基于随机森林误分类处理的 3D 人体姿态估计

蔡轶珩 [1] 王雪艳 [1] 马 杰 [1] 孔欣然 [1]

摘 要 为解决基于随机森林的 3D 人体姿态估计算法容易出现的误分类问题, 提出一种基于自适应融合特征提取和误分类处理机制的改进算法. 该算法利用自适应融合特征提取方法自适应提取深度融合特征, 此特征可表达图像距离信息和部位尺寸信息, 增强特征的表征能力; 针对识别部位误分类问题, 分别从识别部位误分点聚集情况和迭代整合思想出发, 提出误分类处理机制, 改善部位识别结果; 最后提出可进一步处理误分点的改进主方向分析 (Principal direction analysis, PDA) 算法, 自适应计算出部位主方向向量, 实现 3D 人体姿态估计. 结果表明, 该算法能有效去除部位误分点, 并显著改善了 3D 人体姿态估计.

关键词 人体姿态估计, 随机森林, 误分类处理, 主方向分析

引用格式 蔡轶珩, 王雪艳, 马杰, 孔欣然. 基于随机森林误分类处理的 3D 人体姿态估计. 自动化学报, 2020, **46**(7): 1457−1466

DOI 10.16383/j.aas.c180314

3D Human Pose Estimation Based on Random Forest Misclassification Processing Mechanism

CAI Yi-Heng[1] WANG Xue-Yan[1] MA Jie[1] KONG Xin-Ran[1]

Abstract This paper proposed an improved method which can reduce the misclassification in human pose estimation based on random forest and increase the accuracy, included adaptive fusion feature extraction and misclassification processing mechanism. Firstly, we improved the method of feature extraction to adaptive extract deep fusion feature with adaptive feature fusion extractive method, so that, both distance information and part information could enhance feature expression. Furthermore, owing to inspiration from error cluster analysis and iteration thought, the misclassification processing mechanism is proposed to handle misclassi-fication appearance. Finally, we achieved accurate human pose estimation from single depth images by applying the principal direction vector based on the improved principal direction analysis (PDA) algorithm. The experimental results demonstrated that this algorithm can efficiently eliminate several misclassifications and improve the accuracy of the 3D pose estimation.

Key words Human pose estimation, random forest, misclassification processing, principal direction analysis (PDA)

Citation Cai Yi-Heng, Wang Xue-Yan, Ma Jie, Kong Xin-Ran. 3D human pose estimation based on random forest misclassification processing mechanism. *Acta Automatica Sinica*, 2020, **46**(7): 1457−1466

基于图像的人体姿态估计是指获得给定图像中人体各部位在图像中的位置及方向等信息的过程[1], 是计算机视觉领域的重要研究方向, 可应用于视频监控、行为识别[2−3] 和人机交互[4] 等领域. 截至目前, 针对此任务已经有众多的研究算法被提出, 大致可分为两类: 基于模型和基于无模型的人体姿态估计算法. 前者利用人体先验知识对人体进行姿态估计, 通过预先构建人体模型, 将模型和图像中的人体轮廓、梯度等特征对应起来, 求解人体模型参数, 此方法虽然具有较高的识别效率, 但容易受到复杂模型的限制, 只适用于特定的姿态估计环境, 不利推广应用. 而基于无模型的方法, 是通过学习的方式来构建人体特征与人体姿态之间的复杂映射关系[4−8], 由于不需要事先构建复杂人体模型, 使得该方法不受复杂模型的限制, 简化了姿态估计的计算复杂度, 因此近年来, 基于无模型的姿态估计算法得到广泛研究.

其中在基于 RGB 彩色图像的人体姿态估计上, 单人姿态估计或多人姿态估计中均取得了一定成果[9−12], 但由于彩色图像的姿态估计算法受人体的体型、着装、肤色和光照等限制, 算法鲁棒性较弱. 与彩色图像相比, 深度图像记录的是距离信息, 具有颜色无关性, 可在一定程度上应对在彩色图像上遇

收稿日期 2018-05-16 录用日期 2018-10-06
Manuscript received May 16, 2018; accepted October 6, 2018
科技部国家重点研发计划课题 (2017YFC1703302), 北京市教委科技项目 (KM201710005028) 资助
Supported by National Key Research and Development Program (2017YFC1703302), Science and Technology Projects of Beijing Municipal Education Commission of China (KM201710005028)
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 北京工业大学信息学部信号与信息处理研究室 北京 100124
1. Signal and Information Processing Laboratory, Department of Information, Beijing University of Technology, Beijing 100124

到的挑战, 故而许多研究者围绕基于单一深度图像的人体姿态估计算法展开. 在深度图像上, 基于无模型的人体姿态估计算法, 针对学习方法的不同, 可分为基于深度学习的方法和基于传统随机森林的方法.

在基于深度学习的姿态估计方法中, 文献 [12] 利用卷积神经网络 (Convolutional neural network, CNN) 搭建了用于关节坐标回归的网络框架, 并通过级联回归器的方式取得较好的姿态估计结果, 虽然此方法主要从整体的人体部件进行推理, 并未考虑到相邻部件间的局部上下文信息, 但也引起了后续研究者[9−11, 13−14] 的关注. 其中, 在深度图像上, 文献 [14] 采用长短期记忆网络架构 (Long short-term memory, LSTM), 学习局部视点不变特征, 并利用自顶向下的错误反馈机制, 纠正姿态位置; 文献 [13] 则采用了 MatchNet[15] 计算全卷积网络 (Fully neural network, FCN) 预测的关节区域和模板之间的相似度的方法, 并通过相邻关节之间的配置关系, 来达到优化关节点位置的目的.

在对优化姿态的研究方面, 文献 [16−17] 针对人体图像遮挡问题, 提出采用基于范例的方法来纠正最初估计的姿态, 此方法虽能有效降低姿态错误估计, 但不能保证所纠正姿态的规范性, 因此文献 [18] 针对此问题, 将姿态纠正任务视为一种姿态优化问题, 在文献 [16−17] 的纠正结果上, 通过姿态先验模型来优化姿态效果, 保证姿态规范性.

而对于基于随机森林的姿态估计方法, 采用将深度图像像素逐一部位分类的思想, 此方法将姿态估计任务转为了对像素分类的问题, 降低了姿态估计的困难程度[5]. 其中文献 [8] 提出了使用随机森林训练部位模型, 并利用 Mean-shift 方法确定相应部位中关节点位置, 从而完成基于单一深度图像的 3D 人体姿态估计的任务; 文献 [6] 则根据 Mean-shift 方法对识别的身体部位聚类情况以及人体目标尺寸依赖较高等问题, 提出主方向分析算法 (Principal direction analysis, PDA) 来分析识别的身体部位, 通过求取部位主方向向量来估计出图像中的 3D 人体姿态.

