



面向物流分拣的多立体摄像头物体操作系统

张泽坤, 唐 冰, 陈小平*

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

(*通信作者电子邮箱 xpchen@ustc.edu.cn)

摘 要: 为满足物流分拣的低成本和实时性要求, 提出了基于多个立体摄像头的系统获取典型物体的完整立体信息的方法, 并结合机械臂搭建了实验硬件平台。实验采用了 2 个微软 Kinect 摄像头在水平面上实现了约 3 mm 精度的物体定位, 根据物体的立体信息建立立体模型, 并计算了物体的取向、尺寸、含有的平面等多个可用于物体操作的立体特征, 计算速率约为 1 s/帧。根据这些信息, 使用了机械臂成功进行了连续 100 次抓取。实验结果表明, 这套方法和平台无需离线学习即可以实时提取多种尺寸和形状的物体的立体特征, 机械臂可以基于此进行精度较高的物体操作。

关键词: 物流分拣; 立体视觉; 特征提取; 智能抓取; 摄像头校准

中图分类号: TP242.6 **文献标志码:** A

Object manipulation system with multiple stereo cameras for logistics applications

ZHANG Zekun, TANG Bing, CHEN Xiaoping*

(School of Computer Science and Technology, University of Science and Technology of China, Hefei Anhui 230027, China)

Abstract: To meet the low cost and real-time requirements of logistics sorting, a systematic method was proposed to extract complete stereo information of typical objects by using multiple stereo cameras. Combining the cameras with an arm and other hardware, a validation and experiment platform was constructed to test the performance of this method. Two Microsoft Kinect cameras were used to measure the locations of objects in horizontal plane with accuracy of 3 millimeters. The stereo features and models of objects were calculated from the complete stereo information at processing rate of about 1 second per frame. Utilizing these features, the arm continuously picked 100 objects without failure. The experimental results demonstrate that the proposed method can extract the stereo features of objects with various sizes and shapes in real-time without off-line training, and based on which the arm can operate on objects with high accuracy.

Key words: logistics; stereo vision; feature extraction; intelligent picking; camera calibration

0 引言

飞速发展的电子商务对物流行业提出了越来越高的要求, 传统的人工分拣方式已经逐渐无法满足需求。为了节省时间和空间, 每个物流节点需要快速将物件按照大小、形状、目的地进行分拣和装载。机械臂技术的不断发展使得机器人可以以相当高的精度和智能度进行物体操纵, 可以帮助人类分拣物件。典型的机器人参与的物流分拣场景类似于工厂装配流水线。机械臂和传感器安装在平台上, 位置相对固定; 待分拣的物体用传送装置运输到平台当中, 系统对它们进行识别和定位, 然后由机械臂将其分拣到不同的运输设备上。这种场景下, 需要能够实时对平面上所有物体的尺寸、形状、表面特征、适合抓取的位置等进行识别, 使得机械臂能够选取合理的模式进行操作。因为待分拣物件的种类无法预先确定, 识别系统需要较强的泛化性能。同时, 分拣系统的成本需要控制在合理范围内, 而识别与定位精度只需要满足操作要求即可。对于典型大小的物件, 定位精度约为 5 mm 以下就可

以保证操作成功。

普通的彩色摄像头只能得到物体在成像平面上的投影信息, 无法获得物体的真实形状和尺寸。单个立体摄像头可以得到物体在摄像头视角可见的立体结构, 但是不能看到物体的背面, 也不能处理多个物体相互遮挡的场景。这些信息对于物流分拣尤为重要。工业界已经发展出了成熟的使用激光扫描对物体的立体形状进行重建的技术^[1], 并已形成了实用的产品。这些产品可以实现很高的精度, 但是往往需要昂贵的设备和复杂的操作, 无法满足物流分拣场景的实时性和低成本要求。从立体摄像头拍摄的多张图像中重建整个立体场景的方法广泛应用于机器人导航当中^[2]; 使用结合了整体和局部特征的描述子, 可以使用二维摄像头的数据进行机器人定位^[3]。这类机器人导航中使用的方法同样可以用于物体识别, 但是这类方法需要较高的计算量, 且在小尺度下表现不佳。使用多个方向的大量彩色摄像头和立体摄像头得到的互相同步的图像中可以十分精确地重构出场景的三维立体信息^[4], 但是这种方法需要非常复杂的硬件和算法, 一般适用

收稿日期: 2018-02-05; 修回日期: 2018-03-28; 录用日期: 2018-04-16。

基金项目: 国家自然科学基金委-深圳联合基金资助项目(U1613216); 国家自然科学基金资助项目(61573333)。

作者简介: 张泽坤(1994—), 男, 安徽安庆人, 硕士研究生, 主要研究方向: 智能机器人、计算机视觉、机器学习; 唐冰(1991—), 男, 安徽阜阳人, 硕士研究生, 主要研究方向: 智能机器人、多功能手爪、机器人硬件控制; 陈小平(1955—), 男, 北京人, 教授, 博士生导师, 博士, 主要研究方向: 人工智能逻辑、多主体系统、智能机器人。



于训练数据的收集的场合。从人类操作物体的视频中学习物体和工具的物理性质可以赋予机械臂以合理的方式操作工具的能力^[5],但这类方法需要大量数据进行离线训练,而且最适用于人类的工具操作方法并不一定适用于所有类型的机械臂。

