

## 基于空间金字塔池化特征的日常工具分类识别

吴培良<sup>1,2,3†</sup>, 何 犇<sup>1</sup>, 侯增广<sup>2</sup>

(1. 燕山大学 信息科学与工程学院, 河北 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, 河北 秦皇岛 066004)

**摘 要:** 面向人机共融环境下机器智能对工具认知的需要, 为提高家庭服务机器人的工具功能用途认知能力, 设计一种基于深度几何特征空间金字塔池化的工具功用性建模与分类方法. 离线训练阶段, 考虑到各类工具在几何形态上的差异对工具自身更具表征性, 首先, 在各工具的深度图上提取多类几何特征, 并融合形成工具特征图; 然后, 在工具特征图上提取多尺度特征块, 并基于 cciPCA 的方法建立空间池化金字塔, 从而构建最终的工具体征向量; 最后, 在高层语义空间上, 利用 SVM 分类器训练工具分类识别模型. 在线检测阶段, 利用离线训练的工具分类模型对空间池化的样本进行分类测试. 实验结果表明, 所提方法能够实现家庭服务机器人对家庭日常工具的认知及分类识别, 部分工具的识别精度可达 97% 及以上.

**关键词:** 几何特征; cciPCA; 多尺度特征块; 空间金字塔池化; SVM 分类器; 工具分类

中图分类号: TP242.6

文献标志码: A

## Household tools classification recognition based on spatial pyramid pooling features

WU Pei-liang<sup>1,2,3†</sup>, HE Ben<sup>1</sup>, HOU Zeng-guang<sup>2</sup>

(1. School of Information Science and Technology, Yanshan University, Qinhuangdao 066004, China; 2. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 3. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China)

**Abstract:** In view of the need of machine intelligence for tool recognition in man-machine communion environment, in order to improve the ability for home service robot to recognize affordances of tools, a modeling and classification method based on depth geometry feature spatial pyramid pooling is designed. In the off-line training phase, considering that the differences in the geometric shapes of various tools are more indicative for the tools themselves, firstly, the multi-types of geometric features are extracted from each tool depth map and fused to form the tool feature map. Then, the multi-scale feature patches are extracted and the spatial pooled pyramid is built based on cciPCA to construct the final tool feature vector. Finally, in the high-level semantic space, the tools classification model is trained based on the SVM classifier. In the online testing phase, the samples of spatial pooling are tested by a tool model trained offline. Experimental results show that the proposed method realizes recognition and classification of daily tools for home service robots, and the recognition accuracy of some tools reaches 97% and above.

**Keywords:** geometry feature; cciPCA; multi-scale feature patches; spatial pyramid pooling; SVM classifier; tool classification

## 0 引 言

随着人们生活品质的日益提高, 家庭服务机器人越来越多地出现在日常生活中, 人机和谐共融已成为新一代智能服务机器人的发展趋势. 人机共融环境下, 实现自然人机交互, 提高机器人的宜人化服务质量已成为机器智能研究的重要挑战. 在机器人的众

多能力中, 机器人对家庭日常工具的功能用途(功用性)认知无疑是一个关键环节. 传统的工具功用性认知采用标签告知的方式, 机器人通过阅读标签获取功能语义, 属于被动式认知的方式. 只有模拟人类认知工具的过程, 建立工具表观到语义的映射, 才能从高层语义空间上使机器人理解识别工具, 进而更好地使

收稿日期: 2017-12-24; 修回日期: 2018-03-30.

基金项目: 国家自然科学基金项目(61305113); 河北省自然科学基金项目(F2016203358); 燕山大学博士基金项目(BL18007); 中国博士后科学基金项目(2018M631620).

责任编委: 刘民.

†通讯作者. E-mail: peiliangwu@ysu.edu.cn.

用工具完成任务.

