

利用深度图像进行场景的三维重建

何文峰 查红彬

北京大学 信息科学中心, 北京, 100871

e-mail: {wenfeng, zha}@cis.pku.edu.cn

摘要：近年来，利用激光扫描仪获得的深度点云数据来进行场景的 3D 建模受到了广泛的关注。从深度数据获取到最后三维模型的建立涉及到了很多工作，本文就集中讨论了其中的分割、配准和建网。根据真实场景含有大量平面特征的特点，本文提出了利用平面特征来实现配准，这些平面特征是对深度图像进行分割得到的。实验结果显示了这种配准方法的准确性和鲁棒性。另外，本文改进了由深度图像建网的一般方法，从而得到了更好的建网结果。

关键词：深度图像 激光扫描仪 点云 深度图像分割 配准 建网

0 引言

将客观世界在计算机中真实的再现，是计算机视觉和计算机图形学研究的一个热门领域。近年来，激光扫描仪技术得到了很大的发展和商业化，使用它们可以得到真实场景的高精度的 3D 数据，因此利用这些 3D 数据来进行真实世界的三维建模得到了很多关注。

一般的，通过激光扫描仪进行一次扫描得到的是一组二维有序的点阵，其中每一点包含了相应的场景上被扫描点的距离信息，这个点阵被称为深度图像。为了获得完整的场景，往往需要从几个不同的位置对目标场景进行扫描，而得到的不同的深度图像则需要配准到一个坐标系下。另一方面，仅仅用离散的点云数据来描述三维场景是远不够的，构建三角网格则是一个常用而有效的办法。本文就集中完成了这两方面的工作。

由于我们建模的对象的是室内或室外较大型的场景，这类场景往往具有一个典型的特征，即场景中含有大量的平面。我们的工作也充分的利用了这一特征。在获得场景的数据后，我们对点云数据进行了基于平面的分割，在配准中，我们利用了分割提取出的平面特征来完成。

后面部分的安排如下：第 1 部分介绍了一些相关的研究工作，第 2 部分阐述了对深度图像进行平面分割，第 3 部分描述了利用平面特征来进行配准。第 4 部分描述了利用深度图像生成三角网格的过程。第 5 部分是我们的一些实验结果。最后给出了结论。

1 相关的研究工作

利用激光扫描仪获取真实场景的 3D 数据并进而构建出具有真实感的 3D 模型涉及到了很多人的研究工作。这之中，I. Stamos 和 P.Allen[1]，S. F. El-Hakim 等人[2]以及 Y.Yu 等人[3]都做了大量的工作，并各自实现了一个完整的系统。H. Zhao[4]和 C. Frueh[5]则进一步构造了动态的 3D 数据获取系统，他们将激光扫描仪固定在汽车上，从而能够重建更大一级（如街区、城市）的场景模型。

在深度图像分割领域，一直吸引着很多人的研究兴趣。Besl 和 Jain 在[6]中提出了一个基于二次表面拟合的分割算法，具有很好的通用性。I. Stamos 和 P.Allen 在[1]中将上面的算法具体到了平面拟合，得到了相对简洁快速的算法。

在深度图像配准方面，Besl 和 McKay 提出了著名的 ICP (Iterative Closest Point) 算法，这比基于特征的配准算法有更好的鲁棒性和准确性。但它需要一个不错的初始位姿估计。I. Stamos 和 P.Allen 在[1]中使用了直线作为特征来实现配准。Y.Yu 等人在[3] 中则使用了激光扫描仪自

带的靶标和软件[8]进行配准。马和张在[9]中详细描述了利用对应的三维特征进行运动分析。

利用深度图像生成三角网格是一个相对简单的过程,G.Turk 在[10]中用到了一种普遍而通用的算法,但在这里,我们对其做了补充和改进。

2 深度图像的平面分割

我们使用了 CyraTechonogies,Inc.公司[8]的激光扫描仪 Cyrax2500 以及其自带软件 Cyclone3.1 来获得场景的 3D 数据。扫描完后,得到的深度图像中的每一点包含了当前被扫描点在扫描仪空间坐标系下的三维坐标 (x, y, z) 。我们采用了[1]中的方法来完成平面分割。首先对深度图像上的每一点 P , 构造其 $k*k$ 邻域中所有点的协方差矩阵 $A = \sum_{i=1}^N ((v_i - m)^T \cdot (v_i - m))$

