

基于时空方向主成分直方图的人体行为识别

徐海洋¹, 孔军^{1,2}, 蒋敏¹, 笞宝锋¹

¹江南大学物联网工程学院, 江苏 无锡 214122;

²新疆大学电气工程学院, 新疆 乌鲁木齐 830047

摘要 为解决由视角、尺度等变化造成的行为类内差别大的问题, 提出一种基于时空方向主成分直方图(HSTOPC)的人体行为识别方法。首先, 将深度图序列转换为三维(3D)点云序列, 对此序列采用新颖的图像预处理方法获得新的深度图序列, 即在空间和时间维度上对采样获得的深度图序列进行限制, 除去一些动作信息量较少的序列和空间, 从而降低输入数据的冗余减少空间尺度变化的影响; 其次, 为了解决帧间关联性较弱的问题, 采用时空方向主成分方法描述新的点云序列, 获得3D点云序列中每点的方向特征; 然后, 对3D点云序列中所有方向特征进行多层时域重叠分割, 获得HSTOPC特征描述子; 最后, 采用支持向量机分类器进行训练、测试。在3个标准数据库上的实验结果表明, 提出的HSTOPC特征描述子对噪声、运动速度变化、视角变化和时域不对齐具有很好的稳健性, 可以显著提高人体行为识别的准确率。

关键词 图像处理; 人体行为识别; 时空限制; 特征向量和特征值; 多层时域重叠分割; 点云; 支持向量机

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP55.061009

Action Recognition Based on Histogram of Spatio-Temporal Oriented Principal Components

Xu Haiyang¹, Kong Jun^{1,2}, Jiang Min¹, Zan Baofeng¹

¹ School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China;

² College of Electrical Engineering, Xinjiang University, Urumqi, Xinjiang 830047, China

Abstract In order to solve the problem of inter-class difference caused by the angle of view and scale change, we propose a method based on histogram of spatio-temporal oriented principal components of three-dimensional (3D) point clouds for action recognition. Firstly, the depth sequences are converted into 3D point clouds sequences. Then, we use a novel image preprocessing method to get new depth sequences. Namely, the sampled depth sequences are limited in spatio-temporal dimension to remove areas with less information, and reduce the redundancy of the input data and the influence of space scale change. In order to solve the problem of weak correlation between frames, we adopt histogram of spatio-temporal oriented principal components (HSTOPC) method to describe 3D point clouds sequences and obtain the direction of each point of the 3D point clouds in sequences. For all direction of 3D point clouds in sequences, we use multilayer overlapping segmentation method to obtain HSTOPC descriptor. Finally, we use the support vector machine classifier for training and test. Experimental results on three human action recognition datasets show that the proposed HSTOPC feature descriptor has better robust for noise, rate variations, view change and temporal misalignment, and is able to improve the accuracy of human behavior recognition significantly.

Key words image processing; human action recognition; spatio-temporal limitation; eigenvector and eigenvalue; multi-layer overlap segmentation; point clouds; support vector machine

OCIS codes 100.4993; 100.4994; 100.5010

收稿日期: 2017-11-28; 收到修改稿日期: 2017-12-28

基金项目: 国家自然科学基金(61362030, 61201429)、中国博士后科学基金(2015M571720, 2016M606360)、江苏省博士后科学基金(1601416C)、公安部技术研究计划(2014JSYJB007)

作者简介: 徐海洋(1992—), 男, 硕士研究生, 主要从事人体行为识别、机器学习等方面的研究。

E-mail: 1041922650@qq.com

导师简介: 孔军(1974—), 男, 博士, 副教授, 硕士生导师, 主要从事人脸定位、目标跟踪、行为识别等方面的研究。

E-mail: kongjun@jiangnan.edu.cn(通信联系人)