使用多种机器学习方法可以完成对物体的识别与定位^[6-7]。随着基于深度学习的物体识别技术的发展,基于卷积神经网络(Convolutional Neural Network, CNN)的物体抓取方法也在不断提出。桂林电子科技大学的多位研究者利用预先训练好的CNN识别象棋上的文字,结合边缘检测算法,实现了对象棋的位置和旋转角度的检测^[8],并使用机械臂进行了实验。中国科学院自动化研究所的研究者使用基于区域候选的卷积神经网络(Region Convolutional Neural Network, R-CNN)的物体检测和识别算法^[9],从物体上方摄像头的图像出发,定位物体的最佳抓取位置,并使用机械臂进行了实验。这类方法的识别效果较好,如果结合深度学习相关硬件和软件也可以实现较高的运行速度;但是它们都依赖于预先标注好的数据集,需要预先离线训练,如果遇到未知物体,算法可能会失效。

使用机械臂进行成功和不成功的物体抓取的视觉信息进行深度学习训练,可以得到一个端对端的从机械臂和待抓取物体的图像到抓取动作成功率的卷积神经网络,这个网络可以被用于对机械臂进行一系列的连续控制,以完成高成功率的物体抓取^[10]。使用机械臂进行大量的物体抓取尝试可以训练深度卷积神经网络用于从摄像头得到的物体图像中检测适合抓取的点和方向的信息^[11]。这类基于深度学习的方法不依赖于人类经验或人工标注的数据集,可以通过长时间运行采集海量数据,在这些数据上训练得到很高的成功率。但是它们训练得到的神经网络一般与硬件设置以及训练使用的物体相关,当机械臂结构和外观或场景发生及较大的改变,或者出现了与训练物体相差非常大的新物体时,可能需要重新训练。另外这类方法需要机械臂进行大量成功和失败的抓取尝试,而失败的尝试可能伴随着机械臂与其他硬件的碰撞,收集这些数据的过程中可能会对机械臂和其他硬件造成损伤。

由于物体的立体全貌需要由多张图像综合得到,对这些图像进行变换和拼合是立体视觉的核心任务。基于特征点匹配的方法通用性较强,但是需要大量的计算,也不能处理特征较少的物体。物流分拣场景中摄像头和机械臂的安装位置相对固定,可以首先对安装好的摄像头和机械臂进行校准,使得它们可以在同一坐标系中协同工作,之后对于任意物体都可以使用同一套校准参数而不需要每次进行校准。使用多个摄像头从不同方向对物体进行拍摄可以得到物体全貌,并在一定程度上解决物体相互遮挡的问题。从物体的立体全貌可以直接计算有用的特征,如物体尺寸、空间取向、平面法向量等。使用这些信息,结合机械臂的物理结构和机械指标,可以设计出最合适的操作模式。本文设计了一种可以得到物体的立体全貌信息,并使用机械臂进行操作的物流分拣系统。使用二维码标签可以建立多个立体摄像头内部坐标系到平台坐标系的变换关系。将摄像头的三维立体点云图像变换到平台坐标系中后,直接进行拼合即可得到完整的物体立体模型。从这些模型出发,本文计算了尺寸、中心点、取向等立体特征,并提取了物体表面含有的平面。利用这些特征,结合一定的人类

进行物体操作的经验,控制机械臂采用了不同的操作策略进行物体操作,模拟了物流分拣场景中几个典型任务,实验的精度和实时性都达到了应用要求。本文将详细描述摄像头的设置、校准与相关算法,以及手臂的安装和控制方法,讨论在此平台上进行实验的结果并分析误差,最后给出总体结论和下一步工作方向。

1 多立体摄像头物体操作系统

1.1 立体摄像头模型与内部参数

微软公司于2010年为Xbox 360主机发布了第一款立体摄像头Kinect,并很快因其低廉的价格和较高的精度成为机器人科研界广泛应用的立体摄像头。为获取物体的立体信息,Kinect摄像头在普通的彩色(Red Green Blue, RGB)传感器之外还安装了红外线(Infra-Red, IR)投射和接收装置,从而重构出物体表面的立体结构^[12]。大部分技术成熟的立体摄像头都应用了类似的技术。采用一块印有形状和尺寸已知图案的平板,RGB传感器和IR传感器可以在它们接收到的图像的特征点之间建立对应关系,从而完成两个像素空间的对齐^[13],对齐参数称为摄像头的内部参数。在驱动软件中,可以根据内部参数从彩色图像和具有深度信息的IR图案建立三维点云(Point Cloud),相比普通的彩色图像,点云中的每个像素点还具有该点在摄像头坐标系中的位置信息(RGBXYZ)。典型的深度摄像头如图1所示,图1(a)和(b)中分别标出了微软Kinect和后来开发的Kinect 2.0摄像头的主要结构和对应的摄像头坐标系。在物流场景中,物体和摄像头一般置于室内的均匀光照条件下,摄像头和物体之间没有灰尘或水雾的干扰,且物体表面一般不会非常光滑以至于发生镜面反射。这种条件下,商用的立体摄像头可以以较高的精度得到物体的立体点云信息。

