一种基于联合学习的家庭日常工具功用性部件检测算法

吴培良 [1,2,3] 隰晓珺 [1] 杨 霄 [1] 孔令富 [1,3] 侯增广 [2]

摘 要 对工具及其功用性部件的认知是共融机器人智能提升的重要研究方向. 本文针对家庭日常工具的功用性部件建模与检测问题展开研究, 提出了一种基于条件随机场 (Conditional random field, CRF) 和稀疏编码联合学习的家庭日常工具功用性部件检测算法. 首先, 从工具深度图像提取表征工具功用性部件的几何特征; 然后, 分析 CRF 和稀疏编码之间的耦合关系并进行公式化表示, 将特征稀疏化后作为潜变量构建初始条件随机场模型, 并进行稀疏字典和 CRF 的协同优化: 一方面, 将特征的稀疏表示作为 CRF 的随机变量条件及权重参数选择器; 另一方面, 在 CRF 调控下对稀疏字典进行更新. 随后使用自适应时刻估计 (Adaptive moment estimation, Adam) 方法实现模型解耦与求解. 最后, 给出了基于联合学习的工具功用性部件模型离线构建算法, 以及基于该模型的在线检测方法. 实验结果表明, 相较于使用传统特征提取和模型构建方法, 本文方法对功用性部件的检测精度和效率均得到提升, 且能够满足普通配置机器人对工具功用性认知的需要.

关键词 功用性部件检测, 深度几何特征, 联合学习, 条件随机场, 稀疏编码

引用格式 吴培良, 隰晓珺, 杨霄, 孔令富, 侯增广. 一种基于联合学习的家庭日常工具功用性部件检测算法. 自动化学报, 2019, **45**(5): 985−992

DOI 10.16383/j.aas.c170423

An Algorithm for Affordance Parts Detection of Household Tools Based on Joint Learning

WU Pei-Liang[1, 2, 3] XI Xiao-Jun[1] YANG Xiao[1] KONG Ling-Fu[1, 3] HOU Zeng-Guang[2]

Abstract The research for coherent robots to cognize tools and their affordance parts is an important direction to improve their machine intelligence. Aimed at modeling and detecting affordance parts of household tools, a joint learning algorithm for affordance parts detection via both conditional random field (CRF) and sparse coding is proposed. Firstly, geometric features of affordance parts are obtained from depth images of the tools. Secondly, the coupled relationship between CRF and sparse coding is analyzed and described with formulations. Initial CRF model is built by using sparse coded features as latent variables, and both the sparse dictionary and CRF are optimized simultaneously. On one hand, the sparse coded features are considered as the random variable condition and the weight parameter selector of CRF, and on the other hand, sparse dictionary is updated with the modulation of CRF. Then the model is decoupled and solved with the adaptive moment estimation (Adam). Finally, the offline joint learning algorithm for affordance parts modeling and online detection method are given. The experimental results show that, comparing with traditional features extracting and modeling methods, both the accuracy and efficiency of our method are improved, which can satisfy the affordance cognition requirements for robots with common configurations.

Key words Affordance parts detection, depth geometric features, joint learning, conditional random fields (CRF), sparse coding

Citation Wu Pei-Liang, Xi Xiao-Jun, Yang Xiao, Kong Ling-Fu, Hou Zeng-Guang. An algorithm for affordance parts detection of household tools based on joint learning. *Acta Automatica Sinica*, 2019, **45**(5): 985−992

收稿日期 2017-07-31 录用日期 2018-03-24
Manuscript received July 31, 2017; accepted March 24, 2018
国家重点研发计划 (2018YFB1308305), 国家自然科学基金 (61305113), 中国博士后自然科学基金 (2018M631620), 河北省自然科学基金 (F2016203358), 燕山大学博士基金 (BL18007) 资助
Supported by National Key Research and Development Program (2018YFB1308305), National Natural Science Foundation of China (61305113), Postdoctoral Science Foundation of China (2018M631620), Natural Science Foundation of Hebei Province (F2016203358), and Doctoral Fund of Yanshan University (BL18007)
本文责任编委 胡清华
Recommended by Associate Editor HU Qing-Hua
1. 燕山大学信息科学与工程学院 秦皇岛 066004 2. 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 3. 河北省计算机虚拟技术与系统集成重点实验室 秦皇岛 066004
1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004 2. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 3. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004

纵观人类文明史, 社会每一次进步几乎都与使用工具息息相关; 在人的成长过程中, 学习使用工具也是其必备的能力之一. 在机器智能研究领域, 机器人的发展始终都在学习人类智能和技能, 目前机器人可在一定程度上模拟人类的感知能力[1], 而借鉴人类认知方式, 使机器人具备工具及其组成部件的功能用途 (功用性, Affordance) 认知能力, 对机器人从感知到认知的主动智能提升具有重要意义[2].

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, 基于学习的功用性主动认知方法研究刚刚出现. 特别是近年来随着 RGB-D 传感器 (如 Kinect) 的出现, 3D 数据的获取更加方便快捷, 极大地推动了功用性检测领域的研究. Lenz 等学习工具中可供机器人抓取的部位[3], Kjellstrom 等通过学习人手操作来分类所用工具[4], Grabner 等通过 3D 数据检测出可供坐的曲面[5], 文献 [6] 将工具功用性看做相互关联的整体, 通过马尔科夫随机场建模工具与人的操作, 文献 [7] 运用结构随机森林 (Structured random forest, SRF) 和超像素分层匹配追踪 (S-HMP) 方法检测家庭常见工具的 7 种功用性部件 (grasp、cut、scoop、contain、pound、support 和 wrap-grasp), 上述方法均提取彩色图像或深度图像中的特征加以建模, 但没有考虑图像块间的空间上下文信息. 文献 [8] 考虑部件间的空间结构, 针对目标轮廓进行几何特征稀疏表示与分级检测. Redmon 等提出采用卷积神经网络 (Convolutional neural network, CNN) 识别工具[9], 文献 [10] 研究多模特征深度学习与融合方法, 以实现最优抓取判别, Myers 通过双流卷积神经网络 (Two-stream CNN), 将几何特征与材质信息相结合用于功用性检测[11], Nguyen 等以端到端的方式利用深度特征训练 CNN, 并通过 CNN 中的编解码装置保证标签平滑性[12]. 但上述深度学习方法均需较高的硬件配置 (GPU 环境). 文献 [13] 仅利用结构随机森林 (SRF) 训练功用性部件检测模型, 基本实现了无 GPU 配置下的实时检测. Thogersen 等通过联合随机森林与条件随机场 (Conditional random field, CRF) 实现室内各功能区的分割[14], 其中 CRF 的引入有效地整合了空间上下文以描述区域关联性, 但文献 [14] 缺少对特征有效性的判别而文献 [13] 仅依靠经验选取关键特征, 两者均可通过采用更加通用的特征编码方法来提升信息的有效性.