1 引言

在人体行为识别领域,基于深度传感器(如 Kinect^[1])的算法解决了传统方法中数据难以处理的问题,逐渐引起人们的关注。通过深度传感器获得的深度图可以准确地提取出比颜色信息更具有区分度的几何信息,而且,深度图序列不易受光照、纹理和亮度的影响。但是,深度图序列仍存在遮挡、速度变化、传感器噪声,以及摄像视角不同等问题,因此基于深度图序列的人体行为识别依然具有一定的挑战性。

基于深度图序列的人体行为识别的首要任务是设计一个好的特征描述子,其次是选取恰当的分类器。在选取特征描述子方面,最初使用的特征描述子^[2-4]主要是基于 RGB 图像的特征描述,但这种基于颜色信息的方法不能直接应用于深度序列中^[5]。近年来,涌现出很多基于几何特征的行为识别方法^[5-8]。Tang 等^[5]通过深度图序列中的法向量直方图进行行为识别,采用极坐标表示每张深度图,通过计算其空间导数,来获得二维直方图描述子。黄成挥^[6]运用时空特征协方差矩阵表示视频中的人体行为,建立基于 Log-Euclidean 词袋模型和基于 Stein 核系数编码的人体行为识别方法;王军^[7]把分解后的行为看作一个包,把各个动作看作包中的各个示例,结合多示例学习与 AnyBoost 算法提出多示例行为学习算法,通过在多示例框架下对每一类行为进行学习,得到强分类器,用于未知行为包的识别;刘智等^[8]根据深度图序列中的人体关节信息分别提取表示关节之间角度和相对位置的两个特征向量,然后使用 LIBLINEAR-支持向量机(SVM)分类器分别对提取的两个特征向量进行分类识别,最

后通过融合分类结果,得到最终的行为识别结果。在分类器方面,使用较多的是 SVM^[9]、神经网络^[10-11]。张燕君等^[9]将粒子群算法与最小二乘 SVM 相结合,应用于 304 钢板损伤识别研究中,使预测模型具有自诊断功能;刘峰等^[10]将可见光、中波红外和长波红外三个波段的舰船目标特征进行融合识别,性能较单波识别方法有明显提升;毕立恒等^[11]通过径向基函数神经网络建立模型,并采用多环量子算法确定各环量个体选择概率,实现径向基函数神经网络参数优化,从而提高植物叶片图像的认识率。

在上述方法的基础上,提出一种基于三维(3D)点云的时空方向主成分直方图(HSTOPC)的人体行为识别方法。首先,提出一种基于运动集中的时空区域提取方法,对获取到的深度图序列转换为 3D 点云序列,在空间和时间维度上对视频序列进行限制,去除在时空中影响因子较小的序列和空间区域;其次,对获得的运动信息较为集中的 3D 点云序列求出符合条件的三维点序列中的点,对三维点序列中的点的散度矩阵求出其特征向量和特征值,将对应的特征向量映射到柏拉图立体中,从而获得每帧中符合条件的点的时空主成分;然后,通过多层时域重叠分割方法获得最终的 3D 点云的时空方向主成分直方图描述子;最后,采用 SVM 分类器在三个标准数据库进行训练和测试。

2 基本原理

提出的基于 3D 点云的时空方向主成分直方图的行为识别方法包括数据预处理、特征提取、多层时域重叠分割三个步骤。具体的操作步骤如图 1 所示,首先对获取到的深度图序列转换为 3D 点云