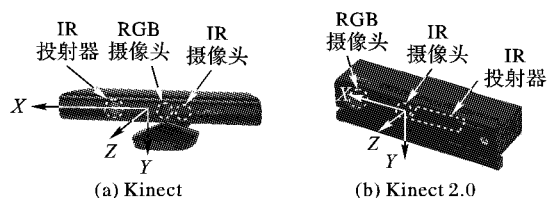


图1 典型的立体摄像头示意图

Fig. 1 Illustration of typical stereo cameras

1.2 摄像头坐标系与外部坐标系变换关系的建立

立体摄像头给出的点云中点的坐标都基于摄像头坐标系,应用中为了能让多个摄像头同时工作,需要将不同摄像头坐标系变换到同一个外部坐标系中。此变换关系可以通过识别位置已知的特征点来建立。在外部坐标系 C 中选取一系列点,其坐标分别为 $\{P_1, P_2, \dots, P_N\}$ 。在这些点处放置摄像头容易识别的图案,根据摄像头识别的结果得到它们在某个摄像头坐标系 C_k 中的位置为 $\{Q_1, Q_2, \dots, Q_N\}$ 。通过最小化误差得到 C_k 到 C 的变换矩阵 T_k :

$$\arg \min_{T_k} \sum_{i=1}^N E(P_i, T_k Q_i) \quad (1)$$

其中:差函数 E 用于计算点的真实坐标 P_i 与摄像头识别到的坐标经过变换矩阵变换后得到的坐标 $T_k Q_i$ 间的误差,为便于优化误差函数一般取为三维欧几里得距离。为了减小误差,选取的点需要有足够的数量,且合理分布在一定范围内。变



换关系建立后,只要摄像头与 C 之间的相对位置没有发生变化,得到的变换矩阵就可以一直使用。多个摄像头将各自的点云存入队列中,从中选取时间戳最接近的点云为 PC_k , 分别使用变换矩阵变换,合在一起即可得到物体的完整点云 PC :

$$PC = \sum_k T_k * PC_k \quad (2)$$

其中: $*$ 表示根据变换矩阵对点云中的每个点的三维坐标进行变换, \sum 表示对点云进行合并。该点云经过处理后即可用于提取物体的立体特征。实际的摄像头设置中,需要合理安排摄像头位置,以尽可能在较小的误差下得到物体的全貌。同一型号多个摄像头发出的红外线之间可能会互相干扰,此时需要使用不同种类的摄像头加以解决。

1.3 物体立体特征的提取

本文使用点云库(Point Cloud Library, PCL)处理点云信息^[14]。从合并得到的点云 PC 可以计算出每个物体的中心点位置、尺寸、取向、含有的平面等立体特征,这些特征可被用于后续的物体操作过程当中。

1.3.1 算法原理

物体的取向采用主成分分析(Principal Component Analysis, PCA)方法进行计算。对一含有 N 个点 $\{x_i, y_i, z_i\}$ 的点云,计算其对应的协方差矩阵 Σ :

$$\Sigma = \begin{bmatrix} \text{cov}(X, X) & \text{cov}(X, Y) & \text{cov}(X, Z) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) & \text{cov}(Y, Z) \\ \text{cov}(Z, X) & \text{cov}(Z, Y) & \text{cov}(Z, Z) \end{bmatrix} \quad (3)$$

其中协方差项 $\text{cov}(X, Y)$ 为点云在 X 和 Y 两个维度上的协方差:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (4)$$

其他各项依此类推。计算 Σ 的三个特征向量,即代表了点云的三个主成分的方向,这三个方向代表了物体在三维空间中的主要取向。

采用随机抽样一致(RANdom Sample Consensus, RANSAC)算法可以从点云中提取平面。使用三维平面模型对点云进行拟合,在设置的误差范围内,采用迭代的方法找到含有最多点的平面模型为:

$$ax + by + cz + d = 0 \quad (5)$$

然后将此平面内的点从点云中移除,对剩余点云继续采用该方法提取平面,直到得到的平面中的点小于一定的下限为止。这样就可以从点云中提取出所有大小符合要求的平面对应的平面方程及其内部的点。

1.3.2 算法流程

1) 点云中除了感兴趣的物体外,还含有其他无关物体的点云。根据硬件系统的结构和尺寸,设置物体可能出现的三维坐标范围,此范围之外的点云属于无关物体,将无关物体裁剪,得到 PC_{crop} , 该部分点云只含有待操作的物体。

2) 不同的摄像头分辨率不同,且同一个物体的不同位置在点云中点的密度不同,直接使用 PC_{crop} 进行后续计算会导致立体特征发生偏差。为解决这一问题,对 PC_{crop} 以采样率 S 进行三维重采样得到点云 PC_{sample} , 该点云中相邻点间的距离基本为 S 。在点云中如有某个平面含有 N 个点,则该平面的面积近似为 $A = S^2 N$ 。

3) 对 PC_{sample} 中的点根据三维欧几里得距离进行聚类。摄像头给出的点云中存在一些并不属于任何物体的噪点,将

只含有很少点的聚类去掉即可去掉这些噪点。剩余的各个点聚类即是代表每个独立物体的点云 $\{PC_{\text{object1}}, PC_{\text{object2}}, PC_{\text{object3}}, \dots\}$ 。

4) 对任一个物体点云 PC_{objecti} :

a) 对所有点的三维坐标计算平均值,得到物体的中心位置 Centroid_i 。

b) 对于点云的所有点的三维坐标值计算协方差矩阵 Σ_i , 该矩阵的三个特征向量按照对应特征值的绝对值的大小排列 $V_i = \{V_i^1, V_i^2, V_i^3\}$ 即为该点云的相互垂直的三个主成分的方向,同时也代表了该物体的空间取向。

c) 将 PC_{objecti} 转换到以 V_i 为三个基向量的空间中,得到 $PC_{\text{objecti}}^{\text{PCA}}$, 该点云在各个坐标轴上的投影长度 $\{L_i^1, L_i^2, L_i^3\}$ 即为该物体在三个空间取向方向的最大尺寸。

d) 根据 $\text{Centroid}_i, V_i, \{L_i^1, L_i^2, L_i^3\}$ 计算物体的最小有方向包围盒 OBB_i (Orientated Bounding Box)。对每个物体的点云进行平面提取得到多个平面对应的点云 $\{Plane_i^1, Plane_i^2, \dots\}$, 计算得到每个平面的中心点位置坐标 $\{\text{Center}_i^1, \text{Center}_i^2, \dots\}$, 平面法向量 $\{n_i^1, n_i^2, \dots\}$, 并根据平面上点的数目计算平面的面积 $\{A_i^1, A_i^2, \dots\}$ 。

