基于深度传感器和TSDF的三维立体重建技术

刘畅

School of Computer Science and Engineering, Beihang University, Beijing, China

Abstract. 本篇文献综述所讲述的主要内容是基于深度传感器，利用TSDF等算法实现由深度图像到三维立体模型的转换。

Keywords: 斯蒂芬是否, Keywords2, Keywords3

1 Introduction

计算机视觉是一门研究如何通过视觉传感器捕捉到的信息，对物体进行识别、分析的学科，而三维立体重建技术正是其子领域。现如今，三维立体重建技术被广泛应用于文物保护、游戏开发、医学研究、地图绘制等领域，例如通过将传感器安装到机器人身体上，机器人可以根据扫描到的场景信息控制移动，完成任务，具有广阔的前景。

三维重建的主要目的是根据传感器对某一物体或场景捕获到的若干图像或一段视频，将物体或场景通过一系列技术手段还原成三维立体模型。模型的重建主要依赖于深度相机，这种相机可以同时获取RGB图像和深度图像，前者记录被拍摄位置的颜色、亮度、饱和度等信息，后者通过某处灰度值大小反映出对应位置与相机间的距离，以此可以得到被拍摄场景的单方向模型。

目前，三维立体重建技术的主要难点是对相机移动的路径进行估算，以确定相邻两幅图形间的位置关系，以及图像的去燥、模型空洞的补全等等。这篇论文主要针对这些难点，论述近年来相关重建方法的种类以及改进方法。

2 三维空间还原

将深度照相机所获取的RGB图像与深度图像相结合，可以重建出每一个像素点在三维空间中的位置。设图像上某像素点$(p_x, P-y)$，通过深度图像I_d可以获得该像素点的深度值

$$z = I_d(p_x, p_y) \tag{1}$$

根据公式1得到的结果，利用图像中心的坐标(c_x, c_y)和相机镜头到图片平面的距离f，可以通过函数ρ得到像素点(p_x, P_y)对应的三维坐标

$$\rho(p_x, p_y, z) = (\frac{p_x - c_x}{f}z, \frac{p_y - c_y}{f}z, z) \tag{2}$$

我们将每组图像中每个像素还原得到的三维点的集合称为“点云(Point Cloud)”。

3 主要的表面生成方法

点云所表现出的结果是离散的，于是接下来的工作是，根据每一组图像得到的点云生成连续的物体表面，并将所有图像所生成的结果合并，得到完整的三维模型。目前两种方法被应用得较为广泛，分别是截断有向距离函数(Truncated Signed Distance Function, TSDF)和八叉树(Octrees)。

3.1 截断有向距离函数

早期一种利用截断有向距离函数(truncated signed distance function, TSDF)生成模型表面的做法 [3]曾被提出，其主要原理是对于单幅深度图像，三维空间中某一点p可用函数$d(p)$表示到表面的距离：当$d(p) > 0$的时候，认为p在表面后方（远离相机位置）；当$d(p) < 0$的时候，认为p在表面前方（靠近相机位置）；而当$d(p) = 0$的时候，即可认为p正好处于表面上。

对于由深度图像像素还原得到的点P_0，可以认为$d(p_0)$接近于0。为了提高结果的准确性，这里对于表面附近不同图像、不同位置的点，分别给予了不同的权重w，利用各个点的权重$d(p)$和距离函数$w(p)$，根据公式3和4对空间中划分的离散体素网格进行权重W和距离D的更新，最终得到一个较为准确的$D(p) = 0$的表面。

$$D(p) = \frac{\sum w_i(p) d_i(p)}{\sum w_i(p)} \tag{3}$$

$$W(p) = \sum w_i(p) \tag{4}$$

但是当这种方法被应用于较大模型重建的时候，由于信息量的增多，会导致内存消耗很大。

3.2 八叉树

八叉树是一种每个父节点拥有八个子节点的树 [13]，树中的每个结点都可用来表示空间中的一个立方体，每个子节点所代表的立方体都是父节点立方体的$\frac{1}{8}$，如图1所示。用八叉树表示空间的方式的优点是可以动态地调整空间的精细度，当需要对物体进行更高精度的描绘时，可以增加八叉树的深度，细化空间单位；当需要节省内存的时候，可以使用较大的立方体表示空间，提高效率。

利用八叉树绘制表面的方式是对立方体进行标记——如果该立方体占有的空间中包含了表面上的点，那么就将该立方体细分并标记为“占有”，否则就将其标记为“空闲”。这种方式更加高效，可以处理规模较大的模型。