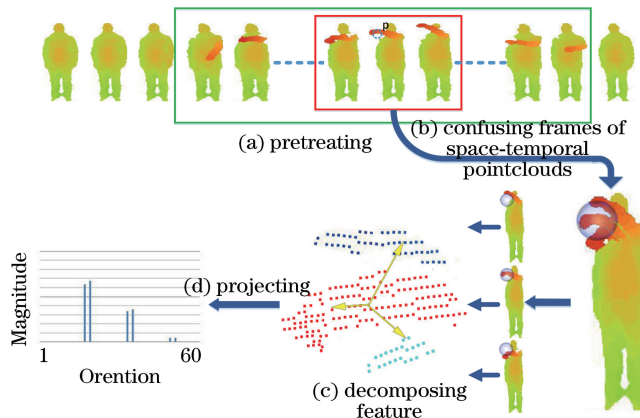


图 1 形成描述子 HSTOPC 的步骤

Fig. 1 Steps of HSTOPC descriptor generation

序列,对此序列在时间和空间维度上进行限制[图 1(a)],提取出运动信息较多的新序列(如图 1 中绿线框标注的序列);然后对点云序列中每个点获取符合条件的点[图 1(b)],计算所有的点对应的散度矩阵的特征值和特征向量,将特征向量映射到柏拉图立体中[图 1(c)],从而获每个点云点的特征信息;最后采用多层时域重叠分割方法获得 HOTSPC 描述子[图 1(d)].

2.1 时空区域中的数据预处理方法

为了获得有效的特征描述子,首先对由深度视频序列转换的 3D 点云序列进行如下预处理.令 $Q = \{Q_1, Q_2, \dots, Q_t, \dots, Q_{n_f-1}, Q_{n_f}\}$ 表示由深度传感器捕获的 3D 点云序列,其中 n_f 表示序列总个数,

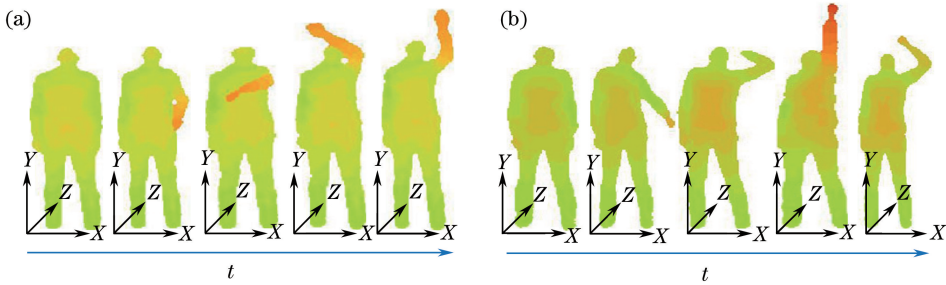


图 2 (a)高抛与(b)画圆动作的 3D 点云序列

Fig. 2 Action sequences of 3D point clouds of (a) high throw and (b) draw circle

2.1.1 LHW 方法

针对已提取前景的点云序列采用 LHW 方法.在时间维上,先提取每帧图像的人体轮廓的最小外接矩形框,计算每帧最小外接矩形框的高度和宽度的比值 α_t .进而计算出帧间的变化比值为

$$\gamma_t^{-1} = \left| \frac{\alpha_{t-1}}{\alpha_t} - 1 \right|. \quad (1)$$

设置一个阈值 η_{LHW} ,找出序列中首次和最后大于 η_{LHW} 的时间点,作为序列的起始时刻 s 和结束时刻 e .在空间范围内,提取在实施阶段内人体占用的最大矩形框.

2.1.2 LEE 方法

针对未提取前景的点云序列采用 LEE 方法.在时间维上,先计算每一帧的散度矩阵 C_t ,其中假设在点云帧 Q_t 中任意一点 $q_i = (x_i, y_i, z_i)^T$ [12], $1 \leq t \leq n_f$.

$$C_t = \frac{1}{n_t} \sum_{q_i \in Q_t} (q_i - \mu')(q_i - \mu')^T, \mu' = \frac{1}{n_t} \sum_{q_i \in Q_t} q_i, \quad (2)$$