上述算法使用点云中点之间的欧几里得距离分割不同的物体,可以适用于物体间有一定程度的相互遮挡,但是不存在相互接触的情况。在物流分拣场景中,这可以通过控制单位时间内通过该系统的待分拣物体的数量来实现。

1.4 机械臂的控制

为了让机械臂具有智能物体操作的能力,本文实验中使用 MoveIt!^[15] 对机械臂的运动进行规划。MoveIt! 是 ROS 中用于控制机械臂的软件模块,使用时,根据机械臂各个部件的尺寸和形状建立物理模型,即可建立各个关节的角度与手臂末端位置间的对应关系。进行运动规划时,通过控制每个关节的转动角度将手臂末端移动到某个特定位置并调整为某个姿态。当手臂的自由度足够多时,可以在一定范围内将手臂运动到直角坐标系中任意位置 (x, y, z) 和任意欧拉角 $(pitch, yaw, roll)$ 代表的方向。进行运动规划时,考虑手臂和整个平台上其他设备发生碰撞的可能性,需要将平台的模型加入整体模型当中。得到物体的相关立体特征后,结合机械臂部件的尺寸判断是否可以进行操作,如果可以操作,根据机械臂的规划结果和形状判断最适合操作的方式。手臂控制使用了手臂坐标系 C_{arm} , 采用变换矩阵 T_{arm} 可将 PC 变换到手臂坐标系当中,手臂就可以对于 C 中的位置进行操作

$$PC_{\text{arm}} = T_{\text{arm}} * PC \quad (6)$$

2 实验平台的搭建与运行

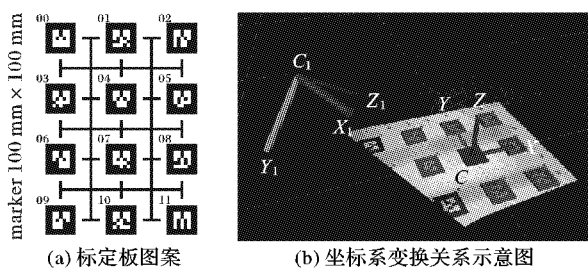
2.1 多立体摄像头系统

本文在平台的不同位置安装两个立体摄像头,分别检测物体在两个相对方向的立体信息。由于 Kinect 采用结构光(Structured Light, SL)信息重构表面立体结构,如果将两台 Kinect 指向同一个表面,它们发出的结构光信息会互相干扰,使得重构失败。为了解决这一问题,本文使用的摄像头除一个 Kinect 摄像头外,另一个摄像头为 Kinect 2.0。Kinect 2.0 摄像头使用了不同的飞行时间(Time-of-Flight, TOF)技术^[16],可以和 Kinect 同时使用。安装摄像头之前,使用一个棋盘图案的标定板对它们的内部参数进行标定。本平台采用



了机器人操作系统(Robot Operating System, ROS)^[15]提供的软件进行摄像头驱动和平台控制,摄像头独立供电,并以USB 3.0连接至计算机。每个摄像头使用一个三自由度云台固定在操作平台的框架上,调整云台的各个关节,使得用于放置待操作物体的平面位于视野中心位置。后续的实验说明了在一定范围内两个摄像头已经足够提供物体的三维全貌。

本文使用虚拟和增强现实库(A Library for Virtual and Augmented Reality, ALVAR)提供的多标签自动校准法^[18]来建立每个摄像头的内部坐标系到统一的外部坐标系之间的变换关系。实验平台整体使用的右手坐标系 C 的原点定义在放置待操作物体的平面上, Z 轴指向上方,与平面垂直。在平面上放置一块印有二维码标签的标定板,选取一个标签作为原点位置。图2(a)是实验中使用的标定板上印刷的标签图案,坐标系原点定义在04号标签的正中。典型的变换关系建立过程如图2(b)所示,图中标记出了摄像头内部坐标系 C_1 以及外部坐标系 C 。



(a) 标定板图案 (b) 坐标系变换关系示意图

图2 立体摄像头变换关系的建立

Fig. 2 Transformation from stereo camera coordinate system to platform coordinate system

2.2 机械臂系统

本文使用了Kinova Robotics公司开发的六自由度手臂MICO²,如图3所示。该手臂由六个转动关节(J1~J6)连接的七个部件(base, shoulder, arm, forearm, wrist1, wrist2, hand)构成,第一个部件固定在基座上,关节的电机可以承受大约2.4 kg的末端重量。最末端的部件上安装有两个可独立开合的手指(finger1, finger2),这两个手指可以对尺寸或一部分的尺寸在0.5~12 cm范围内的物体进行有效的抓取操作。手臂每个关节具有角度和力矩传感器,可以接收物体操作的反馈信息。手臂使用独立供电,并以USB 2.0连接至计算机。控制手臂时采用的手臂坐标系 C_{arm} 在图3中标出,其的原点位于手臂基座的正中。将手臂使用夹具固定安装在平台的一侧,基座略低于平面。使手臂坐标轴 C_{arm} 与平面上的坐标系 C 的坐标轴相互平行,通过简单测量就可以建立 C 到 C_{arm} 的变换关系矩阵 T_{arm} 。

