

基于改进 SIFT-ICP 算法的物体点云建模方法^①

于 灏^② * ** 杜华军 **** * 蔡莹皓 * 鲁 涛 * 王 睿 * 王 硕 * ** ****

(* 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190)

(** 中国科学院大学 北京 100049)

(*** 中国科学院脑科学与智能技术卓越创新中心 上海 200031)

(**** 北京航天自动控制研究所 北京 100854)

(***** 宇航智能控制技术国家级重点实验室 北京 100854)

摘 要 为了实现对任意摆放的物体的识别与位姿估计,需要得到物体的三维模型,提出了一种高效准确的物体三维点云模型构建方法。首先通过 RGB-D 传感器获取物体多角度数据,然后利用尺度不变特征变换(SIFT)特征点匹配和改进的迭代最近点(ICP)算法计算出各角度下传感器的相对位姿,进而生成目标物体所在场景的完整点云,并通过物体分割和点云后处理得到目标物体的三维点云模型。实验结果显示,构建的物体点云模型清晰且不失真,并保留了表面完整的特征信息。

关键词 点云模型,尺度不变特征变换(SIFT),迭代最近点(ICP),物体分割

0 引 言

三维点云模型是一种常见的三维信息表达形式。点云模型中的每一个点对应一个测量点,可包含测量点的空间位置信息和其特征信息,被广泛应用于机器人导航任务中的地图表达^[1]和机器人抓取任务中的目标物体位姿估计^[2]等领域。传统的建模方法是使用专门三维测量设备,如激光三维扫描仪,使用这些设备具有较高的精度,但是复杂度高而且成本昂贵,对噪声敏感,不适用于精度要求不高和物体视野缺损的情况。

随着计算机视觉技术的发展,基于视觉传感器获取的图像进行模型重构的研究,引起国内外的广泛关注,并取得丰硕成果。常见的视觉传感器包括单目相机、双目相机和 RGB-D 深度相机等,对单目和双目相机获取的图像,需要利用三角定理计算获得测量点的空间位置信息,而 RGB-D 深度相机利用相机内参进行坐标系变换可以直接获得测量点的空

间位置信息。基于计算机视觉的模型重构方法,关键在于传感器的位姿估计,通过采集图像中的特征点匹配等方法计算获得采集图像对应时刻的相机位姿估计,实现测量点在三维空间中的表达。文献[3-6]是利用采集数据中的特征,如二维尺度不变特征变换(scale invariant feature transform, SIFT)特征和基于卷积网络的三维特征,通过特征点匹配进行帧间相机的位姿估计,进而实现三维重构。文献[7]中采用双目传感器获取沥青路面的结构图像,通过 SIFT 特征提取和特征点匹配实现三维重构,该方法适用于微观纹理的重构和大场景环境的重构,但是 SIFT 算法对于纹理不足的物体提取的特征点较少,其中可能掺杂的错误匹配对结果造成影响较大。文献[8]设计了一种名为 ElasticFusion 的系统,通过 RGB-D 传感器获取点云数据,通过迭代最近点(iterative closest point, ICP)算法计算相机在不同位置的相对位姿,可以实现对物体的稠密建模和大场景室内环境的建模,但是 ICP 算法容易受到

① 国家自然科学基金(U1713222,61773378)资助项目。
② 男,1992 年生,硕士生;研究方向:计算机视觉;联系人,E-mail: yuhao2016@ia.ac.cn
(收稿日期:2018-10-21)

初始化的影响陷入局部最小,且对相邻帧运动较大的点云配准的误差较大。文献[9]通过移动机器人搭载深度相机进行场景扫描,利用 SIFT 算法和 ICP 算法结合求取相机的位姿变换,但在点云预处理过程中可能会滤掉有用信息,对相机位姿估计造成干扰,影响模型重构的精度。

本文提出了一种可实现对任意姿态放置的物体进行三维点云建模的方法。通过 RGB-D 传感器环绕物体采集 30 帧彩色图像和深度图像,对彩色图像进行 SIFT 特征点提取和特征匹配,估计相邻帧数据对应时刻的相机位姿,将 SIFT 帧间匹配结果作为改进 ICP 算法的初值,通过改进 ICP 算法对帧间相机的位姿进行精确估计,实现场景模型重构,并通过物体分割和点云去噪得到完整的物体三维模型,最后利用主成分分析(principal components analysis, PCA)算法实现物体模型在物体坐标系中的表达。本文的点云重构方法可实现对任意摆放的物体进行点云模型构建,在纹理稀疏和纹理丰富的场景中都有较高的精度,并且对相邻帧运动较大的点云配准具有较强的鲁棒性。

