

# 一种基于结构随机森林的家庭日常工具部件 功用性快速检测算法

吴培良<sup>1,2</sup>, 付卫兴<sup>1</sup>, 孔令富<sup>1,2</sup>

<sup>1</sup>燕山大学信息科学与工程学院, 河北 秦皇岛 066004;

<sup>2</sup>河北省计算机虚拟技术与系统集成重点实验室, 河北 秦皇岛 066004

**摘要** 家庭日常工具的部件功用性主动认知是家庭服务机器人智能提升的重要方面。为满足服务机器人实时自主作业的需要,提出了一种基于结构随机森林(SRF)的工具部件功用性快速检测算法。在离线训练阶段,利用SRF训练功用性边缘检测器与功用性检测器,并通过评估功用性检测结果的 $F_\beta$ 值确定工具各部件功用性对应的先粗糙后逐步精细化(coarse-to-fine)阈值。在线检测阶段,首先使用功用性边缘检测器计算功用性区域边缘的初步概率图,继而加以 coarse-to-fine 阈值滤波得到包含工具部件功用性的外接矩形区域,最后对该区域使用功用性检测器进行检测。实验结果表明,在普通非图形处理器系统下,相较于现有的全局搜索检测方法,本文方法对各功用性部件的检测效率均明显提升,且召回率和精度都有提高。

**关键词** 机器视觉; 功用性检测; 功用性边缘检测; 结构随机森林; 先粗糙后逐步精细化方法

**中图分类号** TP242.6+2      **文献标识码** A

**doi:** 10.3788/AOS201737.0215001

## A Fast Algorithm for Affordance Detection of Household Tool Parts Based on Structured Random Forest

Wu Peiliang<sup>1,2</sup>, Fu Weixing<sup>1</sup>, Kong Lingfu<sup>1,2</sup>

<sup>1</sup> School of Information Science and Technology, Yanshan University, Qinhuangdao, Hebei 066004, China;

<sup>2</sup> The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province,  
Qinhuangdao, Hebei 066004, China

**Abstract** Active cognition of affordance of household tool parts is regarded as an important aspect to improve home service robot intelligence. In order to meet the needs of real-time task of the service robot, a fast algorithm to improve the efficiency of affordance detection is proposed based on structured random forest (SRF). In the offline training phase, SRF is used to train affordance edge detector and affordance detector. Then the corresponding coarse-to-fine threshold of each affordance is determined by evaluating the results  $F_\beta$  of affordance detection. In the online detection phase, the affordance edge detector is used to calculate the initial probability map of the edge of affordance region. Then the coarse-to-fine threshold is used to obtain an outer rectangular including the region of the tool parts of corresponding affordance. Finally, the affordance detector is used to detect affordance of tool parts in the region obtained. The experimental results show that compared with the existing global search detection methods under normal non-graphic processing unit systems, the average detection efficiency of the proposed method increases obviously, and the recall and precision are also improved.

**Key words** machine vision; affordance detection; affordance edge detection; structured random forest; coarse-to-fine method

**OCIS codes** 150.6910; 100.2960; 100.3008; 100.6890

## 1 引 言

使用工具标志人类文明的开始,而认知工具关键部件功用性是使用工具的先决条件。认知科学研究表

**收稿日期:** 2016-08-09; **收到修改稿日期:** 2016-09-27

**基金项目:** 国家自然科学基金(61305113)、河北省自然科学基金(F2016203358)

**作者简介:** 吴培良(1981—),男,博士,副教授,主要从事家庭服务机器人环境认知与工具认知方面的研究。

E-mail: peiliangwu@ysu.edu.cn

明,人类功用性认知是通过以视觉为主的各感觉器官获取工具的结构信息,再经大脑加工处理认知工具各部件功能的过程<sup>[1]</sup>。在机器人研究领域,机器人的发展自始至终都在模仿人类各项能力,通过剖析人类能力的工作机理,能够很好地为机器人研究提供指导思路。目前机器人可在一定程度上模仿人类的感知和识别能力,而借鉴人类认知方式,使机器人具备工具部件功用性认知能力,对机器人主动智能提升具有重要意义。目前,机器人主要通过读取语义标签方式被动获取物品功用性等语义,基于推理学习的功用性主动认知方法研究刚刚出现<sup>[2-3]</sup>。

目前,学者主要采用函数方法进行功用性检测研究,比如物体(如椅子)的三维 CAD 模型<sup>[4]</sup>,此外,许多研究聚焦于基于二维图像的对象抓取点预测<sup>[5-7]</sup>。近年来,随着 RGB-D 传感器的出现,快捷并低成本地获取三维信息成为可能,继而在功用性检测领域催生了一些新的研究成果:Grabner 等<sup>[8]</sup>从三维数据中检测适于坐的曲面,Kjellström 等<sup>[9]</sup>利用人手的操作分类正在使用的工具,Zhu 等<sup>[10]</sup>提出了面向任务的对象模型学习结构模式,Hassan 等<sup>[11]</sup>将人、物和环境三个突出语境作为一组属性,实现了精细和高水平的功用性检测,Kemp 等<sup>[12]</sup>实现可以被机器人抓取的工具尖端的检测,Mar 等<sup>[13]</sup>对可抓取工具的正确抓取方式进行研究,Saxena 等<sup>[14]</sup>运用稀疏自动编码(SAE)方法而 Redmon 等<sup>[15]</sup>则基于卷积神经网络(CNN)方法检测各种物品的抓取位置,Myers 等<sup>[16]</sup>提出了一种关联功用性的局部几何和形状信息结构,并运用结构随机森林(SRF)和超像素分层匹配追踪(S-HMP)方法检测家庭中常见工具的 7 种功用性(grasp、cut、scoop、contain、pound、support 和 wrap-grasp)。上述研究中,文献[14-15]仅针对工具的抓取(grasp)这一功用性展开研究,并未涉及其他功用性检测;文献[14-16]检测精度较高,但由于采用全局搜索策略计算效率较低,在普通性能的计算机或机器人上难以保证实时性。

