

# 基于 kinect 传感器的物体三维重建

郭连朋<sup>1</sup>, 陈向宁<sup>1</sup>, 徐万朋<sup>2</sup>, 刘田间<sup>1</sup>

(1. 中国人民解放军装备学院, 北京 101400; 2. 国家遥感中心航空遥感一部, 北京 100076)

**摘要:**随着计算机视觉的发展及深度传感器的出现, 基于 kinect 传感器的物体三维重建技术得到了广泛的研究。针对现有的基于 kinect 传感器的物体重建算法需要将待重建物体置于转台并随转台转动的局限性, 提出了一种通用的基于 kinect 传感器获取的物体彩色和深度图像完成物体三维重建的方法, 对深度图像进行预处理以平滑噪声并获取三维点云, 提取彩色图像匹配点并根据彩色图像特征点与三维点云之间的关系得到三维特征点, 并通过 SVD 方法完成点云初始配准, 然后利用改进的 ICP 算法完成点云的精确配准, 得到物体表面的三维点云, 重建出物体的三维模型。实验结果表明, 该方法能有效去除深度图像的噪声, 并对各个视角下的点云数据进行精确配准, 完成物体的三维重建并得到较好的重建效果。

**关键词:**kinect 传感器; 双边滤波; 点云配准; ICP 算法; 三维重建

**本文引用格式:**郭连朋, 陈向宁, 徐万朋, 等. 基于 kinect 传感器的物体三维重建[J]. 四川兵工学报, 2014(11): 119 - 123.

中图分类号: TP391.4

文献标识码: A

文章编号: 1006 - 0707(2014)11 - 0119 - 05

## 3D-object Reconstruction Based on Kinect Sensor

GUO Lian-peng<sup>1</sup>, CHEN Xiang-ning<sup>1</sup>, XU Wan-peng<sup>2</sup>, LIU Tian-jian<sup>1</sup>

(1. Equipment Academy of PLA, Beijing 101400, China; 2. No.1 Department of Aerial Remote Sensing of National Remote Sensing Center of China, Beijing 100076, China)

**Abstract:** With the development of computer vision and the emergence of the depth sensor, 3D-object modeling based on kinect sensor has been researched extensively. To overcome the limitations of existing 3D-object modeling algorithm based on kinect that objects to be rebuilt should be placed on the turntable and rotated with it, this paper presents a universal 3D-object modeling method based on the color and depth images of kinect. First, the depth image preprocessing is used to smooth the noise, and using the depth image after processing to obtain 3D point cloud. Then, extracting the matching points of color images, and using the relationship between feature points of color images and 3D points to obtain 3D feature points, the initial transformation matrix is calculated by the method of Singular Value Decomposition (SVD) method. Finally, the improved ICP algorithm is used to accurately registrant the point cloud and obtains 3D points of the surface of object that 3D-object modeling is accomplished. The experiment results shown that the method proposed in this paper can effectively remove the noise in depth images and accurately registrant the point clouds of each perspective to accomplish 3D-object modeling very well.

**Key words:** kinect sensor; bilateral filtering; point cloud registration; ICP; 3D reconstruction

**Citation format:** GUO Lian-peng, CHEN Xiang-ning, XU Wan-peng, et al. 3D-object Reconstruction Based on Kinect Sensor[J]. Journal of Sichuan Ordnance, 2014(11): 119 - 123.

收稿日期: 2014 - 05 - 26

基金项目: 军队探索项目:  $\times \times \times$  成像新方法研究(7131145)。

作者简介: 郭连朋(1990—), 男, 硕士研究生, 主要从事机器视觉和三维重建技术研究。

物体三维重建在一直是计算机视觉领域研究的热点问题,在设计仿真、虚拟现实、3D 电影和文化保护等诸多领域,应用十分广泛。目前物体三维重建方法主要有:基于轮廓线的算法<sup>[1]</sup>,利用物体投影到图像平面的轮廓线来重建出物体的三维结构;基于多深度图像融合的方法<sup>[2]</sup>,在尺度空间融合多幅深度图像得到物体表面的完整信息;基于体素的方法<sup>[3]</sup>,将待重建的物体划分成立方体栅格,然后基于这些体素得到物体的三维模型;基于多视匹配的算法<sup>[4]</sup>,根据序列图像通过匹配算法得到物体的稀疏点云结构,再通过膨胀和过滤等操作完成物体的三维表面重建。

