

基于多描述子特征编码的人体行为识别

宋相法, 姚 旭

(河南大学 计算机与信息工程学院, 河南 开封 475004)

摘 要:针对采用单一描述子和单一特征编码方法导致三维人体骨架序列的行为识别率较低的问题,提出一种基于多描述子特征编码的方法。首先,从三维人体骨架序列中分别提取运动姿态描述子和角度描述子。然后,对每种描述子分别进行向量量化编码、稀疏编码和局部约束线性编码,从而获得六种特征。最后,根据这六种特征分别构造线性分类器,通过投票机制得到最终的识别结果。为了验证所提方法的有效性,在三维人体骨架序列行为数据集 MSR Action3D 上进行了实验,实验结果表明该方法的识别率为 94.9%,并且高于其他方法的识别率。

关键词:人体行为识别;特征编码;运动姿态描述子;角度描述子;三维人体骨架序列

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2018)08-0017-05

doi:10.3969/j.issn.1673-629X.2018.08.004

Human Activity Recognition Based on Multi-descriptor Feature Coding

SONG Xiang-fa, YAO Xu

(School of Computer and Information Engineering, Henan University, Kaifeng 475004, China)

Abstract: Aiming at the problem of low activity recognition rate from 3D human skeleton sequence based on a single descriptor and a single feature coding method, we propose a method based on multi-descriptor feature coding. Firstly, moving pose descriptor and angle descriptor are extracted respectively from 3D human skeleton sequence. Then, vector quantization coding, sparse coding and locality constrained linear coding are employed respectively to get six kinds of feature based on two kinds of descriptor. Finally, linear classifiers are respectively constructed based on these six kinds of feature, and the recognition result is decided by voting strategy. In order to validate the effects of the proposed method, the experiment on MSR Action3D, a public 3D human skeleton sequence activity database, demonstrates that the proposed method achieves 94.9% of recognition accuracy, which is superior to the state-of-art of methods.

Key words: human activity recognition; feature coding; moving pose descriptor; angle descriptor; 3D human skeleton sequence

0 引 言

人体行为识别在人机交互、视频监控、机器人视觉以及体感游戏等领域有着广泛的应用^[1-6]。过去,研究人员主要利用图像和视频研究行为识别问题;如今,研究人员开始利用微软 Kinect 传感器捕获的三维人体骨架序列研究行为识别问题。

人体是链式结构的,行为可由三维人体骨架关节点表示。例如,文献[7]提出了基于三维人体骨架关节点子集和多核学习的行为识别方法;文献[8]提出了基于运动姿态描述子和 KNN 算法的三维人体骨架序列行为识别方法;文献[9]提出了基于离群流形的三维人体骨架序列行为识别方法;文献[10]提出了基于联合学习身体部件行为特征和分类器的三维人体骨

架序列行为识别方法;文献[11]提出了基于分层姿态特征的三维人体骨架序列行为识别方法;文献[12]提出了基于角度描述子协方差特征的三维人体骨架序列行为识别方法;文献[13]提出了一种基于改进的动态时间规整算法的三维人体骨架序列行为识别方法。

在上述研究的基础上,为提高三维人体骨架序列行为识别精度,文中提出了一种基于多描述子特征编码的行为识别方法。该方法利用集成学习的思想,尽可能利用三维人体骨架序列的多个描述子特征编码方法构造不同的基分类器,通过对这些基分类器的结果进行集成来决定最终识别结果。

文中方法为充分利用三维人体骨架序列不同区分能力的描述子,弥补单一描述子提取时存在信息量不

收稿日期:2017-09-28

修回日期:2018-01-09

网络出版时间:2018-04-28

基金项目:国家自然科学基金(U1504611);河南省教育科学技术研究重点项目(15A520010)

作者简介:宋相法(1975-),男,博士,副教授,CCF 会员(15618M),研究方向为计算机视觉与机器学习;姚 旭(1991-),男,硕士研究生,研究方向为人体行为识别。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180427.1644.062.html>