其中 v_i 是邻域中的点, m 为邻域点集的重心。 A 的最小特征值实际上就可以衡量点 P 的平面拟合的好坏,而与之对应的特征向量即为点 P 的拟合平面的法向量。这样就淘汰了那些拟合不好的点并同时求出了拟合成功的点的法向量。另一方面,我们定义了点的相似性条件:1)点的法向一致性;2)点的拟合平面的距离临近性。在算法中,依次对每一个点进行遍历,利用上面的相似性条件,判断该点是否可以和其邻域点合并为同一集合。算法结束后,我们就会得到若干集合,同一集合内的点都处于同一平面上,这样,就实现了对深度图像的平面分割。通过阈值的调整,我们可以得到很好的分割结果。

3 利用平面特征进行深度图像配准

3.1 深度图像配准

由于深度图像中点的三维坐标是就其自己的扫描空间而言,为了得到场景的完整的三维模型,就需要将这些从不同位置获取的深度图像配准到一个坐标系下。简言之,就是要解决一个坐标变换的问题,找出两个坐标系之间的变换关系,这种关系一般用一个 3×3 的旋转矩阵 R 和 3 维平移向量 t 来描述。配准就是要求解出 (R, t) 。

为了求出上述 R 和 t ,一种典型的做法就是在两幅深度图像中找出若干组对应的特征,根据这些对应特征来求解 R 和 t 。在配准中,最为简单和常用的几何特征是点。但我们并没有选用。原因是在两个深度图像中找到对应的点并不容易,假设在实际场景中的某一点 P 是我们认为的特征,在两幅深度图像中,我们分别找到两个点 P' 和 P'' ,并认为这两个点是 P 分别在两幅图像中的位置。但由于扫描仪采样时分辨率的限制以及噪声的存在, P' 和 P'' 往往与实际的 P 存在误差,特别对于大场景而言,这种误差较大。这就会导致求出的 R 和 t 也存在较大的误差。同时,大场景中点的位置范围会很大,这会将即使很小的误差放大,得到的匹配结果也不好。

选择平面作为特征则具有更好的准确性和鲁棒性。在作完平面分割以后,我们已经提取出了一系列平面,由于平面的面积大使得其清晰易辨,因而对应关系很好确定,同时手工选平面也操作简单。另一方面,每一个平面都是对很多扫描点作拟合得到的,因此具有很好的鲁棒性。

3.2 基于平面特征的配准的实现

经过前面的平面分割,我们得到了若干个平面,每一个平面对应了一个平面。我们对每个点集进行平面拟合,得到其法向量。这样,对每一个平面,我们可以用 (n, m) 来描述。其中 n 为该平面的法向量,而 m 为属于该平面的点集的重心。

我们手工的从两组深度点云数据中选取 k (后面会对 k 的取值加以限制)对相对应的平面。不妨设在第一组点云中选取的第 i 个平面的法向和重心分别为 n_i 和 m_i , 与之对应的在第二组点云的第 i 个平面的法向和重心分别是 n'_i 和 m'_i 。同时,我们为每一对平面特征设定一个取值在

$[0, 1]$ 的权值 w_i ，该权值用来设定这组特征在匹配中的重要程度，其取值会在后面有所讨论。

首先，由于法向量对平移是独立的，所以我们可以利用对应的法向量先求解旋转矩阵 R 。

这可以通过最小化 $\sum_{i=1}^k w_i \|n'_i - R n_i\|^2$ 得到，我们利用[9]中描述的四元数法求解出这个旋转矩

阵，不妨设为 \hat{R} 。有了旋转，平移的确定就是要使得坐标系旋转后，对应平面上的任意两点在法线方向上的距离最小。而这两点我们分别选择了这两个对应平面的重心，即通过最小化

$\sum_{i=1}^k w_i (n_i'^T (m'_i - (\hat{R} m_i + t)))^2$ 我们就可以得到平移向量，该式对 t 求导，可得

$$(\sum_{i=1}^k w_i n'_i n_i'^T) t = \sum_{i=1}^k w_i n'_i n_i'^T (m'_i - \hat{R} m_i), \quad \text{求解这个方程组即得到了平移向量 } t。$$

