

基于随机丛林的人体部件分类方法

屈雁秋, 陈 锋

(中国科学技术大学 自动化系, 合肥 230000)

摘 要: 人体部件分类是人体姿态跟踪中的重要前提和关键步骤。传统深度梯度特征下的随机森林分类方法虽然可以达到实时性的要求, 但仍存在准确度不高、对噪声不够鲁棒、内存消耗过大等缺点。为此, 提出传统深度特征与 RGB 边缘特征相融合的一种新的分类方法, 并在特征维度加大的情况下引入随机丛林模型。实验结果表明, 该特征分类方法不仅可以减少 20% 左右的运行时间, 还可以提高 1% 左右的测试准确率。

关键词: 人体部件分类; 深度梯度特征; RGB 边缘特征; 随机丛林; 有向无环图

中文引用格式: 屈雁秋, 陈 锋. 基于随机丛林的人体部件分类方法[J]. 计算机工程, 2018, 44(1): 247-251.

英文引用格式: QU Yanqiu, CHEN Feng. Body Part Classification Method Based on Random Jungle [J]. Computer Engineering, 2018, 44(1): 247-251.

Body Part Classification Method Based on Random Jungle

QU Yanqiu, CHEN Feng

(Department of Automation, University of Science and Technology of China, Hefei 230000, China)

【Abstract】 Body part classification is an important precondition and crucial step in human pose tracking. Although the random forest classical method with gradient feature of depth can get the request of real time, it remains the weakness of low accuracy, not robust enough to the noise and huge memory consumption. This paper proposes a new classification method, which combines the classical depth feature with RGB edge feature, and introduces the random jungle model to solve the increased dimension of the feature. Experimental results show that the proposed method not only saves about 20% running time, but also improves the test accuracy of about 1%.

【Key words】 body part classification; gradient feature of depth; edge feature of RGB; random jungle; Directed Acyclic Graph (DAG)

DOI: 10.3969/j.issn.1000-3428.2018.01.041

0 概述

人体姿态估计是人工智能领域一个很重要的部分, 它隶属于自然人机交互领域。计算机该如何理解人类做出的动作, 不通过如鼠标键盘的媒介, 人类姿态如何控制计算机进行操作, 达到自然交互的目的, 这些问题都是计算机视觉领域研究的热点问题。现阶段有众多人体姿态估计的方法^[1-3], 如文献[4]提出可以由骨架结构近似表示人体姿态, 通过采用人体骨架结构中每个关节的时间序列数据表示人体动作序列。2012 年微软推出了体感游戏设备 xbox360, 其前端深度相机 Kinect 通过获取人体深度数据, 使用随机森林分类器大量训练模拟和真实人体数据, 获得了非常好的效果。

文献[5]提出随机森林模型, 至今仍被认为是最有效的分类及回归算法之一。文献[6]提出通过随机森林训练经过提取的深度特征, 从而完成对人体

全身 31 个部件的大致分类。特征训练样本数量高达 2×10^9 个, 并且随着森林中单个决策树的树深 d 的增加, 内存消耗为 $O(2^d)$ 。

本文在前人的基础上, 为解决部分传统量化深度特征响应不敏感问题而加入新的 RGB 边缘特征, 并与文献[7]中提出的决策丛林算法相结合, 提出优化的人体部件分类流程。根据文献[6]的启发, 本文将人的身体分为 31 个部件, 通过深度数据做特征提取, 采用灰度图像做边缘特征提取。

1 深度特征与 RGB 边缘特征

1.1 深度量级特征

由于深度图像不受光照、衣着、皮肤毛发等因素的影响, 可以很好地保留人体轮廓, 因此利用深度数据作特征提取的案例已经愈发成熟。文献[8]首次提出了一种快速的量化深度特征, 这种特征在 Kinect 骨骼案例中得到了很好的运用, 具体如下:

作者简介: 屈雁秋 (1992—), 男, 硕士研究生, 主研方向为计算机视觉、机器学习; 陈 锋, 副教授、博士。

收稿日期: 2017-01-22 **修回日期:** 2017-02-24 **E-mail:** 451513890@qq.com

$$f_{\theta}^D(I, \mathbf{x}) = d_I(\mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})}) - d_I(\mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})}) \quad (1)$$

其中, $d_I(\mathbf{x})$ 表示在给定图像 I 中像素 $\mathbf{x}$ 处的深度, 参数 $\theta = (\mathbf{u}, \mathbf{v})$ 是自像素点 $\mathbf{x}$ 的 2 个偏移向量, 这里均除以 $d_I(\mathbf{x})$ 以保证深度不变性。这种特征在众多文献中得到使用与优化, 如文献[9-10]。

1.2 RGB 边缘特征

文献[6]指出, 因为深度图像具有低光鲁棒性、大尺度颜色和细节不变性以及可以解决边界模糊等优点, 所以将深度数据用于特征提取比起 RGB 图像具有更高的效率。但是如 Kinect 这样的体感设备往往不仅具有红外设备, 而且同时兼具获取传统 RGB 图像的功能。受文献[11]启发, 结合 RGB 图像与深度图像往往可以获取场景更丰富的信息与更好的分类性能。经典 RGB 特征算子用于关键点匹配^[12]。本文将这种观念进行扩展, 对于灰度 RGB 图像, 从中提取其边缘比较特征 (Edge Comparison Feature, ECF)。

文献[13-14]中均指出, 对于人体识别, 观察人体轮廓是一个非常重要的参考和线索。文献[15]定义了轮廓的详细解释, 轮廓不仅包含物体的外部边界, 而且包括内部边缘信息, 可以描述物体的空间结构。因此, 将人体 RGB 图像的轮廓信息与深度信息结合可以优化分类性能。

为了衡量两幅边缘图像的匹配程度, 文献[16]引入了 Chamfer 匹配算法, Chamfer 距离定义为:

$$d_{CM}(U, V) = \frac{1}{n} \sum_{u_i \in U, v_j \in V} \min |u_i - v_j| \quad (2)$$

其中, $U = \{u_i\}$, $V = \{v_i\}$ 是边缘图像中的边缘点集合, d_{CM} 即 Chamfer 距离为 U 中每一个 u_i 到 V 中最近边缘的距离之和的平均值, $n = |U|$ 。

同式(1)类似, 结合式(2), 引入 ECF 的量级特征:

$$e_I^M(\mathbf{x}) = \min_{\mathbf{x} \in I, \mathbf{x}' \in I_E} \|\mathbf{x} - \mathbf{x}'\| \quad (3)$$

$$f_{\theta}^{EM}(I, \mathbf{x}) = e_I^M(\mathbf{x} + \mathbf{u}) + e_I^M(\mathbf{x} + \mathbf{v}) \quad (4)$$

式(4)用于表示位于图像 I 中像素点 $\mathbf{x}$ 处与其边缘图像 I_E 上一点 $\mathbf{x}'$ 的距离最小值, 式(3)则表示在此像素点 $\mathbf{x}$ 处的 2 个偏移位置偏移的存在性。

当量级信息无法很好地区分人体部件, 比如均为一个很小的值时, 可以尝试引入方向信息特征, 后续的实验证明这种组合可以减少噪声影响, 有效提高准确率。ECF 的方向特征如下:

$$e_I^O(\mathbf{x}) = \text{Orientation}(\arg \min_{\mathbf{x} \in I, \mathbf{x}' \in I_E} \|\mathbf{x} - \mathbf{x}'\|) \quad (5)$$

$$f_{\theta}^{EO}(I, \mathbf{x}) = e_I^O(\mathbf{x} + \mathbf{u}) - e_I^O(\mathbf{x} + \mathbf{v}) \quad (6)$$