稀疏编码已成功应用于图像表示和模式识别等诸多领域, 通过将普通稠密特征转化为稀疏表达形式从而使学习任务得到简化, 使模型复杂度得到降低[15]. 显著性计算领域的研究结果表明, 对 CRF 和稀疏编码的联合学习比两种方法顺序处理性能更好[16]. 借鉴该理论, 本文针对功用性检测问题, 整合 CRF 刻画空间上下文能力和稀疏编码特征约简的优点, 综合考虑两者间的耦合关系, 设计其联合条件概率表示与解耦策略, 继而给出了基于联合学习的算法实现.

1 问题描述与分析

本文研究深度图像中工具部件功用性检测问题, 即给定一幅深度图像, 试图得知其中是否存在某类待检测功用性部件. 针对此问题, 提出了功用性部件字典的概念, 并将稀疏编码用于工具部件功用性特征表示. 此外, 显著性计算和目标跟踪等研究均表明, 如果一个局部块表现了很强的目标特性, 那么其附近的块也可能含有相似的性能[16−17], 遵循这一法则, 针对该功用性字典在描述空间上下文方面的不足, 引入条件随机场 (CRF) 来表征这种空间临域关系, 从而构建出一个自上而下的基于图像块稀疏编码的 CRF 模型. 但分析可知, 在该模型中 CRF 构建和稀疏编码是互相耦合的两个子问题: 一方面, CRF 中节点存储图像块的特征稀疏向量, CRF 权重向量的优化将导致特征字典的更新; 另一方面, 各图像块的特征稀疏向量则被用于计算和优化 CRF 的权重向量.

综合上述分析, 针对不同功用性部件分别训练模型, 将该部件功用性区域视为目标区域, 其他区域视为背景区域, 深度结合 CRF 与稀疏编码, 将稀疏向量作为潜变量构建 CRF, 与此同时, 通过 CRF 的调制更新字典.

2 公式化表示

本文针对深度图像展开功用性部件特征提取, 并针对不同功用性部件分别设置与深度图同尺度的二值标签文件. 深度图中, 假设某局部图像块特征向量 $\boldsymbol{x} \in \mathbf{R}^p$, p 为特征维度, 若在该图像块中存在某功用性部件, 则令该部件二值标签文件中对应位置处的标签 $y = 1$; 否则, 令 $y = -1$[18]. 则可从图像不同位置采样 m 个图像块构建特征集 $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m\}$ 作为观测值, 对应标签集合 $Y = \{y_1, y_2, \cdots, y_m\}$ 记录目标存在与否. 构建字典 $D \in \mathbf{R}^{p\times k}$ 用于存储从训练样本学习得到的最具判别性的 k 个深度特征单词 $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_k\}$, 并引入潜变量 $\boldsymbol{s}_i \in \mathbf{R}^k$ 作为图像块特征 $\boldsymbol{x}_i$ 的稀疏表示, 即有 $\boldsymbol{x}_i = D\boldsymbol{s}_i$. 此稀疏表示可进一步公式化为如下最优化问题:

$$s(\boldsymbol{x}, D) = \arg\min_{\boldsymbol{s}} \frac{1}{2}\|\boldsymbol{x} - D\boldsymbol{s}\|^2 + \lambda\|\boldsymbol{s}\|_1 \qquad (1)$$

其中, λ 为控制稀疏性的参数. 令 $S(X, D) = [s(\boldsymbol{x}_1, D), \cdots, s(\boldsymbol{x}_m, D)]$ 表示所有块的潜变量, 可知 $S(X, D)$ 为关于字典 D 的函数, 且同时包含了字典和图像块特征集信息.

考虑到采样块空间连接特性, 本文创建四连接图 $G = \langle v, \varepsilon\rangle$, 其中 v 表示节点集合, ε 表示边集合, 鉴于 v 中节点只与其周围四邻接节点存在条件概率关系, 而与其他节点无关. 本文以 $S(X, D)$ 作为节点信息, 则可知在 $S(X, D)$ 条件下, 图 G 具有 Markov 性[16], 即可用如下的条件概率作为 CRF 公

式:

$$P\left(Y|S\left(X,D\right),\boldsymbol{w}\right)=\frac{1}{Z}\mathrm{e}^{-E(S(X,D),Y,\boldsymbol{w})} \tag{2}$$

其中, Z 为配分函数, $E\left(S\left(X,D\right),Y,\boldsymbol{w}\right)$ 为能量函数, 其可分解为节点能量项与边能量项[19−20]. 对于每一个节点 $i\in v$, 该节点能量由稀疏编码的总贡献计算得到, 即 $\psi\left(s\left(\boldsymbol{x}_i,D\right),y_i,\boldsymbol{w}_1\right)=-y_i\boldsymbol{w}_1^{\mathrm{T}}s\left(\boldsymbol{x}_i,D\right)$, 其中 $\boldsymbol{w}_1\in\mathbf{R}^k$ 是权重向量. 对于每一条边 $(i,j)\in\varepsilon$, 若只考虑数据间的平滑性, 则有 $\psi\left(y_i,y_j,w_2\right)=w_2\oplus\left(y_i,y_j\right)$, 其中 w_2 表示标签平滑性的权重, $\oplus$ 表示异或运算.