足的缺点,结合运动姿态描述子^[8]和角度描述子^[14]来构造描述子集;为有效地将描述子组织起来提高识别性能,分别采用向量量化编码^[15]、稀疏编码^[16]和局部线性约束编码方法^[17]对运动姿态描述子和角度描述子进行编码,从而得到六种特征,将这六种特征的识别结果进行集成,得到三维人体骨架序列的最终识别结果。

1 多描述子与特征编码

利用微软 Kinect 传感器捕获的 20 个人体骨架关节点位置如图 1 所示。

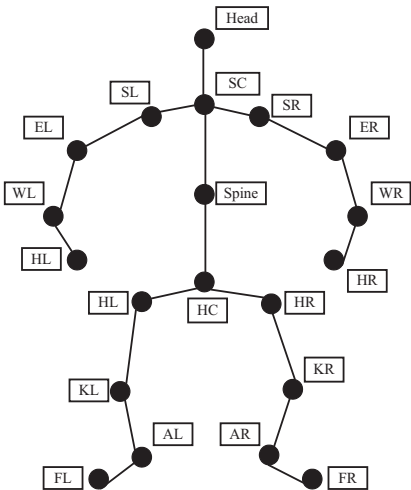


图 1 微软体感传感器捕获的人体骨架关节点位置

1.1 描述子提取

文中分别利用能反映数据不同性质的运动姿态描述子和角度描述子,充分发挥每种描述子在识别性能上的优势和互补性,弥补使用单一描述子时造成的信息量不足的缺点,从而提高识别率。

运动姿态描述子^[8]由每一帧关节点的位置、运动速度和加速度组成,具体实现详述如下。

在三维人体骨架序列中,每一姿态(帧)可表示为所有关节点坐标 $p_i(t) = (p_x, p_y, p_z)$ 构成的向量,其中 $i \in \{1, 2, \dots, n\}$, n 为关节点总数。所以,对于每一姿态(帧),其关节点位置特征向量表示为:

$$\mathbf{P} = [p_1(t), p_2(t), \dots, p_n(t)] \tag{1}$$

由于惯性、肌肉运动延迟等因素会影响人体行为,因此,可以用一个二次函数来近似表示人体行为^[8]。如果把帧表示为人体关节点坐标随时间变化的连续函数,则在以 t_0 时刻为中心的窗口内,其二阶泰勒展开式为:

$$\mathbf{P}(t) \approx \mathbf{P}(t_0) + \delta \mathbf{P}(t_0)(t - t_0) + \frac{1}{2} \delta^2 \mathbf{P}(t_0)(t - t_0)^2 \tag{2}$$

其中,一阶导数 $\delta \mathbf{P}(t_0)$ 表示 t_0 时刻帧的运动速度;二阶导数 $\delta^2 \mathbf{P}(t_0)$ 表示 t_0 时刻帧的运动加速度。

式 2 表明 t_0 时刻的瞬时帧向量及其一阶导数、二阶导数包含了以 t_0 为中心的时间片段内的帧函数信息。所以, t_0 时刻的运动姿态描述子可表示为帧向量 $\mathbf{P}(t_0)$ 、一阶导数 $\delta \mathbf{P}(t_0)$ 和二阶导数 $\delta^2 \mathbf{P}(t_0)$ 的级联。

因为三维人体骨架序列中帧与帧之间存在时间间隔,所以当前帧 $\mathbf{P}(t_0)$ 的一阶导数为:

$$\delta \mathbf{P}(t_0) \approx \mathbf{P}(t_1) - \mathbf{P}(t_{-1}) \tag{3}$$

当前帧 $\mathbf{P}(t_0)$ 的二阶导数为:

$$\delta^2 \mathbf{P}(t_0) \approx \mathbf{P}(t_2) + \mathbf{P}(t_{-2}) - 2\mathbf{P}(t_0) \tag{4}$$

其中, $\mathbf{P}(t_{-1})$ 表示当前帧的前一帧; $\mathbf{P}(t_1)$ 表示当前帧的后一帧,依次类推。

综上所述, t_0 时刻的运动姿态描述子表示为:

$$\mathbf{P}_{t_0} = [\mathbf{P}(t_0), \delta \mathbf{P}(t_0), \delta^2 \mathbf{P}(t_0)] \tag{5}$$

