一种基于RGB-D特征融合的人体行为识别框架

毛 峡[1] 王 岚[1] 李建军[1,2]
(北京航空航天大学电子信息工程学院 北京 100191)[1]
(内蒙古科技大学信息工程学院 内蒙古 包头 014010)[2]

摘 要 人体行为识别是计算机视觉和模式识别领域内一个重要的研究方向。人体行为的复杂性和不同人执行同一动作的差异性,使得行为识别仍然是一个具有挑战性的课题。采用新一代传感技术的RGB-D相机能够同时记录RGB图像和深度图像,并能够实时提取骨骼点信息。充分利用以上信息,成为行为识别领域的研究热点和突破点。文中提出了一种新的基于高斯加权金字塔式梯度方向直方图的RGB图像特征提取方法,并构建了一种多模特征融合的行为识别框架。在UTKinect-Action3D,MSR-Action 3D和Florence 3D Actions 3个数据库上对本研究所提特征和框架进行实验,结果表明,所提框架在3个行为数据库上的识别正确率分别达到了97.5%,93.1%,91.7%,从而证明了该行为识别框架的有效性。

关键词 行为识别,特征融合,高斯加权,梯度直方图,稀疏表示分类器
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.005

Human Action Recognition Framework with RGB-D Features Fusion

MAO Xia[1] WANG Lan[1] LI Jian-jun[1,2]
(School of Electronic and Information Engineering, Beihang University, Beijing 100191, China)[1]
(School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China)[2]

Abstract Human action recognition is an important research direction in the field of computer vision and pattern recognition. The complexity of human behavior and the variety of action performing make behavior recognition still as a challenging subject. With the new generation of sensing technology, RGB-D cameras can simultaneously record RGB images, depth images, and extract skeleton information from depth images in real time. How to take advantages of above information has become the new hotspot and breakthrough point of behavior recognition research. This paper presented a new feature extraction method based on Gaussian weighted pyramid histograms of orientation gradients for RGB images, and built an action recognition framework fusing multiple features. The feature extraction method and the framework proposed in this paper were researched on three databases: UTKinect-Action3D, MSR-Action 3D and Florence 3D Actions. The results indicate that the proposed action recognition framework achieves the accuracy of 97.5%, 93.1%, 91.7% respectively. It shows the effectiveness of the proposed action recognition framework.

Keywords Action recognition, Feature fusion, Gaussian weighted, Histogram of orientation gradients, Sparse representation classifier

1 引言

人体行为识别在新型人机交互、智能监控系统和机器人等方面有着广泛的应用前景,是计算机视觉和模式识别领域的重要研究课题之一。由于人体行为具有复杂性,各类行为存在着较大的类内差异和较小的类间差异,因此人体行为识别目前仍具有较强的挑战性。

较早的行为识别研究对采用普通摄像机获得的RGB动作序列进行识别。由于RGB彩色图像具有噪声大、动作难以分割、缺少完整的结构信息等缺点,导致行为识别技术难以获得突破。采用新一代传感技术的RGB-D相机能够同时记录RGB图像序列和深度图像序列,并实时提取骨骼点信息,同时具有价格低廉、易于获取、获取到的信息噪声小、骨骼点识别准确等优点,因此RGB-D相机逐渐替代彩色摄像机,被广泛应用于行为识别领域。如何更好地融合RGB-D相机获取的多模信息,成为行为识别领域新的突破点。图1为UTKi-

到稿日期:2017-10-24 返修日期:2017-12-19 本文受国家自然科学基金项目(61603013)资助。
毛 峡(1952—),女,博士,教授,博士生导师,主要研究方向为模式识别与人工智能,E-mail:moukyou@buaa.edu.cn(通信作者);王 岚(1992—),女,硕士生,主要研究方向为人体行为识别;李建军(1977—),男,博士生,主要研究方向为行为识别。