因此, 随机能量场可详写为:

$$\begin{aligned}E\left(S\left(X,D\right),Y,\boldsymbol{w}\right)=&\sum_{i\in v}\psi\left(s\left(\boldsymbol{x}_i,D\right),y_i,\boldsymbol{w}_1\right)+\\&\sum_{(i,j)\in\varepsilon}\psi\left(y_i,y_j,w_2\right)\end{aligned} \tag{3}$$

其中, $\boldsymbol{w}=[\boldsymbol{w}_1;w_2]$.

由前面式 (2) 可知, 学习 CRF 权重 $\boldsymbol{w}$ 与字典 D 为两个相互耦合的子问题. 给出 CRF 权重 $\boldsymbol{w}$, 式 (2) 的模型可以看作是 CRF 监督下的字典学习; 给出字典 D , 则可看作是基于稀疏编码的 CRF 调制. 在此模型中, 通过求解下面的边缘概率来计算节点 $i\in v$ 的目标概率[21]:

$$p\left(y_i|s\left(\boldsymbol{x}_i,D\right),\boldsymbol{w}\right)=\sum_{y_{N(i)}}p\left(y_i,y_{N(i)}|s\left(\boldsymbol{x}_i,D\right),\boldsymbol{w}\right) \tag{4}$$

其中, $N\left(i\right)$ 表示图像上结点 i 的邻居节点. 若定义图像块 i 中目标存在的概率为:

$$u\left(s\left(\boldsymbol{x}_i,D\right),\boldsymbol{w}\right)=p\left(y_i=1|s\left(\boldsymbol{x}_i,D\right),\boldsymbol{w}\right) \tag{5}$$

则最终图像中存在某种功用性部件的概率图为:

$$U\left(S\left(X,D\right),\boldsymbol{w}\right)=\{u_1,u_2,\cdots,u_m\} \tag{6}$$

3　模型优化与解耦求解

假设由 N 幅深度图构成的训练样本集为 $\chi=\left\{X^{(1)},X^{(2)},\cdots,X^{(N)}\right\}$, 其对应标签为 $\psi=\left\{Y^{(1)},Y^{(2)},\cdots,Y^{(N)}\right\}$, 本文旨在学习 CRF 参数 $\hat{\boldsymbol{w}}$ 和字典 $\hat{D}$ 来获得训练样本的最大联合似然估计:

$$\max_{\boldsymbol{w}\in\mathbf{R}^{(k+1)},D\in\Omega,S\left(X^{(n)},D\right)}\prod_{n=1}^{N}P\left(Y^{(n)}|S\left(X^{(n)},D\right),\boldsymbol{w}\right) \tag{7}$$

其中, Ω 为满足如下约束的字典集合:

$$\Omega=\left\{D\in\mathbf{R}^{p\times k},\left\|\boldsymbol{d}_j\right\|_2\leq1,\forall j=1,2,\cdots,k\right\} \tag{8}$$

3.1　模型优化

对于上节式 (7), 考虑到从有限的训练样本学习大量参数较为困难, 参考 Max-margin CRF 学习方法[22], 我们将似然最大化转化为不等式约束优化问题以追求最优的 $\boldsymbol{w}$ 和 D, 则对于所有 $Y\neq Y^{(n)},n=1,2,\cdots,N$, 有:

$$P\left(Y^{(n)}|S\left(X^{(n)},D\right),\boldsymbol{w}\right)\geq P\left(Y|S\left(X^{(n)},D\right),\boldsymbol{w}\right) \tag{9}$$

在此约束优化的条件下可将两边的配分函数 Z 去掉, 表示为能量项的形式:

$$E\left(S\left(X^{(n)},D\right),Y^{(n)},\boldsymbol{w}\right)\leq E\left(S\left(X^{(n)},D\right),Y,\boldsymbol{w}\right) \tag{10}$$

若试图使实际的能量 $E\left(S\left(X^{(n)},D\right),Y^{(n)},\boldsymbol{w}\right)$ 比任意 $E\left(S\left(X^{(n)},D\right),Y,\boldsymbol{w}\right)$ 都小[23], 则可令:

$$\begin{aligned}&E\left(S\left(X^{(n)},D\right),Y^{(n)},\boldsymbol{w}\right)\leq\\&\quad E\left(S\left(X^{(n)},D\right),Y,\boldsymbol{w}\right)-\Delta\left(Y,Y^{(n)}\right)\end{aligned} \tag{11}$$

本文中定义 Margin 函数为 $\Delta\left(Y,Y^{(n)}\right)=\sum_{i=1}^{m}I\left(y_i,y_i^{(m)}\right)$. 通过寻求最违反约束来求解:

$$\hat{Y}^{(n)}=\arg\min_{Y}E\left(S\left(X^{(n)},D\right),Y,\boldsymbol{w}\right)-\Delta\left(Y,Y^{(n)}\right) \tag{12}$$

因此, 对式 (7) 中权值 $\boldsymbol{w}$ 和字典 D 的学习可通过最小化如下目标损失函数来实现:

$$\min_{\boldsymbol{w},D\in\Omega}\left\{\frac{\gamma}{2}\|\boldsymbol{w}\|^2+\sum_{n=1}^{N}l^{(n)}\left(\boldsymbol{w},D\right)\right\} \tag{13}$$

其中, $l^{(n)}\left(\boldsymbol{w},D\right)=E\left(S\left(X^{(n)},D\right),\hat{Y}^{(n)},\boldsymbol{w}\right)-E\left(S\left(X^{(n)},D\right),Y^{(n)},\boldsymbol{w}\right)$, 参数 γ 控制 $\boldsymbol{w}$ 的标准化.