综上, 针对深度图像的无模型人体姿态估计算法均取得了大量的研究成果. 其中基于深度学习方法的人体姿态估计需要大量训练数据和训练时间, 才可达到较高的估计精度. 而基于随机森林的方法, 与之相对来说, 可在较少训练样本的情况下, 取得不错的估计效果. 因此本文从实验条件和方法性能两方面考虑, 针对随机森林方法进行研究和改进, 提出新的 3D 人体姿态估计算法.

在以往使用随机森林方法进行 3D 人体姿态估计的算法中, 由于部位分类模型分类准确率限制, 容易出现部位像素误分类的现象, 使得在识别的部位中引入误分干扰点, 这些部位干扰点对后续关节点的准确定位有一定的负面影响, 从而降低姿态估计的准确性. 为此, 本文首先通过特征提取阶段的改进算法, 改善特征的表达性能, 提高部位分类准确率; 随后, 针对此估计算法中存在的像素误分类问题, 分别从识别部位误分点聚集情况和迭代整合思想出发, 提出误分类处理机制来去除识别部位中的干扰点, 以降低对姿态估计的影响; 最终通过改进的 PDA 算法得到更为准确的姿态估计结果.

1 相关工作

为提高基于随机森林的 3D 人体姿态估计结果, 文献 [4, 6−7, 19] 分别提出了改善算法. 其中, 文献 [19] 提出能综合利用深度图像的距离信息和像素部位尺寸信息的改进型特征提取办法, 来改善部位分类准确率, 但由于此特征提取方法利用的部位尺寸信息与训练时抽取的各部位像素样本量有关, 使得所提特征并未充分表达出图像的部位尺寸信息; 文献 [7] 则针对随机森林部位分类时, 容易在相邻部位上出现误分点的情况, 提出了部位融合的思想来改善部位识别结果, 但此方法并未提出能有效去除出现在识别部位其他位置误分点的误分类处理算法; 文献 [6] 则提出了利用 PDA 算法来分析识别的身体部位, 并求取出部位主方向向量, 即部位主轴, 从而利用部位主轴估计出图像中的人体姿态, 但此方法主方向向量的确定对部位识别准确度要求较高, 同时此方法对部位误分点的去除效果也并不明显.

从以往针对随机森林的 3D 姿态估计的改善算法的研究来看, 算法改进主要着手于改善特征提取方法或关节点定位方面, 并未提出有效的误分类处理算法来解决由于部位分类器的误分类问题给部位识别和关节点定位或姿态估计造成的负面问题. 为此, 本文基于合成深度图像数据库, 以上述研究现状为基础, 针对现有算法的局限性[6−7, 19], 提出优化算法: 首先, 通过提出自适应深度融合特征提取办法来改善分类器的分类准确率; 随后, 通过提出的误分类处理机制和改进 PDA 算法的位置权重阈值处理办法, 可极大降低部位识别结果中的误分点, 改善部位的主方向向量, 从而获得更为准确的姿态估计结果.

2 本文方法

本文采用基于随机森林方法完成对单一深度图像的 3D 人体姿态估计. 考虑到现有深度图像数据库, 缺乏本文实验所需的基于部位的颜色标签, 不适用于训练随机森林部位分类器, 为此本文创建了合成深度图像数据库来完成训练任务. 延用目前文献 [4, 6−8, 19−20] 中普遍采用的图像合成思想, 引用 CMU 运动捕获数据库的运动信息, 并基于 Maya

平台[7], 渲染出带部位颜色标签的深度图像数据库. 其中图像分辨率为 450×600, 人体动作包括有打招呼、跳舞、打篮球、洗窗户等共 14 组. 同时考虑到在实际环境拍摄的深度图像, 存在人体离镜头远近不同的情况, 因此本文合成数据库中的人体图像在整幅图像中的占比约为 $0.7 \sim 0.95$ 之间.

使用合成数据库可避免人工逐个对部位标注的大量精力, 也可避免人为标注误差造成的分类不准确的问题. 为使算法具有更好的鲁棒性, 考虑到在实际环境下, 不同人体的高矮胖瘦、着装差异等因素, 本文对模型及部位标注进行以下操作: 1) 调整模型参数来构造不同体型的人体模型, 共构建出 6 个不同的人体模型; 2) 细化模型人工标注部位等方法, 以尽可能地减小由于不同人体各部位深度不同, 或人为因素造成的部位分界误差的影响, 本文将人体分割成 15 个关键部位: 左/右头部、脖子、左/右肩膀、左/右上臂、左/右小臂、左/右手、左/右躯干上部分、左/右躯干下部分. 合成图像如图 1 所示, 其中第一行为深度图像, 第二行为相应身体部位标签图像.

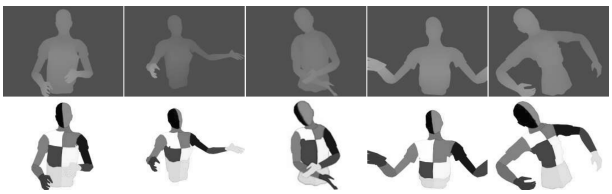


图 1　合成深度图像数据库

Fig. 1　Synthetic depth image dataset

由于实际拍摄的深度图像包含各种复杂背景, 故而本文首先利用背景减除法将图像中的背景移除[21], 只保留深度人体信息, 随后针对深度人体信息进行接下来的特征提取 (具体可见 2.1 节) 以及姿态估计任务. 为估计出更为准确的人体姿态, 针对部位识别误分类问题, 本文通过提出误分类处理机制, 即 Kmeans 分级聚类算法和多级随机森林整合算法 (具体可见第 2.2 节) 以及改进 PDA 算法 (具体可见第 2.3 节) 来完成. 算法整体流程图如图 2 所示.

2.1　身体部位识别

2.1.1　特征提取

图像深度梯度特征作为局部区域特征, 更多关注的是类内、类间不同像素点间的深度值差异, 较少注意到图像像素深度值本身的信息; 而图像深度数据特征代表了像素点的深度值, 关注图像深度信息本身, 但此特征易受到其他相似深度值像素点的干扰[22]. 因此本文基于两种特征的特点, 在文献 [6−8, 19, 23] 中提取深度梯度特征 (Magnitude gradient of depth, MGoD) 思想基础上, 提出自适应深度融合特征提取方法. 此方法融合了深度数据特征和改进的深度梯度特征, 可综合表达图像距离信息和不同身体部位尺寸信息, 自适应确定不同部位在特征提取时的偏移量值, 提高所提特征的表征能力, 改善部位分类模型的准确率. 其特征表达式为

$$F(x,\theta)=\left\{d_I(x), MGoD^{\eta_1}_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)\right\} \tag{1}$$

