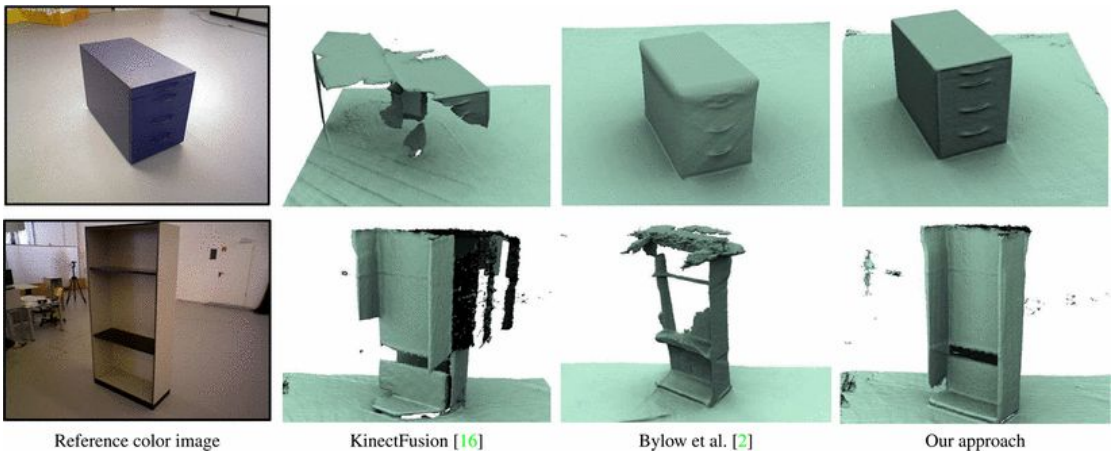

图 5 通过不同的算法在两个 TUM 基准序列重建场景模型的定性比较。彩色图像仅作参考，并没有使用的任何技术。

图 7 TUM 序列中的失败例子。我们的方法和其他方法在如图所示的图像帧都失效了。

实现[20])和 Bylow 等人的算法。每一个方法产生一个相机轨道，使用 Stuem 等人提出的 RMSE 矩阵和地面实况比较。这个结果如表 1 所示。提出的方法产生了最准确的相机轨迹。图 5 显示出不同算法在两个基准序列产生的场景模型。通过我们的方法产生的模型明显比之前技术输出更好。

作为参考，我们也展示了两种除了深度流之外使用彩色图像流方法的性能:Kerl 等人[9]提出的 RGB-D 测量法。结果如表 1 所示。我们的方法没有使用颜色信息，比参考技术更为准确。图 6 表示了通过 Asus Xtion Live 传感器扫描并有提出的方法重建得到的不同的常见对象。在图中展示的图像中，所有的这些序列，通过 KinectFusion 得到的表面排列下滑并且导致了灾难性的轨道丢失。相比较而言，通过我们的方法，轮廓线索的执行可以防止漂移，并始终使运动相机稳定的跟踪和场景几何的重建。

4. 讨论

我们提出了一种使用轮廓线的深度相机跟踪方法，可以稳定地跟踪和重建。这解决了经常在平滑表面中出现的灾难性的漂移。尽管边界法线可能

不好定义，但是我们展示出如何稳定有效地建立和利用广义轮廓约束。我们的相机跟踪方法只使用深度图像，实时运行，显著地提高了跟踪精度。

在将来的工作中，仍然存在限制和机会。当深度图像丢失了大量的数据，例如由于高度数镜面或者半透明表面，这种方法会失效。没有明显的边界，例如当扫描大面积无特征围墙，跟踪会发生漂移。此外，还有甚至在边缘线索不能稳定跟踪退化的情况：例如，在圆桌顶部旋转相机。最终，向之前的情况一样，我们的工作假设场景是静止的。图 7 展示了通过我们的方法和之前的方法得到的关键假设中的 TUM RGB-D 基准，这种方法由于场景移动或者数据缺失而失败了。

致谢

我们感谢 Erik Bylow 和提供方法的同行，还有感谢 Sungjoon Choi 和 Stephen Miller 在试验中的帮助。

参考文献

[1] A. Bachrach, S. Prentice, R. He, P. Henry, A. S. Huang, M. Krainin, D. Maturana, D. Fox, and N. Roy. Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments. International Journal of Robotics Research, 31(11), 2012. 1

[2] 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. 1, 4, 5

[3] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. IJCV, 9(2), 1992. 2, 3

[4] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In SIGGRAPH, 1996.2

[5] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In ICCV, 2003. 1

[6] N. Gelfand, S. Rusinkiewicz, L. Ikemoto, and M. Levoy. Geometrically stable sampling for the ICP algorithm. In 3DIM,2003. 1

[7] J. J. Gibson. The Ecological Approach To Visual Perception. Psychology Press, 1986. 1

[8] K. Karsch, Z. Liao, J. Rock, J. T. Barron, and D. Hoiem. Boundary cues for 3D object shape recovery.In CVPR, 2013.2

[9] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for RGB-D cameras. In ICRA, 2013. 1, 4, 5

[10] K. Khoshelham and S. O. Elberink. Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors, 12(2), 2012. 2