3.2　CRF 权重求解

本文采用 Adam 算法[24] 来优化式 (13) 中的目标损失函数, 从中解耦出 CRF 并计算其权重. 当潜变量 $S\left(X,D\right)$ 已知时, 式 (3) 中能量函数 $E\left(Y,S\left(X,D\right),\boldsymbol{w}\right)$ 对权值 $\boldsymbol{w}$ 是线性的, 则可进一步表示为:

$$E\left(Y,S\left(X,D\right),\boldsymbol{w}\right)=\left\langle\boldsymbol{w},f\left(S\left(X,D\right),Y\right)\right\rangle \tag{14}$$

其中, $f\left(S\left(X,D\right),Y\right)=\left[-\sum_{i\in v}s\left(\boldsymbol{x}_i,D\right)y_i;\sum_{(i,j)\in\varepsilon}I\left(y_i,y_j\right)\right]$, 则可得目标损失函数 (13) 中 CRF 权重向量 $\boldsymbol{w}$ 的梯度函数, 记为:

$$\begin{aligned}g(\boldsymbol{w})=\frac{\partial l^n}{\partial\boldsymbol{w}}=&f\left(S\left(X^{(n)},D\right),\hat{Y}^{(n)}\right)-\\&f\left(S\left(X^{(n)},D\right),Y^{(n)}\right)+\gamma\boldsymbol{w}\end{aligned} \tag{15}$$

对式 (15) 采用 Adam 算法加以求解. 若第 t 次迭代的梯度值记为 $g^{(n)}(\boldsymbol{w}^{(t-1)})$, 有偏的第一时刻向量记为 $\boldsymbol{m}^{(t)}$, 有偏的第二时刻向量记为 $\boldsymbol{v}^{(t)}$, 则有:

$$\begin{aligned}\boldsymbol{m}^{(t)} &= \beta_1 \boldsymbol{m}^{(t-1)} + (1-\beta_1)\cdot g^{(t)}\left(\boldsymbol{w}^{(t-1)}\right),\\ \boldsymbol{v}^{(t)} &= \beta_2 \boldsymbol{v}^{(t-1)} + (1-\beta_2)\cdot \left(g^{(t)}\left(\boldsymbol{w}^{(t-1)}\right)\right)^2\end{aligned} \tag{16}$$

式中, β_1, β_2 分别为某接近 1 的固定参数. 对上式进行偏差校正, 令

$$\hat{\boldsymbol{m}}^{(t)} = \frac{\boldsymbol{m}^{(t)}}{\left(1-{\beta_1}^t\right)}, \quad \hat{\boldsymbol{v}}^{(t)} = \frac{\boldsymbol{v}^{(t)}}{\left(1-{\beta_2}^t\right)} \tag{17}$$

则在第 t 次迭代后的 CRF 权重更新公式如下:

$$\boldsymbol{w}^{(t)} = \boldsymbol{w}^{(t-1)} - \alpha\cdot\frac{\hat{\boldsymbol{m}}^{(t)}}{\sqrt{\hat{\boldsymbol{v}}^{(t)}}} \tag{18}$$

式中, α 为固定参数, 其与 $\hat{\boldsymbol{m}}^{(t)}$, $\hat{\boldsymbol{v}}^{(t)}$ 联合构成可自适应动态调整的学习率函数.

3.3 字典求解

对于字典 D, 本文使用链式法则[25] 来计算 l^n 对 D 的微分:

$$\frac{\partial l^n}{\partial D} = \sum_{i\in v}\left(\frac{\partial l^n}{\partial s\left(\boldsymbol{x}_i, D\right)}\right)^{\mathrm{T}}\frac{\partial s\left(\boldsymbol{x}_i, D\right)}{\partial D} \tag{19}$$

建立式 (1) 的不动点方程:

$$D^{\mathrm{T}}\left(D\boldsymbol{s}-\boldsymbol{x}\right) = -\lambda \mathrm{sgn}\left(\boldsymbol{s}\right) \tag{20}$$

其中 $\mathrm{sgn}(\boldsymbol{s})$ 以逐点的方式表示 $\boldsymbol{s}$ 的符号, 且 $\mathrm{sgn}(0)=0$. 式 (20) 两端分别对 D 求导得:

$$\frac{\partial \boldsymbol{s}_\Lambda}{\partial D} = \left(D_\Lambda^{\mathrm{T}} D_\Lambda\right)^{-1}\left(\frac{\partial D_\Lambda^{\mathrm{T}}\boldsymbol{x}}{\partial D} - \frac{\partial D_\Lambda^{\mathrm{T}} D_\Lambda}{\partial D}\right) \tag{21}$$

其中, Λ 表示 $\boldsymbol{s}$ 的非零编码索引集, $\bar{\Lambda}$ 表示零编码索引集. 为每个 $\boldsymbol{s}$ 引入一个辅助变量 z 来简化式 (19):

$$z_{\bar{\Lambda}} = 0, z_\Lambda = \left(D_\Lambda^{\mathrm{T}} D_\Lambda\right)^{-1}\frac{\partial l^n}{\partial \boldsymbol{s}_\Lambda} \tag{22}$$

其中, $\partial l^n/\partial \boldsymbol{s}_\Lambda = \left(y_i-\hat{y}_i\right)\boldsymbol{w}_\Lambda$, 令 $Z = [z_1, z_2, \cdots, z_m]$, 至此得到目标损失函数 (13) 中字典 D 的梯度为:

$$\begin{aligned}g\left(D\right) = \frac{\partial l^n}{\partial D} = &-DZ\left(S\left(X,D\right)\right)^{\mathrm{T}} + \\ &\left(X - DS\left(X,D\right)\right)Z^{\mathrm{T}}\end{aligned} \tag{23}$$

此处, 同样采用 Adam 算法进行字典的求解, 求解过程与上节相同.

4 算法实现

4.1 几何特征表示与提取

本文所用特征有高斯曲率 (Gaussian curvatures)、方向梯度直方图 (Oriented gradient histograms)、梯度幅值 (Gradient magnitude)、平均曲率 (Mean curvatures)、形状指数 (Shape index)、曲度 (Curvedness) 和表面法向量 (Surface normals)[7]. 其中方向梯度直方图为 4 维特征向量, 表面法向量为 3 维特征向量, 其他特征均为 1 维向量. 将这些特征进行归一化后组合, 得到表征某图像块的工具功用性部件的 12 维特征向量. 上述特征均在家庭日常工具 1/4 下采样的深度图上计算得到, 并经由稀疏编码后作为表征某工具功用性部件的特征向量.