角度描述子^[14]通过计算微软 Kinect 传感器的深度摄像头位置(深度摄像头位置坐标 $\mathbf{ZP} = (0, 0, 0)$)及三维人体骨架关节点位置之间的角度和对应的角速度得到,各角度名称及对应属性,如表 1 所示。

表 1 角度名称及对应属性

特征	属性	特征	属性
θ_1	SC_SR_ER	θ_{19}	AL_HC_AR
θ_2	SR_ER_WR	θ_{20}	WL_HC_AL
θ_3	ER_WR_HR	θ_{21}	WR_HC_AR
θ_4	SC_SL_EL	θ_{22}	WL_HC_AR
θ_5	SL_EL_WL	θ_{23}	WR_HC_AL
θ_6	EL_WL_HL	θ_{24}	WL_SC_ZP
θ_7	HC_HR_KR	θ_{25}	EL_SC_ZP
θ_8	HR_KR_AR	θ_{26}	SL_SC_ZP
θ_9	KR_AR_FR	θ_{27}	AL_HC_ZP
θ_{10}	HC_HL_KL	θ_{28}	KL_HC_ZP
θ_{11}	HL_KL_AL	θ_{29}	HL_HC_ZP
θ_{12}	KL_AL_FL	θ_{30}	WR_SC_ZP
θ_{13}	SL_SC_SR	θ_{31}	ER_SC_ZP
θ_{14}	Head_SC_Spine	θ_{32}	SR_SC_ZP
θ_{15}	SC_Spine_HC	θ_{33}	AR_HC_ZP
θ_{16}	EL_SC_ER	θ_{34}	KR_HC_ZP
θ_{17}	WL_SC_WR	θ_{35}	HR_HC_ZP
θ_{18}	KL_HC_KR		

求解角度首先需要选择相应的位置坐标,以左肘关节点角度 θ_5 为例说明计算方法。首先选取左肩关节点、左肘关节点和左腕关节点,然后通过微软 Kinect 传感器获取到三维坐标数据,分别为左肩 $\text{SL}(\text{SL}_x, \text{SL}_y, \text{SL}_z)$ 、左肘 $\text{EL}(\text{EL}_x, \text{EL}_y, \text{EL}_z)$ 和左腕 $\text{WL}(\text{WL}_x, \text{WL}_y, \text{WL}_z)$, 则关节点方向向量为 $\text{SL_EL}(\text{SL}_x - \text{EL}_x, \text{SL}_y - \text{EL}_y, \text{SL}_z - \text{EL}_z)$ 和 $\text{EL_WL}(\text{EL}_x - \text{WL}_x, \text{EL}_y - \text{WL}_y, \text{EL}_z - \text{WL}_z)$, 最后利用余弦定理计算角度大小,如下所示:

$$\theta_5 = \arccos\left(\frac{|\text{SL_EL}| \times |\text{EL_WL}|}{|\text{SL_WL}|}\right) \tag{6}$$

由于帧与帧之间存在时间间隔,所以当前帧的角度速度表示为:

$$\delta\theta(t_0) \approx \theta(t_1) - \theta(t_{-1}) \quad (7)$$

综上所述, t_0 时刻的角度描述子表示为:

$$\theta_{t_0} = [\theta_1(t_0), \dots, \theta_{35}(t_0), \delta\theta_1(t_0), \dots, \delta\theta_{35}(t_0)] \quad (8)$$

1.2 特征编码

在特征编码过程中,不同的编码方法会导致不同的性能。文中选取常用的向量量化编码、稀疏编码和局部线性编码。

为表述方便,文中使用 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m] \in \mathbb{R}^{d \times m}$ 表示从三维人体骨架序列中提取的一个 d 维的描述子集合,其中 $\mathbf{x}_j \in \mathbb{R}^d$ 表示第 j ($j=1, 2, \dots, m$) 个描述子。相应地,用 $\mathbf{B} = [b_1, b_2, \dots, b_l] \in \mathbb{R}^{d \times l}$ 表示视觉字典,其中 $b_k \in \mathbb{R}^d$ 表示第 k ($k=1, 2, \dots, l$) 个视觉单词, l 表示字典大小。