nect-Action 3D 数据库[1]中“起立”动作的示意图，包括 RGB 图像、深度图像和骨骼点信息。

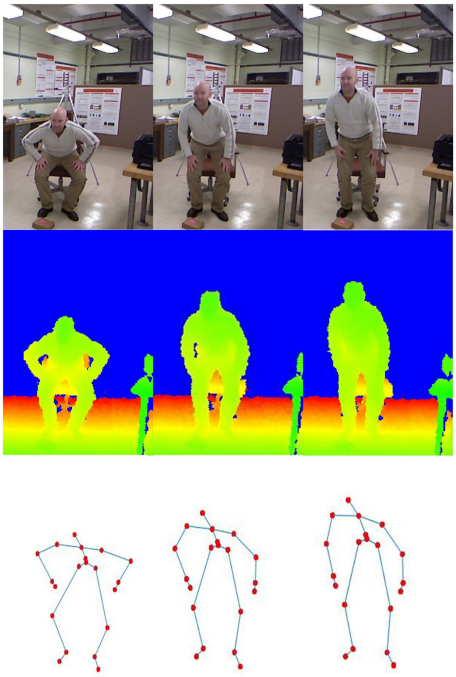

图 1　RGB 图像、深度图像和骨骼点信息示意图

Fig. 1　Information diagram of RGB image, depth image, skeleton

2　相关工作

针对人体行为的 RGB 图像，Dalal 等[2]提出梯度直方图（Histogram of Oriented Gradients，HOG）特征描述子，将图像分块并进行梯度直方图统计，用于人体检测。Bosch 等[3]受 HOG 特征启发，提出金字塔式边缘梯度直方图（Pyramid of Histograms of Orientation Gradients，PHOG）特征，将图像金字塔式分块，并进行边缘梯度直方图统计，用于物体检测分类。

针对从深度图像中提取的骨骼点信息，Xia 等[1]提出了将三维骨骼点进行直方图统计的 3D 骨骼点位置直方图（Histogram of 3D Joints Locations，HOJ3D）特征，并用隐马尔科夫模型进行分类，用于人体行为识别。Luvizon 等[4]提出了基于分组的骨骼点帧间相对位移特征、帧内相对骨骼点距离特征和特征融合方法，并获得了较高的识别率。

Ye 等[5]将 RGB 信息、深度信息和骨骼点信息进行融合，提出了动态时间量化（Dynamic Temporal Quantization，DTQ）算法，在帧对齐的同时保留了序列的时序特征。Zhang 等[6]提出了一种结合梯度信息和稀疏表示的行为识别方法。Aharon 等[7]提出了针对稀疏表示的 K-SVD 字典学习算法，使得稀疏表示分类器获得了更好的效果。

本研究针对 PHOG 特征在直方图分类较多时对人体行为类内差异容忍性不足的缺陷，提出了高斯加权的 PHOG 特征，并在此基础上融合人体行为的多模特征，结合 K-SVD 字典学习算法和稀疏表示分类器，提出了一种新的人体行为识别框架。

3　高斯加权 PHOG 特征

传统 PHOG 特征旨在通过图像的局部形状特征和局部特征空间分布提取图像特征[3]，主要应用于显著物体识别领域。但是，人体行为类内差异较大，比如，不同人做挥手动作，有人向上挥，也有人向前挥。将传统 PHOG 特征应用于行为识别领域时，随着直方图类数的增多，传统 PHOG 特征对人体行为的类内差异容忍度逐渐降低，使得行为识别的准确率也降低。

本文提出的高斯加权 PHOG 特征是对 PHOG 特征的改进，通过改进边缘梯度的直方图映射方式，使得特征对于人体行为类内差异有更好的容忍度。图 2 显示了高斯加权 PHOG 特征的提取过程。通过对图像进行不同金字塔级的分块划分，并连接不同划分级别的直方图统计特征，得到整体特征。