在式(6)中方向的定义可以有多种方式, 本文采取的是 $\mathbf{x}$ 与 $\mathbf{x}'$ 连线与水平轴正方向的夹角值, 因此式(5)中的“-”操作符也充分考虑了角的循环

特性。

同式(1)中的深度偏移参数相同, 本文定义 ECF 的量级与方向特征可以捕捉到图像中特定位置的边缘分布的微弱信号, 可以通过强分类学习器学习到分类间隔参数。

2 决策 DAG 与随机丛林

2.1 决策 DAG

利用有向无环图 (Directed Acyclic Graph, DAG) 做决策的方法已经在众多文献中提到。有向无环图是指, 给定一个有向图 $G(V, E)$, $V = \{v_i\}$, 若从 G 的任意一个顶点 v_i 出发经过若干条边 e_j , 均无法回到该点, 则该图 G 被称为有向无环图。本文将一类特殊的 DAG 称为决策 DAG。决策 DAG 具有以下要求: 1) 只有一个入度为 0 的根结点; 2) 拥有多个入度 ≥ 1 , 出度为 2 的分裂结点; 3) 拥有多个入度 ≥ 1 , 出度为 0 的叶子结点。

由图 1 可以看出, 决策 DAG 可以得到优化的树结构, 把决策 DAG 看作一种有向树。如果检测到两个分裂结点具有类似的分类概率密度分布, 就可以把它们合并起来。当树深较大时, 决策 DAG 可以有效地减少内存消耗, 从后续实验看出, DAG 还可以捕捉到训练数据内在关联的程度, 提高模型的准确率。

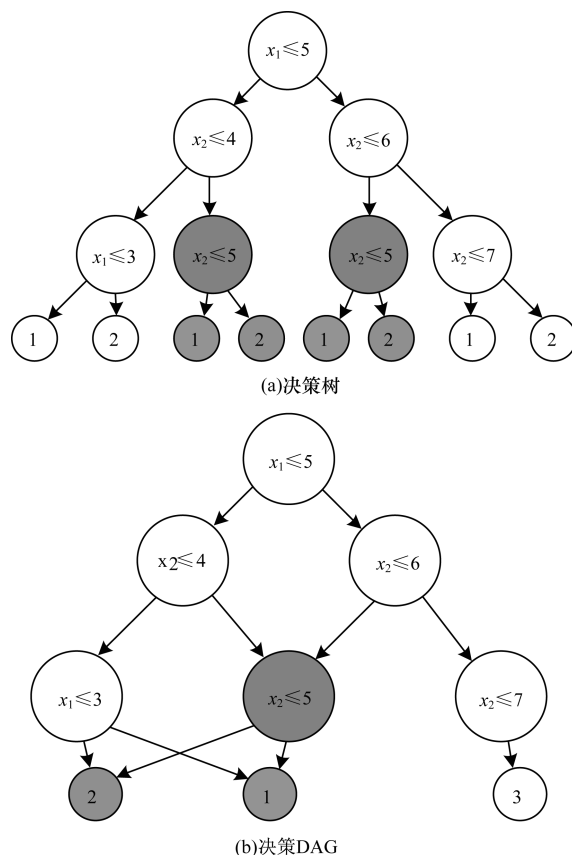


图 1 决策树与决策有向无环图

2.2 随机丛林

同随机森林类型类似,通过集成若干棵决策 DAG,最终可以得到包含多棵 DAG 的整体模型 $J = (G_1, G_2, \dots, G_T)$ 。

设模型 J 中第 i 个决策 $DAGG_i$ 的输出:

$$p_{i=i}(y|v) \quad (7)$$

表示 G_i 输出的标签概率密度分布,这里 y 是分类标签, v 是输入的样本特征。

具有 T 棵决策 DAG 的集成模型 J 的输出定义为:

$$p(y|v) = \frac{1}{T} \sum_{i=1}^T p_i(y|v) \quad (8)$$

由于在样本选取与训练特征筛选时同样采取了随机 bagging 的方式,因此可以有效避免模型出现过拟合。为了将这种整合模型与随机森林相区别,文献[7]将这种集成式 DAG 称为决策丛林 (Decision Jungle),为了体现该模型同样通过随机选取样本与样本特征的方式进行训练的特性,本文将之称为随机丛林 (Random Jungle)。