此外, 考虑到方向梯度直方图、梯度幅值、平均曲率、形状指数和曲度在功用性部件边缘快速检测时的重要作用, 借鉴文献 [13] 中的功用性部件边缘表示方法, 并将这些特征用结构随机森林 (SRF) 进行组织和功用性部件边缘建模, 受篇幅所限, 具体算法不再赘述.

4.2 基于联合学习的模型构建算法

在对 CRF 和稀疏编码耦合分析与求解基础上, 采用联合学习的方法分别对每类功用性部件构建模型, 该模型包括了最宜于表征该功用性部件的字典原子及 CRF 权重向量. 下面给出模型构建的完整算法.

算法 1. 基于联合学习的模型构建算法

输入. χ (训练图像集), ψ (真实标签集), $D^{(0)}$ (初始字典);

$\boldsymbol{w}^{(0)}$ (初始 CRF 权重), λ (在式 (1) 中), T (循环次数);

γ (在式 (13) 中)

输出. $\hat{D}$ 和 $\hat{\boldsymbol{w}}$

1 for $t = 1, \cdots, T$ do

2 /* 依次训练样本集合 (χ, ψ)*/

3 for $n = 1, \cdots, N$ do /* N 为 χ 中深度图像的数量 */

4 通过式 (1) 评估潜变量 $s\left(\boldsymbol{x}_i, D\right), \forall i \in V$;

5 通过式 (12) 解出最违反标签 $\hat{Y}^{(n)}$;

6 采用 Adam 算法通过式 (18) 更新 CRF 权重 $\boldsymbol{w}^{(t)}$;

7 为 $s\left(\boldsymbol{x}_i, D\right)$ 找到有效集 $\Lambda_i, \forall i \in V$;

8 通过式 (22) 计算辅助变量 z_i;

9 采用 Adam 算法更新字典 $D^{(t)}$;

10 通过式 (8) 在 Ω 上对 $D^{(t)}$ 进行正则化;

11　　end for
12　end for
13　$\hat{D} \leftarrow D^{(t)}$, $\hat{\boldsymbol{w}} \leftarrow \boldsymbol{w}^{(t)}$

4.3　功用性部件在线检测

通过前面的离线建模阶段, 得到了最具判别性的特征字典和 CRF 权重向量. 在线检测过程中, 利用工具部件功用性边缘检测器计算功用性的外接矩形区域, 在此区域内以特征稀疏表示作为图像节点信息, 在联合 CRF 图模型与稀疏编码的基础上利用置信度传播算法完成图像的语义分割, 至此得到每个图像块属于目标的概率, 进而产生目标功用性概率图 $U = \{u_1, u_2, \cdots, u_m\}$, 其中, 概率大于某一阈值的区域即为目标区域, 反之则为背景区域.

5　实验及结果分析

5.1　实验数据集

为验证本文理论推导和算法实现的正确性, 使用文献 [7] 中的数据集检测并分类其中的家庭工具功用性部件, 该数据集中包括厨房、园艺和工作间共 17 类 105 种家庭日常工具的 RGB-D 信息, 涵盖了 grasp、wrap-grasp、cut、scoop、contain、pound、support 共 7 种功用性. 图 1 给出了数据集内的部分工具示例, 图 2 给出了示例工具所具有的功用性部件, 可以直观看出, 每类工具都可视为若干功用性部件的集合, 而同一功用性部件则可能出现在不同工具中.

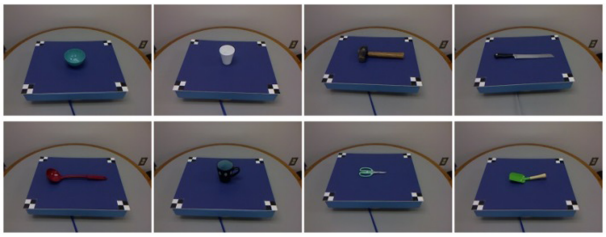


图 1　RGB-D 数据集中部分工具

Fig. 1　Tools in RGB-D data set

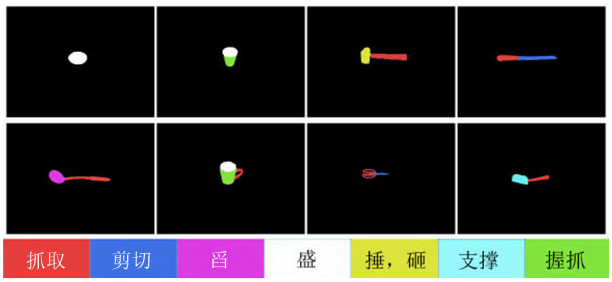


图 2　工具目标部件功用性区域

Fig. 2　Target affordance parts in tools

针对某种功用性部件, 在数据集中选取包含该功用性部件的各类工具的不同角度 Depth 图像以及已标记该功用性部件的二值标签文件作为训练样本. 从功用性角度出发, 图 3 直观地给出了包含功用性部件 “盛 (Contain)” 的工具及其对应的二值标签.

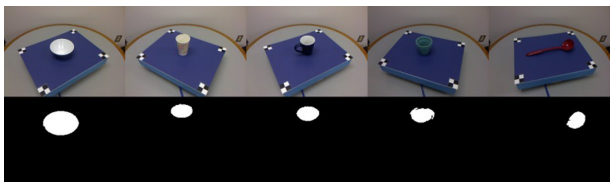


图 3　包含功用性部件 “盛 (Contain)” 的工具及其对应的二值标签

Fig. 3　Tools containing affordance of “contain” and the corresponding labels in binaryzation

5.2　实验条件与配置

将深度图像 1/4 下采样后作为训练样本, 其中每个像素视为一个图像块. 训练过程中, 收集所有块的几何特征, 并使用 K-means 算法初始化字典 $D^{(0)}$. 基于字典计算特征稀疏表示, 并将其作为潜变量与对应标签进行训练得到一个线性 SVM (Support vector machine), 利用此 SVM 初始化 CRF 结点能量权重 $\boldsymbol{w}_1^{(0)}$, 并将边能量权重 $\boldsymbol{w}_2^{(0)}$ 设置为 1. 所有模型训练 3 个周期, 训练得到表征该功用性部件的字典与 CRF 权重向量. 基于该模型进行功用性部件检测和定位, 产生目标功用性存在的概率图, 将概率值大于等于 0.5 的图像块认定为目标块, 将概率值小于 0.5 的块认定为背景块. 本文算法运行于 Windows 7 操作系统, 双核 3.20 GHz CPU, 内存为 8 GB.