边缘检测作为一种重要预处理步骤被应用在对象识别<sup>[17-18]</sup>、图像分割<sup>[19-20]</sup>等任务中,旨在提高任务的执行效率。边缘存在于图像的不规则结构和在不平稳现象中,在进行图像处理确定目标对象时边缘常常提供非常显著的重要特征。文献[21]利用 SRF 实现了二维图像的边缘快速检测。本文借鉴先粗糙后逐步精细化(coarse-to-fine)的思想<sup>[22]</sup>,针对家庭服务机器人完成服务任务时普遍使用的家庭日常工具,设计并构建了工具部件功用性边缘检测器以快速定位工具部件功用性的大致区域,在此区域内进行精确定位及功用性搜索检测,从而有效地提高了家庭日常工具部件功用性的检测效率。

## 2 方法概述

### 2.1 系统框架

工具部件功用性快速检测方法分为离线学习和在线检测两个阶段。

离线学习阶段:分别构建工具部件功用性边缘检测器和工具部件功用性检测器,利用功用性边缘检测器对训练数据集进行检测得到对应概率图,在概率图中用一系列阈值筛选出可能区域,利用工具部件功用性检测器对可能区域进行检测,评估检测结果以确定 coarse-to-fine 阈值。

在线检测阶段:根据待检测功用性及图像深度信息计算相应特征矩阵,利用工具部件功用性边缘检测器检测功用性区域边缘;利用工具部件功用性对应的 coarse-to-fine 阈值筛选出较精确功用性区域;计算选出区域对应的特征矩阵,利用工具部件功用性检测器进行功用性检测。

整体流程如图 1 所示。

### 2.2 结构随机森林

随机森林是结合  $K$  棵决策树( $T_1, T_2, \dots, T_K$ )的集成学习方法<sup>[23]</sup>。为防止过度拟合,随机森林中的每棵决策树在随机排列的数据上训练得到,其输出可以是类的标记(多标记分类)或连续值(回归)。由于仅需通过若干决策树的二值决策函数进行推理,随机森林具有极高的执行速度和灵活性<sup>[24]</sup>。SRF 最先由 Kotschneider 等<sup>[25]</sup>应用于空间约束的场景分割任务中。与一般随机森林相比,SRF 通过对输入输出加以约束,可以学习到更具表现力的信息,如形状、大小甚至抽象关系,同时仍然保留随机森林所固有的效率优势。

SRF 的训练样本  $c$  由特征集  $X$  和标记集  $Y$  组成,局部特征块  $\mathbf{x}_j \in X$  与局部语义标记块  $\mathbf{y}_j \in Y$  属于同一个样本,在图像中对应位置相同,各自维度信息为  $\mathbf{x}_j \in \mathbf{R}^{N \times N \times \alpha}$  和  $\mathbf{y}_j \in \mathbf{R}^{N \times N}$ ,  $N$  为块采样尺寸, $\alpha$  为特征通道个数, $\mathbf{R}^{N \times N \times \alpha}$  和  $\mathbf{R}^{N \times N}$  分别表示  $N \times N \times \alpha$  维和  $N \times N$  维空间。

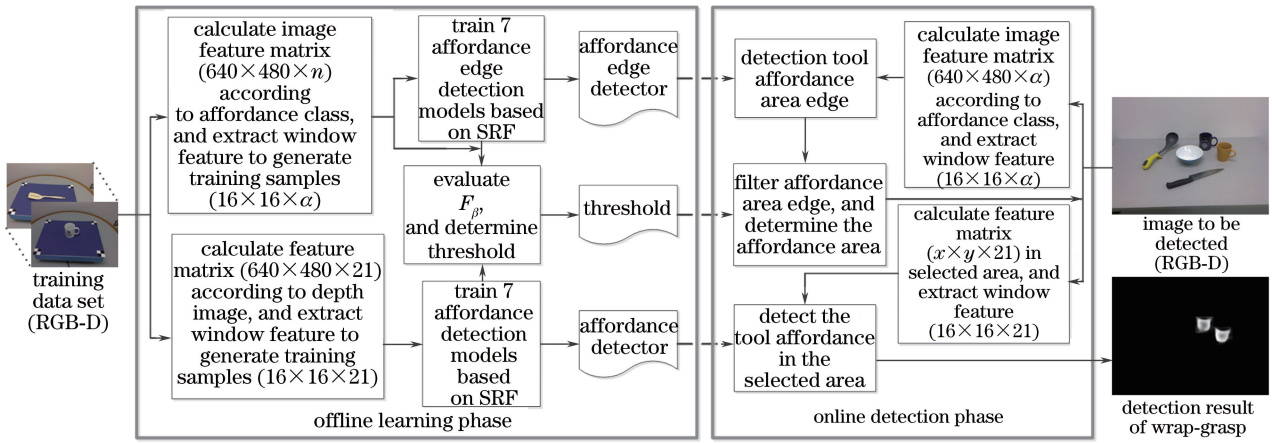


图 1 基于 SRF 的工具部件功用性快速检测整体流程图

Fig. 1 Overall flow chart of the rapid affordance detection of tool parts based on SRF

在构建 SRF 的决策树时,训练样本集  $D$  从 SRF 的根节点进入,每经过一个非叶子节点都被分裂成两个子数据集,分别向当前节点的左孩子和右孩子传递,图 2 直观地给出了单棵决策树的构建过程。需要注意的是,分裂只是样本的分组,不会造成样本的分解,此过程要求分裂结果获得最大信息增益且该增益大于某预设值。信息增益  $G_j$  计算公式为

$$H(D_j) = \sum_y p_y (1 - p_y), \quad (1)$$

$$G_j = H(D_j) - \sum_{c \in \{L,R\}} \frac{|D_j^c|}{|D_j|} H(D_j^c), D_j \subset X \times Y, \quad (2)$$