3.3 基于平面特征进行配准的几点说明

3.3.1 关于 k 的取值

最少需要几对平面特征就可以得到 R 和 t ，这一点可以通过考虑假设 k 对平面完全匹配后物体或场景是否还有空间平移或旋转的自由度来判断。如果有，则显然不能唯一的确定 R 和 t ，也就无法求出他们。当 $k=1$ 和 $k=2$ 时，如下图所示，都还能找到这样的移动自由度。当 $k \geq 3$ 时，无法再找到这样的自由度了，此时，可以唯一的确定变换关系 (R, t) 。

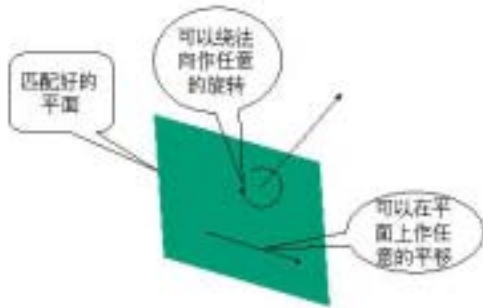


图1 $k=1$ 时，匹配后仍有平移和旋转自由度

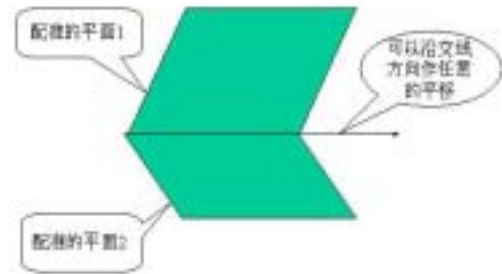


图2 $k=2$ 时，匹配后仍有平移自由度

3.3.2 对选取的平面的要求

显然，当选取的平面特征存在平行的平面时，在求 R 的过程中，这平行的一组平面的法向实际上只是相当于一个法向。结合 3.3.1 可知，至少应找到 3 组两两不平行的平面作为特征，才能完成配准。这也同时成为基于平面特征进行配准的局限，因为在实际场景的不同视点的扫描点云中，有时可能很难找到这样的对应的 3 个平面。

3.3.3 对应平面法向的朝向一致性

在配准阶段求旋转矩阵 R 时，要求两组平面特征中相对应的平面法向的朝向具有一致性，即该法向在两个深度图像中都应该具有一致的朝向，比如都是垂直物体或场景的表面向外，或者都是向内。否则，将导致计算出错误的旋转变换。由于在 3.2 的平面拟合阶段，根据一组点集计算法向时并没有规定朝向问题，因此，可能会出现朝向不一致的情况。我们在匹配前通过交互的手段来调整法向的正负，这样就得到了一致的朝向。

3.3.4 平面特征权值的选择

在配准中，我们为每一对相对应的平面特征设定了一个权值，这使得我们可以调整这对平面在配准中的重要性。比如有的平面本身已属于冗余的，而在拟合时的效果也不是很好，这样，就可以将其权值改小。我们可以根据平面的大小（平面点集所含点数的多少），平面的拟合误差来调整权值的大小，这样可以得到更好地配准效果。

4 由深度图像生成三角网格

由于深度图像本身是一组有序的二维点集，它蕴含了其对应的空间点之间的邻接关系。所以我们只需判断并取舍这些邻接关系，就可以生成相应的三角网格了。我们以相邻 4 点构成的邻域为一个基本单位，在这个基本单位中，我们按照一定的判别条件可分别生成 0 个、1 个或 2 个三角网格，再考虑连接的方向，最终可能生成的网格共有 7 种情况，如下图所示：

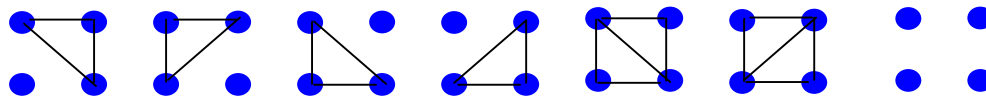


图 3 相邻 4 点的基本单位生成网格的 7 种形式

设深度图像的大小为 $m \times n$ 个点，则当我们处理完 $(m-1) \times (n-1)$ 个基本单位后，整个深度图像的三角网格化就完成了。

这里，问题的核心在于判别条件的定义，即对每一个基本单位，根据这 4 个点的位置关系，应该生成上述 7 种情况中的哪一种网格形式。一个最为简单和直观的判别条件，如 G. Turk 在[10]中所描述的，利用两两点之间的距离，如果超过一个阈值，则不连接，否则，则连接，然后根据连接情况判断是哪一种网格。这个方法简单而有效，但在下面的情况下则会出现问题。如图所示（激光扫描的侧面图，其中，A,...,G 为激光在列方向上依次扫描到的点）