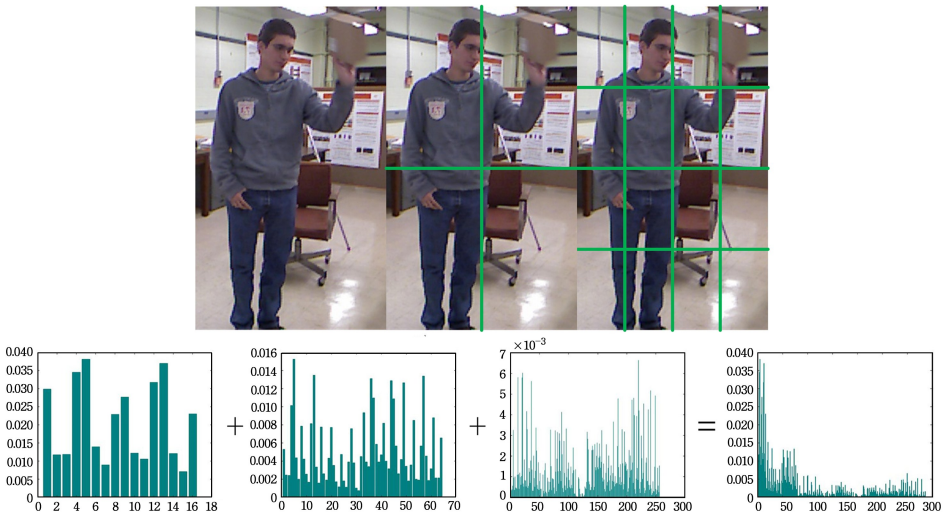

图 2　高斯加权 PHOG 特征示意图

Fig. 2　Diagram of Gaussian weighted PHOG feature

高斯加权 PHOG 特征的具体实现步骤如下：首先，通过 canny 算子对图像进行边缘检测；其次，计算图像的梯度信息；最后，结合边缘信息和梯度信息，可以得到图像在边缘处的梯度幅值和方向。我们将梯度方向分为 n 类，每类的角度

大小为 $360°/n$，如图 3 所示。

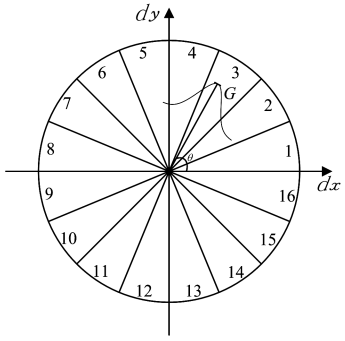

图 3 梯度角度分割

Fig. 3 Segmentation of gradient angle

用 G 表示某像素的梯度幅值，用 θ 表示梯度方向。假设像素梯度方向按高斯分布，以百分之百的概率落入相邻的 3 个直方图统计区域，根据最大似然准则，取 $\mu_\theta=\theta,\sigma=\frac{180}{n}$，则像素梯度的概率分布为：

$$p(x;\theta,\frac{180}{n})=\frac{n}{180\sqrt{2\pi}}e^{-\frac{(x-\theta)^2n^2}{2*180^2}} \tag{1}$$

将该像素映射入某直方图类的权重 W 设为梯度幅值与高斯概率的乘积，如下式所示：

$$W=G*p(\theta_1<\theta<\theta_2;\mu_\theta,\sigma) \tag{2}$$

$$p(\theta_1<\theta<\theta_2;\mu_\theta,\sigma)=\Phi(\theta_2;\mu_\theta,\sigma)-\Phi(\theta_1;\mu_\theta,\sigma) \tag{3}$$

$$\Phi(\theta;\mu_\theta,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\theta}e^{-\frac{(x-\mu_\theta)^2}{2\sigma^2}}dx \tag{4}$$

将金字塔第 L 级的图像划分为 4^L 块，在每个块内对每个边缘像素进行如上所述的直方图类的映射，从而形成该块的直方图统计特征。连接所有分割块的直方图特征，从而形成整个金字塔第 L 级的特征。连接金字塔所有级的直方图特征，从而形成整个图像的高斯加权 PHOG 特征。

4 行为识别框架

在提出高斯加权 PHOG 特征提取方法的基础上，将此特征应用到人体行为识别研究中。本研究提出的行为识别框架如图 4 所示。首先将骨骼点分组[4]并进行特征提取与融合，然后与 RGB 图像的高斯加权 PHOG 特征进行特征级融合，最后通过稀疏表示分类器进行分类。