近年来,在家庭服务机器人的研究背景下,基于计算机视觉的家庭日常工具识别成为了一个非常重要的研究主题.与此同时,在家庭服务机器人工具认知层面上仍存在很大的挑战,这些挑战主要源于各工具的类内变化(颜色、角度、形状大小等),或缺少足够的训练样本.为了得到较高的识别精度,找到有效且具有高鲁棒性的各工具图像表示方法是非常关键的.Zou等<sup>[1]</sup>提出了面向识别任务的3D物体特征提取方法;Kjellstrom等<sup>[2]</sup>利用上下文中的操作,分类正在使用的工具;Zhu等<sup>[3]</sup>提出了学习底层功能的面向任务建模的新的学习框架;Mahmudul等<sup>[4]</sup>将人、物和环境3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;Edsinger等<sup>[5]</sup>研究了工具尖端的检测,从而使机器人获得最佳抓取部位;Tanis等<sup>[6]</sup>对如何以正确的方式抓取工具展开了研究;Lenz等<sup>[7]</sup>运用SAE方法检测各种物品的抓取位置;而Redmon等<sup>[8]</sup>又运用CNN方法对此进行了研究.文献[5-8]仅针对grasp这一被动功用性部件进行研究,而Nguyen等<sup>[9]</sup>提出了基于工具RGB-D图运用CNN方法对7种功用性部件(grasp、cut、scoop、contain、pound、support和wrap-grasp)提取特征,吴培良等<sup>[10]</sup>基于结构随机森林算法实现了上述7种功用性部件的快速检测.文献[9-10]分别针对7种功用性部件进行建模,但忽略了工具的整体性,并未实现工具整体的认知识别.文献[10]利用结构随机森林实现了3D工具部件功用性

快速检测,很大程度上提高了运算速率,保证了实时性,但也属于对工具功用性部件的表示及建模.

从图像分类的角度来看,图像局部特征块已经被广泛应用于图像分类中. BOF(bag-of-features)模型通过对无序局部特征的收集来表示图像<sup>[11-13]</sup>. 空间金字塔匹配方法(SPM)将空间几何信息加入分类,在图像分类中取得了显著成功. SPM首先将图像分割成不同尺度的空间网格,在第 $l$ 层有 $2^l \times 2^l$ 个单元;然后通过预先训练的词典上对每个空间单元内的局部特征(如SIFT特征)进行编码;最后进行平均或最大汇总. 尽管空间池化模型在人脸及遥感图像分类上取得成功,但是尚没有被应用于深度数据及家庭日常工具的功用性识别分类问题上.

本文针对家庭日常工具功用性认知问题展开研究,受深度卷积神经网络及人脸、遥感图像分类中空间金字塔池化方法的启发<sup>[14-16]</sup>,设计了一种基于深度几何特征的工具建模与分类方法. 考虑到不同种功用性工具的形状差异明显,针对深度图像采用空间金字塔池化方法提取表征功用性的形状特征,继而通过增量式主成分分析(cciPCA)方法降维到相同维度,利用工具特征向量训练SVM分类器,最后对工具进行分类识别.

## 1 系统框架

本文提出的家庭日常工具功能认知和分类方法包括工具的特征图构建和工具分类识别两个阶段,具体流程如图1所示.