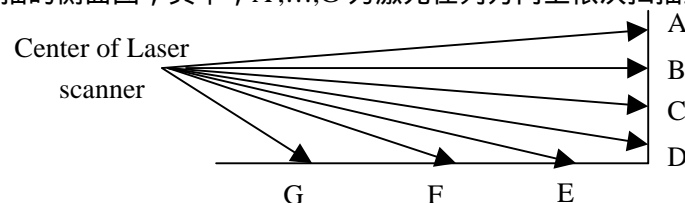


图 4 激光扫描竖直扫描 1 列的侧面图

在上图中，如果仅利用相邻点距离来判断，则 A,B,C,D 间的点距在阈值之内，可以连接，而 E,F,G 这 3 个点的点距则可能超出阈值而不连接。但实际上，他们显然在同一连续表面上，应该连接。如不连接，其结果就会在连续的表面上产生了很多的漏洞。造成这种情况的原因是这 3 点所处的平面与扫描激光方向的夹角较小，甚至近乎平行，因此，相邻两点间产生了较大的间距。

为了解决上述问题，我们削弱了仅考虑相邻点距离的条件，当点距超过阈值时，我们再另外判断这个基本单位里的 4 个点是否处于一个连续的表面，这一点我们是比较各点的法向量（该法向量在对点作局部拟合时已经求得）与其中三点构成的平面的法向的夹角大小来判别的。如果满足条件，则我们依然会作连接并找到一种相应的网格形式。这样就避免了漏洞的产生。

5 实验结果

我们对一系列场景进行了分割实验，图 5 给出了北京大学理科 2 号楼的南侧面的点云及分割结果。在 a) 中显示了通过扫描仪获得的点云数据，在 b) 中显示了分割的结果，不同的颜色表示不同的平面。

在配准实验中，我们从两个不同的视点扫描了一个室内场景，得到了两个点云数据。分割后，人工的选择了 3 组对应的平面对作为特征（天花板、门和一个墙面），实验中对这三个特征

都选用了缺省的权值 1。配准的结果还是令人满意的，参见图 6。

我们用改进前后的方法分别对一个室内场景的深度图像进行了建网，结果如图 7 所示，显然，改进后的方法确实解决了前面提到的问题，并得到了更好的结果。

6 结论

本文集中讨论了通过场景的深度图像重建其三维模型中的几个关键步骤：深度图像的分割、配准和建网。根据室内、室外场景的特点，本文提出了利用平面特征进行深度图像的配准的方法，该方法与手工选取对应特征点进行匹配相比更加准确、鲁棒，同时操作也更为简单方便。在由深度图像生成三角网方面，本文补充并改进了常用的网格生成判断条件，从而得到了更好的建网效果。

参考文献

- [1] I. Stamos, P. E. Allen, 3-D model construction using range and image data. Computer Vision and Pattern Recognition, Hilton Head Island, 2000, p.531-6
- [2] S. F. El-Hakim, C. Brenner, G. Roth, A multi-sensor approach to creating accurate virtual environments, JSPRS Journal of Photogrammetry & Remote Sensing, 1998
- [3] Y. Yu, A. Ferencz, J. Malik, Extracting Objects from Range and Radiance Images, IEEE Trans, Visualization and Computer Graphics, vol.7., 2001
- [4] H. Zhao, R. Shibasaki, A system for reconstructing urban 3D objects using ground-based range and CCD sensors. In Urban Multi-Media/3D Mapping workshop, Institute of Industrial Science(IIS), The University of Tokyo, 1999
- [5] C. Frueh, A. Zakhor, 3D model generation of cities using aerial photographs and ground level laser scans, Computer Vision and Pattern Recognition, Hawaii, USA, 2001, p. II-31-8, vol.2. 2
- [6] P. J. Besl, R. C. Jain. Segmentation through variable-order surface fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(2):167-192, march 1988.
- [7] P. J. Besl, N. D. McKay, A Method for Registration of 3-D Shapes, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 5, pp. 239-256, May 1992.
- [8] CyraTechnologies, Inc. Online documents, <http://www.cyra.com/>
- [9] 马颂德，张正友著 《计算机视觉——计算理论与算法基础》科学出版社 1998
- [10] G. Turk, M. Levoy, Zippered polygon meshes from range images, In Proceedings of SIGGRAPH'94