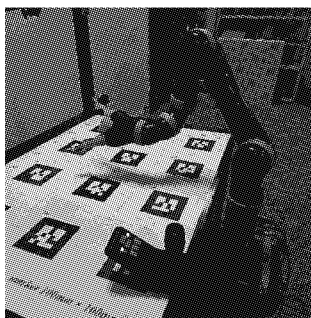


图3 机械臂示意图

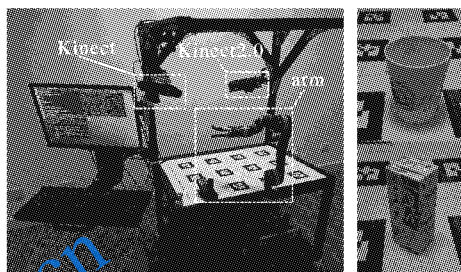
Fig. 3 Illustration of the arm

根据机械臂的物理模型,可以通过每个关节的角度计算

出机械臂的状态;从状态中可以得到机械臂末端的位置,该位置信息可以用于评估机械臂进行特定操作的精度;而物理模型本身的精度可以通过一些外部测量方法进行评估。

3 实验结果与分析

本文通过实验,系统测量了多摄像头视觉系统给出的物体坐标测量值与实际坐标值之间的误差;然后选取了一些常见的形状、尺寸各异的物体进行了特征提取实验;之后对机械臂在实验平台上的可操作的空间范围进行了测试;最后根据视觉系统提取的特征,使用手臂完成了高精度智能抓取。实验平台整体设置如图4(a)所示,主要包括了可移动的整体框架、安装在框架上的两个摄像头、用于放置物体的平面、安装在平面下方的供电系统、在平面一侧安装的手臂、进行计算和控制的计算机。



(a) 实验平台示意图 (b) 实验所用物体

图4 实验平台及实验用物体

Fig. 4 Experiment platform and typical objects

3.1 视觉系统物体定位误差分析

立体摄像头给出的点位置信息往往存在一定的误差,且这部分误差常常随着点远离摄像头及偏离视野中心而增大^[19]。为了测量平台不同位置的位置误差,本文首先比较了视觉系统给出的物体位置测量值与物体实际位置之间的误差。实验选取了两个具有代表性的物体进行位置测量实验:物体一为空心锥台形的水杯,如图4(b)上方所示;物体二为实心棱柱形的饮料,如图4(b)下方所示。这两个物体都具有轴对称性,可以将其稳定放置于平面上,并通过简单的测量,使其中心点在平面上的投影点位于任意指定位置。完成摆放后,使用自动校准后的视觉系统分别测量它们的位置。实验选取了6个有代表性的位置,每个位置进行了连续200次测量。对于每个放置位置 (x_0, y_0) 的 X, Y 坐标的测量结果 (x_i, y_i) ($i \in \{1, 2, \dots, 200\}$),分别计算它们的均值 (\bar{x}, \bar{y}) ,之后计算每个测量值与均值的二维欧几里得距离值的方均根:

$$\sigma = \sqrt{\frac{1}{200} \sum_{i=1}^{200} ((x_i - \bar{x})^2 + (y_i - \bar{y})^2)} \quad (7)$$

σ 即为二维平面内测量值的标准差。每个位置对应的 (\bar{x}, \bar{y}) 相对于真实位置的偏差 $(\Delta x, \Delta y)$ 以及 σ 如图5所示,其中用 \times 标记真实位置 (x_0, y_0) ,用实心圆点标记测量值的均值 (\bar{x}, \bar{y}) 。可以看到对于不同物体的不同位置,测量误差基本是一致的。这可能是由于平面的不同位置虽然距离两个摄像头的距离不同,但是距离其中一个摄像头较远时则距离另一个摄像头较近,使得两个摄像头的误差发生了一定程度的互相抵消。最终表现出的较为一致的偏差是摄像头校准的系统误差所致,



这部分误差随物体和位置的改变变化很小。每次完成自动校准过程后,使用形状规则的物体测量这部分系统误差并在视觉系统中添加修正项即可完成修正,这一修正对于后续测量的其他物体同样适用。使用测量误差的均值 $(-0.010, -0.009)$ 进行修正,修正后共2400次测量的结果在 XY 平面的位置测量值与真实值之间的距离平均误差为 $\bar{\varepsilon} = 0.003\text{ m}$,此误差已经可以满足绝大部分物流操作的精度要求。

分析某项测量值的分布情况可以给出该测量误差来源的信息,本文分析了物体在某个点的200个位置测量值的分布情况。根据实验中物体二放置在 $(-0.1, 0.2)\text{ m}$ 处的测量值绘制二维直方图如图6所示,偏离均值较远的几个测量结果并未显示在图中。可以看到测量值的分布基本成旋转对称模式,偏差在 $\sigma, 2\sigma, 3\sigma$ 内的比例分别为64.5%、96.0%、100.0%,基本符合二维正态分布的特征,其他位置的测量结果也具有类似的特征。这说明误差来源基本为随机测量误差,可以通过一些统计学方法减小。实验中为提高精度,在时间可行的情况下,可以采用多次测量取均值的方法。因为物体在 Z 方向的真实位置不容易测量,且定位中该方向的测量值的重要性低于 X - Y 方向,上述误差分析没有将 Z 方向的坐标测量值计入其中。单独分析物体在 Z 方向的测量值的分布,同样符合正态分布的特征。