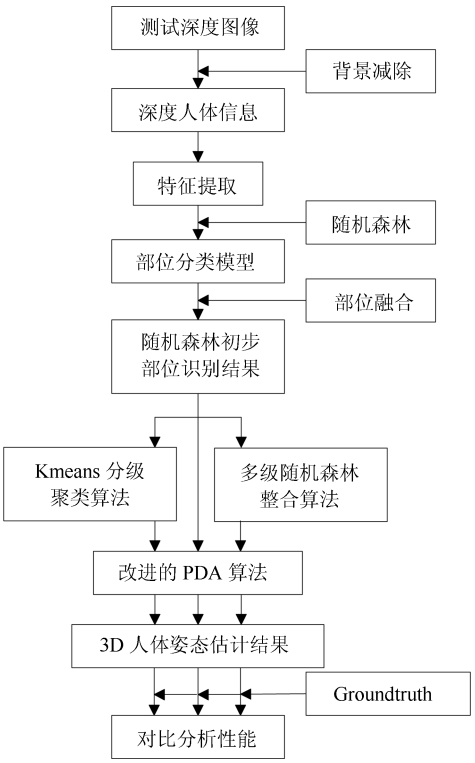


图 2　算法整体流程图

Fig. 2　Overview of proposed technique

深度数据特征 $d_I(x)$ 表示深度图像 I 中的人体部分在像素点 x 上的深度值, $MGoD^{\eta_1}_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 是本文根据 $MGoD_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 改进的自适应深度梯度差分特征, 其中 $MGoD_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 的计算公式为

$$MGoD_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)=d_I(x+\boldsymbol{x_u})-d_I(x+\boldsymbol{x_v}) \tag{2}$$

$\theta=(\boldsymbol{u},\boldsymbol{v})$ 为单位偏移向量对, $\boldsymbol{x_u}=\boldsymbol{u}q$ 和 $\boldsymbol{x_v}=\boldsymbol{v}q$ 为偏移向量, q 为偏移量, 本文中每个像素上含有 8 个偏移向量, 因此可组合偏移向量 36 对[19], 参考文献 [24] 方法, 提取其中 28 对偏移向量进行深度梯度差分特征提取. 图 3 为特征提取中某像素点偏移向量示意图.

$MGoD_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 方法中偏移量的定义在文献 [7] 中为

$$q=\frac{d(\cdot)}{255} \tag{3}$$

$MGoD_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 的偏移量 q 考虑了人体投影在深度图像的区域大小随人体距离深度相机的远近而改变的问题, 但忽略了深度图像中人体各部件的尺寸不同的问题, 使所提取的深度特征缺乏了部位之间的空间局部信息. 文献 [19] 中提到了利用部位尺寸信息的改进方法, 但偏移量的选择与训练样本中各部位的像素样本量有关. 如当人体出现部位遮挡, 或穿着较为厚重时, 如图 4 画框部分, 以文献 [19] 所提方法求取部位尺寸信息时, 得到的相应部位尺寸会出现偏小或偏大的情况, 因而此方法提取的特征并未充分表达出图像的部位尺寸信息.

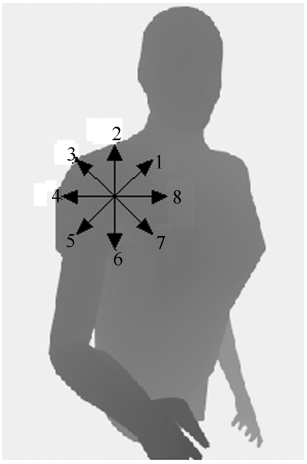
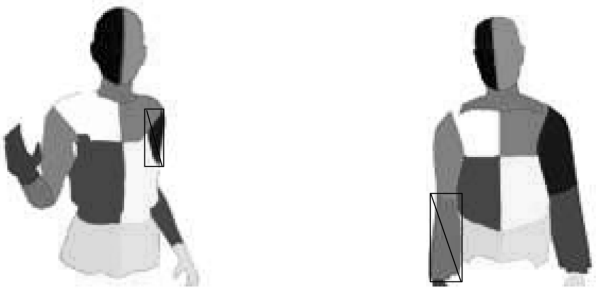


图 3 偏移向量对示意图

Fig. 3 Offset vector pair



(a) (b)

图 4 合成图像部位尺寸示意图

Fig. 4 Part size of the synthetic image

因此本文在文献 [19] 基础上, 对融合特征中的梯度差分特征进行改进, 提出自适应深度梯度差分特征 $MGoD^{\eta_1}_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$, 此特征可综合利用图像距离信息, 自适应确定不同的部位在特征提取时的偏移量值. 由于此方法的偏移量为训练图像上各部位分别对应的特征偏移量值, 因此可更好地学习到深度图像不同像素点的梯度信息, 增强了所提特征的表达能力.

$MGoD^{\eta_1}_{\theta=(\boldsymbol{u},\boldsymbol{v})}(x)$ 的部位尺寸信息确定如图 4 矩形框所示, 采用最小矩形框分别包含身体各部位, 计算最小矩形斜边 $\eta_1(\zeta_1)$, 以此作为偏移量 $q^{(\zeta_1)}$ 计算的部位尺寸信息. 偏移量 q 计算公式如下

$$q^{(\zeta_1)}=\frac{d(\cdot)}{255}\frac{\eta_1(\zeta_1)}{2},\quad \zeta_1=1,2,\cdots,m \tag{4}$$

深度图像 I 包含 m 个身体部位, $q^{(\zeta_1)}$ 表示第 ζ_1 个身体部位对应的偏移量值, $\eta_1(\zeta_1)$ 为第 ζ_1 个身体部位的最小矩形斜边.

2.1.2 分类器训练

本文在提取自适应深度融合特征后, 利用随机森林来训练部位分类器, 并进行部位像素分类. 随机森林由 T 个决策树组成, 每棵决策树分类过程互不影响, 最后的分类结果由所有决策树投票决定.

2.1.3 部位融合

在进行部位像素分类时, 对于较为细小部位 (手、脖子等) 而言, 由于在训练数据中这些较小部位的像素数目所占比例很少, 使该细小部位的正确识别精度不高. 因此, 本文采用了部位融合的思想[7], 将身体部位尺寸较大的部位划分为几个尺寸稍小的部位, 使训练数据中各部位的像素数目比例相近, 以改善部位识别结果. 根据男女人体的生理结构和各部位尺寸大小信息, 将人体分割成 15 个关键部位, 如图 1 所示, 并分别训练部位分类模型, 随后在测试阶段, 再将分类识别到的 15 个部位中的相应部位融合为一, 如分别将左右头和脖子、小臂和手、肩和躯干部分等视为一个身体部位.

采用部位融合的思想不仅可解决不同部位像素比例所占相差太大的问题, 同时由于在部位识别结果中的相邻部位处易发生误分类问题[7, 23], 此方法也可在一定程度上降低部位误分类的问题.