式中 Gini 系数  $H(D_j)$  用于评估样本集  $D_j$  纯度并衡量样本数据的离散程度,  $p_y$  是正样本在样本集中所占比例,  $D_j^c$  是相应分裂子集,  $G_j$  为最大信息增益。最大信息增益  $G_j$  确定后,相应非叶子节点的阈值  $\theta_j$  同时被确定。当子数据集到达某一节点后找不到阈值  $\theta_j$  使得信息增益大于预设值时,样本数据集停止分裂,该节点即为叶子节点。

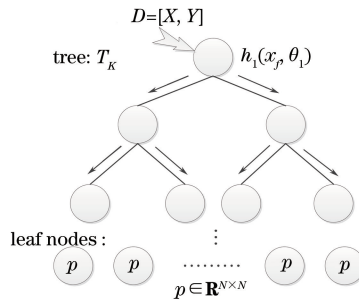


图 2 SRF 中单棵树构建示意图

Fig. 2 Sketch map of the construction of a single tree in SRF

SRF 的决策树叶子节点中存储的是与  $y$  形式相同且距到达叶子节点的所有样本均值最近的标记块。在检测阶段,由待检测图像得到相应特征集  $X'$ ,将其作为  $K$  棵决策树的输入,决定样本  $x \in X'$  到达叶子节点的路径的决策函数为

$$h(x, \theta_j) \Big|_{\theta_j = (f, \rho)} = \begin{cases} 0, & x(f) > \rho \\ 1, & x(f) \leq \rho \end{cases}, \quad (3)$$

式中  $x(f)$  为样本属性值,  $\rho$  为相应节点阈值,当  $h(\cdot) = 1$  [即  $x(f) \leq \rho$ ] 时,样本去向右孩子,反之样本去向左孩子。利用样本所到达叶子节点中存储的分类信息给相应像素点投票,并将所有像素对应投票归一化得到图像检测的概率图。

### 3 模型离线训练

#### 3.1 功用性边缘检测器构建

与根据过完备的几何特征对所有功用性统一建模相比,根据某种功用性的几何特征分别对其建模,并据此从场景中识别该种功用性的方法容错性更强。另外,考虑到不同的功用性部件具有不同的几何结构特性,且在边缘处目标区域与背景形成鲜明对比,故在确定目标功用性形态和位置时,边缘特征的稳健性最好。基于上述考虑,提出了功用性边缘检测的思想,并针对不同功用性选择不同特征构建功用性边缘检测模型,各功用性边缘检测模型联合构成功用性边缘检测器。

##### 3.1.1 特征描述

由于家庭日常工具在不同角度下的某些几何特征可能不同,所以数据采集及特征提取应考虑到对视角变化的稳健性。借鉴文献[23]从图像多通道中提取特征的方式,用到的特征有方向梯度直方图、梯度幅值、平均曲率、形状指数和曲度<sup>[16]</sup>,每个特征通道按照图像原始尺度和1/2原始尺度各取一次得到。 $16 \times 16$ 大小的局部特征块提取的特征矢量  $\mathbf{x} \in \mathbf{R}^{16 \times 16 \times \alpha}$ ,其中 $\alpha$ 为通道数,即为表征某功用性所采用特征在两个尺度下维度之和,其值与功用性类别有关,表1中列出了不同功用性模型所选取的特征及其维度。这里,不同功用性边缘检测选用的特征不尽相同,选取依据是该特征对表征该功用性区域边缘有效且显著。

表1 工具部件各功用性边缘检测模型特征选取

Table 1 Feature selection of each affordance edge detection model of tool parts

Geometric features / dimension	Objective affordance						
	Grasp	Contain	Cut	Pound	Scoop	Support	Wrap-grasp
Oriented gradient histograms /4D	✓	✓	✓	✓	✓	✓	✓
Gradient magnitude /1D		✓	✓	✓	✓	✓	✓
Mean curvatures /1D	✓	✓	✓	✓	✓	✓	
Shape index /1D	✓	✓	✓	✓	✓	✓	✓
Curvedness /1D		✓	✓		✓		✓

根据深度图像计算功用性边缘检测模型对应的几何特征,其中平均曲率为微分几何中反映曲面弯曲程度的内蕴几何量,记为 $f_{MC}$ ,主曲率 $(k_1, k_2)$ , $k_1 > k_2$ ,则 $f_{MC} = (k_1 + k_2)/2$ 。梯度幅值和方向梯度直方图特征是用来进行物体边缘检测的有效特征描述子。形状指数 $S$ 和曲度 $C$ 表征表面在不同方向的弯曲,体现人对形状的感知<sup>[26]</sup>。形状指数 $S$ 和曲度 $C$ 的计算公式为

$$S = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right), C = \sqrt{\frac{k_1^2 + k_2^2}{2}}. \quad (4)$$

工具部件各种功用性边缘检测模型学习用到的特征对应如表1所示,表中“✓”表示在训练对应功用性检测器时选取了相应的几何特征。

##### 3.1.2 功用性边缘检测器构建算法

功用性边缘检测模型基于SRF离线学习得到。训练数据集由 $n$ 幅RGB-D图像及其标记图像组成,其中,深度图像用于计算特征矩阵,标记图像保存对应图像中各工具部件功用性区域边缘标记结果。训练样本由以 $16 \times 16 \times \alpha$ 为单位的特征集及相应 $16 \times 16$ 为单位的标记集组成,标记块中每个像素的值(0或1)对应图像中像素分类结果。用于边缘检测器学习的正样本从功用性区域边缘提取,负样本从背景区域及其他功用性区域边缘提取。