向量量化编码^[15]利用量化思想,使用一个较小的视觉字典对描述子进行表示。向量量化编码只在最近的视觉单词上响应为 1,其余视觉单词上响应为 0,因此又称为硬量化编码,如下所示:

$$\begin{cases} \min_{\mathbf{S}, \mathbf{B}} \sum_{j=1}^m \|\mathbf{x}_j - \mathbf{B}\mathbf{s}_j\|^2 \\ \text{s. t. } \|\mathbf{s}_j\|_{\ell_0} = 1, \|\mathbf{s}_j\|_{\ell_1} = 1, \mathbf{s}_j \geq 0, \forall j \end{cases} \quad (9)$$

其中,视觉字典 \mathbf{B} 使用聚类算法得到, \mathbf{s}_j 表示每个 \mathbf{x}_j 在视觉字典 \mathbf{B} 上量化得到的特征表示, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m] \in \mathbb{R}^{l \times m}$ 。根据 \mathbf{S} 和特征汇聚方法可以得到一个特征向量 \mathbf{F} ,作为三维人体骨架序列的特征表示。

稀疏编码^[16]近年来在计算机视觉领域得到了广泛关注,它在最小二乘重构的基础上加入 ℓ_1 -约束,从而实现在一个过完备视觉字典上响应的稀疏性,得到稀疏表示。稀疏编码减少了重构误差,使得重构后的描述更详细,拥有更强的判别力,如下所示:

$$\begin{cases} \min_{\mathbf{B}, \mathbf{S}} \sum_{j=1}^m (\|\mathbf{x}_j - \mathbf{B}\mathbf{s}_j\|^2 + \lambda \|\mathbf{s}_j\|_1) \\ \text{s. t. } \|\mathbf{s}_j\| \leq 1, \forall j \end{cases} \quad (10)$$

其中, \mathbf{s}_j 表示每个 \mathbf{x}_j 在视觉字典 \mathbf{B} 上的稀疏表示系数, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m] \in \mathbb{R}^{l \times m}$, $\lambda > 0$ 为正则化参数, $\|\mathbf{s}_j\|_1$ 表示系数 \mathbf{s}_j 的 ℓ_1 -范数。

求解式 10 的优化问题,就可以得到输入向量集 \mathbf{X} 的稀疏系数矩阵 \mathbf{S} 和字典 \mathbf{B} ,根据 \mathbf{S} 和特征汇聚方法可以得到一个特征向量 \mathbf{F} 作为三维人体骨架序列的特征表示。

局部约束线性编码^[17]利用局部性比稀疏性更为本质的思想,通过加入局部线性约束,在一个局部流形上对描述子进行编码重构,如下所示:

$$\begin{cases} \min_{\mathbf{B}, \mathbf{S}} \sum_{j=1}^m (\|\mathbf{x}_j - \mathbf{B}\mathbf{s}_j\|^2 + \lambda \|\mathbf{w}_j \odot \mathbf{s}_j\|^2) \\ \text{s. t. } \mathbf{1}^T \mathbf{s}_j = 1, \forall j \end{cases} \quad (11)$$

其中, \mathbf{s}_j 表示每个 \mathbf{x}_j 在视觉字典 \mathbf{B} 上的稀疏系数, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m] \in \mathbb{R}^{l \times m}$, $\lambda > 0$ 为正则化参数,局部正则化项 $\|\mathbf{w}_j \odot \mathbf{s}_j\|^2$ 能够确保相似的特征向量具有相似的编码。 \odot 表示向量内元素两两相乘, $\mathbf{w}_j \in \mathbb{R}^l$ 且

$$w_{jk} = \exp\left[\frac{\text{dist}(x_j, b_k)}{\sigma}\right] \quad (12)$$

其中, dist 函数表示特征向量 \mathbf{x}_j 和视觉单词 b_k 间的欧氏距离; $\mathbf{1}$ 表示全部元素为 1 的向量,约束 $\mathbf{1}^T \mathbf{s}_j = 1$ 保证编码的平移不变性。