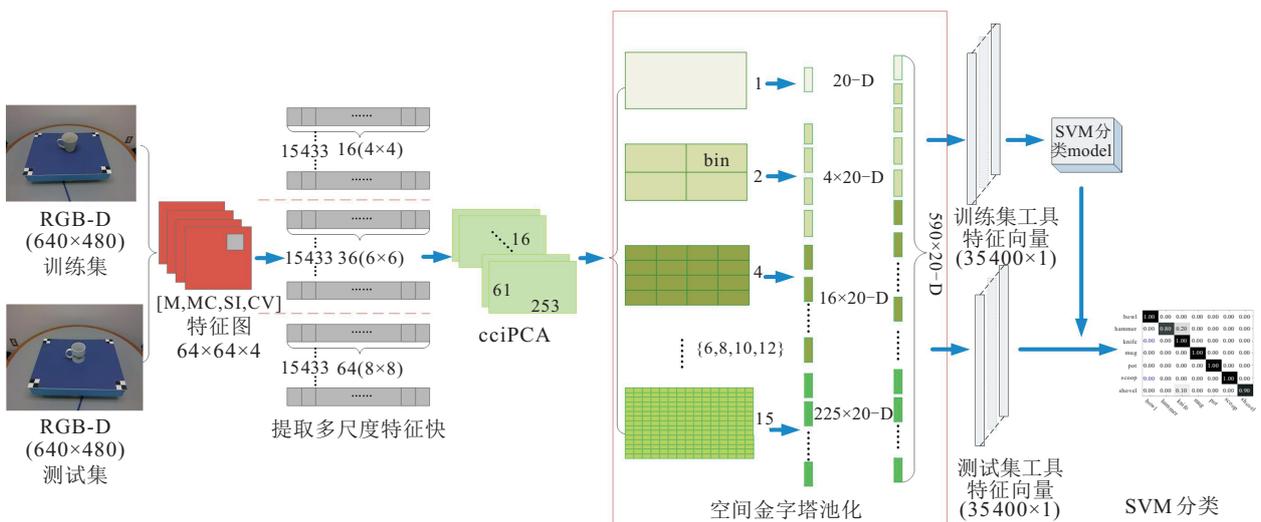


图1 基于空间金字塔池化的工具分类检测整体流程图

工具特征图构建阶段:基于各工具的深度图,分别计算各工具的几何特征形成工具特征图,然后在工具特征图上进行多尺度特征块的密集提取.为了提高分类精度,利用cciPCA对密集提取得到的局部特征块进行处理得到各工具特征图.在经过预处理的

各类工具特征图上,建立空间金字塔池化模型.本文选取了一个8层的空间池化金字塔,并在空间金字塔各层的单元上选用最大池化的方法,对金字塔各层进行池化,继而级联金字塔各层池化结果形成各工具基于空间金字塔池化的特征图.

工具分类识别阶段:采用SVM分类器建立各类工具的分类模型,并将其与mcODM等分类器的分类精度和效率进行对比分析,验证本文方法的有效性.

## 2 工具特征图的空间池化模型构建

### 2.1 工具几何特征描述

由于家庭日常工具在不同视角下的几何结构变化会影响到特征数据的有效性,数据的采集及特征提取需考虑到角度变化的鲁棒性<sup>[10]</sup>.借鉴文献[17]中利用几何特征提取算法从图像多通道中提取特征并进行特征融合的方式,本文针对梯度幅值(Gradient magnitude)、平均曲率(Mean curvatures)、形状指数(Shape index)和曲度(Curvedness)<sup>[17]</sup>4种特征进行特征提取,进而进行组合表示各类工具.

在本文选取的4种几何特征中,梯度幅值能够表示物体的边缘结构特征,描述物体的局部形状信息、位置和方向空间的量化,一定程度上可以抑制因物体视角平移和旋转所带来的影响.令 $I_x$ 和 $I_y$ 分别代表水平和垂直方向上的梯度值,则 $M(x, y) = \sqrt{I_x^2 + I_y^2}$ 即为梯度幅值.平均曲率为微分几何中反映曲面在不同方向上的弯曲程度的内蕴几何量,记为 $f_{MC}$ .在曲面的每个点,一般存在两个互相垂直的切方向,使得它们对应的法曲率 $k_1$ 和 $k_2$ 是该点所有法曲率中的最大和最小值,这两个方向称为曲面在该点的主方向,而 $k_1$ 和 $k_2$ 称为主曲率,则平均曲率可以表示为 $f_{MC} = (k_1 + k_2)/2$ .形状指数(SI)和曲度(CV)表征表面在不同方向的弯曲,映射人对形状的感知<sup>[18]</sup>,计算公式如下所示:

$$\begin{cases} SI = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right), \\ CV = \sqrt{\frac{k_1^2 + k_2^2}{2}}. \end{cases} \quad (1)$$

### 2.2 基于cciPCA的多尺度特征块提取

一个 $h \times w$ 像素的工具图,以 $s$ 像素为步长重叠提取 $r \times r$ 像素的局部块,则 $l = \left\lfloor h - \frac{r}{s} + 1 \right\rfloor$ ,  $j = \left\lfloor w - \frac{r}{s} + 1 \right\rfloor$ ,每一个工具特征图被分成 $l \times j$ 个局部特征块.将每一个局部特征块都转换成一个行向量 $X$ ,通过式 $\hat{x}_i = x_i - \frac{m}{v}$ 对 $X$ 进行归一化.其中: $x_i$ 是 $X$ 的第 $i$ 个元素, $m$ 和 $v$ 是 $X$ 的均值和标准差.密集提取得到的所有局部特征块都进行归一化后,为了去噪及降低后续空间金字塔池化过程中特征向量的维度,通过cciPCA对局部特征块进行降维,将局部特征块序列 $X(n)(n = 1, 2, \dots)$ 作为cciPCA模型的输入样本.算法1中给出了具体的降维过程.