式中 n_t 表示 t^{th} 序列上点的数目.对散度矩阵 C_t 特征分解 $CV = EV$,其中 E 表示以散度矩阵 C 的特征

Q_t 表示由深度视序列中 t 时刻深度图形成的 3D 点云帧(图 2).每个序列中都存在单位信息量少或无用的时空区域,去除上述区域可以减少运算时间.将一组运动序列分为准备阶段、过程阶段和结束阶段,由于准备阶段和结束阶段与其他的运动序列相似,且过程阶段的移动能量占比最大,因此把过程阶段作为有效判断行为类别的数据区域.根据输入的深度序列是否去除背景,采用不同的预处理方法基于高宽比的限制(LHW)和基于特征值和特征向量的限制(LEE)获得新的 3D 点云序列 $Q' = \{Q'_s, Q'_{s+1}, \dots, Q'_t, \dots, Q'_{e-1}, Q'_e\}$,其中 Q'_s 和 Q'_e 分别表示实施阶段的起始时刻 s 的点云帧和结束时刻 e 的点云帧[图 1(a)].

值 $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) 为对角矩阵, V 表示由散度矩阵 C 的三个正交特征向量组成的矩阵 $[v_1, v_2, v_3]$,通过 $v'_j = \sum_{j=1}^3 \lambda_j \times v_j$ 计算每一帧的主向量.然后将向量 v'_j 和向量 v'_{j+1} 进行点乘,从而获得向量之间的相似程度,最后设定一个阈值 η_{EVL} ,找出序列中首次和最后小于 η_{EVL} 的时间点,分别对应起始时刻 s 和结束时刻 e .

2.2 时空方向主成分特征提取

预处理之后的点云序列 $Q' = \{Q'_s, Q'_{s+1}, \dots, Q'_t, \dots, Q'_{e-1}, Q'_e\}$,取时间间隔 $[t - \tau, t + \tau]$,将 3D 点云序列融合形成一个时空累积的 3D 点云帧[图 1(b)].假设在点云帧 Q'_t 中任意一点 $p = (x_t, y_t, z_t)^T$, $s \leq t \leq e$, $\Omega(p)$ 表示一个以点 p 为中心, r 为半径的球体时空邻域(图 1),通过邻域 $\Omega(p)$ 的散度矩阵 C 来描述点 p [图 1(c)].

$$\begin{cases} C = C_{t-\tau} \times C_{t-\tau+1} \times \dots \times C_{t+\tau-1} \times C_{t+\tau} \\ C_i = \frac{1}{n_i^p} \sum_{q_i \in \Omega(p_i)} (q_i - \mu)(q_i - \mu)^T \end{cases}, \quad (3)$$

式中 $\mu = \frac{1}{n_i} \sum_{q_i \in \Omega(p)} q_i$, n_i 表示 $\Omega(p)$ 中点的数目, $\Omega(p_i)$ 表示第 i ($i \in [t - \tau, t + \tau]$) 点云帧在点 p 的空间邻域

内的点, $n_i^p = |\Omega(\mathbf{p}_i)|$ 表示第 i ($i \in [t-\tau, t+\tau]$) 点云帧在点 \mathbf{p} 的空间邻域内的点个数。 \mathbf{p}_i 表示在 $\Omega(\mathbf{p}_i)$ 中的点。将散度矩阵 \mathbf{C} 进行矩阵分解, 即 $\mathbf{C}\mathbf{V} = \mathbf{E}\mathbf{V}$, 其中 \mathbf{E} 表示由散度矩阵 \mathbf{C} 的特征值 $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) 组成的对角矩阵, \mathbf{V} 表示由散度矩阵 \mathbf{C} 的三个正交特征向量组成的矩阵 $[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ 。

在点 \mathbf{p} 的邻域 $\Omega(\mathbf{p})$ 中, HSTOPC 描述子与特征值的顺序存在相关性。文献[12]提出了消除每一个点的模糊特征向量的方法, 其中 δ_{12}, δ_{23} 定义如下:

$$\delta_{12} = \frac{\lambda_1}{\lambda_2}, \delta_{23} = \frac{\lambda_2}{\lambda_3}. \quad (4)$$