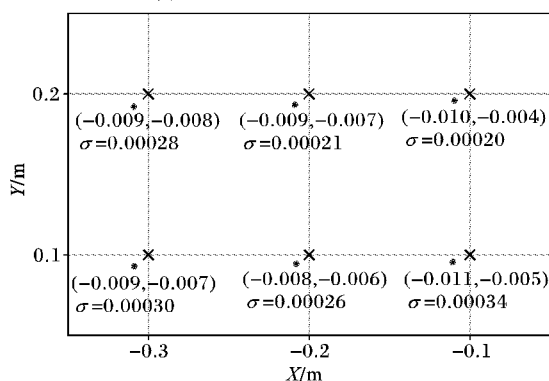
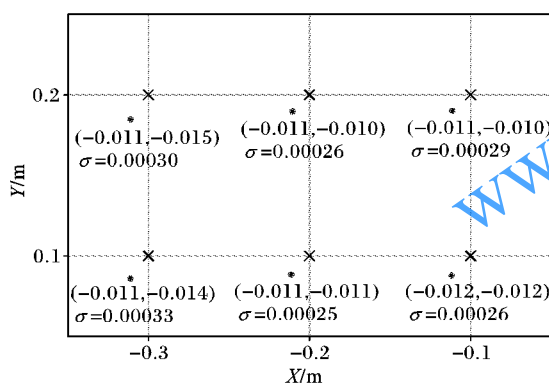


图5 物体在 XY 平面中位置的测量结果

Fig. 5 XY -position measurements of two objects

3.2 常见物体的特征提取结果

为测试多摄像头立体视觉系统在不同条件下的表现,本文选取了一些大小、形状、取向各不相同的物体放置在平面上,如图7左侧所示。使用立体视觉系统计算它们的立体特征,在图形界面中绘制物体的特征向量 V_i 与最小包围盒

OBB_i ,并和物体的点云一起显示,如图7右侧所示。图中标出了操作平台的坐标系 C ,并用白色箭头绘制计算得到的特征向量,箭头起点位于物体中心点,长度为该特征向量对应特征值的绝对值大小。 OBB_i 使用实线标出,并用白色字体在特征向量的箭头末端标注了各个方向的尺寸,单位为 m 。本操作平台使用的计算机安装有一块 Intel Core i7 7700K 处理器,从摄像头原始数据完成一次完整计算并进行显示的时间约为1 s。

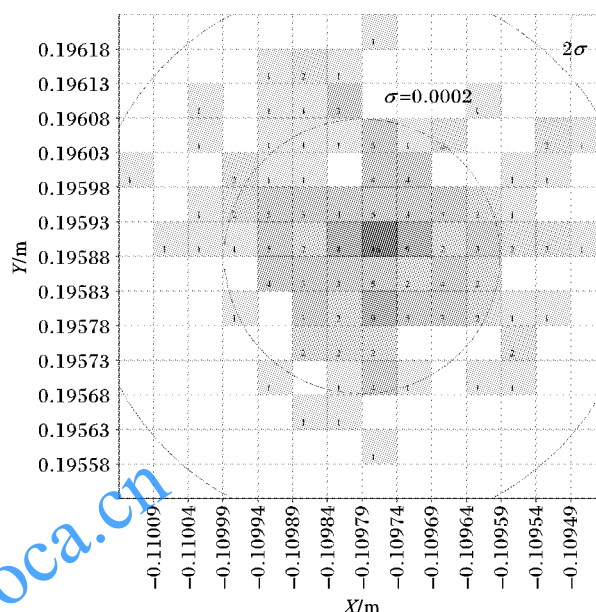
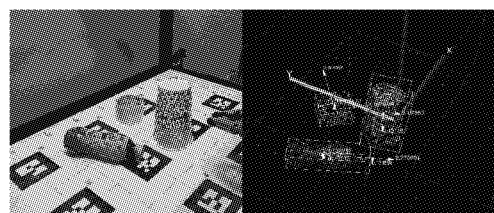
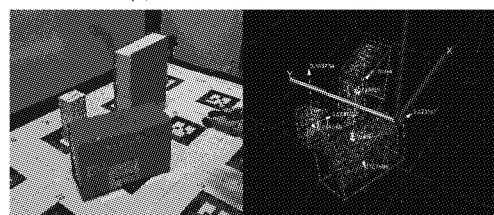


图6 测量值在 XY 平面中的分布

Fig. 6 Distribution of measurements in XY plane



(a) 旋转对称物体的立体特征



(b) 棱柱形物体的立体特征



(c) 不规则物体的立体特征

图7 不同形状物体的立体特征

Fig. 7 Stereo features of different types of objects

图7左侧的图片为普通摄像机拍摄,该摄像机的位置和视角与两个立体摄像头的其中一个基本一致。可以看到在图



7(b)的左侧图像中,物体之间存在一些互相遮挡,物体的图像之间有一定的重叠,单个摄像头无法得到物体的全貌。但是一个摄像头无法看到的部分可以被另一个摄像头捕捉,从而基本看到物体的全貌。旋转对称物体的特征如图7(a)所示,计算得到的物体中心位于旋转轴上,且三个特征向量其中一个与物体的对称轴平行,另外两个对应的特征值大小基本相等。棱柱形物体的特征如图7(b)所示,物体的中心点位于棱柱中心,三个特征向量方向与各条棱平行,包围盒尺寸基本与其各条边长一致,由于存在一定的校准误差和噪点,该尺寸略大于实际尺寸。两个不规则物体的特征如图7(c)所示,其特征向量基本符合人类直觉认为的物体取向,包围盒尺寸与物体尺寸基本一致。实验结果表明,本文使用的算法可以有效地提取物体的立体特征,特征与物体的真实三维形状一致,且符合人类对物体的认知。人类操作物体时也参考了这些特征,如抓物体时选取最细的位置,根据物体的长轴判断物体姿态等,平台安装的机械臂同样可以利用这些特征进行更加有效和智能的物体操作。