**算法1** 基于cciPCA的多尺度特征块降维算法.

输入:局部特征块 $x(n), n = 1, 2, \dots$ ;要降到的维度 $k$ .

输出:前 $k$ 个主成分 $v_1(n), v_2(n), \dots, v_k(n)$ .

- 1) 令 $x_1(n) = x(n)$
- 2) for  $x = 1$  to  $\min(k, n) = 1$  do
- 3) if  $i = n$ .
- 4) 令第 $i$ 个主成分为 $x_i(n)$ ,即 $v_i(n) = x_i(n)$
- 5) else.
- 6)

$$v_i(n) = \frac{n-1-l}{n} v_i(n-1) + \frac{1+l}{n} x_i(n) x_i^T(n) \frac{v_i(n-1)}{\|v_i(n-1)\|}, \quad (2)$$

$$x_{i+1}(n) = x_i(n) - x_i^T(n) \frac{v_i(n)}{\|v_i(n)\|} \frac{v_i(n)}{\|v_i(n)\|}, \quad (3)$$

其中 $l$ 为遗忘因子,给新样本更大权值的同时可以逐步减弱旧样本.不带有遗忘因子的增量表达式为

$$v_i(n) = \frac{n-1}{n} v_i(n-1) + \frac{1}{n} x_i(n) x_i^T(n) \frac{v_i(n-1)}{\|v_i(n-1)\|}. \quad (4)$$

其中: $\frac{n-1}{n}$ 为估算值的权重, $\frac{1}{n}$ 为新样本数据的权重; $v_i$ 为特征值与特征向量乘积 $\lambda_i e_i$ 的估计,即特征向量 $e_i = \frac{v_i}{\|v_i\|}$ ,相对应的特征值 $\lambda_i = \|v_i\|$ , $e_1$ 即为第1个主成分.

- 7) 由式(2)对第1个主成分估算值进行更新.
- 8) 由式(4)得到样本数据与第1个主成分的残差,依此对第2个主成分估算值进行更新.
- 9) 依次进行迭代,对所有待求的主成分估算值进行更新.
- 10) end if.
- 11) end for.

式(3)计算的是正交补空间中的向量 $x_{i+1}(n)$ ,该式可加速算法收敛.本文以保证原特征空间90%的信息被保留为准则,选择将局部特征向量降到10维.

### 2.3 空间池化金字塔的构建

本文所采用的空间金字塔池化的方法与其他特征编码<sup>[13,19]</sup>方法不同之处在于,通常的特征编码方法是通过预先训练的字典对编码的特征进行池化,而本文方法直接在各类别工具图像的局部特征块上建立空间池化金字塔,不必经过稀疏编码预先训练字典.假设设置层的空间金字塔为 $\{c_1, c_2, \dots, c_l\}$ ,这样在第 $l$ 层的网格上的每一维都有 $c_l$ 个单元,则在一幅二维图像上共有 $\sum_{l=1}^L c_l^2$ 个池化单元.在应用cciPCA降维之后,每一个块的维度都变为10,基于无监督的特征学习或者是文献[13,19]中bag-of-features的方法,对于每一个提取的局部特征,编码特征的维度和字典的大小通常都是几千量级.与之不同的是,本文方法通过加大池化金字塔层数建立多层空间池化金

金字塔来保证工具识别精度. 在对各类工具进行空间金字塔池化的过程中, 本文选择最大池化的方法, 在每一个局部特征块池化单元中作求最大值的计算. 最大池化的计算公式如下所示:

$$\max \text{ pooling} : f_i = \max_j x_i^{(j)}. \quad (5)$$

其中:  $x_i^{(j)}$  是当前池化单元中第  $j$  个局部块的第  $i$  个元素;  $m$  是池化单元中块的数量,  $f = \{f_1, f_2, \dots, f_i, \dots\}$  是当前单元的池化特征.