由于学习功用性边缘检测模型的训练数据是目标区域的边缘,这种用局部特征对整体建模时存在信息不完备和不同功用性特征局部信息交叉现象,导致在检测几何特征相似的功用性区域边缘时产生一定的误差,对此本文中边缘检测器借鉴文献[16]的投票机制平抑此类误差。

算法描述如下:

输入:由特征集  $S_f$  和标记集  $S_l$  组成的样本集  $S$ 。

输出:工具部件功用性边缘检测器。

1) 由训练数据集中的深度图像计算各通道特征值,以  $16 \times 16$  为单位在各个特征通道中采用滑动窗口机制随机提取一定数量的特征块及标记图中相应的标记块,分别加入特征集  $S_f$  和标记集  $S_l$ 。

2) 对所有标记块进行主成分分析,判定其对应样本为正样本或负样本。

3) 随机选择  $r$  维特征( $r=M/2, M$  为特征块维度,  $M=16 \times 16 \times \alpha$ )参与构建决策树。

4) 利用样本集  $S$  构建决策树。在每个分裂节点处,从  $r$  维特征中随机选择  $\lfloor \sqrt{r} \rfloor$  维特征作为样本集分裂阈值,  $\lfloor x \rfloor$  表示不超过  $x$  的最大整数。选取对应信息增益最大的特征值为该节点阈值,相应的输入样本集被分裂成两个子样本集作为子节点的输入。

5) 在样本集分裂过程中,当输入样本集取得的最大信息增益小于预设值  $T$  或样本个数不超过 8 个时停止分裂,这个节点就成为叶子节点,所有样本停止分裂。

6) 分析步骤 5) 中叶子节点的输入样本集所对应标记集,得到  $16 \times 16$  大小的标记块作为此叶子节点的内容,所有样本都到达叶子节点则此决策树构建完成,不需要剪枝。

7) 从步骤 1) 到步骤 6) 重复  $K$  次,生成  $K$  棵决策树,这  $K$  棵决策树组合形成一个 SRF,即为该种功用性边缘检测模型。

8) 重复以上步骤,训练其他功用性边缘检测模型,将 7 种功用性边缘检测模型联合构成工具部件功用性边缘检测器。

这里需要指出的是,步骤 1) 中构建决策树的训练集、步骤 3) 中参与训练决策树的样本特征属性和步骤 4) 中决定决策树非叶子节点阈值的样本特征属性均通过随机选择得到,从而有效保证了 SRF 中各决策树间的差异性,这种随机选择机制被证明更加有效<sup>[27]</sup>。根据经验并借鉴文献[16, 21, 28]中做法,随机森林中决策树个数  $K$  取值为 8。

### 3.2 工具部件功用性检测器构建

#### 3.2.1 特征描述

与功用性边缘检测器构建方法类似,功用性检测器的构建同样基于 SRF 并针对每种功用性训练相应的检测模型。不同之处在于,这里所选特征除 3.1.1 节中梯度幅值、平均曲率、方向梯度直方图、形状指数和曲度外,为精确描述工具部件功用性,还选择三维表面法向量和一维高斯曲率。

表面法向量是几何体表面的重要属性。从深度数据恢复三维点云,再从中估计出三维表面法向量,并去除样本块均值,使得视角变化时表面法向量特征稳健性更强。

高斯曲率同平均曲率一样是曲面论中重要的内蕴几何量,记为  $f_{GC}$ ,则  $f_{GC}=k_1 k_2$  ( $k_1$  和  $k_2$  为曲面上一个点的两个主曲率)。联合高斯曲率和平均曲率可以确定 8 种曲面类型:峰、脊、鞍形脊、最小面、平面、阱、谷和鞍形谷<sup>[29]</sup>,有助于识别不同功用性的内部结构。

本文训练工具部件功用性检测模型主要基于上述 7 类特征,提取的  $16 \times 16$  大小的局部特征块矢量  $\mathbf{x} \in \mathbf{R}^{16 \times 16 \times \alpha}$ ,其中  $\alpha$  代表 21 个特征通道:表面法向量 3 个、梯度幅值 2 个、高斯曲率 2 个、平均曲率 2 个、方向梯度直方图 8 个、形状指数 2 个和曲度 2 个,除表面法向量外其余特征均在图像原始尺度和  $1/2$  原始尺度下各取一次得到。

#### 3.2.2 工具部件功用性检测器算法实现

与训练工具部件功用性边缘检测模型方式类似,功用性检测模型同样基于 SRF 离线学习得到,且训练数据集由 RGB-D 图像和标记图像组成,训练样本包含特征集和标记集两部分。与训练功用性边缘检测模型的标记图像不同,训练功用性检测模型的标记图像是对整个功用性区域做标记,目的在于对整个功用性区域进行检测。

算法输入为由特征集  $S'_f$  和标记集  $S'_l$  共同组成的样本集  $S'$ ,输出为工具部件功用性检测器。由于功用性检测模型与功用性边缘检测模型都是基于 SRF 构建,其 SRF 学习过程相同,这里不再赘述,算法描述见 3.1.2。其中,样本特征维数为  $M=16 \times 16 \times 21$ 。在完成 7 种功用性检测模型的训练后,将其联合起来构成

工具部件功用性检测器。

### 3.3 Coarse-to-fine 阈值选取

Coarse-to-fine 阈值是指边缘检测器对图像检测得到的概率图中目标区域与干扰性区域的临界值,用以从概率图中筛选出目标区域。阈值过低,检测过程易受到图片噪声影响而多选非目标区域;反之,阈值过高,在处理复杂场景时则会误滤除部分目标区域。鉴于此,提出 coarse-to-fine 阈值,对功用性边缘检测结果区域进行阈值滤波,旨在尽可能准确地得到目标功用性区域。