随着 kinect 传感器的出现,作为一种 3D 传感器<sup>[5]</sup>,能够廉价、可靠和快速地获取彩色和深度图像进行三维测量,在三维场景恢复和物体重建等领域具有广泛的应用前景和实际应用价值。国内外许多专家学者对利用 kinect 传感器进行物体三维重建的方法进行了深入的研究,并取得了一系列的研究成果。在国外,Shahram Izadi、Richard A. Newcombe 等人 2011 年构建了 KinectFusion<sup>[6]</sup> 系统,对小范围静态场景和单独物体进行三维重建,得到相对精细的重建效果,但重建效率有待进一步提高。在国内,中国科学院自动化研究所模式识别国家重点实验室刘鑫等<sup>[7]</sup>实现了一种基于 Kinect 传感器的快速物体重建方法,并通过 GPU 并行加速,实现对转台上转动的物体的三维重建,但重建过程依赖于转台的速度,限制了其应用范围。

本文提出一种通用的基于 kinect 传感器的物体三维重建方法,利用 Kinect 传感器来获取物体的彩色和深度图像,对深度图像进行预处理以平滑噪声并获取三维点云,提取彩色图像匹配点并根据彩色图像特征点与三维点云之间的关系,得到三维特征点,对点云通过 SVD 方法完成点云初始配准,然后利用改进的 ICP 算法完成点云的精确配准,得到物体的三维点云,完成物体的三维重建,重建过程如图 1 所示。

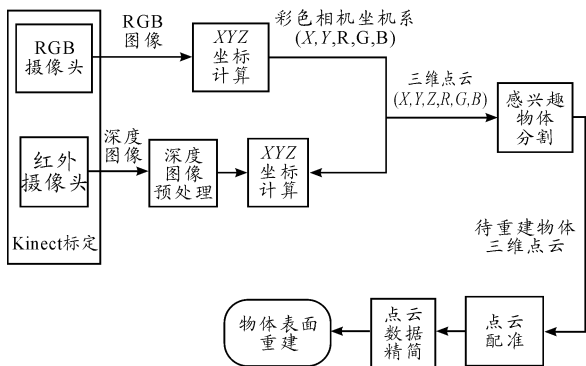


图 1 三维物体重建流程

## 1 kinect 标定

kinect 传感器标定的目的是矫正相机畸变、获取彩色和深度相机之间的对应关系以及获取相机几何模型。根据文献<sup>[8]</sup>,对 kinect 传感器进行标定,得到:

1) 标定彩色摄像机的内参数矩阵和深度相机参数矩阵

彩色相机的内参数矩阵为

$$K_{RGB} = \begin{bmatrix} f_{uRGB} & 0 & u_{0RGB} \\ 0 & f_{vRGB} & v_{0RGB} \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

其中:  $(f_{uRGB}, f_{vRGB})$  为彩色相机的焦距;  $(u_{0RGB}, v_{0RGB})$  为彩色相机的主点坐标。

$$K_{IR} = \begin{bmatrix} f_{uIR} & 0 & u_{0IR} \\ 0 & f_{vIR} & v_{0IR} \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

其中:  $(f_{uIR}, f_{vIR})$  为红外相机的焦距;  $(u_{0IR}, v_{0IR})$  为红外相机的主点坐标。根据红外相机的内参数矩阵,便可根据深度图像生成三维点云。令深度图像一点  $(x', y')$  处的深度值为  $depth(x', y')$ , 则该点在红外相机坐标系下的三维坐标  $(X_d, Y_d, Z_d)$  为

$$\begin{aligned} X_d &= (x' - u_{0IR}) \times depth(x', y') \times \frac{1}{f_{uIR}} \\ Y_d &= (y' - v_{0IR}) \times depth(x', y') \times \frac{1}{f_{vIR}} \\ Z_d &= depth(x', y') \end{aligned} \quad (3)$$

2) 标定彩色相机和红外相机的变换关系

$$T_{RGB-IR} = (R_{RGB-IR}, t_{RGB-IR}) \quad (4)$$

根据彩色相机与红外相机的变换关系和彩色相机及深度相机的内参数矩阵,可得到彩色图像坐标系和红外相机坐标系的对应关系,把彩色图像和三维点云联系起来。令  $(X_d, Y_d, Z_d)$  为红外相机坐标系下的三维坐标,则其在彩色相机坐标系下的三维坐标  $(X_{RGB}, Y_{RGB}, Z_{RGB})$  为

$$\begin{bmatrix} X_{RGB} \\ Y_{RGB} \\ Z_{RGB} \end{bmatrix} = R_{D-RGB} \begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} + t_{D-RGB} \quad (5)$$