2.3 随机丛林训练与参数优化

决策 DAG 训练参数时采用逐层 (level-wise) 的方式, DAG 图结构与特征选择、阈值训练需要同时优化,这一点与决策树分裂准则不尽相同。在每个 DAG 结点 v 处,需要优化 4 个参数:选择特征 d_v , 阈值 θ_v , 左孩子结点 l_v , 右孩子结点 r_v 。

假设 s 是某种合并方案,如 $s(D) = \min(2^D, 128)$, 这里 D 代表 DAG 的层数。则 DAG 分裂算法如下:

1. $G \leftarrow (\{\text{root}\}, \emptyset)$
2. for $D = 1, 2, \dots$ do
3. 将 $s(D)$ 个新结点加入 G 作为孩子结点
4. 初始化上层父结点参数
5. 优化上层父结点参数
6. end for

DAG 的逐层参数优化如图 2 所示。其中, p_i 是第 i 个父结点, c_j 是相对于 p 的第 j 个子结点。

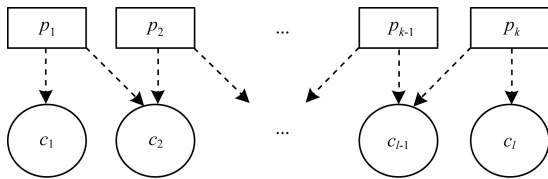


图2 决策 DAG 某层结构示意图

本文首先给出信息熵的定义,再给出寻找父结点最优参数的目标函数 E :

给定 $X \subset \mathbb{R}^n \times \{1, 2, \dots, C\}$ 为训练样本集,则信息熵 $H(X)$ 定义为:

$$H(X) = - \sum_{i=1}^C p(i) \log_a p(i) \quad (9)$$

这里:

$$p(i) = \frac{|\{(x, y) \in X : y = i\}|}{|X|} \quad (10)$$

定义目标函数 E :

$$E(\Theta_1, \Theta_2, \dots, \Theta_k) = \sum_{i=1}^l |S_{c_i}| H(S_{c_i}) \quad (11)$$

这里:

$$\Theta_i = (d_{p_i}, \theta_{p_i}, l_{p_i}, r_{p_i}) \quad (12)$$

是父结点 p_j 的待优化参数, S_{p_j} 、 S_{c_i} 分别是落入结点 p_j 、 c_i 的训练样本集合。

在式(11)中, S_{c_i} 与 $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_k)$ 的关系由下式可以看出:

$$S_{c_i} = \bigcup_{j=1, 2, \dots, k : l_{p_j} = c_i} \{(x, y) \in S_{p_j} : x_{d_{p_j}} \leq \theta_{p_j}\} \cup \bigcup_{j=1, 2, \dots, k : r_{p_j} = c_i} \{(x, y) \in S_{p_j} : x_{d_{p_j}} > \theta_{p_j}\} \quad (13)$$