1 相机运动姿态估计

本文整体算法流程如图 1 所示。

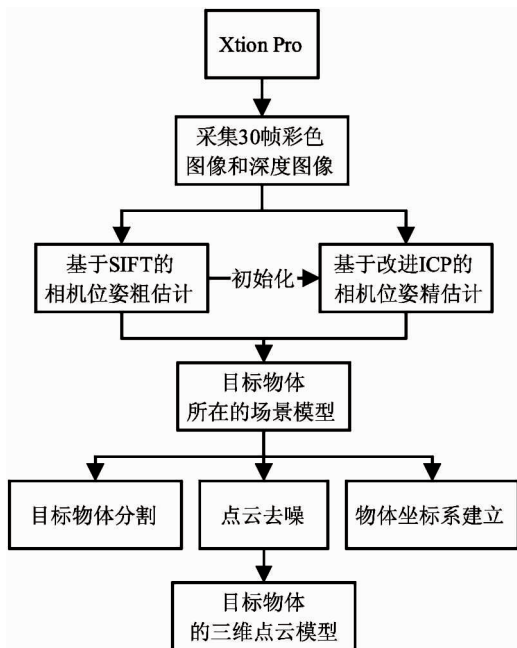


图1 算法流程图

从图 1 可知,为了得到物体的三维模型,首先需要构建实验场景的全局模型。传统的 SIFT 特征匹配实现模型重构,适用于纹理丰富场景,对纹理不足的环境难以找到足够的特征,而通过 ICP 匹配的模型重构方法对纹理不敏感,但是受到初始化的影响,对相邻帧图形运动较大的点云配准效果不佳。本文通过 RGB-D 深度相机环绕物体采集 30 帧彩色图像和深度图像,利用 SIFT 特征提取匹配得到相机位姿估计的初始值,再进行 ICP 迭代得到精确的相机相对位姿,实现对场景的三维重构。

1.1 基于 SIFT 的相机运动位姿粗估计

SIFT 算法^[10]的思想是对不同尺度空间中的极值点进行检测形成特征点,分析特征点附近的灰度变化,建立特征描述子,通过描述子进行特征匹配。本文通过深度相机 Xtion Pro 环绕目标物体采集 30 帧彩色图像和深度图像,对彩色图像进行 SIFT 特征提取和特征匹配,结合匹配特征点的深度信息,实现对相邻帧数据对应时刻的深度相机的位姿粗估计。具体算法流程如表 1 所示。

表 1 基于 SIFT 算法的相机位姿估计

基于 SIFT 的相机位姿估计算法流程
1:对相邻帧 i 和 j 的彩色图像进行 SIFT 特征匹配,得到对应点集 P 和 Q ;
2:点集 P 和 Q 中各点结合其深度值 z ,通过式(1)变换,得到对应的三维点集 $\{P_i\}$ 和 $\{Q_i\}$,其中 M_{in} 是深度相机的内参;
3:点集 $\{P_i\}$ 和 $\{Q_i\}$ 中各点分别减去其质心坐标,得到变换后的坐标 P_i^{k*} 和 Q_i^{k*} ;
4:最小化目标函数 J ,构造矩阵 H , H 如式(3),对矩阵 H 进行奇异值分解,求得旋转矩阵 R_j^i 和平移矩阵 t_j^i ,得到相邻帧 i 和 j 的对应时刻的相机位姿齐次变换矩阵 T_j^i 。

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z M_{in}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, M_{in} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$J = \sum_{k=1}^N \| P_i^k - (R_j^i \cdot Q_j^k + t_j^i) \|^2 \quad (2)$$

$$H = \sum_{k=1}^N P_i^{k*} \cdot Q_i^{k*} \quad (3)$$

$$\mathbf{T}_j^i = \begin{bmatrix} R_j^i & t_j^i \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (4)$$

通过上述方法能有效计算出相邻视角下深度相机的相对位姿变换,但在 SIFT 特征点匹配的选取过程中可能存在错误匹配,对深度相机的位姿估计精度造成影响。为了减小 SIFT 特征点错误匹配的影响,将 SIFT 特征匹配与随机抽样一致性(RANSAC)算法^[11]相结合。RANSAC 算法核心是针对一组包含局外点的观测数据,其分布可以通过数学模型来表示,假设可以通过 k 个数据估计模型参数,通过迭代方式将观察数据分为内点和外点,再用内点重新估计模型参数,尽可能地消除外点的干扰。RANSAC 算法结合 SIFT 特征匹配对深度相机的位姿估计计算流程如表 2 所示。