其中,  $R_{D-RGB}$  和  $t_{D-RGB}$  分别为红外相机坐标系到彩色相机坐标系的旋转矩阵和平移向量,进一步可得其在彩色图像坐标系下的坐标位置  $(u_{RGB}, v_{RGB})$

$$\frac{1}{Z_{RGB}} \begin{bmatrix} u_{RGB} \\ v_{RGB} \\ 1 \end{bmatrix} = K_{RGB} \begin{bmatrix} X_{RGB} \\ Y_{RGB} \\ Z_{RGB} \end{bmatrix} \quad (6)$$

其中,  $K_{RGB}$  为彩色相机的内参数矩阵。

## 2 深度图像预处理及点云生成

如图 2 所示,由于 Kinect 传感器设备自身的限制、测量环境和待测物体表面特性等因素的影响,深度图像往往具有很大的噪声,根据文献<sup>[9]</sup>,利用 C. Tomasi 和 R. Manduchi 提出的双边滤波算法,对深度图像进行平滑去噪处理。

一般图像滤波可用式(4)表示为

$$I'(x, y) = \frac{1}{w_p} \sum_{i,j \in \Omega} w(i, j) * I(i, j) \quad (7)$$

其中:  $I(i, j)$  为输入图像;滤波器在  $(x, y)$  处的权值为  $w(i, j)$ ;像素  $(x, y)$  为领域范围为  $\Omega$ ;  $w_p$  为归一化参数;  $I'(x, y)$  为滤

波后的图像。

双边滤波的权值系数由空间域权值  $w_s$  和图像灰度域权值  $w_r$  组成,即:

$$w = w_s \times w_r \quad (8)$$

$$w_s = \exp\left(-\frac{(i-x)^2 + (j-y)^2}{2\delta_s^2}\right) \quad (9)$$

$$w_r = \exp\left(-\frac{(I(i,j) - I(x,y))^2}{2\delta_r^2}\right) \quad (10)$$

式(9)和式(10)中,  $\delta_s$  和  $\delta_r$  表示基于高斯函数的标准差。

经过去噪后的深度图像,根据式(3)生成场景及目标物体的三维点云。

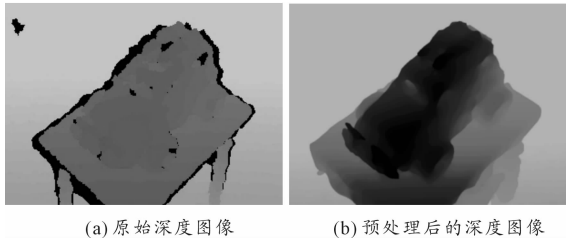


图2 Kinect 传感器深度图像

### 3 彩色图像特征点提取及匹配

常用的图像特征点检测方法有: Harris, SIFT 和 SURF 等,但由于 SIFT 算子时间复杂度高,计算量大,为了提高特征点检测效率及检测的鲁棒性和稳定性,根据文献[10],本文采用具有尺度不变性和旋转不变性的 SURF 特征检测方法提取 kinect 传感器彩色图像的特征点,估计对应两彩色图像之间的最佳变换矩阵,从而实现彩色图像的配准,如图3所示。

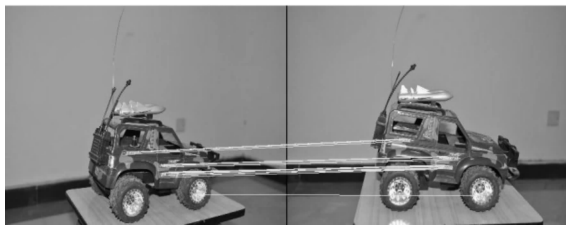


图3 Kinect 彩色图像配准结果

### 4 点云配准及数据精简

通过 RANSAC 算法<sup>[11]</sup>可得到比较稳定的变换矩阵,实现两片点云数据之间的粗匹配,要实现点云数据的精确匹配,可利用 ICP 算法<sup>[12]</sup>,在迭代过程中不断改变采样点的相对位置以提高配准精度。