参考图 2, 给出 DAG 分裂算法中步骤 5 的参数优化算法如下所示。

1. function Optimization($\Theta_{p_1}, \Theta_{p_2}, \dots, \Theta_{p_k}$):
2. while Θ_{p_i} 发生变化 do
3. for $i = 1, 2, \dots, k$ do
4. $\zeta \leftarrow$ 随机特征选择集合
5. $(d_{p_i}, \theta_{p_i}) \leftarrow \arg \min_{d \in \zeta, \theta \in \mathbb{R}} E(\dots, \Theta_{p_{i-1}}, (d, p, l_{p_i}, r_{p_i}), \Theta_{p_{i+1}}, \dots)$
6. end for
7. for $i = 1, 2, \dots, k$ do
8. $l_{p_i} \leftarrow \arg \min_{l=c_1, c_2, \dots, c_i} E(\dots, \Theta_{p_{i-1}}, (d_{p_i}, \theta_{p_i}, l, r_{p_i}), \Theta_{p_{i+1}}, \dots)$
9. $r_{p_i} \leftarrow \arg \min_{r=c_i, c_{i+1}, \dots, c_k} E(\dots, \Theta_{p_{i-1}}, (d_{p_i}, \theta_{p_i}, l_{p_i}, r), \Theta_{p_{i+1}}, \dots)$
10. end for
11. end while
12. return $\Theta_{p_1}, \Theta_{p_2}, \dots, \Theta_{p_k}$
13. end function

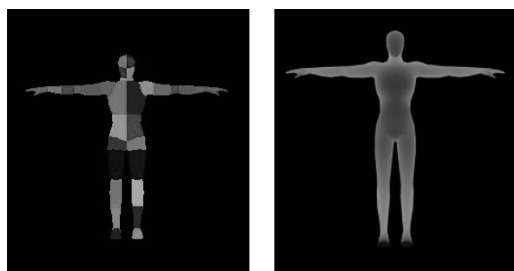
上述参数优化算法是一种坐标梯度下降算法,在第 1 步中,算法依次遍历每个父结点 k 并试图找到最优分裂参数 (d_{p_i}, θ_{p_i}) 最小化目标函数 E , 同时保持左右子结点指向以及其他父结点的分裂参数不变。在第 2 步中,算法重定向由父结点指向子结点的分支,以此最小化目标函数 E 。当该层参数不再发生变化时,算法终止,上层父结点参数优化完毕。

3 实验结果与分析

本文总样本数为 4 000 个, 取其中的 20% 作为测试样本。横向比较了随机丛林与随机森林在不同训练样本数、弱分类器个数下完成训练所需的空间与时间消耗, 也比较了单独使用经典深度特征、RGB 边缘特征, 以及两者融合特征下分别使用随机丛林

和随机森林模型进行分类的平均正确率。

训练数据的制作基于 CMU 的 MotionBuilder 骨骼数据库。图像如图 3 所示。



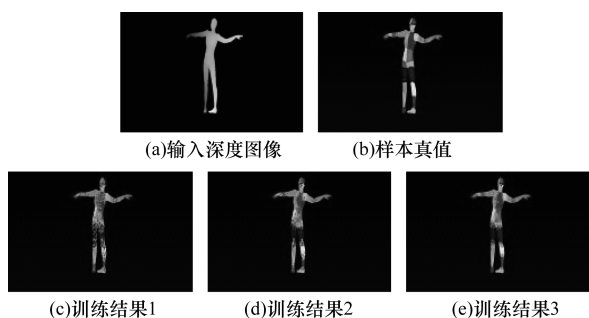
(a) 真值标注图像

(b) 深度图像

图 3 训练样本

3.1 测试样本对比

图 4 为输入图像与输出图像对比。训练结果 1 为带有经典深度梯度特征的随机森林训练结果,训练结果 2 为深度特征与 RGB 边缘特征融合的随机森林训练结果,训练结果 3 为深度特征与 RGB 特征融合的随机丛林训练结果。通过对比可以发现,引入 RGB 特征可以较好地解决误分类中集簇误分类的情况,即不会在某个部件中心出现大规模误分的情况,而使用随机丛林模型可以提高每类的平均准确率。



(a) 输入深度图像

(b) 样本真值

(c) 训练结果1

(d) 训练结果2

(e) 训练结果3

图 4 输入图像与输出图像

3.2 训练过程耗时对比

本文的训练采用 8 棵决策树或决策 DAG,偏移向量模值取水平和垂直方向上 0 ~ 100 的高斯采样,决策 DAG 的合并策略为:

$$s(D) = \min(128, 2^{\min(5, D)} \times 1.2^{\max(0, D-5)})$$

为了比对实验效果,分别使用 500、1 000、2 000、4 000 个训练样本进行训练,测试了使用深度梯度特征和深度与 RGB 融合特征下算法迭代完毕所消耗的时间,结果如表 1、表 2 所示。