假设是 3D 对称曲面, 则 δ_{12} 或 δ_{23} 的值等于 1; 假设曲面的主方向发生变化, 则 δ_{12} 和 δ_{23} 的值必大于 1, 即存在判别性点。鉴于存在噪声等干扰因素, 设置阈值 $\theta > 1 + \epsilon$, 其中 ϵ 表示噪声临界值, 如果 $\delta_{12} < \theta, \lambda_1, \lambda_2$ 值设为 0; 如果 $\delta_{23} < \theta, \lambda_2, \lambda_3$ 值设为 0。

特征向量表示 3D 空间中相应点的最大变化方向, 但存在方向歧义性问题。为了消除该歧义性, 采用特征向量 \mathbf{v}_j 与 \mathbf{p} 空间邻域内所有向量的内积符号作为特征向量 \mathbf{v}_j 的符号: $\text{sign}\left\{\sum_{\mathbf{q} \in \Omega(\mathbf{p})} \text{sign}(\mathbf{o}^T \mathbf{v}_j)(\mathbf{o}^T \mathbf{v}_j)^2\right\}$, 其中 $\mathbf{o} = \mathbf{q} - \mathbf{p}$ 。

将散度矩阵 \mathbf{C} 对应特征值非 0 的特征向量投影到正 m 面体的 m ($m=20$) 个方向上[图 1(c)]。 $\mathbf{U}\{\{\mathbf{u}_i\}_{i=1}^m\}$ 表示正 m 面体中心到面中心的所有方向向量构成的矩阵。

对于一个以中心为原点的正二十面体, 其方向向量归一化表示为 $\left(\frac{\pm 1}{L_u}, \frac{\pm 1}{L_u}, \frac{\pm 1}{L_u}\right), \left(0, \frac{\pm \varphi^{-1}}{L_u}, \frac{\pm \varphi}{L_u}\right),$

$\left(\frac{\pm \varphi^{-1}}{L_u}, \frac{\pm \varphi}{L_u}, 0\right), \left(\frac{\pm \varphi}{L_u}, 0, \frac{\pm \varphi^{-1}}{L_u}\right)$, 其中 $\varphi = \frac{1+\sqrt{5}}{2}$,

$L_u = \sqrt{\frac{\varphi^2+1}{\varphi^2}}$ 表示向量 \mathbf{u}_i ($1 \leq i \leq m$) 的长度。

然后将每一个特征向量 \mathbf{v}_j 映射到 \mathbf{U} 上:

$$\mathbf{b}_j = \mathbf{U}^T \mathbf{v}_j \in \mathbf{R}^m, 1 \leq j \leq 3. \quad (5)$$

若特征向量 \mathbf{v}_j 和 $\mathbf{u}_i \in \mathbf{U}$ 的方向一致, 那么 \mathbf{v}_j 的幅值应该完全投影到第 i 个 bin。但由于 \mathbf{u}_i 之间不存在正交, \mathbf{b}_j 在其他 bin 中的投影值可能是非零值。为了克服上述影响, 采用由任意两个相邻向量 \mathbf{u}_k 和 \mathbf{u}_l 间的投影计算得到的阈值 $\phi^{[12]}$ 对 \mathbf{b}_j 进行量化处理。

$$\phi = \mathbf{u}_k^T \mathbf{u}_l = \frac{\varphi + \varphi^{-1}}{L_u^2}, \mathbf{u}_k, \mathbf{u}_l \in \mathbf{U}. \quad (6)$$

量化后的向量表示为

$$\hat{\mathbf{b}}_j(z) = \begin{cases} 0, & \mathbf{b}_j(z) \leq \phi \\ \mathbf{b}_j(z) - \phi, & \mathbf{b}_j(z) > \phi \end{cases}, \quad (7)$$

式中 $1 \leq z \leq m$ 表示向量分量下标。对于第 j 个特征向量, \mathbf{h}_j 表示对相应 $\hat{\mathbf{b}}_j$ 进行尺度归一化, 具体表示为

$$\mathbf{h}_j = \frac{\lambda_j \hat{\mathbf{b}}_j}{\|\hat{\mathbf{b}}_j\|_2} \in \mathbf{R}^m, 1 \leq j \leq 3. \quad (8)$$