#### 4.1 ICP 算法基本原理

ICP 算法是基于最小二乘的最优匹配方法,通过不断寻找对应点集和求解变换关系的过程,直到找到目标点集和参考点集之间的旋转矩阵  $R$  和平移向量  $t$ ,使两片点云满足某种条件下的最优匹配。设待匹配的两片点云数据的集合:

$$P = \{p_i\}_{i=1}^{N_p}, X = \{x_i\}_{i=1}^{N_x}$$

点集  $P$  中的每个点  $p_i$ ,搜索其在点集  $X$  上最近邻点作为对应点,设  $P = \{p_i\}_{i=1}^{N_p}$  的对应点组成模型点集:  $Y = \{y_i\}_{i=1}^{N_p}$ ,点云配准的目的是寻找点集  $P$  和点集  $Y$  之间的最优变换关系,即寻找最小的旋转矩阵  $R$  和平移向量  $t$ ,使目标函数最小

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \|y_i - (Rp_i + t)\|^2 \quad (11)$$

具体步骤如下:

1) 在点集  $X$  中寻找与点集  $P$  中对应点

$$Y = \{y_i\}_{i=1}^{N_p} \quad (12)$$

2) 求解旋转矩阵  $R$  和平移向量  $t$ ,使

$$\sum_{i=1}^{N_p} \|y_i - (Rp_i + t)\| \rightarrow \min \quad (13)$$

3) 更新模型点云

$$Y^{k+1} = \{y_i^{k+1} \mid y_i^k = R^k p_i^k + t^k, p_i^k \in P\} \quad (14)$$

4) 计算对应点集间的距离

$$d^{k+1} = \sum_{i=1}^n \|y_i^{k+1} - x_i^k\| / n \quad (15)$$

5) 若  $d^{k+1} < l$ ,则算法结束,否则返回步骤1)。

#### 4.2 点云粗匹配

如图4所示,将彩色图像中的匹配点映射到点云数据上得到点云数据的特征点,并利用 SVD 方法估计对应点云之间的变换矩阵,从而实现点云数据的粗匹配,为 ICP 算法的精确匹配提供良好的初始位置。

#### 4.3 改进的 ICP 算法

根据 ICP 的原理可知,直接对两片点云进行配准,在查找两幅点云之间的对应点时,时间复杂度为  $O(N_x N_p)$ ,速度非常慢,考虑将点感兴趣的物体从场景中分离出来进行点物体点云配准,可大幅减少配准点云的数量,提高点云配准效率。根据文献[13],利用基于谱聚类的点云模型分割算法把待重建的物体区域从场景中分割出来,如图5所示。

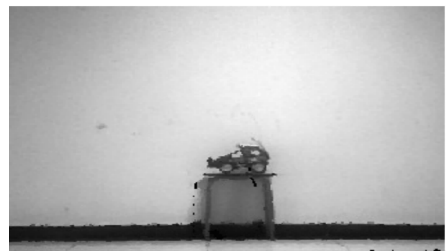


图4 单个场景三维点云

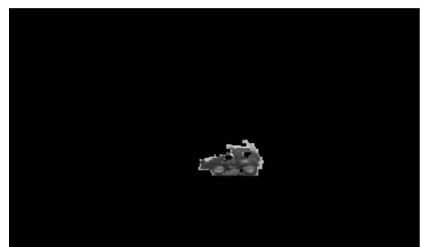


图5 分离出的待重建物体

4.4 数据精简

点云配准完成之后,会产生大量冗余的特征点,直接对精确配准后的点云进行三角化会消耗大量的时间和内存,因此需要对点云进行合并完成数据的精简。

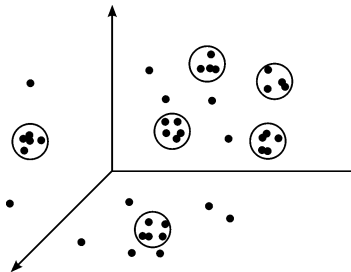


图6 三维特征点合并示意图

如图6所示,特征点合并的过程就是近邻点聚类过程。在三维空间中,一个三维向量是一个样本点 $s'$ ,这些样本点组成样本集 $E$ ,样本点 $s'$ 的最近邻就是任意样本点 $s \in E$ 满足 $Nearest(E, s', s)$ 。