对功用性检测结果进行评估可以区分不同阈值的筛选质量。采用文献[30]中介绍的  $F_{\beta}^w$ -measure 评估方法,该方法综合考虑噪声点的概率值大小和到正确目标功用性区域的距离,对筛选出目标区域(功用性检测结果区域)的质量评估更为客观,有助于找到更为准确的阈值。

如前所述,每种功用性分别对应一个功用性边缘检测模型,同样地,针对不同的边缘检测模型选取不同的 coarse-to-fine 阈值,其算法描述如下:

输入:训练数据集中的  $m$  幅 RGB-D 图像。

输出:阈值  $t = (t_1, \dots, t_7)$ 。

- 1) 利用功用性边缘检测器依次对  $m$  幅图像进行某种功用性的边缘检测得到其概率图,在一系列阈值(取值从 0 到 1,每次增量为 0.01)下获取目标区域。
- 2) 用工具部件功用性检测器对步骤 1)得到的区域进行功用性检测。
- 3) 对步骤 2)的检测结果进行  $F_{\beta}^w$ -measure 评估,取  $m$  幅图像  $F_{\beta}^w$ -measure 评估平均值记为  $F_{\beta}^w$ ,这样得到不同阈值与  $F_{\beta}^w$  值的对应关系,取最大  $F_{\beta}^w$  值所对应的阈值即为此种功用性区域检测的阈值  $t_i$ 。
- 4) 重复步骤 1)~3),完成其他功用性区域检测的阈值选取。

## 4 家庭日常工具部件功用性在线检测

由图 1 所示,离线阶段训练得到工具部件功用性边缘检测器、coarse-to-fine 阈值及工具部件功用性检测器,在线阶段,将其分别应用于检测目标功用性区域、筛选精确目标功用性区域和在选出区域检测目标功用性。

在线检测过程算法描述如下:

输入:待检测 RGB-D 图像,待检测功用性类别。

输出:概率图  $P$ ,其中每个像素的值代表该像素点属于目标功用性区域的概率。

- 1) 根据待检测功用性种类选用相应边缘检测模型,继而确定对应几何特征种类,根据深度图像计算得到特征矩阵。
- 2) 从特征矩阵读入检测样本,检测样本只包含特征集,并且记录样本在图像中的位置信息。
- 3) 功用性边缘检测模型中对待检测样本分类,利用叶子节点中内容信息对样本在图像中相应像素位置点进行投票,综合所有决策树结果得到功用性边缘检测的概率图  $P'$ 。
- 4) 选择目标功用性对应 coarse-to-fine 阈值,对步骤 3)得到的概率图  $P'$  进行处理,滤除噪声,筛选出精确功用性区域边界,确定目标功用性矩形区域。
- 5) 计算步骤 4)选定区域对应的特征矩阵,采用滑动窗口机制读入检测样本,并记录样本在图像中的位置信息。
- 6) 功用性检测模型对所有检测样本进行分类,利用叶子节点中内容信息对样本在图像中相应像素位置点进行投票,综合所有决策树结果得到最终功用性检测的概率图  $P$ 。

## 5 实验及结果分析

### 5.1 实验数据集

实验选用文献[16]中数据集,该数据集是目前比较完备的工具部件功用性数据集,采集了包含厨房、园艺等共计 17 大类 105 种家庭日常工具的 RGB-D 信息,涵盖了 grasp、cut、scoop、contain、pound、support 和

wrap-grasp 共 7 类功用性。每种工具在近 300 个不同视角下进行采集,产生了超过 30000 组的 RGB-D 数据,其中有 1/3 的数据进行了功用性标记。在实验过程中,利用标记的数据完成离线训练和在线测试,相应训练数据和测试数据比例约为 4:1。图 3(a)中列举了部分工具及其最主要功用性语义描述,图 3(b)列举了不同工具的不同部件所对应的功用性标记,表 2 给出了 7 种工具部件功用性的描述及其举例<sup>[16]</sup>。除了单一物品数据信息,此数据集还提供了 3 个系列的多种物品随意摆放的复杂场景各 1000 多组数据。

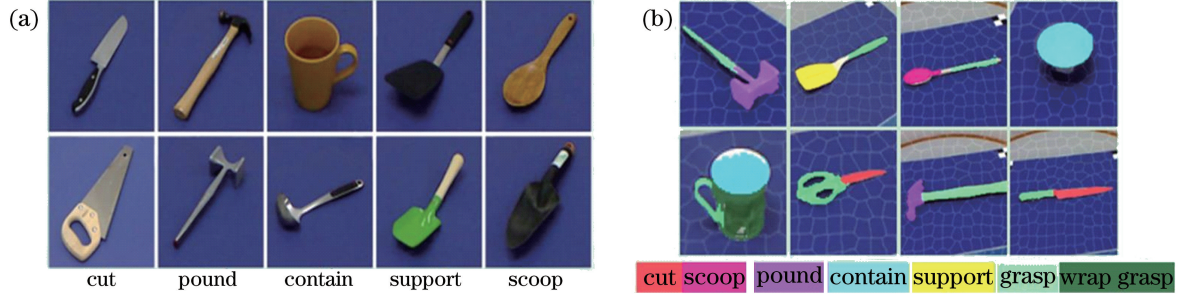


图 3 RGB-D 数据集中部分对象。(a) 对应用用性; (b) 目标部件功用性检测结果  
Fig. 3 Part of the object in the RGB-D data set. (a) Corresponding main affordance;

(b) affordance detection results of the target parts

表 2 工具部件功用性描述及举例

Table 2 Affordance description and examples of tool parts

Affordance	Description	Examples
Grasp	Can be enclosed by a hand for manipulation	Handles of tools
Wrap-grasp	Can be held with the hand and palm	The outside of a cup
Cut	Used for separating another object	The blade of a knife
Contain	With deep cavities to hold liquid	The inside of bowls
Support	Flat parts that can hold loose material	Turners
Scoop	With a curved surface and mouth for gathering and holding soft material	Trowels
Pound	Used for striking other objects	The head of a hammer