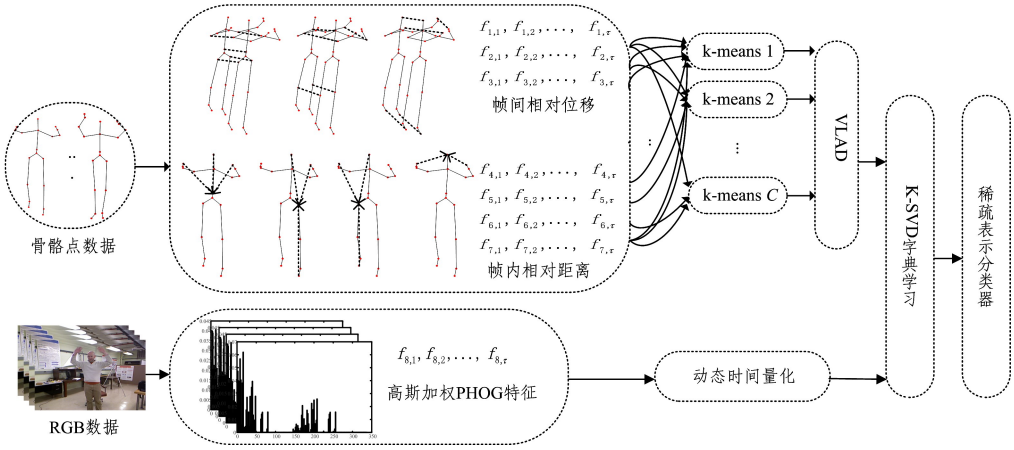

图 4 行为识别框架

Fig. 4 Action recognition framework

在行为识别框架的第一通道，即 3D 骨骼点通道，根据表 1 对骨骼点分组，并按照式(5)提取帧间相对位移特征。

$$v_i^s=\frac{p_i^{s+1}-p_i^{s-1}}{\Delta T}\,|\,1<s<\tau \tag{5}$$

其中，p_i^s 表示第 s 帧中骨骼点 i 的坐标 (x,y,z)，ΔT 表示第 $s+1$ 帧和第 $s-1$ 帧之间的时间差，τ 表示该动作序列的帧数。

表 1 骨骼点帧间相对位移特征分组

Table 1 Feature grouping of relative shift of skeleton between frames

帧间相对位移特征	骨骼点分组
f_1	头，右手，左手，右脚，左脚
f_2	颈，右肘，左肘，右膝，左膝
f_3	脊椎，右肩，左肩，右臀，左臀

按照表 2 对骨骼点分组，按照式(6)提取帧内骨骼点的相对距离信息：

$$\omega_{i,k}^s=p_i^s-p_k^s\,|\,i\neq k \tag{6}$$

其中，p_i^s 和 p_k^s 表示第 s 帧中骨骼点 i 和骨骼点 j 的坐标 (x,y,z)。因此，我们共得到 7 组骨骼点特征。通过 C 组 k 均值(k-means)算法，将每组特征划分为 k 类。最后，通过局部聚合向量描述子(Vector of Locally Aggregated Descriptors，VLAD)算法[8]，将 C 组、每组 k 类的骨骼点特征进行融合，从而实现了骨骼点特征融合。

表 2 骨骼点帧内相对距离特征分组

Table 2 Feature grouping of relative shift of skeleton within frames

帧内相对距离特征	骨骼点分组	相对骨骼点
f_4	头，左手，右手	脊椎
f_5	头，左手，左脚	右臀
f_6	头，右手，左手	左臀
f_7	左手，右手	头

在行为识别框架的第二通道，对 RGB 图像提取高斯加权 PHOG 特征，并使用动态时间量化算法[5]进行帧对齐。动态时间量化算法在进行帧对齐的同时，能够保留动作序列的时序特征。第一通道中的特征融合方法忽略了动作的时序特

征，而动态时间量化算法的使用则弥补了第一通道中无法对空间特征相似、时序特征相反的动作（如“起立”和“蹲下”）进行检测的缺陷，从而进一步提高了行为识别的正确率。

最后，将两个通道的骨骼点特征和 RGB 特征进行特征级融合，得到动作序列的整体特征。我们使用稀疏表示分类器[9]（Sparse Representation Classifier，SRC）进行动作分类。