3D Reconstruction of Scenes Using Range Images

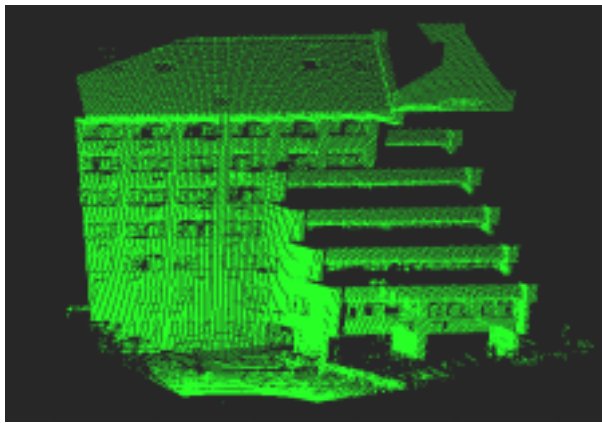
Wenfeng He, Hongbin Zha

Center for Information Science, Peking University, Beijing, 100871

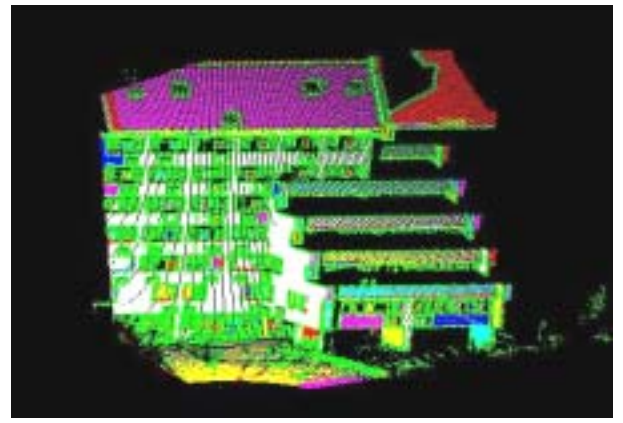
e-mail: {wenfeng, zha}@cis.pku.edu.cn

Abstract: Recently, there has been a large interest in the 3D modeling of real scenes using range data scanned by laser sensors. After acquisition of range data, lots of work should be covered to create a final 3D model of a scene, and among them we focus on segmentation, registration and meshing in this paper. Considering that there often exist a great number of planes in real scenes, we present a range registration method based on planar features, which are extracted when we segment range images. The results of our experiments indicate that the method is accurate and robust. In addition, we improve the general method of creating meshes from range images, consequently we get better results of meshing.

Keywords: Range image Laser scanner Point cloud Segmentation Registration Meshing