## 5.2 评价方法

采用两种方法对本文方法和文献[16]方法的功用性检测结果进行评价。

方法一:统计每种功用性精度召回率对比和阈值  $F_\beta$  值对比。阈值在 0 与 1 之间选取,对应每个阈值计算精度  $P_r$ 、召回率  $R_c$  和  $F_\beta$  值为

$$P_r = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad R_c = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (5)$$

$$F_\beta = (1 + \beta) \frac{P_r \cdot R_c}{\beta \cdot P_r + R_c}, \quad \beta = 0.3, \quad (6)$$

式中  $N_{TP}$ 、 $N_{FP}$ 、 $N_{TN}$  和  $N_{FN}$  的含义分别为

$N_{TP}$ :检测结果中被判定为正样本,事实上也是正样本的点的总数;

$N_{FP}$ :检测结果中被判定为正样本,但事实上是负样本的点的总数;

$N_{TN}$ :检测结果中被判定为负样本,事实上也是负样本的点的总数;

$N_{FN}$ :检测结果中被判定为负样本,但事实上是正样本的点的总数。

方法二:为了体现对噪声的抑制效果,本文还采用  $F_\beta^w$ -measure 方法对实验结果进行评价。 $F_\beta^w$ -measure 方法由 Margolin 等<sup>[30]</sup>提出,通过给背景像素分配权重的方法来实现,其权重分配原则是像素到真实标记距离越小所分配权重值越大,反之所分配权重值越小。 $F_\beta^w$ -measure 方法评价结果为

$$F_\beta^w = (1 + \beta^2) \frac{P_r^w \cdot R_c^w}{\beta^2 \cdot P_r^w + R_c^w}, \quad \beta = 1, \quad (7)$$

式中当  $\beta=1$  时,  $P_r^w$  和  $R_c^w$  是标准精度  $P_r$  和召回率  $R_c$  的加权扩展,  $F_\beta^w$ -measure 被定义为  $P_r^w$  和  $R_c^w$  的加权

调和平均。

### 5.3 实验结果分析

依次对表 2 中的 7 种功用性进行实验。

离线训练阶段,学习得到某功用性的边缘检测模型及其功用性检测模型(均为由 8 棵决策树构成的随机森林),继而由这两个模型学习该功用性的 coarse-to-fine 阈值。依次训练得到各功用性的 coarse-to-fine 阈值分别为 grasp 0.57、cut 0.38、scoop 0.43、contain 0.51、pound 0.72、support 0.59、wrap-grasp 0.53。

在线检测阶段,工具部件边缘检测器对各种工具功用性区域边缘进行检测,效果如图 4(b)所示;利用离线学习得到的各类工具相应的 coarse-to-fine 阈值加以滤波,筛选出精确功用性区域,效果如图 4(c)所示;工具部件功用性检测器对筛选出区域进行功用性检测,效果如图 4(d)所示。图 4(e)是文献[16]基于 SRF 方法的功用性检测模型对相同图像的功用性检测结果。对比可见,本文方法在背景滤除方面效果显著。



图 4 本文方法和文献[16]方法在单一场景下对不同工具 7 种功用性检测效果。(a) 待检测单一场景图像;  
(b) 功用性边缘检测器检测结果;(c) coarse-to-fine 阈值滤波结果;(d) 本文最终检测结果;(e) 文献[16]检测结果

Fig. 4 7 kinds affordance detect results of different tools in single scene using the method and that in ref. [16]. (a) Single scene image to be detected; (b) affordance edge detector results; (c) coarse-to-fine threshold filtering results; (d) final detection results of the proposed method; (e) detection results of the method in ref. [16]

图 5 给出了复杂场景下本文方法[图 5(d)]和文献[16]方法[图 5(e)]对不同功用性的检测效果。对比图 5(d)和图 5(e)可以很容易看出,本文方法结果过滤掉了文献[16]结果中的噪声干扰,直接找到了目标功用性部件。在抓取检测中,本文方法和文献[16]方法均未有效地检测出杯子把手,究其原因主要在于近距离观察物体可以清晰地分辨物体的轮廓结构,而距离较远时物体轮廓结构甚至整个物体都变得模糊,导致边缘检测及功用性检测效果不佳。总体而言,针对复杂场景本文所提方法具有更好的功用性检测效果。

图 6 是使用传统方法对本文提出方法和使用 SRF 方法<sup>[16]</sup>进行功用性检测的评价统计结果,精度  $P_r$ 、召回率  $R_c$  和  $F_\beta$  值依据(3)、(4)式计算得到。从图中可以看出在精度和召回率方面,本文所提方法较文献[16]中 SRF 方法均有不同程度的提高,在  $F_\beta$  值的对比上本文方法的优势更加明显。

表 3 中给出了在相同数据集下采用 S-HMP、SRF 和本文方法的评估结果对比,采用  $F_\beta^w$ -measure 方法的评估结果。从表中可以看出在所有功用性检测结果的  $F_\beta^w$  平均值的比较中,采用本文方法比单纯使用 SRF 功用性检测方法高出了 7.5%,其原因主要是由于本文的方法能够提高目标功用性检测的精度,同时本文方法与 S-HMP 方法在  $F_\beta^w$  平均值的比较中差别不大。

表 4 给出了本文算法与已有算法的效率对比,该结果数据在配有 16 GB 内存,4 核 1.8 GHz CPU 的设



备上,使用 Matlab 软件实验得到。对比算法采用文献[16]的 SRF 方法,此方法在当前功用性检测方面实时性最佳,但对比可见,本文方法无论是对单一场景,还是对复杂场景的检测速度均明显提升。此外,在本文实验设备上文献[14]的 SAE 方法对单一场景下对象抓取位置的检测耗时约 40 min,而文献[15]的 CNN 方法则运行于 NVIDIA Tesla K20 GPU 环境下,在普通配置的 CPU 上无法运行。