与基于深度学习的方法相比,本文的算法不需要预先标注的数据集进行训练。本文的物体定位与特征提取基于点云中点的坐标信息;与文献[8]中的方法相比,不需要物体表面具有任何便于识别的颜色和纹理,只需要物体表面能够被立体摄像头精确成像即可;与文献[9]中的方法相比,本文的方法使用立体视觉分割物体,不需要物体和背景的颜色具有较大的区别。另外,本文提取的物体取向信息是三维的,而不是限制在水平面内,也即抓取点具有6个自由度,分别是3个空间位置和3个取向角度,这为后续的抓取提供了更大的灵活性。但是由于立体摄像头的精度低于传统的二维相机,本文在物体定位精度方面略有下降。处理速度方面,本文的方法与文献[7]中不使用GPU加速神经网络计算的情况下接近,都为1 s/帧左右。

3.3 机械臂可达范围测试结果

本文使用了OptiTrack公司开发的多摄像头动作捕捉系统(Motion Capture System, MCS)精确测量手臂末端位置来评估机械臂运动到特定位置的精度。实验中在手臂最末端粘贴一个由3个动作捕捉系统可以定位的标记点构成的标记物,系统可以以亚毫米精度测量该标记物的三维位置坐标。控制手臂末端沿直线移动一小段距离,并计算动作捕捉系统中标记物移动的实际距离。实验发现手臂移动的精度在2 mm以下。此精度已经足够完成物流分拣场景中的物体操作。

为测试机械臂在实验平台上的可操作范围,本文选取了物体操作的一个典型手臂末端姿态,即手臂末节与水平面平行,在此姿态下测试了手臂末端可以精确到达的位置范围。实验在手臂坐标系中 $X = [-0.25, 0.25] \text{ m}$, $Y = [-0.65, -0.45] \text{ m}$, $Z = [0.2, 0.45] \text{ m}$ 的空间范围内,每个坐标轴方向以0.05 m为间隔均匀选取测试点,控制机械臂末端到达该点,并根据关节最终角度计算实际到达的位置。实验结果表明,机械臂末端在选取的姿态下可以在 $X = [-0.25, 0.25] \text{ m}$, $Y = [-0.6, -0.45] \text{ m}$, $Z = [0.2, 0.4] \text{ m}$ 范围内以低于0.005 m的位置误差到达目标点。通过调整机械臂与物

体放置平面的相对位置,可以让此可精确到达的范围与前述的摄像头可以精确定位的范围相互重合,此时可以让机械臂与摄像头系统以5 mm的精度协同运行。

3.4 机械进行物体操作实验结果

本文模拟物流分拣场景的实际需求,根据计算得到的物体的立体特征,根据机械臂的结构和人类进行物体操作的经验,采用不同的策略控制手臂进行抓取和其他操作。如果平面上有多个物体,优先操作距离机械臂最近的物体以避免可能发生的碰撞。一次操作完成后,因为操作过程中可能会碰到其他物体使得其位置和取向发生改变,所有立体特征需要重新计算。若物体的尺寸在可以被机械臂抓取的范围内,控制手臂前端垂直于物体最长轴,即最大的特征值对应的特征向量方向,平行于尺寸适于抓取的方向,抓取物体的中心点,如图8(a)、(b)、(c)所示。如果物体的质量分布均匀,这种方法可以实现最稳定的抓握。如果抓取方向不同,可能由于物体在该方向的尺寸太大无法抓握。若抓取点偏向物体一侧,则可能导致抓取点两侧质量分布不一致,使得抓取过程中物体发生旋转甚至掉落。对于具有较大平面的物体,控制手指触碰该平面的中心点,模拟物流中机械臂为物体贴标签或识别标签的情况,如图8(d)所示。请访问<http://t.cn/R8MIQWM>观看机械臂进行连续物体操作过程的视频。虽然计算得到的物体特征、手臂坐标系的变换、手臂末端位置和姿态的控制都存在一定的误差,在本文的实验条件下这些误差并不会导致物体操作的失败。本文的实验选取了各种不同的物体进行了100次抓取,未出现抓取失败的情况。

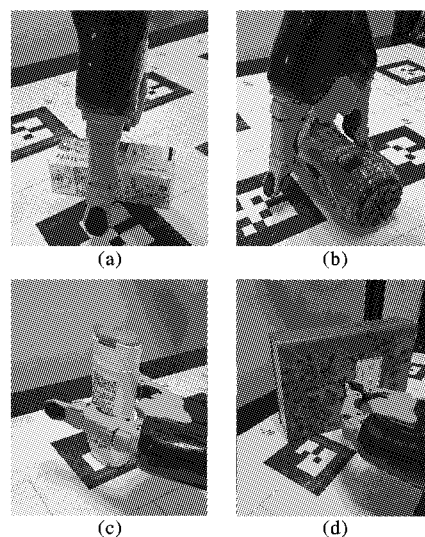


图8 针对不同特征物体的操作

Fig. 8 Manipulation strategies for different objects

4 结语

本文通过对多个深度摄像头和机械臂进行联合校准,建立各个摄像头坐标系和手臂坐标系到一个统一坐标系间的变换关系,提出了一套适用于物流分拣场景的物体立体特征提取方法和机械臂操作,并搭建了实验平台进行了实验验证和物体操作测试。提前校准的方法避免了对于每个新物体进行调整。从物体立体全貌计算的立体特征则代表了物体的真实



信息,提高了后续操作的成功率。本文选取了一些常见的典型物体进行实验,平台可以在可接受的误差范围内实时运行,机械臂根据视觉系统给出的信息连续进行了100次成功抓取。这套方法应用在物流行业中,具有成本较低、搭建简单、无需预先收集数据进行训练等优势。

