

文章编号:1672-6413(2017)05-0028-03

基于 RGB-D 图像的三维物体检测与抓取^{*}

何若涛, 陈龙新, 廖亚军, 陈十力

(广东工业大学 机电工程学院, 广东 广州 510006)

摘要:提出了一种基于 RGB-D 图像的三维物体检测与抓取方法。该方法主要实现 RGB-D 图像的模板匹配, 对匹配结果进行聚类、评估和非极大值抑制, 物体定位, 并在三维点云中分割出平滑曲面, 从而计算机器人的抓取位置与姿态。实验结果表明:该方法可在杂乱环境下对目标进行定位, 并引导机器人抓取, 抓取成功率达到 89%。

关键词:三维物体检测; 机器人抓取; 计算机视觉

中图分类号: TP391.76 **文献标识码:** A

0 引言

三维物体检测对于机器人方面的应用有着极大的意义,因为此类技术是机器人进行工件抓取、工件装配等任务的前提条件。对三维物体检测的研究,目前的方法可分为基于图像局部描述子、基于模板匹配和基于三维点云描述子三种。但是,环境的不确定性和不规则性仍然对目前的算法构成挑战。基于图像局部描述子的算法^[1-2]不能检测纹理信息缺失的物体;基于模板匹配^[3-4]的方法容易在杂乱环境下生成误匹配(False Positives);基于三维点云描述子的方法^[5-7]容易受点云噪声、点云密度的影响。本文改进了模板匹配算法 LINEMOD^[8],使其能够在图像中同时检测多个物体,并根据图像检测结果,在不需计算物体三维位姿的情况下,基于三维点云生成机器人的抓取位姿。

1 物体检测

本文提出的基于模板匹配的物体检测算法包含三个阶段:首先,采用 LINEMOD 产生初步识别结果;然后,基于模板的空间位置对匹配模板进行聚类;最后,对所有聚类进行评估,并利用非极大值抑制去除重复识别结果,实现多个物体的检测。

1.1 模板匹配

首先利用模板匹配算法 LINEMOD 在图像中进行模板匹配。LINEMOD 对模板的定义为 $T = (\{O_m\}_{m \in M}, P)$ 。其中, O 为模板特征,表示彩色图像中的梯度方向或深度图像中的法向量方向; M 为模态,表示彩色图像或深度图像; P 为所有的模板特征在模板图像中位置 r 的集合。通过滑动窗口的方式,每张

模板在图像中的位置 c 与图像 I 进行相似度计算,即:

$$\epsilon(I, T, c) = \sum_{r \in P} \max_{t \in R(c+r)} f_m(O_m(r), I_m(t)) \quad (1)$$

其中:函数 f_m 为梯度夹角或法向量夹角的余弦值; R 为一个以 $c+r$ 为中心的矩形区域; t 为该矩形区域中的位置。当模板的相似度 ϵ 高于阈值 τ_s 时,该模板被匹配上。

在进行匹配之前,本文利用 OpenGL 对物体的 CAD 模型进行渲染,得到模板图像。为了采集物体在多个视角下的模板,我们以一个虚拟球体包围模型,移动相机到球面上,对物体进行各个视角下的图像采集。本文以三通管为检测目标,其部分模板图像如图 1 所示。同时为了实现尺度不变性,我们通过改变包围球半径的大小,模拟相机在不同距离下对物体的采集。在采集图像的过程中,物体相对于相机的位姿 $\{R, t\}$ 和距离 d 被保存下来。



图1 三通管在多个视角下的模板图像

1.2 模板聚类

LINEMOD 匹配后的结果如图 2 所示,其中三通管是检测目标,检测结果以包围盒标示。由图 2 我们可以看出,检测结果中包含了重复检测结果和错误检测结果。因此,本文提出了一种模板聚类的算法,对具有相似空间位置的模板进行聚类。模板在图像位置 (r, c) (其中 r 和 c 分别表示行和列)被匹配上,而物体相对于相机的距离 d 在渲染该模板时已知,因此,我

^{*}广东省重大科技专项(2014B090919002, 2016B0911006)

收稿日期:2017-03-20; 修订日期:2017-06-29

作者简介:何若涛(1991-),男,广东清远人,在读硕士研究生,研究方向为计算机视觉、机器人。

们定义模板的空间位置 $F=(r,c,d)$ 。

模板聚类的过程分为两个阶段:首先,利用步长 S 对 r 和 c 进行采样,得到 r' 和 c' ,因此拥有相同 r' 和 c' 的模板会分配到一个聚类中去;然后,在每一个聚类内部,我们把拥有相同 d 的模板再分配到同一个聚类中去。完成上述两个步骤后,得到一组聚类,每个聚类包含拥有相似图像位置 (r,c) 和相同训练距离 d 的模板。每个聚类表示一个检测结果,其在图像中的位置 (r_{obj}, c_{obj}) 是聚类内所有模板位置的均值,表示为:

$$r_{obj} = \frac{1}{n} \sum_{i=1}^n r_i \quad (2)$$

$$c_{obj} = \frac{1}{n} \sum_{i=1}^n c_i \quad (3)$$

其中: n 为聚类中模板的数量。

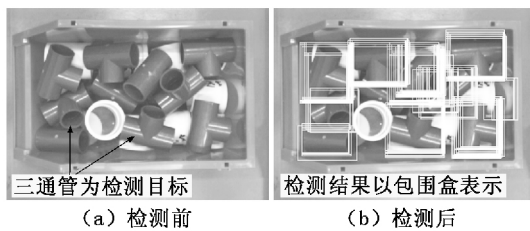


图2 LINEMOD 匹配检测前后对比

通过实验观察,我们发现误检测结果所属的聚类的模板数量与正确检测结果所属聚类的模板数量相差甚远,因此可以设定阈值 τ_c ,剔除模板数量少于 τ_c 的聚类。

1.3 聚类评估及抑制

对于1.2节得到的每个聚类,本文设计了一个评价函数 γ 对其进行评分,即:

$$\gamma = a \times \beta \quad (4)$$

深度相似度 a 计算公式为:

$$a = \frac{1}{e^{\frac{1}{n} \sum_{i=1}^n \sigma_i}} \quad (5)$$

其中: σ_i 为模板深度图像和对应场景深度图像的深度差。

法向量相似度 β 计算公式为:

$$\beta = \frac{1}{e^{\frac{1}{n} \sum_{i=1}^n \theta_i}} \quad (6)$$

其中: θ_i 为模板图像和对应场景图像的法向量夹角差。

上述评价函数的计算需要模板的深度图像和对应的场景深度图像。为了得到模板的深度图像,本文利用渲染模板图像时所保存的物体相对于相机的位姿 $\{R, t\}$ 对物体模型进行渲染,得到模板深度图像,同时得到掩膜图像。而为了得到对应的场景深度图像,我们将模板图像的掩膜图像移动到了其匹配位置 (r, c) 截取场景深度图。

由于匹配模板中包含重复识别的结果,因此可能会出现多个聚类属于同一个物体的情况。为了解决这个问题,在图像中采用非极大值抑制移除重复识别结

果。对于每一个聚类,在图像中搜索 (r_{obj}, c_{obj}) 邻域内的聚类,并比较聚类的评分 γ 。邻域的范围由半径为 r 的圆确定,如果一个聚类的评分 γ 并不是局部最大的,该聚类会被剔除。至此,认为每一个聚类表示一个检测结果。每个检测结果在图像中以一个包围盒显示,而包围盒的中心为 (r_{obj}, c_{obj}) ,包围盒的宽和高分别由聚类内模板图像的宽和高均值计算而得,最终检测结果如图3所示。

2 抓取位姿

对于每个检测结果,首先使用其掩膜图像在深度图像中截取对应的三维点云 P ,如图4所示。

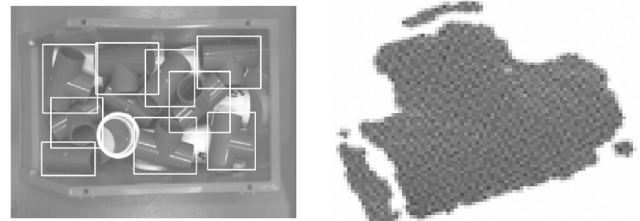


图3 最终检测结果 图4 物体在场景中的三维点云

本文以吸盘夹持器为适用对象。考虑到吸盘对于曲率小的曲面具有更可靠的吸附力,从 P 中分割出平滑的曲面作为吸附作用面。为了实现这个目的,用区域生长法^[9]对 P 进行分割。区域生长法首先创建一个空集合 P_k ,通过比较参考点和邻域点的法向量夹角差值 θ_d ,把 θ_d 小于阈值 θ_{th} 的邻域点加入到 P_k 中,同时计算新加入点的曲率 C ,若 C 小于阈值 C_{th} ,该算法把该点作为新的参考点。当所有参考点处理完后,该算法对点云中剩余的点继续进行上述步骤。利用区域生长法可将 P 分割为一组点云 $\{P_k\}$,其中,每一个 P_k 所包含的点云表示一个平滑的曲面。从 $\{P_k\}$ 中选取点云数量最多的 P_{max} 作为吸附作用面。对于抓取位置,首先计算点云 P_{max} 的质心 p_c ,然后在 P_{max} 中搜索离 p_c 最近的点 p_{mc} 作为吸附位置。这样保证了吸取点 p_{mc} 一定存在于物体表面中,确保了吸附位置的可行性。