最后按照特征值 λ 的降序将三个特征向量的时空主成分直方图相连接形成点 \mathbf{p} 的描述子:

$$\mathbf{h}_p = [\mathbf{h}_1^T, \mathbf{h}_2^T, \mathbf{h}_3^T]^T \in \mathbf{R}^{3m}. \quad (9)$$

2.3 多层时域重叠分割

为了获取相同维度的特征, 对时空区域块中所有点的描述子 \mathbf{h}_p 进行叠加, 其中时空区域块是时空域上采用不同分割方式获得的。首先, 对空间人体运动区域分割, 从而捕捉到人体躯干的运动情况, 不仅减少了待处理的数据量, 而且在一定程度上减小了背景影响。其次, 在时域分割方面, 保持分割处信息之间的相关性, 以增强对噪声因素的稳健性。本文采用多层时域重叠分割模型, 在局部时间维度上聚集方向信息。模型的构建思路: 第一层是对整个序列在时间维度上进行均匀分割; 第二层和第三层是在第一层时域分割处的基础上前后重叠一定的帧数^[13], 鉴于预处理之后的点云序列总帧数不同, 重叠帧数 o 可由总帧数 $n = e - s + 1$ 计算得到, $o = n \times \delta$, 其中重叠系数 $\delta = 0.11$; 最后将这三层按空间和时间顺序连接成 HSTOPC 向量。模型如图 3 所示, 其中, $\lfloor n/3 \rfloor$ 表示 n 和 3 数字相除之后, 对 $n/3$ 向下取值, 即取小于等于这个数字的最大整数, o 表示第二层的重叠数目。

3 实验结果和讨论

为了验证本文算法的识别效果, 分别采用 MSR Action3D 数据库^[14]、ActionPairs3D 数据库^[15] 和 UWA3D 数据库^[12] 进行测试。

为了便于性能比较, 本文遵循文献[12]中的实验设置(所有的实验使用固定的球半径 r 和时间跨度 τ , 未使用文献[12]中的自适应方法)。在本实验中, 使用 $\theta = 1.12, \eta_{\text{WHL}} = 0.11, \eta_{\text{EVL}} = 1.4$, 每个 3D 点云序列按照 X, Y 和 T 维度(空间和时间维度)被分为 $6 \times 7 \times 3$ 的时空体。

3.1 MSR Action3D 数据库

MSR Action3D 数据库^[14]是由 10 个人完成 20 类动作的深度图序列构成, 每个人的每类动作采集 2~3 次, 深度图片的分辨率为 320×240 。虽然此数