5.3　实验结果及分析

本文依次构建上文提到的 4 种功用性 contain、scoop、support 与 wrap-grasp 的部件检测模型. 仅使用文献 [15] 的稀疏编码并分别采用 SIFT (Scale invariant feature transform) 特征和深度特征得到的检测结果如图 4 (c) 和图 4 (d) 所示, 使用文献 [16] 的联合学习方法并分别采用 SIFT 特征和深度特征得到的检测结果如图 4 (e) 和图 4 (f) 所示, 采用深度特征并使用文献 [7] 方法和文献 [13] 方法得到的检测结果如图 4 (g) 和图 4 (h) 所示, 使用本文方法的检测结果如图 4 (i) 所示. 通过对比可以直观看出, 相较于 SIFT 特征, 深度特征能够更加有效地表征工具的功用性部件, 且相较于仅采用稀疏编码方法、SRF 方法以及传统的 CRF 与稀疏编码结合的方法, 本文通过对多类深度特征进行稀疏编码, 同时采用 CRF 表征特征空间关系, 使得检测效果获得了不同程度的提升.

为了进一步定量评定本文方法的性能, 图 5 给出了采用不同特征及不同方法所得到的精度召回率曲线. 可以看出, 采用 SIFT 特征表征功用性部件时,

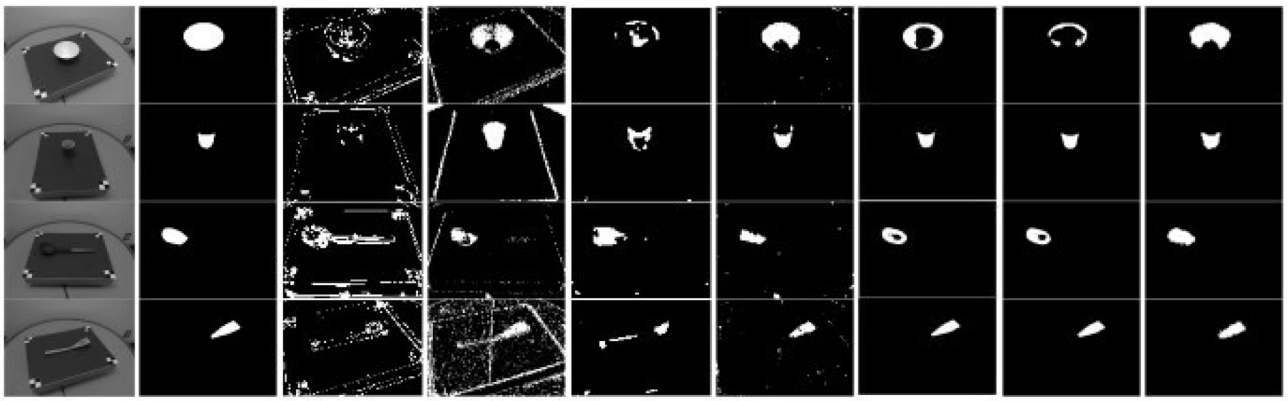
其精度和召回率普遍低于采用深度特征表征功用性部件. 本文算法采用深度特征及性能更优的 Adam 优化算法, 对 4 种功用性部件的检测效果普遍都较好, 总体性能优于现有方法.

为了评判不同算法的效率, 表 1 给出了本文方法与其他已有方法的用时对比. 实验过程中, 文献 [7] 方法需先将深度数据做较为费时的平滑预处理, 再提取深度特征并交由训练好的 SRF 模型进行功用性判别; 文献 [13] 中采用功用性部件边缘检测器快速定位目标区域, 有效提升了检测效率; 文献 [15−16] 方法本用于处理 SIFT 特征和显著性检测, 但针对功用性部件建模深度特征较 SIFT 特征更具优势, 在 Depth 图像中多类深度特征的提取速度稍慢于在 RGB 图像中 SIFT 特征. 本文从 Depth 图像中提取多类深度特征, 采用功用性部件边缘检测器快速定位目标区域, 加之采用能够快速收敛的 Adam 算法, 因此取得了较为理想的检测效率.

此外, 需要说明的是, 深度学习方法已被用于功用性部件的学习和检测, 并取得了与本文相当的识别准确率, 但该类方法的运行均需 GPU 支持, 如文献 [9] 的 CNN 方法运行于 NVIDIA Tesla K20 GPU 环境下, 文献 [12] 的 CNN 方法运行于 NVIDIA Titan X GPU 环境下, 两者的识别速度均达到毫秒级, 但在普通配置的 CPU 上无法运行. 文献 [3] 的 SAE (Sparse auto-encoder) 方法虽可运行于 CPU 环境, 但算法运行耗时较长 (如功用性部件 grasp 的检测用时约几十分钟), 无法满足服务机器人任务的实时性要求.

6 结论

机器人与人的共融, 将成为下一代机器人的本质特征. 事实上, 功用性语义频繁出现在人们的日常思维和交互中, 功用性认知也已成为了人机和谐共融的必然要求. 本文利用工具的多类深度特征, 结合



(a) (b) (c) (d) (e) (f) (g) (h) (i)

图 4 本文方法与其他方法的检测结果对比图 ((a) 为单一场景下的待检测工具图, 由上到下分别为碗 (bowl)、杯子 (cup)、勺子 (ladle)、铲子 (turner); (b) 为待检测目标功用性部件的真实值图, 由上到下分别为盛 (contain)、握抓 (wrap-grasp)、舀 (scoop)、支撑 (support); (c) SIFT + 文献 [15] 方法检测结果; (d) 深度特征 + 文献 [15] 方法检测结果; (e) SIFT + 文献 [16] 方法检测结果; (f) 深度特征 + 文献 [16] 方法检测结果; (g) 深度特征 + 文献 [7] 方法检测结果; (h) 深度特征 + 文献 [13] 方法检测结果; (i) 本文方法检测结果)