表 2 基于 SIFT 和 RANSAC 的相机位姿估计

基于 SIFT 和 RANSAC 的相机位姿估计
1:从点集 $\{P_i\}$ 和 $\{Q_i\}$ 中随机抽取 3 对对应点,用上述方法计算相邻帧 i 和 j 对应时刻的相机位姿齐次变换矩阵 $\mathbf{T}_j^i$;
2:计算点集 $\{P_i\}$ 和 $\{Q_i\}$ 中所有点对的误差 e_k , 设阈值 ε , 当 $e_k < \varepsilon$ 时,判定该点是模型的内点,统计内点数目;
3:重复步骤 1 和 2 至最大迭代次数,选择内点数最多的变换矩阵为最优解;
4:利用步骤 2 确定最优解时的内点,重新计算得到相机位姿齐次变换矩阵 $\mathbf{T}_j^i$ 。

$$e_k = \sum_{k=1}^N \|P_i^k - (R_j^i \cdot Q_j^k + t_j^i)\|^2 \quad (5)$$

1.2 基于 ICP 的相机运动位姿精估计

利用上述方法得到的帧间深度相机位姿齐次变换矩阵 $\mathbf{T}_j^i$, 作为 ICP 迭代的初始值,当迭代次数达到指定次数或误差小于设定阈值时,停止迭代,可以得到更加精确的相机位姿估计。经典的 ICP 算法^[12]是不断迭代最小化目标函数,计算两组点云之间的变化关系,其具有较大的局限性,容易受到初始化的影响陷入局部最小值,利用 SIFT 特征匹配计算结果作为初值,可以减小初始化的影响,但是经典

ICP 算法针对相机运动较快的点云配准误差较大,实验数据是环绕物体采集的 30 帧图像,相邻图像的点云配准存在类似相机运动较快的情况。文献[13]中提出的改进 ICP 算法可适用于相机运动较快的点云拼接,其中点对的权重设为

$$Weight = 1 - \frac{distance(p_i, q_j)}{distance(max)} \quad (6)$$

其中, $distance(p_i, q_j)$ 是两组点云 $\{P_i\}$ 和 $\{Q_i\}$ 中对应点 p_i, q_j 的欧式距离, $distance(max)$ 是两组点云中对应点的欧式距离最大值,设计的权重与对应点之间的欧式距离成反比,使近点对结果的影响最大化,减小错误匹配的影响。改进的 ICP 算法采用单位四元数法计算相邻帧对应的深度相机位姿变换矩阵。

设旋转向量是单位四元数 $\mathbf{q}_R = [q_w \ q_v]^T$, 其中 $\mathbf{q}_v = [q_1 \ q_2 \ q_3]^T$, $q_w \geq 0$, 并且 $q_w^2 + q_1^2 + q_2^2 + q_3^2 = 1$, 可以得到旋转矩阵 $\mathbf{R}(\mathbf{q}_R)$:

$$\mathbf{R}(\mathbf{q}_R) = (q_w^2 - \mathbf{q}_v^T \mathbf{q}_v) \mathbf{I}_{3 \times 3} + 2\mathbf{q}_v \mathbf{q}_v^T + 2q_w [\mathbf{q}_v]_{\times} \quad (7)$$

设平移向量是单位四元数 $\mathbf{q}_t = [q_4 \ q_5 \ q_6]^T$, 可得齐次变换向量 $\mathbf{q}_T = [\mathbf{q}_R \ | \ \mathbf{q}_t]^T$ 。设定目标函数如式(8), 求取最小化时对应的齐次变换向量 $\mathbf{q}_T$ 。

$$f(\mathbf{q}_T) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|q_i - \mathbf{R}(\mathbf{q}_R)p_i - \mathbf{q}_t\|^2, \quad p_i \in \{P_i\}, q_i \in \{Q_i\} \quad (8)$$