4 相机追踪与模型生成

所谓相机追踪，是指在三维空间还原中，每幅图像生成的点云所属的坐标系是以相机所处位置为基准确定的，但相机在拍摄的过程中是保持移动的，因此不同点云采用的是不同的坐标系。为了将不同点云进行融合，必须要统一坐标系，因此必须确定相机所处的位置。

拍摄相邻两幅图像时相机的位置变化可以用旋转量R和位移量t来表示，目前有很多方法可以实现对R和t的估算。在确定了相机的位置之后，就可以综合之前生成表面的方法，对被拍摄物体或场景进行模型的生成。

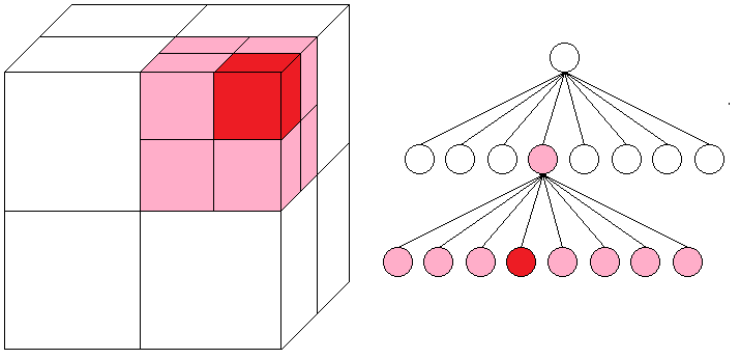

Fig. 1. 递归分割空间的八叉树

4.1 迭代最近点(Iterated Closest Point, ICP)

迭代最近点的主要工作是对两组相邻点云进行点与点之间的配对，然后计算出旋转量R和位移量t，将其中一组点云转化为另一组点云，并使得花费E最小 [1]。其中，E的计算方式是转换后两组点云中对应点的位置差异之和。

我们假设原点云为$x_1, x_2, ..., x_n$，目标点云为$y_1, y_2, ..., y_n$，则需要最小化的花费可表示为

$$E(R,t) = \sum_{i=1}^{n} ||Rx_i + t - y_i||^p \tag{5}$$

对于已经确定配对关系(x_i, x_i)的两组点云，通过调整R和t的值，使得公式5的值降到最低。对于指数p的取值，通常为$p = 1$或$p = 2$。

然而，找到最符合实际变化的旋转量R和位移量t的前提是配对(x_i, x_i)的正确。为了提高配对的准确性，这里采用迭代的方式不断优化点云间的配对关系。主要步骤如下：

1. 决定初始的配对关系
2. 利用公式5找到合适的旋转量R和位移量t
3. 采用此种转换方式对原点云x进行转换，得到转换后的原点云
4. 重复步骤1~3直到$E(R,t)$的值不发生明显变化

当原点云与目标点云差异较小的时候，这种方法有较大概率找到全局最优解；但是当原点云和目标点云差异较大的时候，往往会使结果陷入局部最优解。

在此基础上，一种采用结合了法线的改进方法被提出 [2]，这种方法以最小化原点云中x_i的法线n_i与x_i和y_i间连线$L(x_i, y_i)$投影之和的方式代替了原来的单纯地最小化两点间距离之和的方法，即

$$E'(R,t) = |(Rx_i + t - y_i)^T n_i|^p \tag{6}$$

ICP方法具有它的局限性，对于点云上所有点位于同一平面的情况，用ICP方法求解得到旋转量R和位移量t并不唯一。

2011年，微软提出了一种经典的重建模型方法——KinectFusion。KinectFusion [9]的伟大之处在于它实现了通过传感器实时生成3D模型的功能，具有极高的实用性。它使用ICP方法进行相机移动路径的追踪，使用TSDF方法有效地对中等大小的模型进行表面构造，并使用 [3]的方法建立3D模型。

有很多类似的方法曾被提出 [8,10]，但是前者使用的测试数据是人造的，且缺乏合理的评估方法；后者的SDF函数是已知的，且缺乏实时性。KinectFusion可以进行可靠的即时渲染，并同时生成两组点云——一组是根据当前深度图像生成的，另一组是之前所有深度图像生成点云的合并，当所有图像都被处理完毕后，后者即为全局的点云结果。