$Nearest(E, s', s) \Leftrightarrow \forall s'' \in E \mid s' - s \leq \mid s' - s'' \mid$  (16)  
其中:

$$\mid s' - s \mid = \sqrt{\sum_{i=1}^k (s_i - s'_i)^2} \tag{17}$$

这里, $s_i$ 是向量 $s$ 的第 $i$ 维度。由于点云的数据量很大,直接逐点搜索会消耗大量的时间和内存,根据文献[14],采用八叉树算法将样本进行剖分,然后进行最近邻搜索能有效解决该问题。

5 实验结果与分析

为了验证本文算法的有效性,选用长为30 cm,宽为15 cm的玩具车作为重建对象,玩具车结构复杂,表面纹理丰富,有凹凸的细节,具有一定的代表性。本文的实验环境为微软kinect传感器一台及配套设备;操作系统为64位Microsoft Windows7的笔记本电脑一台,处理器为Intel 酷睿 i5 3230M,主频为2.6 GHz,内存4 G,硬盘1T,显卡型号为NVIDIA GeForce GT 750M,显存容量为2 GB。

实验中,用Kinect传感器绕玩具车模型一周拍摄24幅彩色和深度图像,对深度图像的滤波去噪去除深度图像噪声并修补深度图像中的孔洞,根据kinect传感器的标定结果得到单个场景的三维点云,通过基于谱聚类的点云模型分割算法把场景中感兴趣的待重建物体分割出来。提取彩色图像匹配特征点并同步映射到三维点云得到三维点云中的匹配特征点,对物体表面三维匹配特征点利用SVD方法得到初始变换矩阵实现物体点云的初始配准,并通过ICP算法对其进行精确配准,结果如图7所示。