基于改进 ICP 算法的相机位姿估计的计算流程如表 3 所示。

$$\mu_p = \frac{1}{N_p} \sum_{i=1}^{N_p} p_i \quad (9)$$

$$\mu_q = \frac{1}{N_q} \sum_{i=1}^{N_q} q_i \quad (10)$$

$$U = cov(P, Q) = \frac{1}{N_q} \sum_{i=1}^{N_q} (p_i - \mu_p)(q_i - \mu_q)^T \quad (11)$$

$$\mathbf{T} = \begin{bmatrix} tr(U) & \Delta^T \\ \Delta & U + U^T - tr(U)\mathbf{I}_{3 \times 3} \end{bmatrix}, \quad \Delta = [A_{23} \ A_{31} \ A_{12}]^T, A_{ij} = (\mathbf{R} - \mathbf{R}^T)_{i,j} \quad (12)$$

$$q_t = \mu_q - \mathbf{R}(q_R)\mu_p \quad (13)$$

表 3 基于改进 ICP 的相机位姿估计

基于改进 ICP 算法的相机位姿估计

- 1:计算相邻帧采集数据 i 和 j 对应的点云 $\{P_i\}$ 和 $\{Q_i\}$ 的均值和协方差,分别为 μ_p, μ_q 和 U ,如式(9)、(10)、(11);
- 2:根据协方差矩阵构造对称矩阵 T ,如式(12);
- 3:计算对称矩阵 T 的特征值和特征向量,最大特征值对应的特征向量是最优旋转向量 q_R ,可得最优平移向量 q_t ,如式(13)。

通过上述方法可以实现相邻帧采集数据对应时刻的相机位姿的精准估计。将第一帧采集数据对应时刻的相机坐标系设为世界坐标系,利用 SIFT 特征点匹配和 ICP 匹配,每一帧都对齐到世界坐标系,最终得到目标物体所在场景的完整三维模型。

2 目标物体三维建模

2.1 目标物体分割

为了得到目标物体的三维点云模型,需将物体模型从所在场景模型中提取。实验场景是物体放置在平整桌面上,利用深度阈值分割,提取深度值小于阈值的前景部分,进而将目标物体与桌面分割,需要对桌面进行平面模型参数估计。选取点云模型中不共线的 3 点,其齐次坐标为 $p_i = [x_i, y_i, z_i, 1]^T, i = 1, 2, 3$, 通过选取的 3 点表示的平面方程是 $\Pi = [a, b, c, d]^T$, 则有:

$$\begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} \cdot \Pi = 0 \tag{14}$$

设 C 表示上述阈值分割后得到的点云, C 中的点数远大于 3, 为了减小平面参数估计的误差, 选用平面拟合和 RANAC 算法相结合的方法对平面模型进行参数估计, 具体计算流程如表 4 所示。

计算 C 中所有点到平面的距离 $|X^T \cdot \Pi|$, 判断其是否属于平面模型, 将平面模型与目标物体模型分割。

表 4 基于 RANSAC 的平面模型拟合

基于 RANSAC 算法的平面模型拟合

- 1:初始化 $N_{\max} = 0$, 随机选取 C 中不共线的三点 $P_i, i = 1, 2, 3$, 根据式(14)求得这三点确定的平面 Π^* , 归一化 $\sqrt{a^2 + b^2 + c^2} = 1$;
- 2:对 C 中所有的点, 判断其是否属于平面 Π^* , 统计内点数目 N , 当 $N > N_{\max}$ 时, $N_{\max} = N$;
- 3:重复步骤 1 和 2 至最大迭代次数, 选取内点最多的 Π^* , 并重新确定其内点, 根据式(14)计算平面模型的最优参数估计。

2.2 目标物体三维点云模型构建

首先对物体的三维点云模型去噪。对物体点云中的所有点, 统计其半径 r 内的三维空间中的点数 N , 当 N 小于设定阈值时, 判断该点为噪点并滤除。通过上述方法获得的物体三维点云模型的坐标系是基于世界坐标系, 为了完成后面更加复杂的任务, 需要将物体模型表达转换到自身的物体坐标系上。采用主成分分析(PCA)算法建立物体坐标系。物体坐标系建立流程如表 5 所示。