表 1 不同深度梯度特征下的训练消耗时间比较 s

训练样本数	决策丛林模型	随机丛林模型
500	267	259
1 000	633	587
2 000	1 897	1 210
4 000	4 192	2 505

表 2 不同深度梯度与 RGB 特征的训练消耗时间比较 s

训练样本数	决策丛林模型	随机丛林模型
500	357	355
1 000	1 062	797
2 000	2 467	2 311
4 000	6 219	3 185

从表 1 与表 2 中可以看出,当训练样本较少时,随机森林与随机丛林模型消耗时间基本相同,随着训练样本的增多,消耗时间开始出现明显差异,具体原因是随机丛林中非叶子结点数目开始大大减少,待优化的参数数目显著降低。

3.3 测试过程耗时对比

本文选取训练样本的 20% 作为测试样本,即使用 100、200、400、800 张深度图像作为测试样例,对以 3.2 节所述参数训练完成的模型分别计算完成一幅图像分类所消耗的平均时间,结果如表 3、表 4 所示。

表 3 不同深度梯度特征下的测试消耗时间比较 s

训练样本数	决策丛林模型	随机丛林模型
100	0.107	0.067
200	0.091	0.062
400	0.098	0.073
800	0.091	0.061
均值	0.096	0.065

表 4 不同深度梯度与 RGB 特征下训练消耗时间比较 s

训练样本数	决策丛林模型	随机丛林模型
100	0.076	0.064
200	0.098	0.066
400	0.099	0.072
800	0.104	0.069
均值	0.094	0.067

由表 4 可以看出,使用随机丛林结合深度与 RGB 特征的方法每秒钟可以处理 15 帧左右的图像,基本可以满足实时性的要求。

3.4 测试样本平均分类正确率比较

本文考察测试样本的平均分类正确率,这里给出弱分类器数量分别为 4 和 8 的 2 种情况。在同一坐标系下分别比较了单深度特征下随机森林(RFD)与随机丛林模型(RJD)、单 RGB 边缘特征下随机森林(RFRGB)与随机丛林模型(RJRGB)、深度特征与 RGB 特征融合下随机森林(RFD + RGB)与随机丛林模型(RJD + RGB)的测试分类正确率与训练样本数量的关系,如图 5 所示。

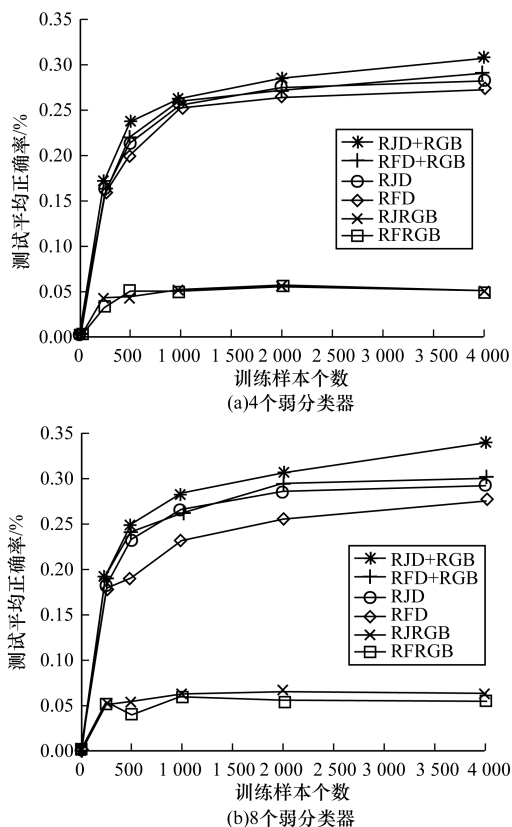


图5 测试样本分类正确率比较

实验结果表明,单独的 RGB 边缘特征几乎无法作为分类主特征进行人体部件分类,但是与传统深度特征的融合可以提高分类正确率。通过相同特征的横向比较也可以发现,使用决策 DAG 作为弱分类器相比较传统决策树也可以提高整体模型的泛化性能。