4.2 图像一致性比对

对于每组图像中的深度图像I_d^n和RGB图像I_c^n，其下一组图像中深度图像和RGB图像分别为I_d^{n+1}和I_c^{n+1}。图像一致性比对方法的主要工作是先利用公式2，根据(I_d^n, I_c^n)还原出点云P^n，之后选取适当的旋转量R和位移量t对P^n进行变换，并映射到平面π上 [11]。在理想情况下，映射到平面π上的RGB图像应当与I_c^{n+1}完全一致。但由于误差的存在以及(R,t)的选取，两图像间应当存在差异，这时需要动态调整(R,t)的选值，尽可能减少I_c^{n+1}上各点(i,j)和平面上对应点$\pi(p)$的色度差异，其差异可用公式7表示：

$$E(R,t) = \sum_{i=1}^{m}\sum_{j=1}^{n} ||I_c^{n+1}(i,j) - I_c^n(\pi(R\rho(i,j,z_{ij}) + t))||^2 \tag{7}$$

其中，z_{ij}是根据公式1得到的图像I_d^n中点(i,j)的深度值，m和n表示图像的高度和宽度。除了色度差异外，这种方法也可通过比较深度差异进行(R,t)的计算 [6]。和ICP不同的是，ICP是对两组图像生成的点云进行比较，而图像一致性比对是对二维图像进行比较，其优点在于有效利用了图像的色度信息，并且不需要对两组点云中的点进行匹配。

有很多基于图像一致性比对的模型生成技术 [5,7]，但是它们共性的缺点是生成的效果与实际误差较大，这是由于对相机移动路径的错误估计造成的。在图像一致性比对方法中，每一次计算出的旋转量和位移量具有一定误差，这种误差会随着图像组数的增多而累积，最终导致生成的模型与实际具有较大的偏差。目前，有很多减小误差的方法，例如loop-closure [4]和bundle-adjustment [12]。

5 3D模型生成

5.1 KinectFusion

5.2 基于图像一致性比对的方法

5.3

6 Related Work

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (No. 61332018), the National Department Public

Benefit Research Foundation (No. 201510209), and the Fundamental Research Funds for the Central Universities.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2), 239–256 (1992), https://doi.org/10.1109/34.121791
2. Chen, Y., Medioni, G.: Object modelling by registration of multiple range images. Image & Vision Computing 10(3), 145–155 (1992)
3. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, New Orleans, LA, USA, August 4-9, 1996. pp. 303–312 (1996)
4. Kähler, O., Prisacariu, V.A., Murray, D.W.: Real-time large-scale dense 3d reconstruction with loop closure. In: Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII. pp. 500–516 (2016)
5. Kerl, C., Sturm, J., Cremers, D.: Robust odometry estimation for rgb-d cameras. In: 2013 IEEE International Conference on Robotics and Automation. pp. 3748–3754 (May 2013)
6. Kerl, C., Cremers, D.: Large-scale multi-resolution surface reconstruction from rgb-d sequences. In: IEEE International Conference on Computer Vision. pp. 3264–3271 (2013)
7. Kerl, C., Sturm, J., Cremers, D.: Dense visual SLAM for RGB-D cameras. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, November 3-7, 2013. pp. 2100–2106 (2013)
8. Kubacki, D.B., Bui, H.Q., Babacan, S.D., Do, M.N.: Registration and integration of multiple depth images using signed distance function. In: Computational Imaging X, part of the IS&T-SPIE Electronic Imaging Symposium, Burlingame, California, USA, January 23-24, 2012, Proceedings. p. 82960Z (2012)
9. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.W.: Kinectfusion: Real-time dense surface mapping and tracking. In: 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, Basel, Switzerland, October 26-29, 2011. pp. 127–136 (2011)
10. Ren, C.Y., Reid, I.D.: A unified energy minimization framework for model fitting in depth. In: Computer Vision - ECCV 2012. Workshops and Demonstrations - Florence, Italy, October 7-13, 2012, Proceedings, Part II. pp. 72–82 (2012)
11. Steinbrücker, F., Sturm, J., Cremers, D.: Real-time visual odometry from dense RGB-D images. In: IEEE International Conference on Computer Vision Workshops, ICCV 2011 Workshops, Barcelona, Spain, November 6-13, 2011. pp. 719–722 (2011)
12. Urban, S., Wursthorn, S., Leitloff, J., Hinz, S.: Multicol bundle adjustment: A generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems. International Journal of Computer Vision 121(2), 234–252 (2017)
13. Wurm, K.M., Hornung, A., Bennewitz, M., Stachniss, C., Burgard, W.: Octomap: A probabilistic, flexible, and compact 3d map representation for robotic systems.

In: Proc. of the ICRA Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation (2010)

UTF8song