Fig. 4 Comparison of detection results between our method and others ((a) Tools in a single scene, from the top to the bottom: bowl, cup, ladle and turner; (b) Ground truth of object affordances, from the top to the bottom: contain、wrap-grasp、scoop、support; (c) Detection result with SIFT + Paper [15]; (d) Detection result with Depth + Paper [15]; (e) Detection result with SIFT + Paper [16]; (f) Detection result with Depth + Paper [16]; (g) Detection result with Depth + Paper [7]; (h) Detection result with Depth + Paper [13]; (i) Detection result with our method)

表 1 本文方法与其他方法的效率对比 (秒)

Table 1 Comparison of efficiency between our method and others (s)

功用性部件	SIFT 特征 + 文献 [15]	SIFT 特征 + 文献 [16]	深度特征 + 文献 [15]	深度特征 + 文献 [16]	深度特征 + 文献 [13]	深度特征 + 文献 [7]	Ours
盛	6.46	8.00	9.41	10.95	1.25	16.29	1.13
舀	6.09	7.09	8.60	10.67	1.18	16.34	1.33
支撑	5.94	6.93	10.40	10.98	1.53	16.28	1.56
握抓	5.93	6.99	10.65	11.73	1.27	15.52	1.24

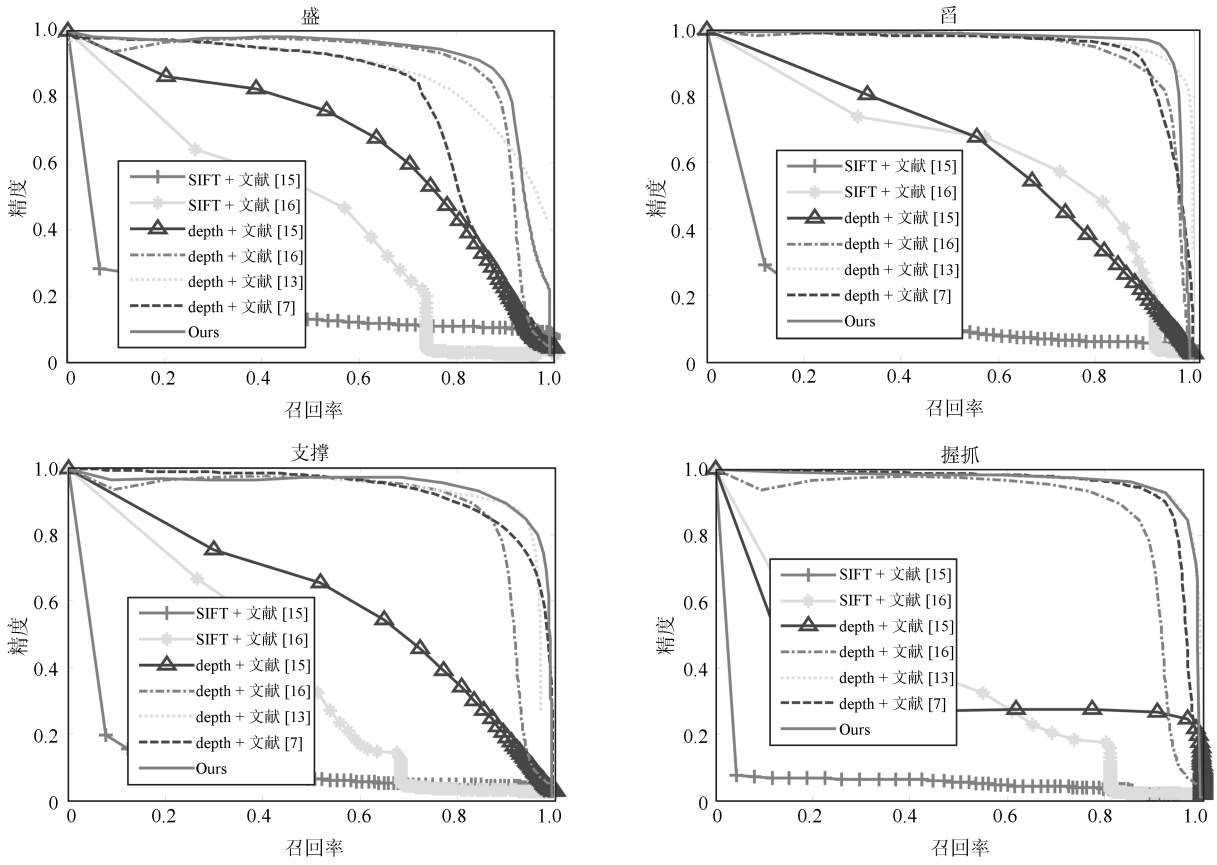


图 5　本文方法与其他方法的精度召回率曲线对比

Fig. 5　Comparison of precision recall curves between our method and others

稀疏编码与 CRF 优势训练家庭日常工具功用性部件的检测模型, 通过与利用 SIFT 特征表示图像信息和传统联合 CRF 与稀疏编码训练模型的算法进行比较, 由精度召回率曲线可知本文模型对工具部件的目标功用性检测效果良好, 为机器人工具功能认知及后续人机共融和自然交互奠定基础.