使用 $A_i=[v_{1i},\cdots,v_{mi}](v\in R^m)$ 表示第 i 类动作的训练数据，其中 $v_{mi}(1\leqslant i\leqslant n)$ 表示第 i 类训练数据中的第 m 个样本。若共有 k 类动作，则所有训练数据可以表示为 $A=[A_1,\cdots,A_k]$，构成稀疏表示的字典。对于测试数据 $y\in R^m$，我们通过最小 L1 范数来获取测试样本的稀疏表示：

$$y=Ax \tag{7}$$

$$\hat{x}_1=\arg\min\|x\|_1,\ \|Ax-y\|\leqslant\varepsilon \tag{8}$$

则测试样本可表示为：$y\approx\sum_{i=1}^{k}A_i\hat{x}_{1i}$。定义重构残差 $\gamma_i=\|y-A_i\hat{x}_{1i}\|^2$，则测试样本属于重构残差最小的类。

图 5 为某一测试样本的稀疏表示系数示意图。其字典共有 10 类，每类含 16 个字典元素，该测试样本属于第 3 类，即字典元素 33－48 所属的类。图 5 中，稀疏表示系数在第 3 类，即黑框内的绝对值明显大于其他类，显示了稀疏表示分类器的效果。

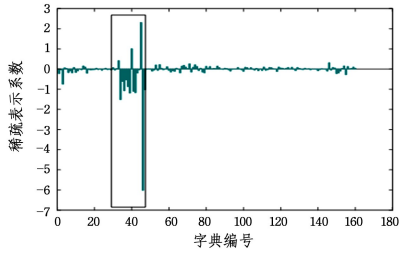

图 5　稀疏表示示意图

Fig. 5　Diagram of sparse representation

在稀疏表示分类器中，字典 A 的构造显得至关重要。K-SVD 算法[7]是一种针对稀疏表示，逐列更新字典的字典学习算法。因此，我们首先通过 K-SVD 算法对特征级融合后的动作特征进行字典学习，再使用稀疏表示分类器进行分类，以有效提高识别率。

综上所述，本文提出的行为识别框架具有以下优势。

1）第一通道中，骨骼点数据所提取出的特征为骨骼点距离特征和位移特征；第二通道中，所提取的高斯加权 PHOG 特征为形状特征。两个通道的特征相互补充，使得特征信息更为丰富。

2）第一通道所采用的特征融合与帧对齐算法打乱了动作序列的时序特征，关注于动作序列中每一帧各自的特点；第二通道所采用的帧对齐算法完整保留了动作的时序特征。两个通道的特征融合与帧对齐算法相互补充，使得融合后的特征包含了更丰富的信息。

3）本文提出的行为识别框架采用 K-SVD 字典学习算法与 SRC 分类器相结合的分类算法，与其他分类器相比，能够在分类之前进一步提炼出更有代表性的行为特征，从而进一步优化分类效果。

5　实验

本节在 3 个公开数据库上测试了本文所提特征和行为识别框架。MSR-Action3D 数据库[10]是最常用的 3D 人体行为识别数据库，包含了 10 个表演者录制的 20 个游戏动作，每个动作重复 3 次。这个数据库包含了十分相似的动作，因此非常具有挑战性。UTKinect-Action3D 数据库[3]包含了 10 个表演者录制的 10 个动作，其中包含 9 位男性和 1 位女性，每个动作重复 2 遍。每个表演者在不同的视角执行同一动作，导致了很大的类内差异。Florence3D Actions 数据库[11]由 10 个人录制 9 个动作，每个动作重复 2～5 遍。

5.1　实验参数的选择

本研究仅使用 MSR-Action3D 数据库进行参数优化，在 UTKinect-Action3D 和 Florence3D Actions 数据库上使用相同的参数设置。参数优化的方式为，使用不同的参数设置进行多次实验，选取实验动作识别率最高的参数组合作为最终实验参数。