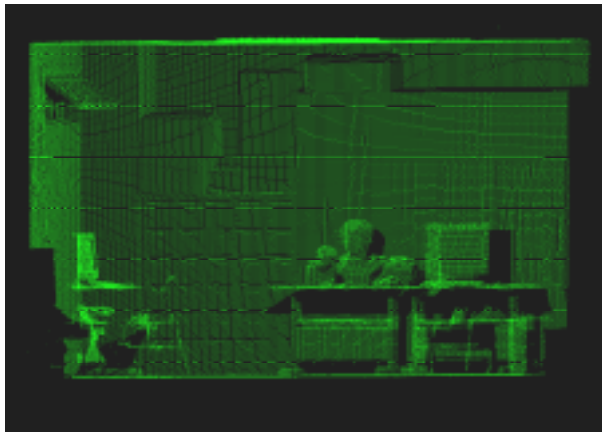


a) 点云

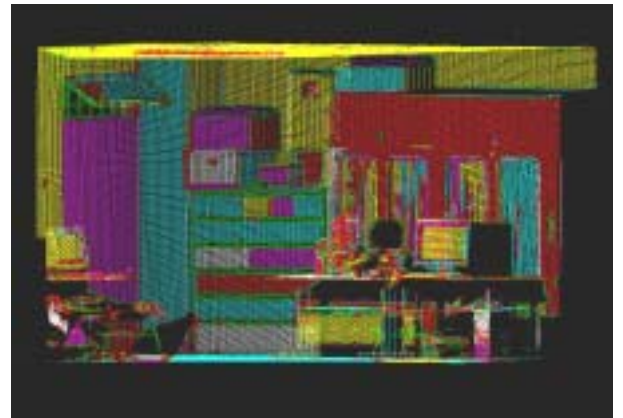


b) 分割后

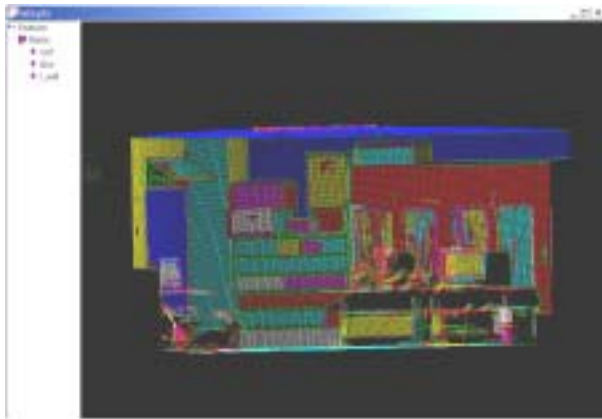
图 5 北大理科 2 号楼南侧面点云数据及分割结果



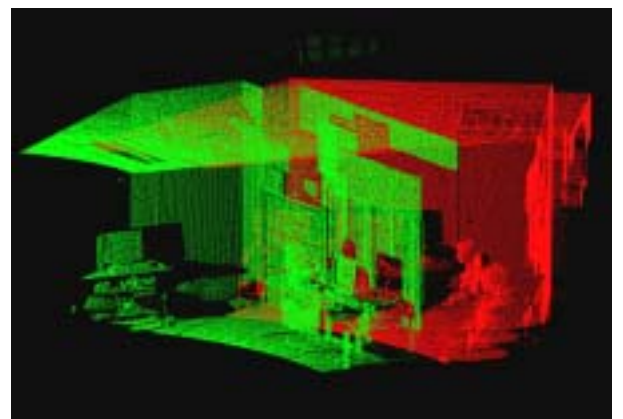
a) 从一个视点得到的扫描点云



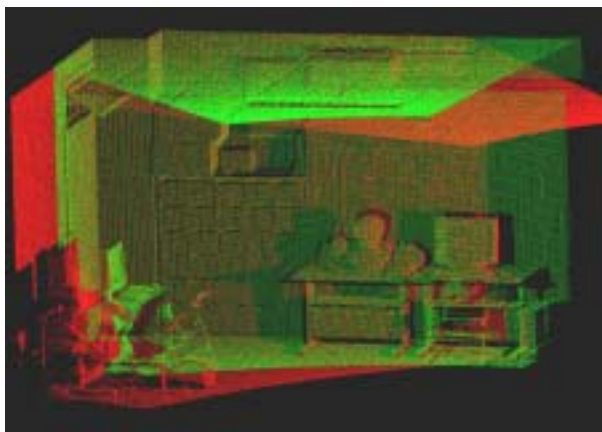
b) 对该点云分割的结果



c) 选择了相应的 3 个平面特征(蓝色)