2.2 误分类处理机制

融合后的部位初步识别结果在 xy 方向上的投影如图 5(a1) 所示, 误分类点不只在相邻部位, 还离散或聚集于正确分类周围, 其中以误分点的小集群居多, 因此本文在部位融合算法基础上, 从部位误分点聚集情况, 以及受文献 [12] 的迭代回归思想启发, 分别提出 Kmeans 分级聚类算法和多级随机森林整合算法两种误分类处理算法.

其中, Kmeans 分级聚类算法针对部位分类结果中的小型错误聚集点, 以分级的方式, 将错误聚集点去除; 多级随机森林整合算法则采用多级偏移量的方式, 分别进行随机森林分类, 获得多种随机森林初步分类结果, 随后将相应部位分类结果中的相同分类点两两整合, 去除不同分类点, 从而达到降低误分点的目的.

2.2.1 Kmeans 分级聚类

本文将随机森林分类得到的身体部位以 3D 点

云形式表示, 其中图 5 (a) 中画圈部分为误分类聚集点在 xy 方向的投影. Kmeans 分级聚类算法根据识别的部位中正确分类聚集点远大于错误聚集点的特点, 首先将聚类点数最少的类别去除, 随后再从剩余聚类中去除聚类点数次之的类别, 最后保留剩余聚类点, 即部位的正确分类点.

图 5 为将部位点云利用 Kmeans 聚类算法, 在不同总聚类下处理后的效果. 其中图 5 (a) ∼ (d) 为随机森林识别的人体右小臂在 xy 方向上的投影结果, 图 5 (e) ∼ (h) 分别是将部位点云聚为 4 ∼ 7 类的 Kmeans 分级聚类处理后的部位投影结果. 从图 5 (e) ∼ (h) 可看出, 此算法可有效去除部位误分点, 但同时在总聚类较少时, 也同时去除了部分正确分类点, 如图 5 (e); 而在图 5 (f) ∼ (h) 中, 误分类去除效果随总聚类增多而变小. 因此本文将识别部位聚为 5 类, 即图 5 (f) 所示, 依据此两级聚类算法, 去除聚类点数较少的聚类, 保留聚类点数较多的聚类, 以此方法将各个部位误分类聚集点去除, 从而获得较为准确的部位识别结果.

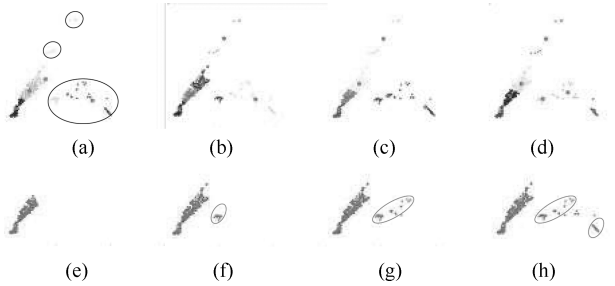


图 5　基于 Kmeans 算法在不同总聚类下的部位识别结果

Fig. 5　Part recognition results based on Kmeans algorithm under different total clusters

2.2.2　多级随机森林整合算法

基于梯度特征的人体部位识别中, 偏移量的选择密切影响分类结果的准确率. 由于在测试阶段, 分类识别时采用的偏移量不可能做出有效的自适应调整, 一般将测试图像的偏移量在延用距离信息基础上, 将身体部位信息选为所有训练样本中部位尺寸的平均值. 在实验中发现, 在部位分类结果中, 尤其是部位误分点, 随偏移量的选择不同而变化明显, 但大部分的正确分类还是基本保持一致的. 为获得更为准确的部位识别结果, 本文受文献 [12] 的迭代回归思想启发, 提出多级随机森林整合算法, 应用于测试识别阶段.

本文以所有训练样本部位偏移量的均值为基础, 通过等差方法前后选择多个偏移量, 分别进行自适应融合特征提取, 并利用随机森林分类模型获得部位分类结果. 其中每个偏移量对应的特征提取及其部位分类, 都是独立进行的. 由于分类结果与偏移量的选择有关, 特别是误分类点, 基于此, 本文利用多次分类结果, 将其两两整合, 保留相同分类点, 去除不同分类点的方式, 获得较为准确的部位识别结果, 实现过程如图 6, 具体算法如下.

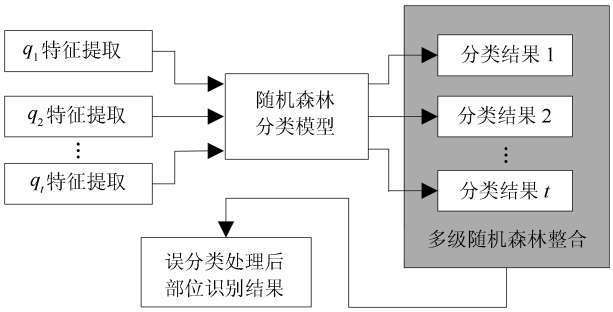


图 6　多级随机森林整合算法流程图

Fig. 6　The flowchart of the multi-level random forest integration algorithm

本文多级随机森林整合算法分为两个阶段, 第一阶段采用排列组合思想, 对各分类结果进行两两求与运算, 保留相同分类点, 去除不同分类点; 而第二阶段则因考虑到各分类结果的差异性, 针对第一阶段的整合结果, 依次求或运算, 最终得到更为准确的部位识别结果.

本文算法共进行了 $C_t^2+(C_t^2-1)$ 即 $t\,(t-1)-1, t\geq 2$ 次整合运算. 由于选择的偏移量越多, 算法的复杂度越高, 因而本文在此阶段基于算法性能和时间复杂度的考量, 共进行了 3 次偏移量的选择, 即能获得 3 种随机森林初步部位分类结果, 随后在多级随机森林整合算法中进行了 5 次整合运算, 从而获得更为准确的部位识别结果.

2.3　改进的 PDA 算法

在对识别的身体部位进行姿态估计时, 文献 [6] 提出利用 PDA 算法来分析识别的身体部位, 并利用求取出部位的主方向向量, 估计出图像中的人体姿态的方法. 但此方法求取部位主方向向量时, 对部位识别的准确度要求较高, 也并未对部位误分点提出有效的去除算法. 而利用上述误分类处理机制的两种算法, 分别处理随机森林初步分类结果后, 虽能有效降低识别部位中存在的错误分类点, 但也还会在处理后的部位周围残留一些离散点, 如图 5 (f), 影响后续姿态估计准确率. 为此本文在文献 [6] 算法基础上, 提出权重阈值处理算法, 进一步去除部位识别结果中的误分点.

在文献 [6] 中, PDA 算法首先利用逻辑函数计算每个像素点的位置权重, 计算公式为

$$w\,(t_i)=\frac{C}{1+\mathrm{e}^{\alpha(t_i-t_0)}} \tag{5}$$

其中, C 是限定输出值 (此时 $C=1$), t_0 和 α 根据部位尺寸大小选择, t_i 为马氏距离.