表 5 基于 PCA 算法的物体坐标系建立

基于 PCA 算法的物体坐标系建立

- 1:设物体点云为 C , 将 C 中所有点坐标减去质心坐标 $\bar{P}$, 变为 P^* , 构造系数矩阵 Q , 如式(17);
- 2:计算矩阵 Q 的特征值 λ 和特征向量 X , 得到物体坐标系相对世界坐标系的位姿变换矩阵 T , 如式(20);
- 3:根据式(21), 对点云 C 中各点进行坐标系变换, 将物体点云模型表达转换到物体坐标系上。

$$\bar{P} = \frac{1}{N} \sum_N P_i, i = 1, 2, \dots, N \tag{15}$$

$$A = [P_1^*, P_2^*, \dots, P_n^*] = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ z_1 & z_2 & \dots & z_n \end{bmatrix} \tag{16}$$

$$Q = A \cdot A^T \tag{17}$$

$$Q \cdot X_i = \lambda \cdot X_i, i = 1, 2, 3 \tag{18}$$

$$R = [X_1, X_2, X_3] \tag{19}$$

$$T = \begin{bmatrix} R & \bar{P} \\ 0_{1 \times 3} & 1 \end{bmatrix} \tag{20}$$

$$P_i^* = T^{-1} \cdot P_i \tag{21}$$

3 实验结果分析

3.1 物体三维点云建模

为了验证所设计方法的有效性,通过手持华硕 Xtion Pro 深度相机环绕目标物体,采集 30 帧彩色图像和深度图像,对采集到的彩色图像进行 SIFT 特征提取和相邻帧的特征点匹配,结合深度图像中匹配特征点的深度信息,实现相邻帧所对应的深度相机的位姿估计。SIFT 特征匹配为后续的 ICP 匹配提供良好的初始化条件,利用改进的 ICP 算法进一步精确地对准相邻帧的点云,通过 SIFT 特征匹配和 ICP 匹配结合的方式对采集的 30 帧数据进行处理,可得到目标物体所在实验场景的完整点云模型。对场景模型进行阈值分割、物体分割、点云去噪和物体坐标系建立,最终得到目标物体的三维点云模型。本文采用的实验环境是 Windows 系统,编程环境为 Visual Studio 2013,实验结果如图 2 所示。

图 2(a)表示从采集数据中抽取的 2 张相邻帧的 SIFT 特征匹配结果,SIFT 特征匹配点对数是 97 对。图 2(b)表示利用 SIFT 特征匹配实现相邻两帧图像的的点云配准,结果显示两组点云配准出现错位,因为匹配的特征点对是 97 对,相对较少,其中可能包含的错误匹配对结果造成影响,但可以为后面的 ICP 匹配提供良好的初始化条件。图 2(c)表示通过改进 ICP 算法精配准两帧点云,设定配准精度是 $1e-8$,结果显示修正了 SIFT 特征匹配的错位。图 2(d)表示对采集的 30 帧数据分别进行 SIFT 特征匹配和 ICP 匹配,再通过阈值分割后的场景模型。图 2(e)表示通过物体分割和点云去噪处理后,得到物体的三维点云模型,通过 PCA 算法将物体的点云模型表示在物体坐标系下,以红、绿、蓝标示的 X 、 Y 、 Z 轴分别与物体点云的短轴、次短轴、长轴对齐。