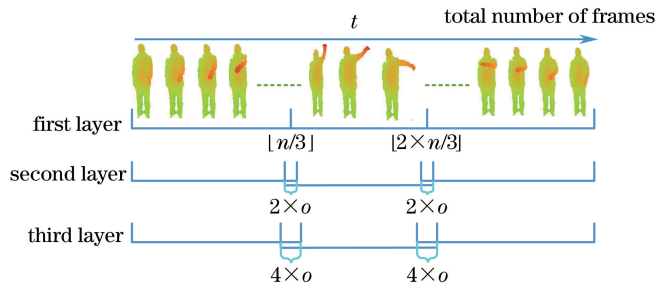


图3 HSTOPC的多尺度时域重叠分割

Fig. 3 Multi-scale temporal overlap segmentation method of HSTOPC

数据库的背景经过预处理,但是20种动作中存在许多非常相似的行为,且类内差别大。

本文选择近几年比较典型的识别方法作为比较对象,实验中使用相同的实验设置^[12]。测试1和测试2的识别精度对比如表1和表2所示,其中5/5表示数据库中受试者编号{1,3,5,7,9}的到受试者用于训练,而编号为{2,4,6,8,10}的受试者用于测试。。表1中将HSTOPC描述子与其他识别方法进行比较,HSTOPC+LHW方法比基于3D面的直方图(H3DF)方法^[16]高5.06%,比姿势集合方法(Pose Set)方法^[17]高4.51%,比3D梯度方向直方图(HOG3D)+基于位置的线性编码方法(LLC)方法^[18]高3.61%,比多层3D核描述子(HKD)方法^[19]高1.78%,比深度运动图-局部二值模式-决策层融合(DMM-LBP-DF)方法^[18]高1.51%。Pose Set方法由于不规则的遮挡导致关节容易连接错误,文献^[17]选择最好的 k 个关节点组合模型,在很大程度上消除了由不准确的关节点造成的影响。本文方法优于文献^[16]和文献^[17]方法,原因是点云对遮挡不敏感,并且提供额外的类似于骨骼关节的形状信息,更好地描述了帧间的前后连续信息,弥补了分割处的信息。

表1 MSR Action3D数据库的识别率比较

Table 1 Recognition rate comparison on MSR Action3D dataset

Method	Recognition rate /%
H3DF ^[16]	89.45
Pose Set ^[17]	90.00
HOG3D+LLC ^[18]	90.90
HKD ^[19]	92.73
DMM-LBP-DF ^[20]	93.00
HOPC ^[12]	91.64
HOPC+LHW	93.77
HOPC+LEE	93.04
HSTOPC+LHW	94.51
HSTOPC+LEE	94.14

在主成分方向直方图(HOPC)方法基础上增加LHW方法,则识别率提高了2.13%,对应地,增加EIL方法,则识别率提高了1.4%。可以看出,图像预处理方法LHW和LEE均可以提高识别精度。这是由于加上了对时空区域的限制,可以提取更多的有效区域。HSTOPC+LHW方法使得识别率提高2.87%;HSTOPC+LEE方法使得识别率提高2.5%。以上比较结果表明,HSTOPC描述子可以很好地表示行为。这是由于多尺度时域重叠分割的计算和时间域的增强会提供更多可用于判别的信息。

表2采用一个个体用于测试其他个体用于训练的方式,表中的数据分别表示10次分类识别的平均值(Mean)、均方差(STD)、最大识别率(Max)和最小识别率(Min)。与HOPC对应项比较,HSTOPC的结果均提高。这是由于通过数据的预处理,获得了信息集中的时空域,增强了时域特性,这也进一步证明了时空限制和时间信息增强的重要性。

表2 不同方法的实验结果比较

Table 2 Comparison of experimental results with different methods

Method	Mean±STD	Max /%	Min /%
HOPC ^[12]	87.06±10.63	96.55	66.67
HOPC+LHW	89.55±9.55	100	68.21
HOPC+LEE	89.71±9.80	100	68.42
HSTOPC+LHW	90.41±9.95	100	69.81
HSTOPC+LEE	90.56±9.05	100	71.93

图4所示为混淆矩阵,大多数识别错误为手抓、画叉、画圆。因为大多数动作涉及到相同的身体部位,容易产生一些相似的动作特性,所以这个数据库难以准确地提取序列中几何数据与动作的关联信息。