### 3 基于线性分类器的工具功用性分类

SVM在线性核下能够取得较好的结果, 故本文将SVM用于空间金字塔池化后的工具特征向量, 工具分类识别的完整算法如下所示.

**算法2** 联合SVM线性分类器和空间金字塔池化特征的工具分类识别.

输入: 基于各类工具RGB-D深度图的训练和测试样本;

输出: 测试工具图的类别标签.

1) 分别对训练集测试集进行几何特征提取, 形成各工具特征图.

2) 定义各工具特征图的局部块及空间池化金字塔尺寸.

3) for 每一个尺寸的局部特征块 do.

4) 分别对训练集和测试集进行局部特征块的密集提取.

5) 对每一个特征块进行归一化处理.

6) 在归一化后的每一个局部特征块上使用cciPCA进行降维.

7) end for.

8) 建立8层金字塔.

9) 在金字塔每层的每个单元中进行max pooling.

10) 级联池化向量形成最后的工具特征向量  $x_i \in R^n$ .

11) 对所得到的各工具特征向量进行标准化.

12) 训练集工具的特征向量作为训练SVM分类器的样本集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ,  $x_i \in R^n$ , 类标记  $y_i \in \{+1, -1\}$ ,  $i = 1, 2, \dots, N$ .

13) 通过间隔最大化找出各类别特征样本与其他特征样本的最优分类超平面  $y(x) = w^T \Phi(x) + b$ ,  $\Phi(x)$  为特征空间转换函数. 得到相应的分类决策函数  $f(x) = \text{sign}(w^T \Phi(X) + b)$ .

14) 将测试集工具的特征向量作为分类决策函数的输入, 利用分类判决函数得出分类结果.

## 4 模型选择及实验结果

### 4.1 实验数据集

本文实验选用文献[17]中所采用的UMD part affordance数据集, 该数据集包含的工具图像数据完

整、功用性部件种类丰富、标记清晰, 其收集了包含厨房、园艺等共计17个大类的家庭日常工具(铲子(turner)、泥刀(trowel)、铁锤(tenderizer)、汤匙(spoon)、铁锹(shovel)、大剪刀(shear)、铲斗(scoop)、剪刀(scissors)、锯子(saw)、陶罐(pot)、马克杯(mug)、木槌(mallet)、勺子(ladle)、锤子(hammer)、杯子(cup)、碗(bowl)、刀(knife))的RGB-D四个通道的信息数据, 每一个大类中又包含多种颜色和形状各异同类工具, 每种工具在近300个旋转视角下进行采集, 产生了超过30000组的彩色和深度数据, 其中约三分之一的工具都进行了部件功用性的清晰标记, 是目前研究家庭日常工具功能认知和分类识别的理想数据集. 考虑到UMD Part Affordance数据集共提供了contain、pound、cut、scoop、support五种主动功用性部件以及grasp、wrap-grasp两种被动功用性部件, 本文选择了家庭常见的且涵盖上述各种功用性部件的7类工具, 即bowl、hammer、knife、mug、pot、scoop和shovel, 组成工具分类的训练集和测试集. 根据基于cciPCA的空间金字塔池化算法计算各类工具的特征向量, 进而基于各类工具的特征向量模型训练线性分类器, 实现工具有效识别及分类. 图2列举了本文选用的7种工具的类型示例, 每一种工具代表一类具有相同功能属性的工具, 忽略其外在形状和颜色差异. 在能完成同种任务的多类工具中, 本文选择一类工具, 并在其中为实验进行随机样本的选取.