4 结束语

本文在传统深度梯度特征的基础上,加入 RGB 边缘特征,提出一种新的分类方法。由此带来的特征维度相应变高,为了解决传统随机森林在处理高维特征时内存消耗的指数级增长,结合文献[7,9]中提出决策 DAG 模型,目的在于降低内存成本消耗,从而降低时间复杂度。实验结果表明,本文方法不仅可以减少算法的运行时间,而且可以提高模型的分类正确率。今后将提出改进的随机丛林模型,使得分类耗时和准确率上能有进一步提高。

参考文献

- [1] DANTONE M, GALL J, LEISTNER C, et al. Human Pose Estimation Using Body Parts Dependent Joint Regressors[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013:3041-3048.
- [2] GIRSHICK R, SHOTTON J, KOHLI P, et al. Efficient Regression of General-activity Human Poses from Depth Images[C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2011:415-422.
- [3] YE M, WANG X, YANG R, et al. Accurate 3D Pose Estimation from a Single Depth Image [C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2011: 731-738.
- [4] 汪成峰,王 庆,梅树立,等. 基于插值小波关键帧提取的动作评价算法[J]. 计算机工程, 2017, 43(1): 309-315.
- [5] BREIMAN L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [6] SHOTTON J, FITZGIBBON A, COOK M, et al. Real-time Human Pose Recognition in Parts from Single Depth Images[J]. Communications of the ACM, 2013, 56(1): 116-124.
- [7] SHOTTON J, NOWOZIN S, SHARP T, et al. Decision Jungles: Compact and Rich Models for Classification[C]//Proceedings of International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2013: 234-242.
- [8] LEPETIT V, LAGGER P, FUA P. Randomized Trees for Real-time Keypoint Recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 775-781.
- [9] 郭天楚,吴晓雨,杨 磊,等. 基于改进的随机森林的人体部件识别[J]. 中国传媒大学学报(自然科学版), 2014, 21(5): 32-38.
- [10] 徐岳峰,周书仁,王 刚,等. 基于深度图像梯度特征的人体姿态估计[J]. 计算机工程, 2015, 41(12): 200-205.
- [11] JIU Mingyuan, WOLF C, BASKURT A. Integrating Spatial Layout of Object Parts into Classification Without Pairwise Terms—Application to Fast Body Parts Estimation from Depth Images[C]//Proceedings of International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Berlin, Germany: Springer, 2013.
- [12] LEPETIT V, PILET J, FUA P. Point Matching as a Classification Problem for Fast and Robust Object Pose Estimation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2004: 244-250.
- [13] WU Jianxin, GEYER C, REHG J M. Real-time Human Detection Using Contour Cues[C]//Proceedings of IEEE International Conference on Robotics & Automation. Washington D. C., USA: IEEE Press, 2011: 860-867.
- [14] ZHANG Huigang, WANG Junxiu, BAI Xiao, et al. Object Detection via Foreground Contour Feature Selection and Part-based Shape Model [C]//Proceedings of the 21st International Conference on Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 2524-2527.
- [15] SHOTTON J, BLAKE A, CIPOLLA R. Multiscale Categorical Object Recognition Using Contour Fragments[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 30(7): 1270-81.
- [16] BARROW H G, TENENBAUM J M, BOLLES R C, et al. Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching[C]//Proceedings of the 5th International Joint Conference on Artificial Intelligence. New York, USA: ACM Press, 1977: 659-663.