为加速编码过程,文献[17]提出一种简单的近似 LLC 编码方法。该方法首先采用聚类算法学习视觉字典 \mathbf{B} ,然后对任意一个待编码的特征向量 \mathbf{x}_j ,选取视觉字典 \mathbf{B} 中距离其最近的 t 个视觉单词,形成子字典 $\mathbf{B}_j = [b_{[1]}, b_{[2]}, \dots, b_{[t]}]$;最后使用子字典 \mathbf{B}_j 重构特征向量 \mathbf{x}_j ,即优化式 13 得到编码 $\mathbf{s}_j \in \mathbb{R}^l$,如下所示:

$$\begin{cases} \min_{\mathbf{S}} \sum_{j=1}^m (\|\mathbf{x}_j - \mathbf{B}_j \mathbf{s}_j\|^2) \\ \text{s. t. } \mathbf{1}^T \mathbf{s}_j = 1, \forall j \end{cases} \quad (13)$$

式 13 可以使用拉格朗日乘子法求得解析解。将最终编码 $\mathbf{s}_j^* \in \mathbb{R}^l$ 的第 $([1], [2], \dots, [t])$ 维的值设置为 $(s_{[1]}, s_{[2]}, \dots, s_{[t]})$,令其他维的值为 0 即可得到 \mathbf{s}_j 的解。根据 \mathbf{S} 和特征汇聚方法可以得到一个特征向量 \mathbf{F} ,作为三维人体骨架序列的特征表示。

1.3 文中方法

文中方法的实现框架如图 2 所示。

对于三维人体骨架序列训练集,首先提取运动姿态描述子和角度描述子,其次对每种描述子通过聚类学习和字典学习的方法得到不同的视觉字典,将运动姿态描述子和角度描述子分别采用与之对应的视觉字典进行特征编码,并在向量量化编码、稀疏编码和局部约束线性编码的处理下得到 6 种特征。最后将训练样本集的这 6 种特征分别采用线性分类器 (<http://www.csie.ntu.edu.tw/~cjlin/liblinear/>) 进行训练,得到 6 个基分类器。

对于三维人体骨架序列测试样本,首先分别提取运动姿态描述子和角度描述子,每种描述子在与之对应的视觉字典上分别进行向量量化编码、稀疏编码和局部约束线性编码,最终将得到 6 种不同的特征;然后用训练时得到的 6 个对应分类器进行识别,得到 6 种不同的识别结果;最后通过投票策略将 6 个基分类器的识别结果集成起来,得到最终的识别结果。

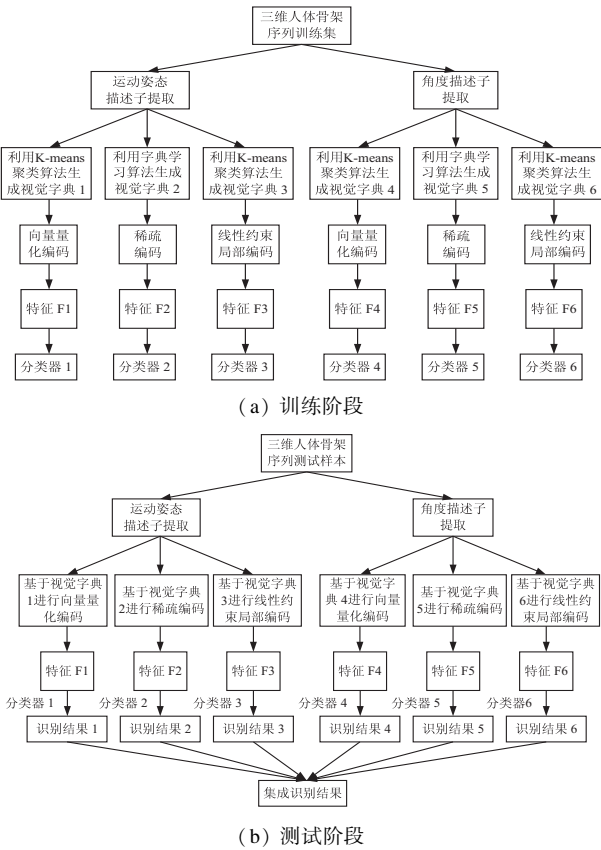


图 2 基于多描述子特征编码的人体行为识别框架

2 实验结果与分析

为验证文中方法的有效性,在最常用的三维人体骨架序列行为数据集 MSR Action3D 上进行实验。

MSR Action3D 数据集包含 20 种人体行为,这些人体行为包含手部行为、腿部行为、躯干行为以及复杂行为。由于数据集中存在许多非常相似的行为,彼此之间容易混淆,这使得该数据库识别起来很具有挑战性,每种类型的人体行为由 10 个表演者重复表演 2 至 3 次。