应用本文提出的方法,为 12 种物体建立了三维模型,如图 3 所示。表 6 列出了各种物体点云模型

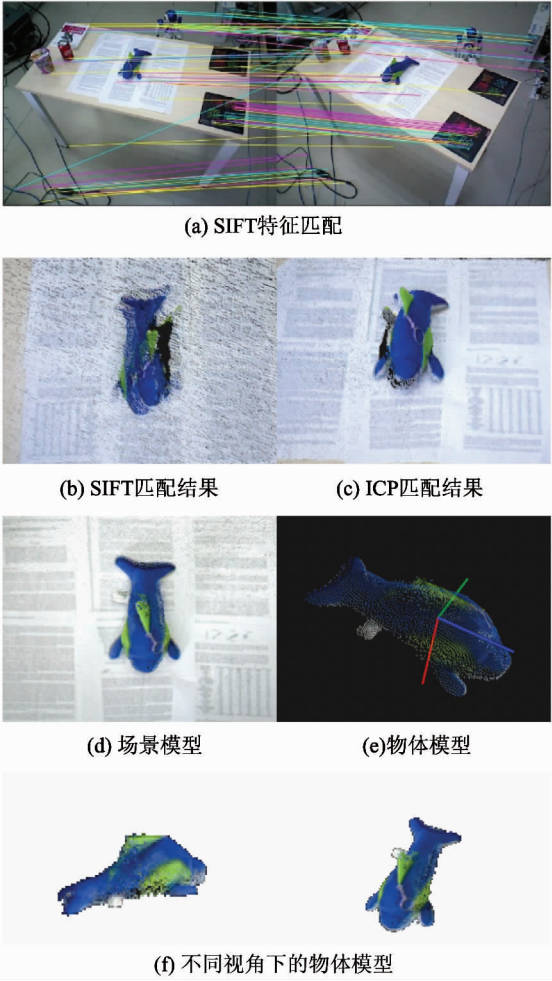


图 2 物体三维点云模型构建

所包含的点数、帧间 SIFT 特征匹配过程中的平均匹配点数以及帧间相机位姿估计的平均运行时间,其中物体编号对应图 3 中的从左到右、从上到下的各种物体。由于点云数据大多数超过 10 000 点,ICP 匹配过程耗时较长,可以通过 GPU 加速 SIFT 特征匹配和 ICP 匹配过程,缩短运行时间。

如表 7 所示,对于不同数据量的点云数据,采用传统 ICP 算法、基于全局优化的点云配准 Go-ICP^[14]算法和本文算法进行点云配准实验对比,从表中可知,在同样的误差收敛条件下,本文算法的效率明显优于传统 ICP 算法和 Go-ICP 算法。

值得一提的是,由于深度相机 Xtion Pro 采用红外测距,本文采用的方法适合于带有纹理的物体,无法适用于透明或表面近似镜面的物体的建模,基于 SIFT 特征匹配的相机位姿粗估计可以用 SURF、ORB 等特征提取算法代替,基本框架无需改变。



图3 物体的三维点云模型

表6 各物体模型的相关数据

物体 编号	三维 点数	平均匹配 特征点数	平均运行 时间(s)
1	6 787	85	17.153
2	8 623	94	21.371
3	9 153	107	23.783
4	8 934	114	23.014
5	12 754	136	26.841
6	13 394	148	27.612
7	15 271	162	30.782
8	14 327	154	29.673
9	4 713	73	12.645
10	23 185	169	41.752
11	10 285	118	24.642
12	11 839	126	25.207

3.2 物体点云模型误差分析

通过计算采集数据中相邻帧点云配准过程中的对应点间的平均欧式距离,对最终构建的物体点云

表7 不同数据规模的点云配准对比

点云 数量	平均配准时间(s)		
	传统 ICP	Go-ICP	本文算法
4 749	4.681	14.596	1.318
8 915	9.572	21.715	2.524
13 853	24.627	50.151	7.825
39 616	67.957	107.207	15.719
61 825	174.762	271.856	24.682

模型的误差进行估计,其中对应点通过 ICP 算法中两点间欧式距离最小的方法确定。本文针对纹理稀疏和纹理丰富两种环境,对 SIFT 算法、ICP 算法和本文改进 SIFT-ICP 算法进行比较,结果如图 4 和图 5 所示。

图 4 是在纹理稀疏的场景中对物体点云模型进行构建,SIFT 算法、ICP 算法以及改进 SIFT-ICP 算法的配准误差分别为 9.26 cm、3.98 cm 和 0.74 cm,改进 SIFT-ICP 算法相对于 SIFT 算法和 ICP 算法的配准精度分别提高了 92% 和 81.4%。图 5 是在纹理丰富的场景中对物体点云模型进行构建,三种算法的配准误差分别为 5.73 cm、4.18 cm 和 0.68 cm,改进 SIFT-ICP 算法相对于前两种算法的配准精度分别提高了 88.1% 和 83.7%。从误差对比中可知,改进 SIFT-ICP 算法在纹理稀疏和丰富的场景下对物体点云模型的构建精度都远高于 SIFT 和 ICP 算法。