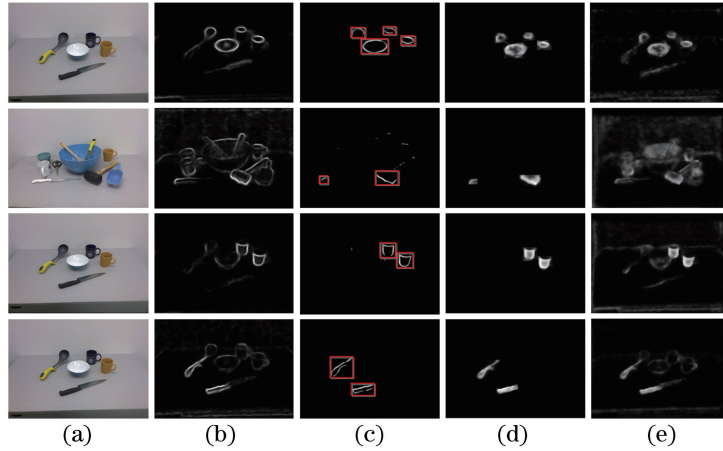


图 5 本文方法和文献[16]方法在复杂场景下对不同功用性的检测效果。(a) 待检测复杂场景图像; (b) 功用性边缘检测器检测结果; (c) coarse-to-fine 阈值滤波结果; (d) 本文最终检测结果; (e) 文献[16]方法检测结果

Fig. 5 Different affordance detect results in cluttered scene using the method proposed and that in ref. [16].

(a) Cluttered scene image to be detected; (b) affordance edge detector results;

(c) coarse-to-fine threshold filtering results; (d) final detection results of the proposed method;

(e) detection results of the method in ref. [16].

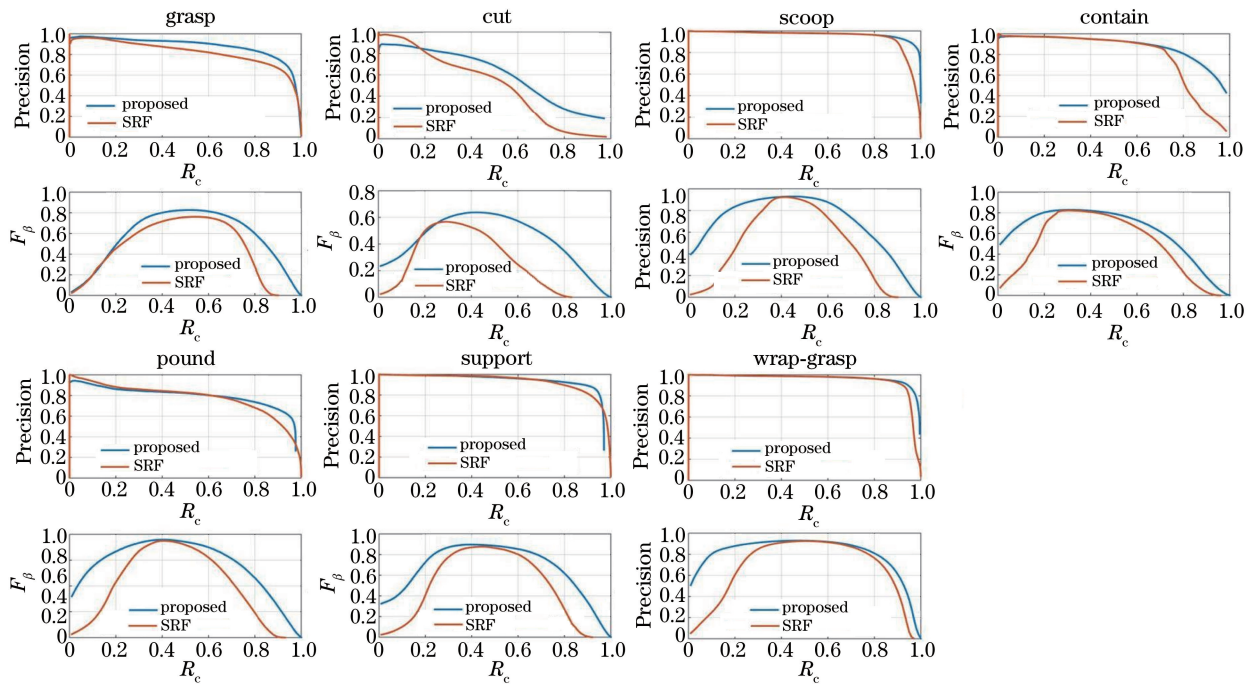


图 6 本文方法和文献[16]SRF 方法对各种功用性检测结果的精度、召回率和  $F_\beta$  对比图

Fig. 6 Precision, recall rate and  $F_\beta$  contrast of various affordance detect results using the method proposed and that in ref. [16].

表 3 不同功用性检测方法的  $F_{\beta}^{\omega}$ -measure 方法评估结果Table 3  $F_{\beta}^{\omega}$ -measure evaluation results of different affordance detection methods

Affordance	$F_{\beta}^{\omega}$ -measure (single objects)		
	S-HMP <sup>[16]</sup>	SRF <sup>[16]</sup>	Proposed method
Grasp	0.367	0.314	0.554
Cut	0.373	0.285	0.224
Scoop	0.415	0.412	0.573
Contain	0.810	0.635	0.605
Pound	0.643	0.429	0.511
Support	0.524	0.481	0.489
Wrap-grasp	0.767	0.666	0.787
Mean	0.557	0.460	0.534

表 4 本文方法与其他方法的效率对比

Table 4 Comparison between the proposed method and the other methods on the execution efficiency