由于计算的身体部位各像素点的位置权重 w 随位置不同而变化, 本文根据离部位点云均值越远的像素点的位置权重值越小的性质, 提出位置权重阈值处理算法, 此算法能利用识别的身体部位 3D 点云的尺寸信息, 来设定位置权重的保留阈值, 并将低于保留阈值的位置权重对应的部位像素点去除, 使其自适应去除部位干扰点. 此算法在进一步去除部位干扰点的同时, 计算出部位的主方向向量, 提高姿态估计结果, 具体实现算法如下.

考虑到识别的身体部位还存在许多离散点, 为保证 $\phi^{(\zeta_2)}$ 阈值选择的有效性, 同时便于部位尺寸计算, 本文假设人体各部位为标准正方形, 然后统计出各识别部位所包含的像素点总数, 并以此作为即将构建的正方形的面积, 随后取此正方形斜边 $\eta_2(\zeta_2)$ 作为该部位的几何尺寸值[19], 以此计算 $\phi^{(\zeta_2)}$ 阈值大小, 计算公式如下

$$\phi^{(\zeta_2)}=\frac{\beta\eta_2(\zeta_2)}{\max(\eta_2(\cdot))},\quad \zeta_2=1,2,\cdots,L \tag{6}$$

在经过部位融合处理后, 此时深度人体信息中共包含 L 个身体部位, β 为初始设定阈值, $\beta=0.45$.

随后利用上述处理后的部位点云求取主方向向量 V_d[6], 估计出单一深度图像的 3D 人体姿态, 计算公式如下:

$$V_d(E_k)=\arg\max_{||E_k||_{k=1}^3}\left(E_k^{\mathrm{T}}S^*E_k\right) \tag{7}$$

其中, S^* 是部位像素点位置权重 w 的协方差, E_k 是协方差矩阵 S^* 的特征向量.

3 实验及结果分析

3.1 分类器分析

在进行下述实验之前, 本文首先针对不同分类器在此多分类任务中的性能优劣进行探讨, 探究随机森林算法在多分类任务中的性能.

本文首先使用 200 张合成图像, 其中从每幅图像中平均抽取 2 000 个像素样本提取深度梯度特征样本, 并基于不同的分类器, 如强分类器 (Ababoost)、K 最近邻分类器 (K-nearest neighbor, KNN) 和随机森林 (Random forest, RF) 分别训练部位分类模型, 其中分类准确率结果如表 1. 从表 1 中可以看出针对本文部位多分类任务中, Ababoost 和 KNN 在识别准确率上则明显弱于随机森林分类器, 并且分类器的训练时间也均远大于随机森林, 在同等训练样本条件下, 针对本文多分类任务, 可见随机森林分类器具有明显的优势, 因此本文选用随机森林算法训练部位分类模型具有可行性.

表 1 不同分类器的部位平均识别准确率结果

Table 1 Average recognition accuracy results for different classifiers

方法	训练时间 (s)	平均识别准确率 (%)
Ababoost	2 377.93	52.58
KNN	977.46	66.62
RF	187.97	70.29

3.2 身体部位识别性能

本文利用随机森林分类器来识别身体部位, 共使用 2 000 张训练图像和 300 张测试图像, 图像分辨率为 450×600. 通过多次实验总结, 在不影响分类准确率基础上, 将图像分辨率降为 225×300, 设置随机森林分类模型最佳参数配置如下: 决策树 30 棵, 树深度为 15, 从每幅图像中平均抽取训练采样点 2 000 个, 保证所提像素点均匀遍布全身各个部位, 每个像素点含有 28 个深度梯度特征属性和 1 个深度数据特征属性, 每次随机选取 6 个特征属性训练随机数中分类节点的最佳分类属性[6, 19]. 为探讨本文提出算法的鲁棒性, 分别以合成数据、ITOP (Invariant-top view dataset) 深度数据库[14] 以及实际的深度图像作为测试图像进行算法评估.

为探究方法的可靠性, 本文针对不同深度特征提取方法对部位分类模型准确率的影响进行研究. 部位识别对比结果可见表 2 和图 7, 其中文献 [19] 方法的识别准确率结果, 是基于该方法, 利用本文 2 000 张合成深度图像训练部位分类模型的再现结果. 从表 2 可以看出, 本文仅使用深度数据特征的平均识别准确率较低, 而仅使用自适应深度梯度特征及融合方法的部位平均识别准确率, 较其他随机森林相关方法均有所改善, 其中融合特征比文献 [19] 所提方法准确率提高 3.58 %, 说明本文提出的自适应融合特征增强了深度特征的表达能力, 提高了识别准确率.

表 2 不同特征方法的部位平均分类准确率结果

Table 2 Classification accuracy results for different feature methods

方法	平均识别准确率
深度梯度差分特征	0.7046
文献 [19] 改进型特征	0.8245
文献 [20] FCN 方法	0.8417
本文深度数据特征	0.6215
本文自适应深度梯度特征	0.8405
本文融合特征	0.8603

本文以文献 [20] FCN 方法的部位识别结果作为算法衡量对象, 该方法结果也是基于合成深度图

像数据库的针对上肢肢体分类识别的结果, 此算法属于深度学习领域, 需要大量的训练数据和训练时间, 在文献 [20] 中使用了 10 076 张训练图像, 并在 Geforce gtx tianx 下训练了 22 天, 因此本文仅直接引用了文献结果, 来衡量基于随机森林的部位识别性能. 从表 2 中可以看出, 本文相对文献 [20] 而言, 在少量训练样本基础上获得了与之相近的识别准确率.

在图 7 中, 本文融合特征与相关文献相比, 各部位识别准确率较为稳定, 未出现其他算法存在的部位识别准确率不平衡的问题. 并且由于人体手臂部位灵活性很高, 在基于图像的姿态估计任务中, 手臂部位的姿态估计精确性更能说明姿态估计算法的性能, 而本文算法手臂部位的识别结果明显优于深度梯度差分特征和文献 [19] 方法.

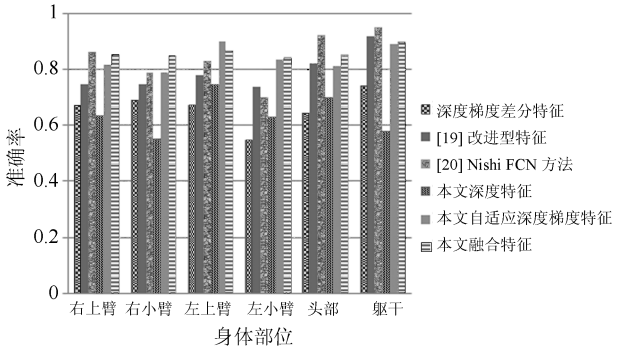


图 7　不同特征提取方法的部位分类结果对比

Fig. 7　Results of different feature extraction methods in part classification

3.3　误分类处理机制性能

为探究本文提出的误分类处理机制的算法实现性能, 本文可视化了由第 2.2 节误分类处理算法处理后的识别效果图, 如图 8 所示. 图 8 (a) 中部位识别结果周围存在有许多离散点和离群点, 在分别经过 Kmeans 分级聚类和多级随机森林整合算法处理后, 图 8 (b) 和图 8 (c) 上分别可看出, 部位识别结果周围已经去除了一部分误分点, 特别在多级随机森林整合算法处理后的图 8 (c) 上的去除效果更加明显, 但仅在正确分类部位周围还存在一些误分点.