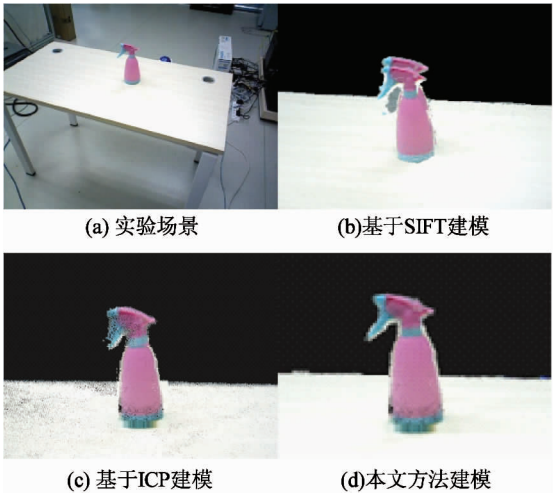


图4 纹理稀疏场景中物体建模对比

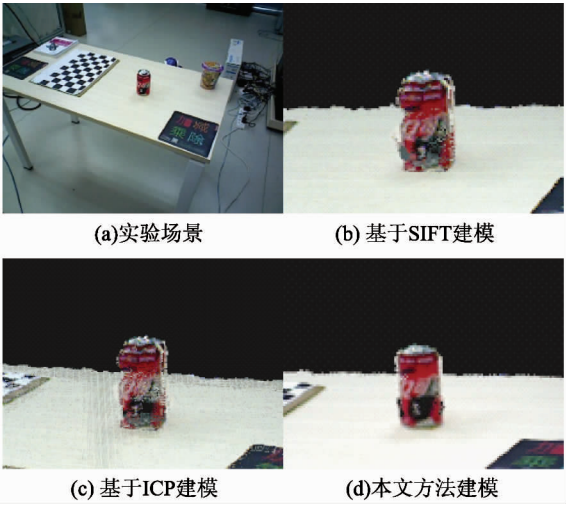
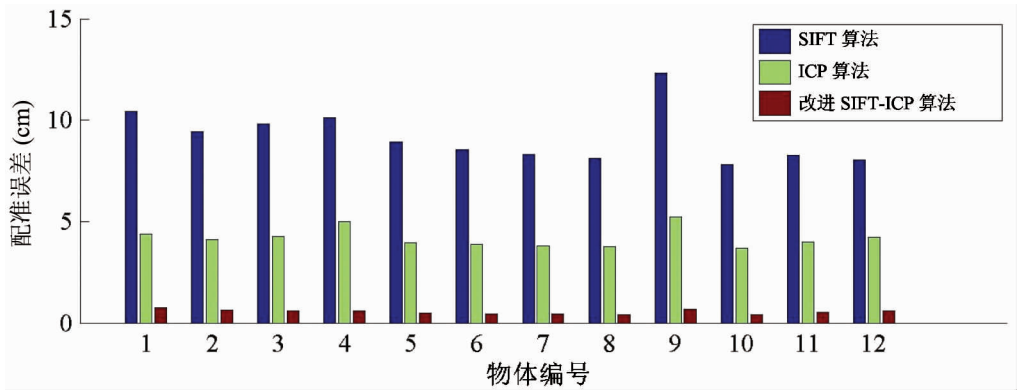


图 5 纹理丰富场景中物体建模对比

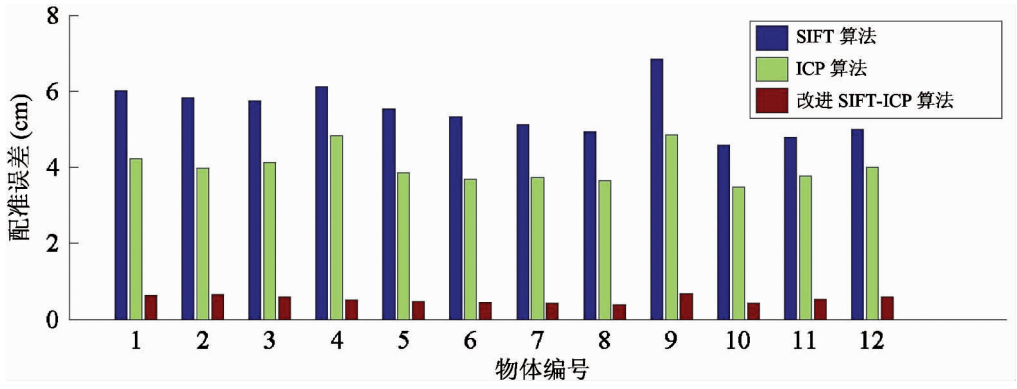
为验证本算法的可靠性,在纹理丰富和纹理稀疏的场景中分别对 12 种已建模型的物体,采用 SIFT 算法、ICP 算法和改进 SIFT-ICP 算法计算其相邻帧的平均配准误差,结果如图 6 所示。对比实验结果显示,在纹理稀疏和纹理稠密这两种场景下,本文算法的误差都远小于 SIFT 算法和 ICP 算法,同一种物体在两种场景下的配准精度近似,误差均小于 8 mm,可证本文算法的可靠性。