定义 O_g 为吸盘夹持器的坐标系,如图5所示。在机器人抓取时,只有当 O_g 的 Z 轴与 p_{mc} 的法向量 n_{mc} 平行时,吸盘才能形成稳定的真空。因此,需要计算一个旋转矩阵 R 来转换 O_g ,使其 Z 轴与 n_{mc} 平行。本文利用等效轴角表示法来计算旋转矩阵 R ,旋转轴 K 的计算公式为:

$$K = Z \times n_{mc} \quad (7)$$

旋转角度 θ 的计算公式为:

$$\theta = \tan^{-1}(\|Z \times n_{mc}\|_2, Z \cdot n_{mc}) \quad (8)$$

等效轴角 $R_k(\theta)$ 可转换为旋转矩阵 R ,若吸盘夹持器当前姿态为 B_0R ,则吸附姿态为 B_0RR 。

3 机器人抓取实验

为了验证所提出的物体检测算法及抓取位姿计算方法的有效性,本文以三通管和直通管为检测物体进行了机器人抓取实验,实验场景如图6所示。

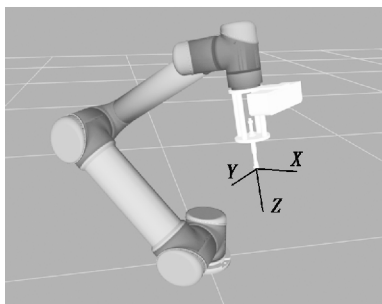


图5 吸盘夹持器坐标系



(a) 抓取三通管

(b) 抓取直通管

图6 机器人物体抓取实验场景

采用3D相机和2D彩色相机采集RGB-D图像,且相机和机器人的位姿关系已通过手眼标定实验确定。抓取过程可以分为感知—规划—执行3个阶段。在感知阶段,利用本文所提出的算法在RGB-D图像中检测出目标物体,并在对应的点云中计算抓取位姿;在规划阶段,机器人根据当前吸盘夹持器的位姿和抓取位姿进行轨迹规划;在执行阶段,机器人执行规划好的轨迹,到达目标点后吸附起工件,并把工件放置在固定的位置,随后回到准备位姿,开始新一次的物体检测。本文对两种物体分别进行了100次的抓取,成功率达到89%。部分实验截图如图7所示。

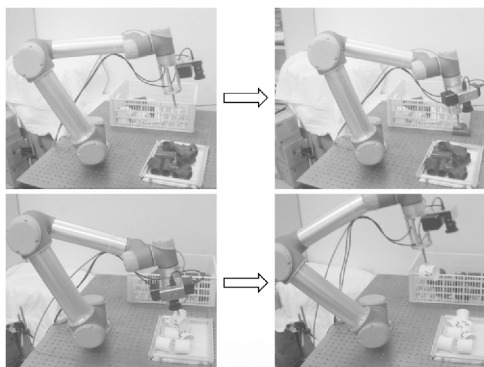


图7 机器人抓取实验

4 结束语

本文提出了一种基于RGB-D图像的三维物体检测与抓取算法,通过在RGB-D图像中进行模板匹配,实现物体图像位置的检测,然后,利用物体位置对应的三维点云计算出机器人抓取位置和姿态。实验结果表明,本文提出的算法可有效地引导机器人进行物体抓取。

参考文献:

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [2] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer Vision & Image Understanding, 2008,110(3):346-359.
- [3] Park I K, Germann M, Breitenstein M D, et al. Fast and automatic object pose estimation for range images on the GPU[J]. Machine Vision and Applications, 2010,21(5):749-766.
- [4] Cai H, Werner T, Matas J. Fast detection of multiple textureless 3-D objects[C]// International Conference on Computer Vision Systems. [s. l.]:[s. n.], 2013:103-112.
- [5] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration[C]// IEEE International Conference on Robotics and Automation. [s. l.]: IEEE, 2009:3212-3217.
- [6] Rusu R B, Bradski G, Thibaux R, et al. Fast 3D recognition and pose using the viewpoint feature histogram[C]// IEEE/RSI International Conference on Intelligent Robots and Systems. [s. l.]: IEEE, 2010:2155-2162.
- [7] Drost B, Ulrich M, Navab N, et al. Model globally, match locally: efficient and robust 3D object recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. [s. l.]: IEEE, 2010:998-1005.
- [8] Hinterstoisser S, Holzer S, Cagniart C, et al. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes[C]// IEEE International Conference on Computer Vision. [s. l.]: IEEE Computer Society, 2011:858-865.
- [9] Rabbani T, Heuvel F A V D, Vosselman G. Segmentation of point clouds using smoothness constraint[C]// International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences. [s. l.]:[s. n.], 2006.

3D Object Detection and Grasping Based on RGB-D Images

HE Ruo-tao, CHEN Long-xin, LIAO Ya-jun, CHEN Shi-li

(School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: This paper proposes a 3D object detection and grasping pose generation method based on RGB-D images. The approach starts with template matching and clusters the matched templates of similar spatial location. Then each cluster is evaluated based on a scoring function, and non-maximum suppression is adopted to remove duplicate results based on the scores. Then the feasible grasping poses is computed by extracting smooth surface from 3D point clouds. The experimental results show the approach can robustly detect objects in cluttered environment and guides robot to accomplish pick-and-place tasks, achieving an average success rate of 89%.

Key words: 3D object detection; robotic grasping; computer vision