本行为识别框架的参数设置如下，在第一通道特征融合中，通过 5 组 k-means 分类，每组 k-means 将每组骨骼点特征分为 23 类。在第二通道高斯加权 PHOG 特征中，选择金字塔级数为 3，直方图类数为 20。动态时间量化算法将帧数对齐为 14 帧。在 K-SVD 字典学习算法中，每类训练数据学习得到 16 个字典元素。

对于 UTKinect-Action3D 和Florence3D Actions 数据库，本文采用数据库提出者所建议的交叉验证方式进行实验。对于 MSR-Action3D 数据库而言，由于数据库提出者所建议的分组验证方式目前已不再具有挑战性，因此我们采用目前较有挑战性且较为常用的验证方式，即利用第 1，3，5，7，9 个录制人的动作序列进行训练，利用第 2，4，6，8，10 个录制人的动作序列进行测试[12]。本文所采用的验证方式均为各数据库的常用验证方式，以便于与其他文献的实验结果相对比。

5.2　UTKinect-Action3D 数据库

UTKinect-Action3D 数据库的提出者建议采用交叉验证的方式进行实验[1]，即取某录制人的动作样本为测试数据，其他人的动作样本为训练数据，并依次循环，直到所有人的动作样本都既充当过测试数据又充当过训练数据。我们采用这种交叉验证的方式进行实验，实验结果如表 3 第 2 列所示。

表 3　本文框架与同类框架的识别率对比

Table 3　Comparison of recognition ratio between proposed framework and similar frameworks in references

（单位：%）

框架	UTKinect-Action3D	MSR-Action3D	Florence3D Actions
文献[1]框架	90.92	78.97	—
文献[11]框架	—	—	82.00
文献[12]框架	—	88.20	—
文献[13]框架	97.08	89.48	90.88
文献[14]框架	91.50	92.10	87.04
本文框架	97.50	93.10	91.70

由表 3 可知，本文所提出的行为识别框架的识别率达到

了 97.5%，明显高于其他文献的识别正确率。由图 6 的混淆矩阵可以看出，本文所提框架在大多数类上达到了 100% 的识别率，其余类识别率多为 95%。

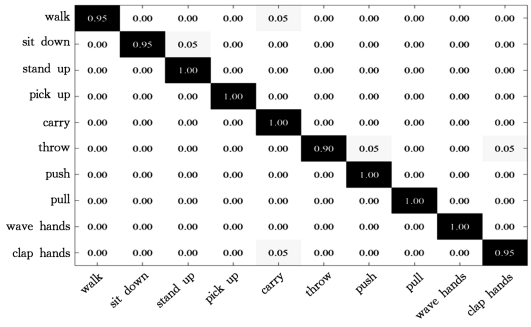

图 6 UTKinect-Action3D 混淆矩阵

Fig. 6 Confusion matrix of UTKinect-Action3D

5.3 MSR-Action 3D 数据库

本文采用第 1,3,5,7,9 个录制人的动作序列进行训练，利用第 2,4,6,8,10 个录制人的动作序列进行测试[12]。如表 3 第 3 列所示，本文所提行为识别框架的识别率达到了 93.1%，明显高于其他文献的识别率，且比目前的最高识别率高 1%。

5.4 Florence3D Actions 数据库

与 UTKinect-Action3D 数据库的实验相似，采用交叉验证的方式进行实验。如表 3 第 4 列所示，本文所提行为识别框架的识别率高于其他文献的识别率。

5.5 行为识别框架

首先，在本文所提行为识别框架的第二通道上使用原 PHOG 特征替换本文所提出的高斯加权 PHOG 特征进行实验，以验证特征改进的效果。实验结果如表 4 所列。可以看出，使用本文提出的高斯加权 PHOG 特征使得本文行为识别框架的识别率平均提高了 1.17%。

表 4 高斯加权 PHOG 特征的改进对识别率的提升效果

Table 4 Improvement of recognition ratio modified by Gaussian weighted PHOG feature

（单位：%）

特征	UTKinect-Action3D	MSR-Action3D	Florence3D Actions
PHOG	96.5	91.7	90.6
高斯加权 PHOG	97.5	93.1	91.7