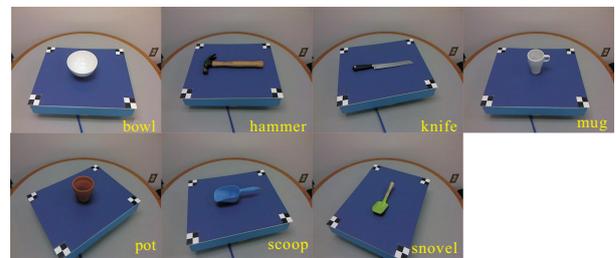


图2 实验中选用的7类工具示例

### 4.2 实验结果及分析

本文依次对图2中的7种工具进行训练和测试实验. 工具特征图构建阶段, 从7类工具的深度图上, 分别计算梯度幅值、平均曲率、形状指数、曲度4种表观几何特征并级联. 进而基于各工具的表观特征, 提取包括  $4 \times 4$ ,  $6 \times 6$ ,  $8 \times 8$  的多尺度的局部特征块, 并为了提高识别精度, 使用cciPCA的方式对各局部特征降维到10维. 接着, 对各局部特征块建立8层的空间池化金字塔, 并级联每一层金字塔的池化结果, 最后得到各工具图像的特征图.

分类识别阶段, 基于工具特征图, 训练SVM分类模型. 从数据集中任选每类工具20个, 构成共计140个待测样本, 利用训练好的分类模型进行检测, 最后

得到7类工具的分类混淆矩阵及分类准确率. 本文中, 分类准确率定义如下:  $Accuracy = \frac{\text{正确分类的样本数}}{\text{总样本数}}$ .

**实验1** 取不同空间池化金字塔层数及池化方法时的各工具分类精度实验. 在包括bowl、hammer、knife、mug、pot、scoop、shovel七种工具的深度图上进行实验. 8层的金字塔池化在 $1 \times 1, 2 \times 2, \dots, 15 \times 15$ 的常规网格单元上进行. 8层金字塔为[1, 2, 4, 6, 8, 10, 12, 15]. 实验中默认局部块的大小为 $4 \times 4$ , 密集提取的步长为1像素, 每一幅图像共提取15433个局部特征块. 表1给出了选取不同的池化金字塔层数时对工具分类精度的影响. 由此可以清楚地看出, 8层的金字塔取得了较好的分类精度. 从最大池化和平均池化方法在不同层的空间金字塔上分类准确率看: 当空间金字塔层数小于5时, 平均池化的方法对于工具分类取得了较好的效果, 这是因为当池化金字塔层数较少时, 最大池化会比平均池化损失较多的特征信息; 当金字塔层数大于5时, 选择最大池时的分类精度略高于平均池. 因此, 8层池化金字塔加max pooling的方法使得工具分类效果较好.

表1 选取不同空间池化金字塔层数时的工具分类准确率 %

方法	2层	3层	4层	5层	6层	7层	8层
max pooling	57.14	81.3	90.71	90	93.57	96.43	97.14
average pooling	79.28	81.42	90.71	94.28	84.28	92.14	91.43

**实验2** 对采用cciPCA的方法进行局部特征块降维对工具分类的影响进行实验. 实验中对从各工具的几何特征图中提取到的多尺度特征块进行降维, 将不同尺度的特征块归一化到相同的维度空间. 表2给出了分别使用PCA和cciPCA两种降维方法对特征块进行降维, 降到不同维度时工具分类准确率的变化情况. 从表2可以看出, cciPCA用于工具特征块降维时, 分类准确率一般高于PCA. 此外可以看出, 采用cciPCA降维处理后分类准确率普遍优于不做降维处理的分类方法. 因此, 采用本文的cciPCA方法进行局部特征块降维, 对于工具分类准确率的提高起着非常重要的作用.

表2 cciPCA和PCA用于 $8 \times 8$ 的局部特征块降维时工具分类准确率 %

方法	1维	5维	10维	20维	40维	60维	不降维
cciPCA	93.57	93.57	96.42	98.57	95	96.42	93.57
PCA	92.85	91.42	95.71	97.85	96.42	96.42	93.57