[11] G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In ISMAR, 2007. 1

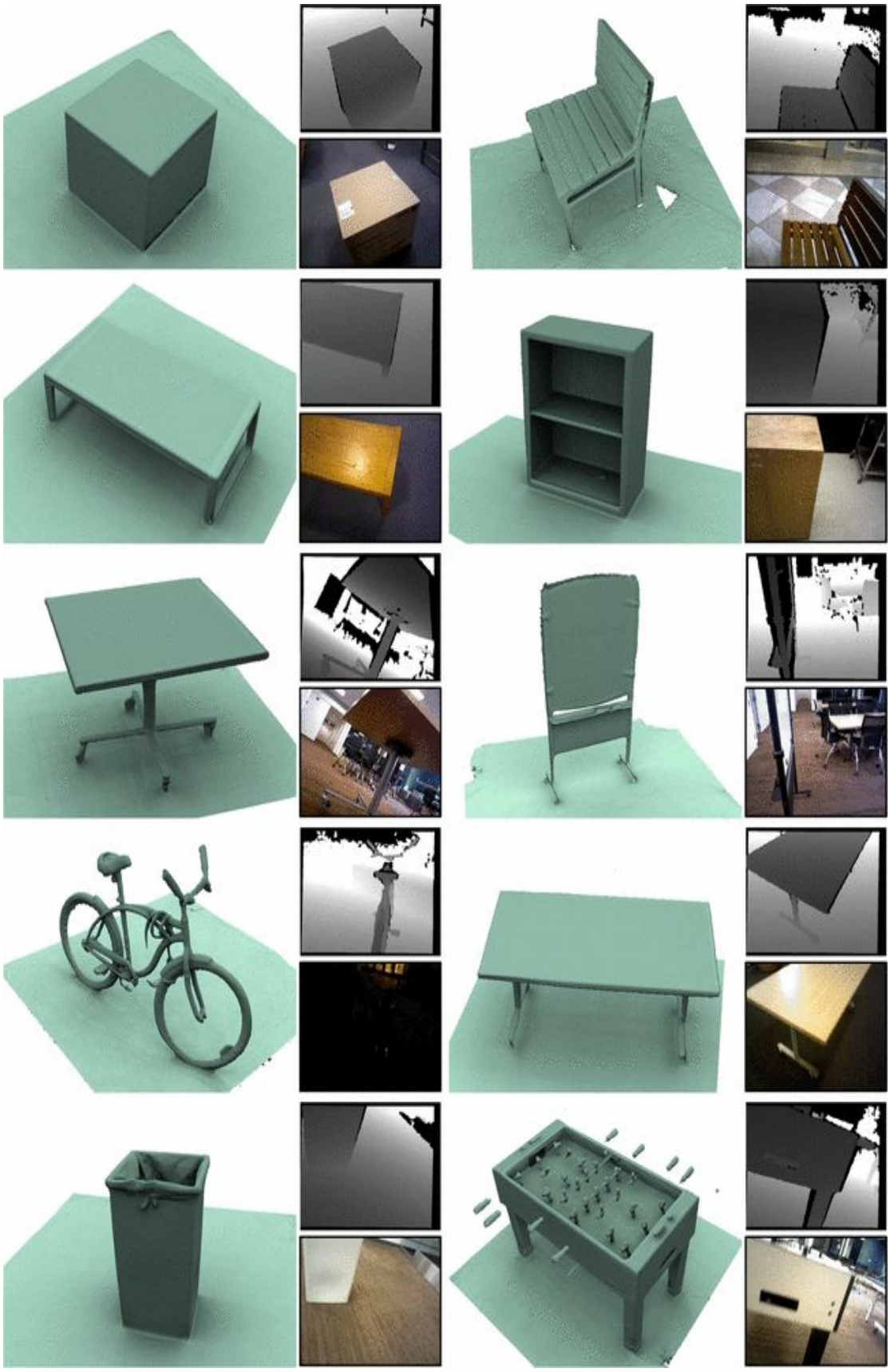

图 6 通过提及的方法得到的不同的重建对象。对于所有这些序列，KinectFusion 在插图中显示的图像帧中完全失效。我们提出的方法稳定了相机跟踪并且能够成功地重建。

[12] J. J. Koenderink. What does the occluding contour tell us about solid shape? Perception, 13(3), 1984. 2

[13] K. Konolige and P. Mihelich. Technical description of Kinect calibration,2012. http://wiki.ros.org/kinect_calibration/technical . 2

[14] D. Marr. Analysis of occluding contour. Proceedings of the Royal Society of London, 197, 1977. 2

[15] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J. Frahm, R. Yang, D. Nistér, and M. Pollefeys. Real-time visibility based fusion of depth maps. In ICCV, 2007. 2

[16] 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, 3, 4, 5

[17] C. V. Nguyen, S. Izadi, and D. Lovell. Modeling Kinect sensor noise for improved 3D reconstruction and tracking. In 3DIMPVT, 2012. 3

[18] D. Nistér, O. Naroditsky, and J. R. Bergen. Visual odometry.In CVPR, 2004. 1

[19] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In 3DIM, 2001. 1

[20] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In ICRA, 2011. 4

[21] Q. Shan, B. Curless, Y. Furukawa, C. Hernandez, and S. M.Seitz. Occluding contours for multi-view stereo. In CVPR, 2014. 2

[22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In IROS, 2012. 2, 4, 5

[23] H.Vu, P.Labatut, J.Pons, and R.Keriven. Highaccuracyand visibility-consistent dense multiview stereo. PAMI, 34(5), 2012. 2

[24] R. Wang, J. Choi, and G. G. Medioni. 3D modeling from wide baseline range scans using contour coherence. In CVPR, 2014. 2

[25] W. H. Warren. Self-motion: Visual perception and visual control. In Perception of Space and Motion. Academic Press,1995. 1

[26] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald. Robust real-time visual odometry for dense RGB-D mapping. In ICRA, 2013. 1, 4, 5