为了进行公平的比较,参照文献[7-12]的设置,将编号为 1、3、5、7、9 的表演者的骨架序列数据集作为训练数据,编号为 2、4、6、8、10 的表演者的骨架序列数据集作为测试数据,实验中视觉字典的大小 $l = 4\ 096$ 。

表 2 给出了文中方法和其他方法^[7-12] 的识别结果

方法	识别精度/%
Actionlet Ensemble ^[7]	88.2
Lie Group ^[9]	89.5
Angles Covariance ^[12]	91.1
Moving Pose ^[8]	91.7
Hierarchical Pose ^[11]	93.0
Moving Poselets ^[10]	93.6
万友数据	94.9

果。由表 2 可知:文中方法的识别率达到了 94.9%,相比于其他 6 种方法,识别精度提高了 1.3% ~ 6.7%;相比于 Angles Covariance 使用基于角度描述子协方差特征方法进行识别,文中方法的识别精度提高了 3.8%;相比于 Moving Pose 使用运动姿态描述子和 KNN 方法进行识别,文中方法的识别精度提高了 3.2%。上述结果说明了该方法能够提高识别率。

为考察视觉字典大小对文中方法识别性能的影响,分别采用 6 种不同大小的视觉字典,即 256、512、1 024、2 048、4 096 和 8 192,在 MSR Action3D 数据集上进行了实验,识别性能对比情况如图 3 所示。从图 3 可以看出,视觉字典大小为 4 096 时获得的识别准确率最高,而 256 时最低,字典大小为 1 024、2 048 和 8 192 时获得的识别准确率相近,但都低于 4 096 时获得的识别准确率,因此,文中将视觉字典的大小设置为 4 096。

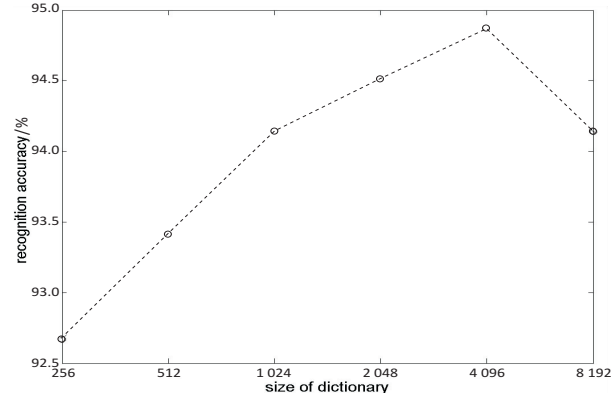


图 3 视觉字典大小对文中方法识别性能的影响

为验证文中方法能有效解决使用单一描述子提取的信息量不足,及单一特征编码方法对三维人体骨架序列造成信息丢失等问题,同时为验证该方法具有性能提升的潜力,进一步对描述子和特征编码的各种组合对 MSR Action3D 数据集的识别率性能的影响进行了实验,内容如下。

组合 1:角度描述子和向量量化编码组合识别人体行为。

组合 2:角度描述子和稀疏编码组合识别人体的行为。

组合 3:角度描述子和局部约束线性编码组合识别人体行为。

组合 4:运动姿态特征描述子和向量量化编码组合识别人体行为。

组合 5:运动姿态描述子和稀疏编码组合识别人体行为。

组合 6:运动姿态描述子和局部约束线性编码组合识别人体行为。

组合 7:角度描述子与文中的 3 种编码方法集成

识别人体行为。

组合 8:运动姿态描述子与文中的 3 种编码方法集成识别人体行为。

组合 9:运动姿态描述子和角度描述子与文中的 3 种编码方法集成(即本文方法)识别人体行为。

各种组合的实验结果如表 3 所示。从中可以看出,单独使用角度描述子和特征编码方法的识别精度比文中方法下降了 9.2%~12.1%,单独使用运动姿态描述子和特征编码方法的识别精度比文中方法下降了 1.5%~3.7%,角度描述子和文中的 3 种编码方法集成的识别精度比文中方法下降了 6.6%,运动姿态描述子和文中的 3 种编码方法集成的识别精度比文中方法下降了 0.4%。上述结果均说明,文中方法具有识别性能更好的优点。