Affordance	Computation time /s				
	Single objects			Cluttered scene	
	SAE <sup>[14]</sup>	SRF <sup>[16]</sup>	Proposed method	SRF <sup>[16]</sup>	Proposed method
Grasp	~2400	14.60	1.12	15.54	1.61
cut		14.70	1.08	15.43	1.25
Scoop		15.34	0.95	14.84	1.09
Contain		14.29	0.95	14.90	1.76
Pound		14.80	1.04	15.01	1.53
Support		15.28	1.29	15.06	1.07
Wrap-grasp		14.52	0.97	15.93	1.05
Mean	2400	14.79	1.06	15.24	1.34

## 6 结 论

提出了一种基于 SRF 的家庭日常工具部件功用性快速检测方法。借鉴 coarse-to-fine 思想,构建了一种基于 SRF 表示的工具部件功用性边缘检测模型,并将其与功用性检测模型结合,从而提出了一种满足服务机器人在线任务需要的工具部件功用性快速检测算法。在普通非 GPU 系统下,本文方法对各功用性部件的检测效率均明显优于现有算法,且精度和召回率都有提高,为机器人快速检测工具继而提供高效智能服务奠定基础。

## 参 考 文 献

- [1] Gibson J J. The theory of affordances [M].//Shaw R, Bransford J. Perceiving, acting, and knowing: toward an ecological psychology. Hilldale: Lawrence Erlbaum Associates, 1977: 67-82.
- [2] Zhu Y, Fathi A, Li F F. Reasoning about object affordances in a knowledge base representation [C]. European Conference on Computer Vision, 2014: 408-424.
- [3] Koppula H S, Saxena A. Physically grounded spatio-temporal object affordances [C]. European Conference on Computer Vision, 2014: 831-847.
- [4] Stark L, Bowyer K. Function-based generic recognition for multiple object categories [J]. CVGIP: Image Understanding, 1994, 59(1): 1-21.
- [5] Bohg J, Kragic D. Grasping familiar objects using shape context [C]. International Conference on Advanced Robotics, 2009.
- [6] Saxena A, Driemeyer J, Ng A Y. Robotic grasping of novel objects using vision [J]. The International Journal of Robotics Research, 2008, 27(2): 157-173.
- [7] Stark M, Lies P, Zillich M, *et al.* Functional object class detection based on learned affordance cues [C]. International Conference on Computer Vision Systems, 2008: 435-444.

- [8] Grabner H, Gall J, Van Gool L. What makes a chair a chair?[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2011: 1529-1536.
- [9] Kjellström H, Romero J, Kragić D. Visual object-action recognition: inferring object affordances from human demonstration[J]. Computer Vision and Image Understanding, 2011, 115(1): 81-90.
- [10] Zhu Y X, Zhao Y B, Zhu S C. Understanding tools: task-oriented object modeling, learning and recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2855-2864.
- [11] Hassan M, Dharmaratne A. Attribute based affordance detection from human-object interaction images [C]. Proceedings of the Image and Video Technology Workshops, 2015: 220-232.
- [12] Kemp C C, Edsinger A. Robot manipulation of human tools: autonomous detection and control of task relevant features [C]. International Conference on Intelligent Manipulation and Grasping, 2006: 1-8.
- [13] Mar T, Tikhonoff V, Metta G. Multi-model approach based on 3D functional features for tool affordance learning in robotics[C]. IEEE-RAS International Conference on Humanoid Robots, 2015: 482-489.
- [14] Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps[J]. International Journal of Robotics Research, 2015, 34(4-5): 705-724.
- [15] Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks [C]. IEEE International Conference on Robotics and Automation, 2015: 26-30.
- [16] Myers A, Teo C L, Fermuller C, *et al.* Affordance detection of tool parts from geometric features [C]. IEEE Conference on Robotics and Automation, 2015: 1374-1381.
- [17] Ferrari V, Fevrier L, Jurie F, *et al.* Groups of adjacent contour segments for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(1): 36-51.
- [18] Ullman S, Basri R. Recognition by linear combinations of models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(10): 992-1006.
- [19] Arbeláez P, Maire M, Fowlkes C, *et al.* Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.
- [20] Wachinger C, Fritscher K, Sharp G, *et al.* Contour-driven atlas-based segmentation[J]. IEEE Transactions on Medical Imaging, 2015, 34(12): 2492-2505.
- [21] Dollar P, Zitnick C L. Fast edge detection using structured forests [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(8): 1558-1570.
- [22] Pedersoli M, Vedaldi A, Gonzalez J, *et al.* A coarse-to-fine approach for fast deformable object detection[J]. Pattern Recognition, 2015, 48(5): 1844-1853.
- [23] Ho T K. Random decision forests [C]. International Conference on Document Analysis and Recognition, 1995: 278-282.
- [24] Criminisi A, Shotton J. Decision forests for computer vision and medical image analysis [C]. Advances in Computer Vision and Pattern Recognition, 2013: 201-212.
- [25] Kotschieder P, Bulo S R, Bischof H, *et al.* Structured class-labels in random forests for semantic image labelling [C]. IEEE International Conference on Computer Vision, 2011: 2190-2197.
- [26] Koenderink J J, Van Doorn A J. Surface shape and curvature scales [J]. Image and vision computing, 1992, 10(8): 557-564.
- [27] Criminisi A, Shotton J, Konukoglu E. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning [J]. Foundations and Trends in Computer Graphics and Vision, 2012, 7(2-3): 81-227.
- [28] Lei Qin, Shi Chaojian, Chen Tingting. Structured random forests for target detection in sea images [J]. Opto-Electronic Engineering, 2015, 42(7): 31-35.  
雷 琴, 施朝健, 陈婷婷. 可见光海面目标检测的结构随机森林方法 [J]. 光电工程, 2015, 42(7): 31-35.
- [29] Besl P J, Jain R C. Segmentation through variable-order surface fitting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10(2): 167-192.
- [30] Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps? [J]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014: 248-255.