图 9 (c) 为改进的 PDA 算法进一步去除误分点的效果. 从图 9 中可以看出本文改进的 PDA 算法 (c) 可最大化的将部位的误分点去除, 并且保留部位的骨骼走向; 与文献 [7] 中的 PDA 算法处理后的识别结果 (b) 相比, 改进的 PDA 算法 (c) 误分点去除效果更加明显. 由此可见, 本文提出的改进的 PDA 算法在不影响部位骨骼走向的前提下, 更好地去除部位周围的误分点, 保留正确分类点, 提高后续姿态估计的准确性.

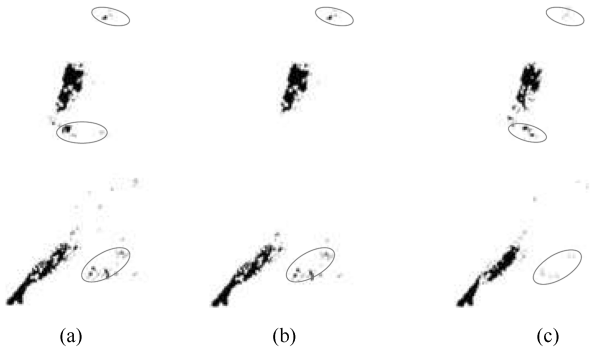


图 8　本文误分类处理机制处理后的部位分类结果图 ((a) 为随机森林初始识别 + 膨胀的结果; (b) 为分级聚类 + 膨胀的结果; (c) 为多级随机森林整合 + 膨胀的结果)

Fig. 8　Part classification result based on misclassification processing mechanism ((a) ∼ (c) representing the results of random forest, Kmeans, and multi-level random forest integration algorithm, respectively)

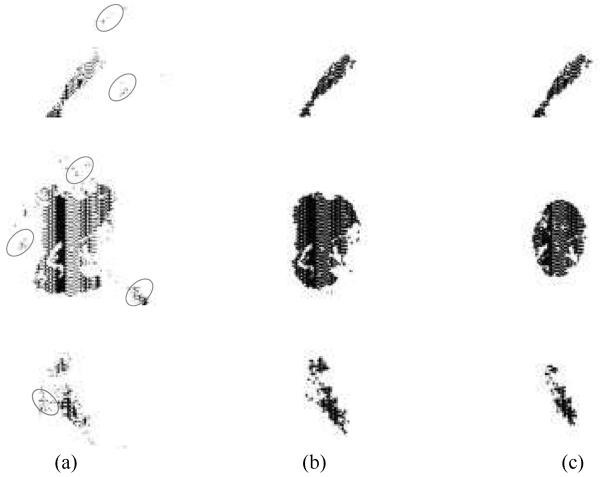


图 9　本文改进 PDA 算法和 PDA 算法的对比识别结果图 ((a) 多级随机森林整合算法的识别结果; (b) PDA 算法处理 + 膨胀的识别结果; (c) 改进的 PDA 算法处理 + 膨胀的识别结果)

Fig. 9　Contrast recognition results for improved PDA algorithm and PDA algorithm ((a) multi-level random forest integration algorithm, (b) ∼ (c) representing the results of PDA algorithms, and improved PDA algorithms, respectively)

3.4　人体姿态估计

为定量验证提出的误分类处理机制算法的有效性, 本文随机选取了 300 张除训练图像之外的合成深度图像进行相关算法的姿态评估, 考虑到人体手臂部位姿态更加灵活的问题, 本文以人体肘部角度误差作为评定标准[6], 此评定策略可表达出估计姿态与真实标定姿态在肘部角度之间的差异, 以此来评价算法性能.

在不同合成图像数据库中, 由于存在动作不同而造成肘部角度误差不同这一问题, 为避免此情况, 本文表 3 中不同算法的角度误差结果, 是基于本文合成深度图像, 并利用相关算法的再现结果对比. 从表 3 中可以看出, 将本文提出的自适应深度梯度特征和改进的 PDA 算法分别与文献 [6] 相比, 左右手肘角度误差均有所降低; 而对本文提出的自适应融合特征的识别结果和仅使用自适应深度梯度特征的识别结果进行对比发现, 在融合特征下, 左右手肘角度误差均低于自适应深度梯度特征的识别结果, 可见融合了深度数据特征的自适应深度梯度特征的部位识别结果更准确, 证明了添加深度数据特征的可行性. 而对融合特征下的部位识别结果分别使用了 Kmeans 和多级随机森林整合算法后的姿态肘部角度进行对比, 发现本文误分类点处理算法后的姿态更加准确, 其中使用多级随机森林整合算法处理后的平均误差在 8.5° 左右, 不过使用 Kmeans 分级聚类的运算速度要快于多级随机森林整合算法.

图 10～12 分别给出了利用本文算法在合成数据、公开深度数据集 ITOP[14] 和实际拍摄的深度图像上的部分实验结果. 从图中可以看出, 本文针对合成数据的姿态估计比较稳定, 而对于 ITOP 和实际深度图像的姿态估计, 基于误分类处理算法后的姿态, 特别是基于多级随机森林整合算法后的姿态更加接近于图像人体姿态, 证明了本文误分类处理方法能有效提高姿态估计结果, 但本文方法也存在以下不足: 首先, 在姿态复杂的图像上, 能有效针对部位遮挡情况的估计方法还有待研究, 如在图 11 的 ITOP 数据库的第 4 组数据中, 可以发现本文算法在对于有明显遮挡的背侧式图像的姿态估计效果不理想, 将有遮挡的人体姿态左右判断错误; 其次, 提高姿态估计的实时性和准确率还有待更多研究, 这将成为我们下一步的工作目标.