4 结 论

本文采用改进SIFT-ICP算法对目标物体进行



(a) 纹理稀疏场景下物体模型误差对比



(b) 纹理丰富场景下物体模型误差对比

图 6 不同纹理场景下物体模型误差对比

三维点云模型构建,实现了对 12 种不同物体的建模,实验结果验证了本文设计的算法的可行性和鲁棒性。此外,针对纹理程度不同的场景,将本文的改进 SIFT-ICP 算法与 SIFT 算法、ICP 算法进行比较,对比结果表明本文提出的算法的点云配准误差更

小,大幅度提高了物体点云模型的精度。

参考文献

[1] McCormac J, Handa A, Davison A, et al. SemanticFu-sion: dense 3D semantic mapping with convolutional neu-ral networks[C]. In: IEEE International Conference on

- Robotics and Automation, Singapore, 2017. 4628-4635
- [2] Zeng A, Yu K T, Song S, et al. Multi-view self-supervised deep learning for 6D pose estimation in the Amazon picking challenge[C]. In: IEEE International Conference on Robotics and Automation, Singapore, 2017. 1386-1383
- [3] Dias A S, Brites C, Ascenso J, et al. SIFT-based homographies for efficient multiview distributed visual sensing[J]. *IEEE Sensors Journal*, 2015, 15(5):2643-2656
- [4] Flitton G. Object recognition using 3D SIFT in complex CT volumes[C]. In: Proceedings of the British Machine Vision Conference, Aberystwyth, UK, 2010. 1-12
- [5] Ni D, Chui Y P, Qu Y, et al. Reconstruction of volumetric ultrasound panorama based on improved 3D SIFT[J]. *Computerized Medical Imaging & Graphics the Official Journal of the Computerized Medical Imaging Society*, 2009, 33(7):559-566
- [6] Zeng A, Song S, Nießner M, et al. 3DMatch: learning local geometric descriptors from RGB-D reconstructions[C]. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017. 199-208
- [7] Ran M, Xiao S, Zhou X, et al. Asphalt pavement texture 3D reconstruction based on binocular vision system with SIFT algorithm[C]. In: International Conference on Smart Grid and Electrical Automation, Changsha, China, 2017. 213-218
- [8] Whelan T, Leutenegger S, Moreno R S, et al. ElasticFusion: dense SLAM without a pose graph[C]. In: Robotics: Science and Systems, Roma, Italy, 2015. 413-421
- [9] Lin S, Wen C. Preprocessing and modeling for visual-based 3D indoor scene reconstruction[C]. In: International Conference on Computer Science & Education, Colombo, Sri Lanka, 2013. 1324-1328
- [10] Lowe D G. Distinctive image features from scale-invariant key points[J]. *International Journal of Computer Vision*, 2004, 60(2):91-110
- [11] Fischler M A. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Readings in Computer Vision*, 1987, 24(6):726-740
- [12] Besl P J, McKay N D. A method for registration of 3-D shapes[J]. *IEEE Computer Society*, 1992:239-256
- [13] Yuan C, Yu X, Luo Z. 3D point cloud matching based on principal component analysis and iterative closest point algorithm[C]. In: International Conference on Audio, Language and Image Processing, Shanghai, China, 2017. 404-408
- [14] Yang J, Li H, Campbell D, et al. Go-ICP: a globally optimal solution to 3D ICP point-set registration[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2016, 38(11):2241-2254

Object point cloud modeling method based on improved SIFT-ICP algorithm

Yu Hao^{***}, Du Huajun^{****}, Cai Yinghao^{*}, Lu Tao^{*}, Wang Rui^{*}, Wang Shuo^{***}

(^{*} State Key Laboratory of Management and Control for Complex System, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)

(^{**} University of Chinese Academy of Sciences, Beijing 100049)

(^{***} Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031)

(^{****} Beijing Aerospace Automatic Control Institute, Beijing 100854)

(^{*****} National Key Laboratory of Science and Technology on Aerospace Intelligent Control, Beijing 100854)

Abstract

To realize the recognition and pose estimation of arbitrary objects, a three-dimensional model of the object is needed, and an efficient and accurate 3D point cloud model building method is proposed. First, the multi view data of the object is obtained by the RGB-D sensor, and the relative pose of the sensor at various angles are calculated by using scale invariant feature transform (SIFT) feature point matching and the improved iterative closest point (ICP) algorithm to generate the complete point cloud of the scene where the target object is located, and the 3D point cloud model of the object is further obtained by the object segmentation and the point cloud after processing. The experiment results show that the point cloud model is clear and undistorted, retaining the complete feature information of the surface of the object.

Key words: point cloud model, scale invariant feature transform (SIFT), iterative closest point (ICP), object segmentation