为进一步验证高斯加权 PHOG 特征与直方图分类数的关系，以 MSR-Action3D 数据库为例，在直方图分类数分别为 8,12,16,20,24 时，分别使用 PHOG 特征和高斯加权 PHOG 特征在本文行为识别数据框架下进行实验，实验结果如表 5 所列。

表 5 MSR-Action3D 数据库中识别率与直方图类数的关系

Table 5 Relationship between recognition ratio and bin numbers in MSR-Action3D dataset

（单位：%）

特征	8	12	16	20	24
PHOG	90.4	90.1	91.1	91.7	90.8
高斯加权 PHOG	90.0	90.6	92.2	93.1	91.1

可以看出，使用高斯加权 PHOG 特征的识别率高于使用原 PHOG 特征的识别率，且直方图类数越多时识别率提升得越多。图 7 以折线图的方式给出了表 5 中识别率的增加值与直方图类数的关系。由图 7 可知，识别率的提高与直方图类数呈正相关，这也验证了我们改进 PHOG 特征的初衷，即改进后的 PHOG 特征对类内差异的容忍度更高。

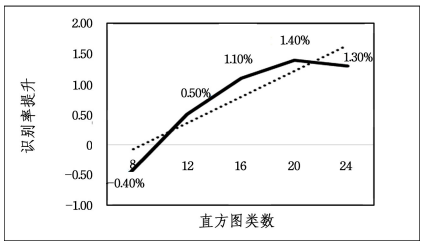

图 7 识别率提升与直方图类数的关系

Fig. 7 Relationship between improvement of recognition ratio and bin numbers

最后，使用 k-NN 和 SVM 分类器替代本文所用的 SRC 分类器进行分类器对比实验。k-NN 分类器和 SVM 分类器均采用与本文研究框架相同的参数优化方式，即在 MSR-Action3D 数据库上进行参数优化，在 UTKinect-Action3D 和 Florence3D Actions 数据库上采用相同的参数设置。参数优化的结果为，k-NN 分类器将 k 设为 1，SVM 分类器选用 sigmoid 核函数，参数 $gamma$ 和 C 分别为 1 和 10。实验结果如表 6 所列，使用 SRC 分类器的识别率平均高于 k-NN 分类器 5.23%，平均高于 SVM 分类器 1.87%，从而证明了 K-SVD 字典学习算法与 SRC 分类器相结合在行为识别中产生了很好的实验效果。

表 6 各分类器的识别率对比

Table 6 Comparison of recognition ratio between different classifiers

（单位：%）

分类器	UTKinect-Action3D	MSR-Action3D	Florence3D Actions
SRC	97.5	93.1	91.7
k-NN	84.9	91.0	90.7
SVM	93.8	91.7	91.2

5.6 时间效率

本文所提框架运行于台式机上，使用 Intel® CoreTM i5-4590 处理器。各数据库测试样本的平均时间消耗如表 7 所列。由表 7 可知，本研究框架的时间消耗较大，效率较低。这是由于在 SRC 分类器中需要对每一个测试样本通过最小 $L1$ 范数求取稀疏表示系数，导致该行为识别框架的时间消耗过大。我们拟在后续研究中通过稀疏正则化判别分析[15]（Sparse Regularization Discriminant Analysis，SRDA）来改进本文算法的时间效率问题。

表 7 各数据库测试集的平均时间消耗

Table 7 Average time comsumption of test cases of different dataset

（单位：ms）

数据库	UTKinect-Action3D	MSR-Action3D	Florence3D Actions
时间消耗	66.9	62.3	54.7

结束语　本研究的贡献有以下几点：

1）改进传统的 PHOG 特征，提出了高斯加权 PHOG 特征，且通过实验表明所提方法较原有方法获得了更好的类内差异容忍度；

2）基于 RGB 图像和骨骼点信息，提出多模特征融合的行为识别框架，并将 k-SVD 字典学习算法和 SRC 分类器应用于行为识别中；

3）在 MSR-Action3D，UTKinect-Action3D，和 Florence3D Actions 3 个数据库上进行实验，分别达到了较高的识别率 93.1%，97.5%和 91.7%，充分证明了该方法的有效性。

本研究的主要缺陷在于时间复杂度较高，未来拟通过稀疏正则化判别分析（Sparse Regularization Discriminant Analysis，SRDA）[15]进行改进。

参 考 文 献

[1] XIA L，CHEN C C，Aggarwal J K. View Invariant Human Action Recognition Using Histograms of 3D Joints[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence：IEEE Press，2012：20-27.