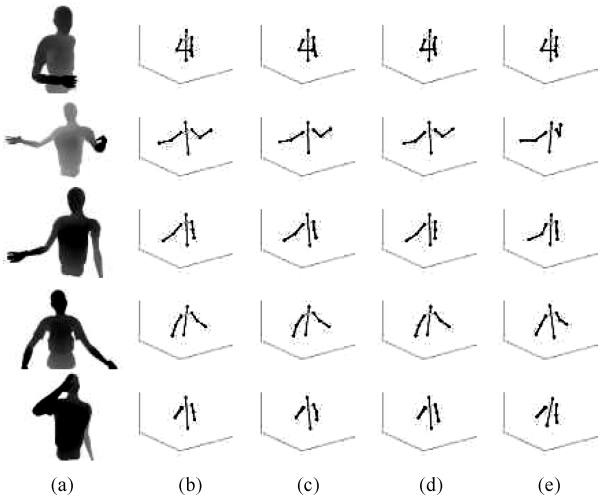


图 10 合成数据集上的姿态估计结果 ((a) 深度图像; (b) 误分类处理前的结果; (c) Kmeans 处理后的结果; (d) 多级随机森林整合后的结果; (e) groundtruth)

Fig. 10 Pose estimation on the synthetic dataset ((a) depth image, (b) ∼ (d) representing the results of random forest, Kmeans, multi-level random forest integration algorithm, respectively, (e) ground truth)

4 结论

基于深度图像的人体姿态估计是当前计算机视觉的难点之一. 首先, 本文为提高随机森林分类准确率, 提出改进型自适应深度融合特征提取办法, 提高特征表征能力, 改善部位识别精度; 然后, 对于随机森林误分类现象, 为避免其对后续姿态估计的影响, 提出误分类处理机制来处理误分点, 获得更为准确的部位识别结果; 最后, 利用识别部位的尺寸信息, 提出位置权重处理办法再次去除部位误分点, 从而得到较优的 3D 人体姿态估计. 实验表明结合改进的 PDA 算法, 多级随机森林整合算法获得的 3D 人体姿态估计较分级聚类算法更具有鲁棒性.

表 3 合成深度图像上的肘部角度误差结果

Table 3 Elbow angle error results on synthetic depth images

算法	左肘角度误差	右肘角度误差
深度梯度特征 + PDA (文献 [6])	14.5575°	13.5241°
自适应 (本文) + PDA (文献 [6])	12.7654°	13.3342°
自适应 + 改进的 PDA (本文)	12.2893°	13.1284°
融合特征 + 改进的 PDA (本文)	11.8462°	12.0331°
自适应 + Kmeans + 改进的 PDA (本文)	11.9879°	12.7443°
融合特征 + Kmeans+ 改进的 PDA (本文)	10.2546°	10.6436°
自适应 + 多级整合 + 改进的 PDA (本文)	9.9637°	9.6216°
融合特征 + 多级整合 + 改进的 PDA (本文)	8.4581°	8.6824°

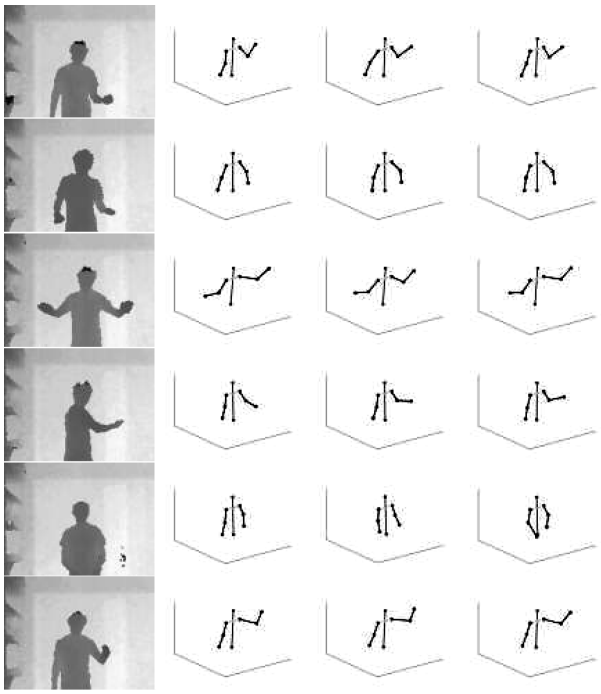


图 11 ITOP 数据集上的姿态估计结果 ((a) ∼ (d) 算法同图 10 (a) ∼ (d))

Fig. 11 Pose estimation on the ITOP dataset ((a) ∼ (d) same as Fig. 10 (a) ∼ (d))

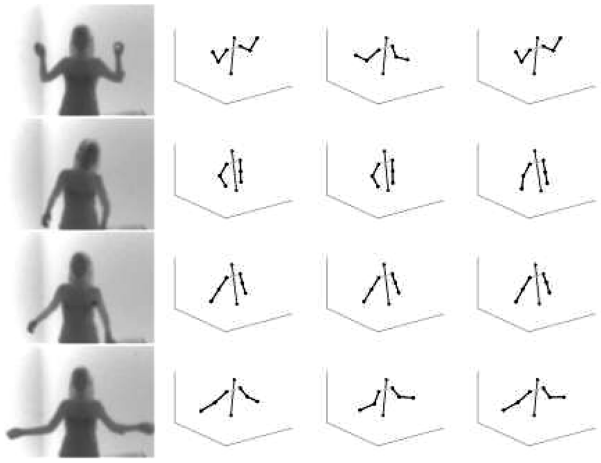


图 12 实际拍摄的深度图像上的姿态估计结果 ((a) ∼ (d) 算法同图 10 (a) ∼ (d))

Fig. 12 Pose estimation on the actual captured depth image ((a) ∼ (d) same as Fig. 10 (a) ∼ (d))

References

1 Shi Qing-Xuan, Di Hui-Jun, Lu Yao, Tian Xue-Dong. A medium granularity model for human pose estimation in video. *Acta Automatica Sinica*, 2018, **44**(4): 646−655
(史青宣, 邸慧军, 陆耀, 田学东. 基于中粒度模型的视频人体姿态估计. 自动化学报, 2018, **44**(4): 646−655)

2 Li You-Jiao, Zhuo Li, Zhang Jing, Li Jing-Feng, Zhang Hui. A survey of person re-identification. *Acta Automatica Sinica*, 2018, **44**(9): 1554−1568
(李幼蛟, 卓力, 张菁, 李嘉锋, 张辉. 行人再识别技术综述. 自动化学报, 2018, **44**(9): 1554−1568)

3 Zhu Yu, Zhao Jiang-Kun, Wang Yi-Ning, Zheng Bing-Bing. A review of human action recognition based on deep learning. *Acta Automatica Sinica*, 2016, **42**(6): 848−857
(朱煜, 赵江坤, 王逸宁, 郑兵兵. 基于深度学习的人体行为识别算法综述. 自动化学报, 2016, **42**(6): 848−857)

4 Shotton J, Girshick R, Fitzgibbon A, Sharp T, Cook M, Finocchio M, et al. Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(12): 2821−2840

5 Du Xiao-Peng, Hao Jian-Ping, Li Xing-Xin, Yang Jun. Human pose recognition research based on single depth images. *Computer and Modernization*, 2012, **1**(4): 192−195
(杜霄鹏, 郝建平, 李星新, 杨俊. 基于单一深度图像的人体姿态实时识别技术研究. 计算机与现代化, 2012, **1**(4): 192−195)