References

1 Aly A, Griffiths S, Stramandinoli F. Towards intelligent social robots: current advances in cognitive robotics. *Cognitive Systems Research*, 2017, **43**: 153−156

2 Min H Q, Yi C A, Luo R H, Zhu J H, Bi S. Affordance research in developmental robotics: a survey. *IEEE Transactions on Cognitive and Developmental Systems*, 2016, **8**(4): 237−255

3 Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, 2015, **34**(4−5): 705−724

4 Kjellström H, Romero J, Kragić D. Visual object-action recognition: inferring object affordances from human demonstration. *Computer Vision and Image Understanding*, 2011, **115**(1): 81−90

5 Grabner H, Gall J, Van Gool L. What makes a chair a chair? In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011. 1529−1536

6 Koppula H S, Gupta R, Saxena A. Learning human activities and object affordances from RGB-D videos. *The International Journal of Robotics Research*, 2013, **32**(8): 951−970

7 Myers A, Teo C L, Fermüller C, Aloimonos Y. Affordance detection of tool parts from geometric features. In: Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Seattle, WA: IEEE, 2015. 1374−1381

8 Li Yu-Dong, He Hong-Jie, Chen Fan, Yin Zhong-Ke. A rigid object detection model based on geometric sparse representation of profile and its hierarchical detection algorithm. *Acta Automatica Sinica*, 2015, **41**(4): 843−853
(林煜东, 和红杰, 陈帆, 尹忠科. 基于轮廓几何稀疏表示的刚性目标模型及其分级检测算法. 自动化学报, 2015, **41**(4): 843−853)

9 Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks. In: Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Seattle, WA: IEEE, 2015. 1316−1322

10 Zhong Xun-Gao, Xu Min, Zhong Xun-Yu, Peng Xia-Fu. Multimodal features deep learning for robotic potential grasp recognition. *Acta Automatica Sinica*, 2016, **42**(7): 1022−1029
(仲训杲, 徐敏, 仲训昱, 彭侠夫. 基于多模特征深度学习的机器人抓取判别方法. 自动化学报, 2016, **42**(7): 1022−1029)

11 Myers A O. From form to function: detecting the affordance of tool parts using geometric features and material cues [Ph. D. dissertation], University of Maryland, 2016

12 Nguyen A, Kanoulas D, Caldwell D G, Tsagarakis N G. Detecting object affordances with Convolutional Neural Networks. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. Daejeon: IEEE, 2016. 2765−2770

13 Wu Pei-Liang, Fu Wei-Xing, Kong Ling-Fu. A fast algorithm for affordance detection of household tool parts based on structured random forest. *Acta Optica Sinica*, 2017, **37**(2): 0215001
(吴培良, 付卫兴, 孔令富. 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. 光学学报, 2017, **37**(2): 0215001)

14 Thogersen M, Escalera S, González J, Moeslund T B. Segmentation of RGB-D indoor scenes by stacking random forests and conditional random fields. *Pattern Recognition Letters*, 2016, **80**: 208−215

15 Bao C L, Ji H, Quan Y H, Shen Z W. Dictionary learning for sparse coding: algorithms and convergence analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(7): 1356−1369

16 Yang J M, Yang M H. Top-down visual saliency via joint CRF and dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(3): 576−588

17 Yang E, Gwak J, Jeon M. Conditional random field (CRF)-boosting: constructing a robust online hybrid boosting multiple object tracker facilitated by CRF learning. *Sensors*, 2017, **17**(3): 617

18 Liu T, Huang X T, Ma J S. Conditional random fields for image labeling. *Mathematical Problems in Engineering*, 2016, **2016**: Article ID 3846125

19 Lv P Y, Zhong Y F, Zhao J, Jiao H Z, Zhang L P. Change detection based on a multifeature probabilistic ensemble conditional random field model for high spatial resolution remote sensing imagery. *IEEE Geoscience & Remote Sensing Letters*, 2016, **13**(12): 1965−1969

20 Qian Sheng, Chen Zong-Hai, Lin Ming-Qiang, Zhang Chen-Bin. Saliency detection based on conditional random field and image segmentation. *Acta Automatica Sinica*, 2015, **41**(4): 711−724
(钱生, 陈宗海, 林名强, 张陈斌. 基于条件随机场和图像分割的显著性检测. 自动化学报, 2015, **41**(4): 711−724)

21 Wang Z, Zhu S Q, Li Y H, Cui Z Z. Convolutional neural network based deep conditional random fields for stereo matching. *Journal of Visual Communication & Image Representation*, 2016, **40**: 739−750

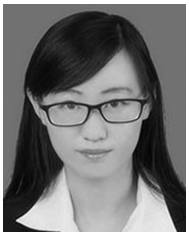
22 Szummer M, Kohli P, Hoiem D. Learning CRFs using graph cuts. In: Proceedings of European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 5303. Berlin, Heidelberg: Springer, 2008. 582−595

23 Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2004, **26**(2): 147−159

24 Kingma D P, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference for Learning Representations. San Diego, 2015.

25 Mairal J, Bach F, Ponce J. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, **34**(4): 791−804

吴培良 燕山大学副教授. 2010 年获得燕山大学博士学位. 主要研究方向为家庭服务机器人智能提升, 功用性认知, SLAM. 本文通信作者.
E-mail: peiliangwu@ysu.edu.cn
(**WU Pei-Liang** Associate professor at Yanshan University. He received his Ph. D. degree from Yanshan University in 2010. His research interest covers intelligence promotion home service robot, affordance cognition, and SLAM. Corresponding author of this paper.)



隰晓珺 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为 RGB-D 数据处理, 工具功用性认知.
E-mail: xixiaojun@ysu.edu.cn
(**XI Xiao-Jun** Master student at the School of Information Science and Engineering, Yanshan University. Her research interest covers RGB-D data processing and tools affordance cognition.)

杨 霄 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为 RGB-D 数据处理, 行为建模与学习.
E-mail: yangxiao@ysu.edu.cn
(**YANG Xiao** Master student at the School of Information Science and Engineering, Yanshan University. His research interest covers RGB-D data processing, human behavior modeling and learning.)

孔令富 燕山大学教授. 1995 年获得哈尔滨工业大学博士学位. 主要研究方向为家庭服务机器人, 机器视觉, 智能信息处理, 并联机器人及自动控制.
E-mail: lfkong@ysu.edu.cn
(**KONG Ling-Fu** Professor at Yanshan University. He received his Ph. D. degree from Harbin Institute of Technology in 1995. His research interest covers home service robot, machine vision, intelligent information processing, parallel robotics, and automatic control.)

侯增广 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研究员. 主要研究方向为机器人与智能系统, 康复机器人与微创介入手术机器人.
E-mail: zengguang.hou@ia.ac.cn
(**HOU Zeng-Guang** Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interest covers intelligent robotic systems, rehabilitation and surgery robots.)