[2] DALAL N，TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego：IEEE Press，2005：886-893.

[3] BOSCH A，ZISSERMAN A，MUMOZ X. Representing Shape with a Spatial Pyramid Kernel[C]// Proceedings of the Sixth ACM International Conference on Image and Video Retrieval. New York：ACM，2007：401-408.

[4] LUVIZON D C，TABIA H，PICARD D. Learning Features Combination for Human Recognition from Skeleton Sequences[J/OL]. (2017-02-02)[2017-07-06]. http://dx. doi. org/10. 1016/j. patrec. 2017. 02. 001.

[5] YE J，LI K，QI G，et al. Temporal Order-Preserving Dynamic Quantization for Human Action Recognition from Multimodal Sensor Streams[C]// Proceedings of the Annual ACM International Conference on Multimedia Retrieval. Shanghai：ACM，2015：99-106.

[6] ZHANG H L，ZHONG P，HE J L，et al. Combining Depth-Skeleton Feature with Sparse Coding for Action Recognition[J]. Neurocomputing，2016，230(C)：417-426.

[7] AHARON M，ELAD M，BRUCKSTEIN A. K-SVD：An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. IEEE Transactions on Signal Processing，2006，54(11)：4311-4322.

[8] JEGOU H，DOUZE M，SCHMID C，et al. Aggregating Local Descriptors into a Compact Image Representation[C]// Proceedings of the 23th IEEE Conference on Computer Vision and Pattern Recognition. California：IEEE Press，2010：3304-3311.

[9] WRIGHT J，YANG A Y，GANESH A，et al. Robust Face Recognition via Sparse Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2009，31(2)：210-227.

[10] LI W，ZHANG Z，LIU Z. Action Recognition Based on a Bag of 3D Points[C]// Proceedings of the 23th IEEE Conference on Computer Vision and Pattern Recognition. San Francisco：IEEE Press，2010：9-14.

[11] SEUDEBARU L，VARANO V，BERRETTI S，et al. Recognizing Actions from Depth Cameras as Weakly Aligned Multi-part Bag-of-Poses[C]// Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Portland：IEEE Press，2013：479-485.

[12] WANG J，LIU Z，WU Y，et al. Mining Actionlet Ensemble for Action Recognition with Depth Cameras[C]// Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition. Providence：IEEE Press，2012：1290-1297.

[13] VEMULAPALLI R，ARRATE F，CHELLAPPA R，et al. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group[C]// Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition. Columbus：IEEE Press，2014：588-595.

[14] DEVANNE M，WANNOUS H，BERRETTI S，et al. 3-D Human Action Recognition by Shape Analysis of Motion Trajectories on Riemannian Manifold[J]. IEEE Transactions on Systems，Man，and Cybernetics，2015，45(7)：1340-1352.

[15] YIN F，JIAO L C，SHANG F，et al. Sparse Regularization Discriminant Analysis for Face Recognition[J]. Neurocomputing，2014，128(5)：341-362.

（上接第 16 页）

[13] SATTLER T，TORII A，SIVIC J，et al. Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization? [C]// CVPR 2017-IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[14] TRIGGS B，MCLAUCHLAN P F，HARTLEY R I，et al. Bundle-adjustment—a modern synthesis[C]// International Workshop on Vision Algorithms. Springer Berlin Heidelberg，1999：298-372.

[15] WU C. Towards linear-time incremental structure from motion[C]// 2013 International Conference on 3DTV-Conference. IEEE，2013：127-134.

[16] WU C，AGARWAL S，CURLESS B，et al. Multicore bundle adjustment[C]// 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE，2011：3057-3064.

[17] STRECHA C，HANSEN W V，GOOL L V，et al. Thoennessen on Benchmarking Camera Calibration and MultiView Stereo for High Resolution Imagery[C]// CVPR 2008. 2008：1-8.