表 3 各种组合在 MSR Action3D 数据库上的实验结果

各种组合方法	识别精度/%
组合 1	82.8
组合 2	85.7
组合 3	85.7
组合 4	91.2
组合 5	92.7
组合 6	93.4
组合 7	88.3
组合 8	94.5
组合 9	94.9

3 结束语

为解决利用单一描述子和单一特征编码方法进行三维人体骨架序列行为识别导致识别率较低的问题,提出一种多描述子特征编码的三维人体骨架序列行为识别方法,在 MSR Action3D 数据集上的实验结果表明该方法能够提高识别精度。

参考文献:

[1] 胡 琼,秦 磊,黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报,2013,36(12):2512-2524.

[2] 黄菲菲,曹江涛,姬晓飞. 基于多通道信息融合的多人交互动作识别算法[J]. 计算机技术与发展,2016,26(3):58-62.

[3] 陈万军,张二虎. 基于深度信息的人体动作识别研究综述[J]. 西安理工大学学报,2015,31(3):253-264.

[4] 田国会,尹建芹,韩 旭,等. 一种基于关节点信息的人体行为识别新方法[J]. 机器人,2014,36(3):285-292.

[5] HAN Fei,REILY B,HOFF W,et al. Space-time representation of people based on 3D skeletal data;a review[J]. Com-

puter Vision & Image Understanding,2017,158:85-105.

[6] 单言虎,张 彰,黄凯奇. 人的视觉行为识别研究回顾、现状及展望[J]. 计算机研究与发展,2016,53(1):93-112.

[7] WANG Jiang,LIU Zicheng,WU Ying,et al. Mining actionlet ensemble for action recognition with depth cameras[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Providence, RI, USA: IEEE, 2012: 1290-1297.

[8] ZANFIR M,LEORDEAN M,SMINCHISESCU C. The moving pose:an efficient 3D kinematics descriptor for low-latency action recognition and detection[C]//Proceedings of IEEE conference on computer vision. Sydney, NSW, Australia:IEEE,2013:2752-2759.

[9] VEMULAPALLI R,ARRATE F,CHELLAPPA R. Human action recognition by representing 3d skeletons as points in a lie group[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Columbus, OH, USA: IEEE, 2014:588-595.

[10] TAO Lingling,VIDAL R. Moving poselets: a discriminative and interpretable skeletal motion representation for action recognition[C]//Proceedings of IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 61-69.

[11] LILLO I,NIEBLES C J,SOTO A. A hierarchical pose-based approach to complex action understanding using dictionaries of actionlets and motion poselets[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Las Vegas,NV,USA:IEEE,2016:1981-1990.

[12] SHARAF A,TORKI M,HUSSEIN M E,et al. Real-time multi-scale action detection from 3D skeleton data[C]//IEEE winter conference on applications of computer vision. Waikoloa, HI, USA:IEEE,2015:998-1005.

[13] 张友梅,常发亮,刘洪彬. 基于 3D 人体骨架的动作识别[J]. 电子学报,2017,45(4):906-911.

[14] NOWOZIN S,SHOTTON J. Action points: a representation for lowlatency online human action recognition[R]. Cambridge, England:Microsoft Research Cambridge,2012.

[15] SIVIC J,ZISSERMAN A. Video Google: a text retrieval approach to object matching in videos[C]//IEEE international conference on computer vision. Nice, France: IEEE, 2003: 1470-1477.

[16] YANG Jianchao,YU Kai,GONG Yihong,et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Miami,FL,USA:IEEE,2009: 1794-1801.

[17] 李宗民,蒋 迪,刘玉杰,等. 结合空间上下文的局部约束线性特征编码[J]. 计算机辅助设计与图形学学报,2017,29(2):254-261.