d) 配准前,将两个点云直接放在一个坐标系下



e) 配准后

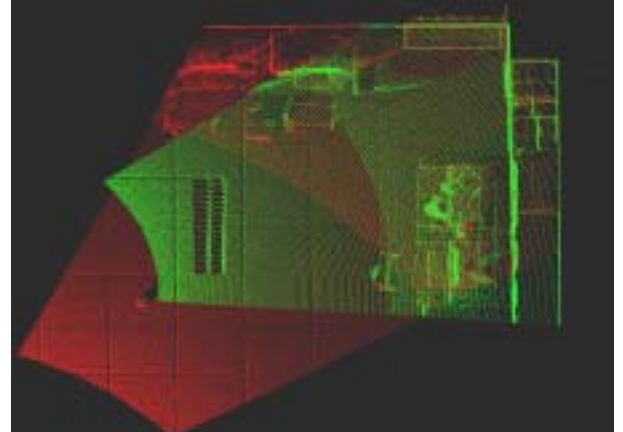
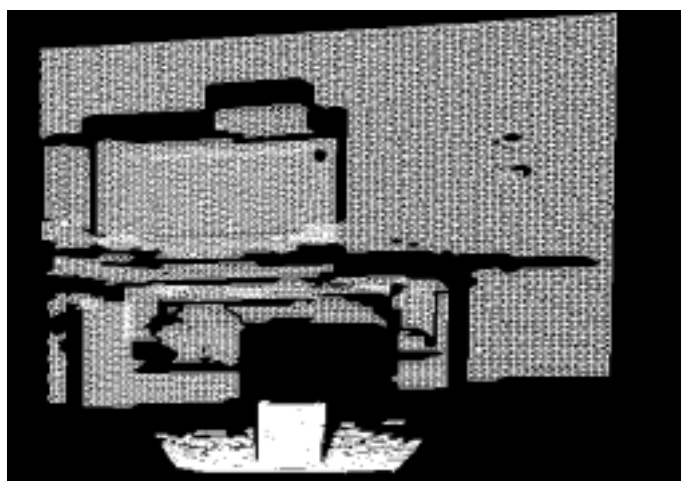
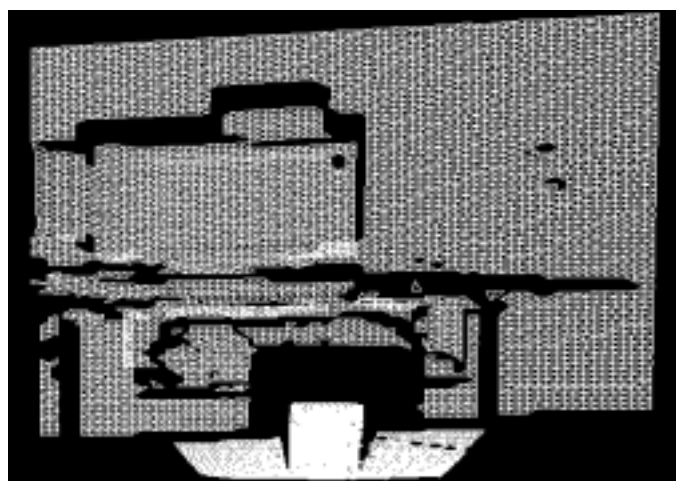


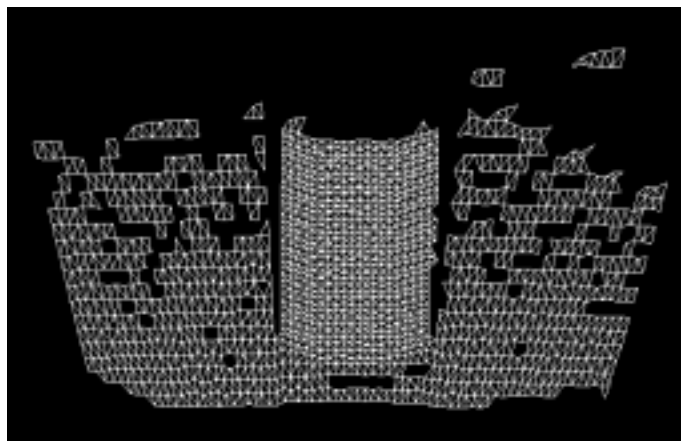
图 6 室内场景的配准实验结果



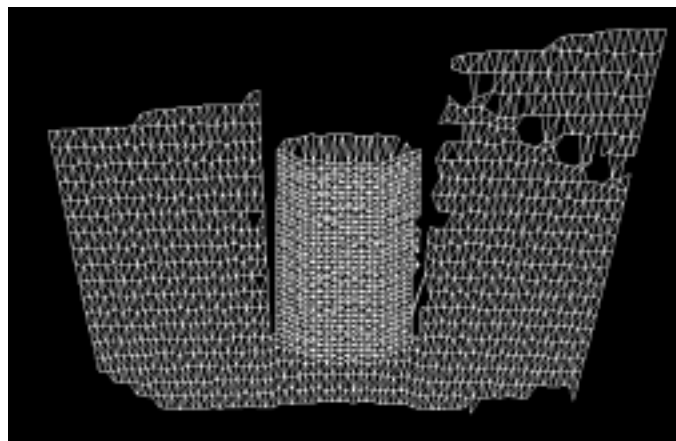
a) 一般方法的建网结果



b) 改进后方法的建网结果



c) 对 a) 中细节放大后的结果



d) 对 b) 中相应细节放大后的结果

图 7 改进前后的深度图像建网的结果比较