此时,得到的仅仅是离散的物体表面的三维特征点,为了重建出物体表面,还要对其进行三角化。由于精确配准后会产生大量的三维点云,直接进行三角化消耗大量的时间和

内存,因此,本文先对精确配准生的点云进行简化,如图8所示。

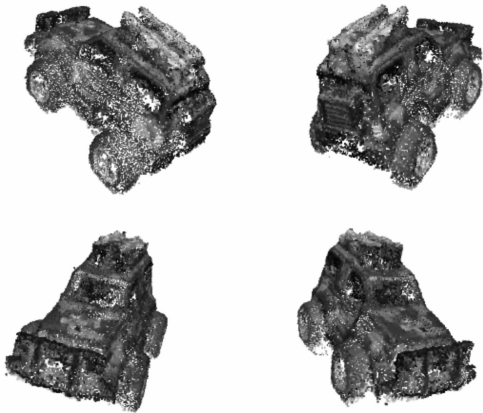


图7 玩具车模型点云精确配准结果

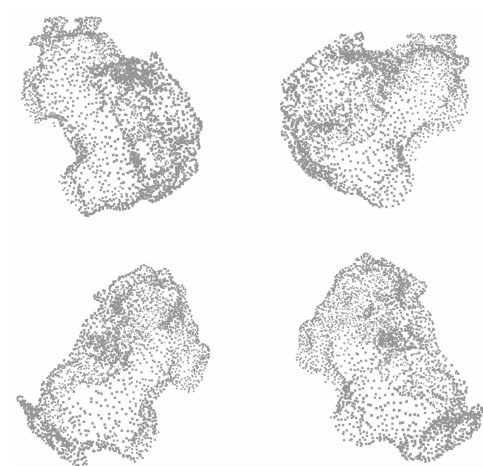


图8 精简后的玩具车散乱点云

对简化后的点云进行Delaunay三角化<sup>[15]</sup>,如图9所示。

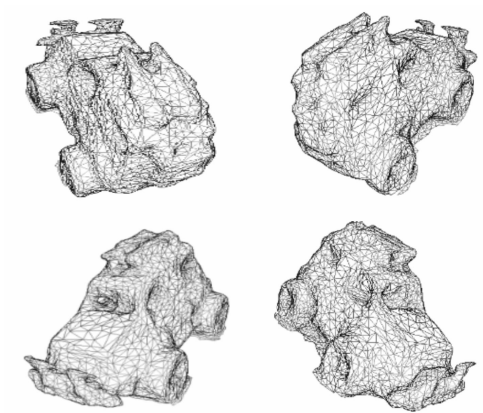


图9 玩具车点云模型的三角化结果

对精简后的点云采用基于泊松(Poisson)方程的曲面重建,如图10所示。

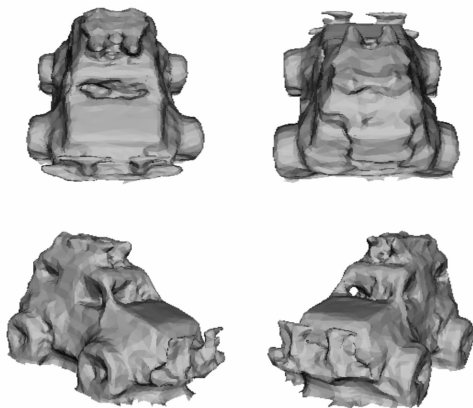


图 10 玩具车表面重建结果

## 6 结论

本文提出了一种通用的基于 kinect 传感器的物体三维重建方法,该方法可利用 Kinect 传感器获取的物体彩色和深度图像,通过对深度图像预处理及点云精确配准等步骤得到物体的三维点云,重建出物体的三维表面模型,并得到相对比较精细的细节特征,而不必接触物体或将物体放于转台令其转动。但该算法的局限在于:①点云初始配准依赖于彩色图像特征点的匹配结果,所以对于彩色纹理较为单一的物体,重建效果不好;②kinect 传感器深度信息获取精度较差,仅为 1 cm,而且深度获取的有效距离为 0.5 ~ 3.5 m,所以对于体积较好小的物体,无法完成三维表面重建。此外,本文的重建结果不包含彩色纹理信息,因此,接下来的工作需要彩色图像纹理映射,得到带彩色纹理的三维模型。

## 参考文献:

- [1] Zaharescu A, Boyer E, Horaud R. Transformesh: a topology-adaptive mesh-based approach to surface evolution [C]//Computer Vision-ACCV 2007. Springer Berlin Heidelberg, 2007: 166 – 175.
- [2] Bradley D, Boubekeur T, Heidrich W. Accurate multi-view reconstruction using robust binocular stereo and surface meshing[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1 – 8.
- [3] Paris S, Sillion F X, Quan L. A surface reconstruction method using global graph cut optimization[J]. International Journal of Computer Vision, 2006, 66(2): 141 – 161.
- [4] Lhuillier M, Quan L. A quasi-dense approach to surface reconstruction from uncalibrated images[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005, 27(3): 418 – 433.
- [5] Arieli Y, Freedman B, Machline M, et al. Depth mapping using projected patterns [P]. U. S. Patent: 8, 150, 142, 2012 – 04 – 03.
- [6] Newcombe, Richard A., Izadi, Shahram, Hilliges, Otmar et al. KinectFusion: Real-time dense surface mapping and tracking [C]//2011 10th IEEE International Symposium on Mixed and Augmented Reality. [S. l.]: [s. n.], 2011: 127 – 136.
- [7] 刘鑫,许华荣,胡占义,基于 GPU 和 Kinect 的快速物体重建[J]. 自动化学报, 2012, 38(8): 1288 – 1297.
- [8] Herrera C, Kannala J, Heikkilä J. Joint depth and color camera calibration with distortion correction [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(10): 2058 – 2064.
- [9] Lai P L, Tian D, Lopez P. Depth map processing with iterative joint multilateral filtering [C]//Picture Coding Symposium (PCS), 2010. IEEE. [S. l.]: [s. n.], 2010: 9 – 12.
- [10] 杨鸿,钱堃,戴先中,等. 基于 Kinect 传感器的移动机器人室内环境三维地图创建[J]. 东南大学学报: 自然科学版, 2013, 43(A01): 183 – 187.
- [11] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381 – 395.
- [12] Liu Y. Automatic registration of overlapping 3D point clouds using closest points[J]. Image and Vision Computing, 2006, 24(7): 762 – 781.
- [13] 马腾,龙翔,冯路等. 点云模型的谱聚类分割[J]. 计算机辅助设计与图形学学报, 2013, 24(12): 1549 – 1558.
- [14] 宋涛,欧宗瑛,王瑜,等. 八叉树编码体数据的快速体绘制算法[J]. 计算机辅助设计与图形学学报, 2005, 17(9): 1990 – 1996.
- [15] Akio D, Koide A. An efficient method of triangulating equi-valued surfaces by using tetrahedral cells[J]. IEICE TRANSACTIONS on Information and Systems, 1991, 74(1): 214 – 224.

(责任编辑 杨继森)