本文的方法适用于尺度在分米级且形状较为规则的物体,尺寸过大或者过小的物体都会使定位精度下降。由于定位和特征提取都基于物体的整体点云,如果物体形状较为不规则,可能会出现计算出的抓取点不在物体上的情况,使得后续操作出现问题。在物流分拣场景中,大部分包裹都是尺寸适中的方形物体或是柱形物体,它们的尺寸和形状在本文提出方法的适用范围内。

本文方法对物体的尺寸要求较高,如果物体较大,很容易超出视觉系统的可见范围。为了让这套方法适用于更大尺度下的物流分拣应用场景,后续工作中可以适当增加立体摄像头的数目和它们之间的距离,以在更大范围内实现物体定位和特征提取。本文基于点云之间的距离分割物体,为了实现紧密贴合物体的分割,可以采用基于图像特征的视觉方法,这样可以处理物体间存在更大面积内互相遮挡的情况。为提高机械臂进行物体操作的智能度,可以使用测量得到的物体立体模型结合机械臂和手爪模型进行计算,从而实现更有效的操作;或者使用机器学习的方法,实现更加智能的物体操作。

参考文献 (References)

- [1] CURLESS B. From range scans to 3D models [J]. *Computer Graphics*, 1999, 33(4): 38–41.
- [2] HENRY P, KRAININ M, HERBST E, et al. RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments [J]. *The International Journal of Robotics Research*, 2012, 31(5): 647–663.
- [3] ZHAO Z, FENG X, WEI F, et al. Learning representative features for robot topological localization [J]. *International Journal of Advanced Robotic Systems*, 2013, 10(4): 215.
- [4] JOO H, LIU H, TAN L, et al. Panoptic studio: a massively multi-view system for social motion capture [C]// ICCV 2015: Proceedings of the 2015 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2015: 3334–3342.
- [5] ZHU Y, ZHAO Y, ZHU S. Understanding tools: task-oriented object modeling, learning and recognition [C]// CVPR 2015: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 2855–2864.
- [6] WANG Z, TENG S, LIU G, et al. Hierarchical sparse representation with deep dictionary for multi-modal classification [J]. *Neurocomputing*, 2017, 253, 65–69.
- [7] WANG Zh, ZHAO Z, WENG S, et al. Incremental multiple instance outlier detection [J]. *Neural Computing and Applications*, 2015, 26(4), 957–968.
- [8] 伍锡如, 黄国明, 孙立宁. 基于深度学习的工业分拣机器人快速视觉识别与定位算法[J]. *机器人*, 2016, 38(6): 711–719. (WU X R, HUANG G M, SUN L N. Fast visual identification and location algorithm for industrial sorting robots based on deep learning [J]. *Robot*, 2016, 38(6): 711–719.)
- [9] 杜学丹, 蔡莹皓, 鲁涛, 等. 一种基于深度学习的机械臂抓取方法[J]. *机器人*, 2017, 39(6): 820–837. (DU X D, CAI Y H, LU T, et al. A robotic grasping method based on deep learning [J]. *Robot*, 2017, 39(6): 820–837.)
- [10] LEVINE S, PASTOR P, KRIZHEVSKY A, et al. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection [C]// ISER 2016: Proceedings of the 2016 International Symposium on Experimental Robotics, SPAR 1. Cham: Springer, 2017: 173–184.
- [11] PINTO L, GUPTA A. Supersizing self-supervision: learning to grasp from 50K tries and 700 robot hours [C]// ICRA 2016: Proceedings of the 2016 IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2016: 3406–3413.
- [12] ZHANG Z. Microsoft Kinect sensor and its effect [J]. *IEEE Multi-Media*, 2012, 19(2): 4–10.
- [13] ZHANG Z. A flexible new technique for camera calibration [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330–1334.
- [14] RUSU R, COUSINS S. 3D is here: Point Cloud Library (PCL) [C]// ICRA 2011: IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2011: 1–4.
- [15] SUCAN I, CHITTA S. MoveIt! [OL]. [2017-12-15]. <http://moveit.ros.org>.
- [16] SARBOLANDI H, LEFLOCH D, KOLB A. Kinect range sensing: structured-light versus time-of-flight Kinect [J]. *Computer Vision and Image Understanding*, 2015, 139: 1–20.
- [17] QUIGLEY M, CONLEY K, GERKEY P, et al. ROS: an open-source robot operating system [C/OL]// ICRA 2009: Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE, 2009 [2018-02-05]. <http://www.willowgarage.com/sites/default/files/icra09-ROS.pdf>.
- [18] SILTANEN S, HAKKARAINEN M, HONKAMAA P. Automatic marker field calibration [C]// Proceedings of the Second Virtual Reality International Conference. Berlin: Springer-Verlag, 2007: 261–267.
- [19] CHOO B, LAUDAU M, DEVORE M, et al. Statistical analysis-based error models for the Microsoft Kinect? depth sensor [J]. *Sensors*, 2014, 14: 17430–17450.

This work is partially supported by the National Natural Science Foundation of China and Shenzhen Union Fund (U1613216), the National Natural Science Foundation of China (61573333).

ZHANG Zekun, born in 1994, M. S. candidate. His research interests include intelligent robot, computer vision, machine learning.

TANG Bing, born in 1991, M. S. candidate. His research interests include intelligent robot, multi-functional gripper, robot hardware control.

CHEN Xiaoping, born in 1955, Ph. D., professor. His research interests include logic-based artificial intelligence, multi-agent systems, intelligent robot.