**实验3** 分别在各工具原像素上和深度几何特征图上进行实验. 实验过程中, 几何特征图均由梯度幅值、平均曲率、形状指数和曲度4种几何特征级联形成, 局部特征块的提取包括 $4 \times 4, 6 \times 6, 8 \times 8$ 三种尺度, 8层的金字塔池化均在 $1 \times 1, 2 \times 2, \dots, 15 \times 15$ 的常规网格单元上进行. 表3统计了不同的特征提取方法及分类器选择下的7种工具随机3次实验的分类结果. 其中: 通过第2行和第3行的分类结果对比可以看出, 在工具的深度几何特征图上进行空间金字塔池化的分类结果比在原像素上直接使用空间金字塔的分类精度有明显提高, 故深度几何特征比原像素有着更优的表征性. 通过第2行、第4行和第5行的分类结果对比可以看出, 基于工具的深度几何特征及空间金字塔池化的特征提取方法比传统BOW模型分类准确率有明显的提高. 由于高层多尺度空间池化金字塔模型的建立, 获得了各工具更为精准的特征向量表示, 为线性分类器的训练提供了较丰富的信息. 空间池化金字塔算法的加入使得工具分类准确率提升约8%. 通过第2行和第6行的分类结果对比可以看出, SVM分类效果较mcODM分类效果更好. 算法运行时间方面, 文献[9]的CNN方法无法运行于普通CPU环境下, 此外, 根据表3对实验中各类方法的耗时统计可知, 本文方法基本满足普通配置家庭服务机器人在执行任务过程中的实时性要求.

表3 采用不同工具分类方法的分类情况

方法	准确率/%			时间/ms
	第1次	第2次	第3次	
本文方法	97.14	97.85	96.42	129.34
原像素 SPP + SVM	88.57	89.28	89.28	63.82
几何特征 BOW + SVM	72.14	75.00	75.00	102.57
几何特征 + SVM	89.28	87.86	85.71	28.45
几何特征 SPP + mcODM	85.71	90.00	86.42	128.14

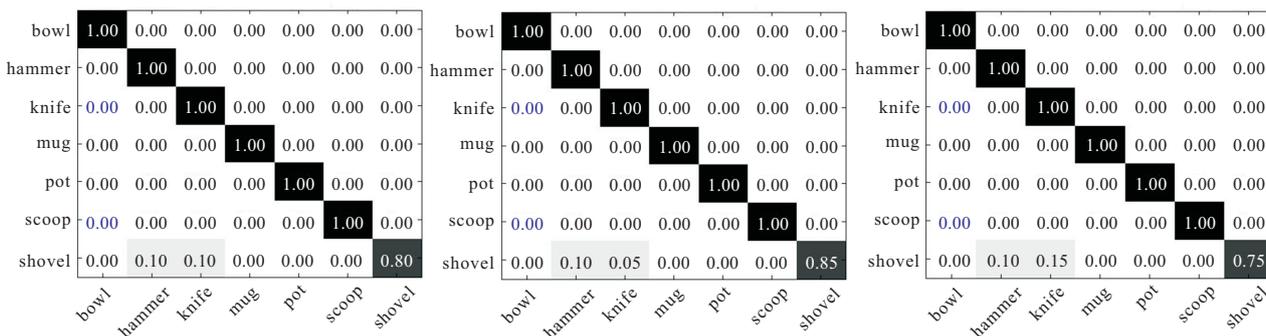


图3 3次实验的工具分类混淆矩阵

图3所示为本文方法3次实验的工具分类混淆矩阵,通过此分类混淆矩阵可以清楚地看出每类工具的具体分类情况,其中只有shovel类工具的分类结果被混淆到hammer和knife类中,其他类工具的分类结果没有出现混淆.可见,本文方法取得了较好的工具分类识别效果,能够满足对工具分类的需求.

## 5 结论

本文提出了一种基于工具深度图像几何特征和空间金字塔池化建模的家庭日常工具分类方法.本文方法对家庭日常工具取得了较好的分类效果,能够在一定程度上满足家庭服务机器人在识别家庭日常工具上的要求,为家庭服务机器人免标签环境下的工具认知,进而提供高效智能的服务奠定了基础.此外,与主流的深度学习方法相比,本文分类方法在普通CPU配置的机器上即可实现,有效降低了机器人对设备配置的要求.本文仅对单一场景下的工具进行了分类研究,考虑到家庭服务机器人的实际工作场景较为杂乱,针对客厅、厨房等实际复杂场景下的工具识别与分类是下一步的研究方向.