6 Dinh D L, Han H S, Jeon H J, Lee S, Kim T S. Principal direction analysis-based real-time 3D human pose reconstruction from a single depth image. In: Proceedings of Symposium on Information and Communication Technology. New York, USA: ACM, 2013. 206−212

7 Yin Hai-Yan. Human body pose recognition from the depth image [Master thesis]. Beijing University of Technology, China, 2013
(殷海艳. 基于深度图像的人体姿态识别 [硕士学位论文]. 北京工业大学, 2013)

8 Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, et al. Real-time human pose recognition in parts from single depth images. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington, D. C., USA: IEEE, 2011. 1297−1304

9 Park S, Hwang J, Kwak N. 3D Human pose estimation using convolutional neural networks with 2D pose information. In: Proceedings of the 2016 IEEE Conference on European Conference on Computer Vision (ECCV). Netherlands, Amsterdam: IEEE, 2016. 156−169

10 Wei S E, Ramakrishna V, Kanade T, Sheikh Y. Convolutional pose machines. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 4724−4732

11 Cao Zhe, Simon T, Wei S E, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 7291−7299

12 Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 1653−1660

13 Wang Ke-Ze, Zhai Sheng-Fu, Cheng Hui, Liang Xiao-Dan, Lin Liang. Human pose estimation from depth images via inference embedded multi-task learning. In: Proceedings of the 2016 ACM on Multimedia Conference. New York, USA: ACM, 2016. 1227−1236

14 Haque A, Peng Bo-Ya, Luo Ze-Lun, Alahi A, Yeung S, Li Fei-Fei. Towards viewpoint invariant 3D human pose estimation. In: Proceedings of European Conference on Computer Vision (ECCV). Netherlands, Amsterdam: IEEE, 2016. 160−177

15 Han Xu-Feng, Leung T, Jia Yang-Qing, Sukthankar R, Berg A C. MatchNet: Unifying feature and metric learning for patch-based matching. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. 3279−3286

16 Tu Zhuo-Wen. Exemplar-based human action pose correction and tagging. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE, 2012. 1784−1791

17 Shen Wei, Deng Ke, Bai Xiang, Leyvand T. Exemplar-based human action pose correction. *IEEE Transactions on Cybernetics*, 2014, **44**(7): 1053−1066

18 Shen Wei, Lei Rui, Zeng Dan, Zhang Zhi-Jiang. Regularity guaranteed human pose correction. In: Proceedings of the 12th Asian Conference on Computer Vision (ACCV). Singapore, 2014. 242−256

19 Zhang Yue-Feng, Zheng Yi, Fu Chao. Improved depth comparison feature for the recognition of human parts. *Microcomputer Its Applications*, 2015, **34**(14): 54−57
(张乐锋, 郑逸, 傅超. 用改进的深度差分特征识别人体部位. 微型机与应用, 2015, **34**(14): 54−57)

20 Nishi K, Miura J. Generation of human depth images with body part labels for complex human pose recognition. *Pattern Recognition*, 2017, **71**(6): 402−413

21 Lv Jie, Liu Ya-Zhou, Han Qing-Long, Du Jing. Method of locationg human body joints based on depth-images. *Naval Aeronautical and Astronautical University*, 2016, **31**(5): 538−546
(吕洁, 刘亚洲, 韩庆龙, 杜晶. 基于深度图像的人体关节点定位方法. 海军航空工程学院学报, 2016, **31**(5): 538−546)

22 Wu Min, Yang Yuan, Zhang Yuan-Qiang, Ku Tao, Zha Yu-Fei, Zhang Sheng-Jie. An infrared target tracking algorithm based on the fusion of deep feature and gradient feature. *Air Force Engineering University (Natural Science Edition)*, 2017, **18**(6): 76−82
(吴敏, 杨源, 张园强, 库涛, 查宇飞, 张胜杰. 深度融合特征与梯度特征的红外目标跟踪算法. 空军工程大学学报 • 自然科学版, 2017, **18**(6): 76−82)

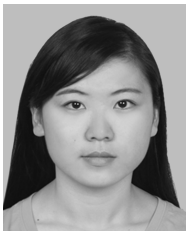
23 Xu Yue-Feng, Zhou Shu-Ren, Wang Gang, She Kai-Sheng. Human body attitude estimation based on gradient feature of depth images. *Computer Engineering*, 2015, **41**(12): 200−205
(徐岳峰, 周书仁, 王刚, 佘凯晟. 基于深度图像梯度特征的人体姿态估计. 计算机工程, 2015, **41**(12): 200−205)

24 Li Hong-Bo, Ding Lin-Jian, Ran Guang-Yong. Human body recognition based on Kinect depth image. *Digital Communication*, 2012, **39**(4): 21−26
(李红波, 丁林建, 冉光勇. 基于 Kinect 深度图像的人体识别分析. 数字通信, 2012, **39**(4): 21−26)

蔡轶珩 北京工业大学信息学部副教授. 美国罗切斯特大学访问学者. 1998 年获得合肥工业大学精密仪器专业硕士学位. 2007 年获得北京工业大学智能化信息处理专业博士学位. 主要研究方向为医学图像信息处理, 光度立体三维表面重建, 视觉感知信息处理. 本文通信作者.
E-mail: caiyiheng@bjut.edu.cn
(**CAI Yi-Heng** Associate professor in the Department of Information, Beijing University of Technology. Visiting scholar in the University of Rochester at USA. She received her master degree in precision instruments from Southeast University in 1998, and Ph. D. degree in intelligent information processing from Beijing University of Technology in 2007. Her research interest covers medical image information processing, photometric three dimensional surface reconstruction, and visual perception information processing. Corresponding author of this paper.)

王雪艳 北京工业大学信息学部研究生. 2016 年获得河北工程大学信息与电气工程学院学士学位. 主要研究方向为图像与视频处理.
E-mail: xinxiY23@126.com
(**WANG Xue-Yan** Master student in the Department of Information, Beijing University of Technology. She received her bachelor degree from the College of Information and Electrical Engineering, Hebei University of Engineering in 2016. Her research interest covers image and video processing.)

马 杰 北京工业大学信息学部研究生. 2016 年获得北京工业大学信息学部学士学位. 主要研究方向为图像与视频信号处理. E-mail: 13241247924@163.com
(**MA Jie** Master student in the Department of Information, Beijing University of Technology. He received his bachelor degree from the College of Department of Information, Beijing University of Technology in 2016. His research interest covers image and video signal processing.)



孔欣然 北京工业大学信息学部研究生. 2016 年获得北京工业大学信息学部学士学位. 主要研究方向为图像与视频处理.
E-mail: duzouran@163.com
(**KONG Xin-Ran** Master student in the Department of Information, Beijing University of Technology. She received her bachelor degree from the College of Department of Information, Beijing University of Technology in 2016. Her research interest covers image and video processing.)