## 参考文献(References)

- [1] Zou Fuhao, Wang Yunfei, Yang Yang, et al. Supervised feature learning via l2-norm regularized logistic regression for 3D object recognition[J]. *Neurocomputing*, 2015, 151(2): 603-611.
- [2] Kjellstrom H, Romero J, Kragi D. Visual object-action recognition: Inferring object affordances from human demonstration[J]. *Computer Vision and Image Understanding*, 2011, 115(1): 81-90.
- [3] Zhu Y X, Zhao Y B, Zhu S C. Understanding tools: Task-oriented object modeling, learning and recognition[C]. *Proc of the IEEE Conf on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 2855-2864.
- [4] Mahmudul H, Dharmaratne A. Attribute based affordance detection from human-object interaction images[C]. *Proc of the Image and Video Technology-PSIVT Workshops*. New Zealand: Springer Verlag, 2015: 220-232.
- [5] Edsinger A, Kemp C C. Manipulation in human environments[C]. *IEEE-RAS Int Conf on Humanoid Robots*. Genoa: IEEE, 2006: 13-19.
- [6] Tanis M, Vadim T, Giorgio M. Multi-model approach based on 3D functional features for tool affordance learning in robotics[C]. *IEEE-RAS Int Conf on Humanoid Robots*. Seoul: IEEE, 2015: 482-489.
- [7] Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps[J]. *Int J of Robotics Research*, 2015, 34(4/5): 705-724.
- [8] Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks[C]. *IEEE Int Conf on Robotics and Automation*. Seattle: IEEE, 2015: 26-30.
- [9] Nguyen A, Kanoulas D, Caldwell D G, et al. Detecting object affordances with convolutional neural networks[C]. *Proc of IEEE/RSG Int Conf on Intelligent Robots and Systems*. IEEE, 2016: 2765-2770.
- [10] 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[J]. *光学学报*, 2017(2): 155-165.  
(Wu P L, Fu W X, Kong L F. A fast algorithm for affordance detection of household tool parts based on structured random forest[J]. *Acta Optica Sinica*, 2017(2): 155-165.)
- [11] Grauman K, Darrell T. The pyramid match kernel: Discriminative classification with sets of image features[C]. *The 10th IEEE Int Conf on Computer Vision*. Beijing: IEEE, 2005: 1458-1465.
- [12] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]. *IEEE Computer Society Conf on Computer Vision and Pattern Recognition*. New York: IEEE, 2006: 2169-2178.
- [13] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]. *IEEE Computer Society Conf on Computer Vision and Pattern Recognition*. Miami: IEEE, 2009: 1794-1801.
- [14] Shen F, Shen C, Zhou X, et al. Face image classification by pooling raw features[J]. *Pattern Recognition*, 2016, 54(C): 94-103.
- [15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2014, 37(9): 1904-1916.
- [16] Han X, Zhong Y, Cao L, et al. Pre-trained alexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification[J]. *Remote Sensing*, 2017, 9(8): 848.
- [17] Myers A, Teo C L, Fermüller C, et al. Affordance detection of tool parts from geometric features[C]. *IEEE Int Conf on Robotics and Automation*. Seattle: IEEE, 2015: 1374-1381.
- [18] Koenderink J J, Van D A J. Surface shape and curvature scales[J]. *Image and Vision Computing*, 1992, 10(8): 557-564.
- [19] Coates A, Ng A. The importance of encoding versus training with sparse coding and vector quantization[C]. *Int Conf on Machine Learning*. Bellevue: Association for Computing Machinery, 2011: 921-928.

## 作者简介

吴培良(1981—),男,副教授,博士,从事家庭服务机器人智能提升、功用性认知、SLAM等研究, E-mail: peiliangwu@ysu.edu.cn;

何犇(1992—),女,硕士生,从事家庭服务机器人工具功用性认知及其应用的研究, E-mail: ben.he@ia.ac.cn;

侯增广(1969—),男,研究员,博士生导师,从事机器人与智能系统、康复机器人与微创介入手术机器人等研究, E-mail: zengguang.hou@ia.ac.cn.

(责任编辑: 齐 霖)