3.2 ActionPairs3D数据库

ActionPairs3D数据库^[15]是由10个人完成6对动作的深度序列构成,每个人每类动作重复3次采样。其中成对的动作具有高度相似的轮廓和运

动。按照文献[12]的实验设置,实验结果如表3所示,将HSTOPC描述子与现有的典型方法进行比较,本文方法比HOPC方法的识别率提高了

1.11%。这是因为本文方法提取运动信息集中区域,并且强调运动的先后顺序,同时对运动序列具有一定的对齐效果。

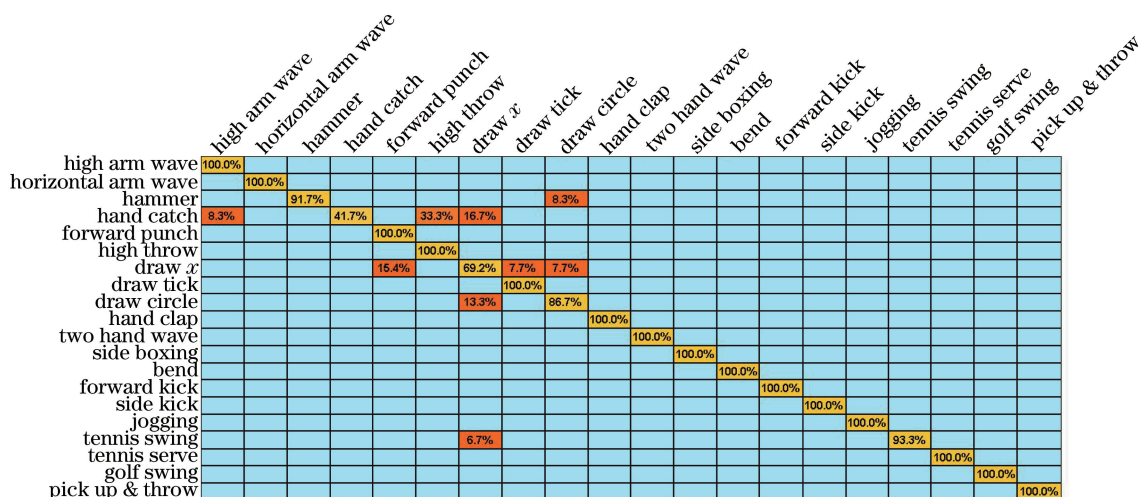


图4 HOPCT+WHL算法在MSR Action3D数据集测试5/5实验结果的混淆矩阵

Fig. 4 Confusion matrix of the HOPCT+WHL algorithm by experiment setting of 5/5 on MSR action 3D dataset

表3 ActionPairs3D数据库上的识别率比较

Table 3 Accuracy comparison on MSR Action3D dataset

Method	Mean±STD	Max / %	Min / %	5/5
STPC ^[15]	-	-	-	98.33
HOPC ^[12]	97.15±2.21	100	88.89	97.22
HSTOPC+LEE	98.18±2.75	100	90.56	98.33

3.3 UWA3D数据库

UWA3D数据库^[12]是由10个个体在不同尺度下完成30种行为的深度序列构成,并以分辨率 $640 \times 480, 30 \text{ frame} \cdot \text{s}^{-1}$ 设置的。采集的样本主要针对如下影响因子:1)个体间尺度差异;2)视角变化;3)所有动作以一种连续且没有断开或停顿的方式进行,对于相同的动作,身体的开始和结束位置不同。本实验主要针对个体间尺度差异。数据库是从前面视角捕捉到深度视频序列,其中每个个体以随机排列的方式完成30种行为中2~3个连续的动作,然后将数据库中连续序列按行为类别分段。实验结果如表4所示。由于此数据库经过背景预处理,所以采用HSTOPC+LHW方法。实验结果显示,识别精度的均值比HOPC方法高1.34%,并且高于HON4D^[21]方法5.83%,对应的最小值和最大值都有提高。识别结果不高主要是由于UWA3D数据库中存在遮挡、相似性高等动作。例如,喝酒和接听电话是非常相似的动作,只有手的位置略有不同;躺下来和坠落是非常相似的动作,只是行动执行的速度不同。

表4 UWA3D数据库的识别率比较

Table 4 Accuracy Comparison with other description on UWA3D datasets

Method	Mean±STD	Max / %	Min / %
HON4D ^[20]	79.28±2.68	88.89	70.14
HDG ^[3]	75.54±3.64	85.07	61.90
HOPC ^[12]	83.77±3.09	92.18	74.67
HSTOPC+LHW	85.11±3.21	94.04	75.67

4 结论

通过分析已有的基于深度视频序列的人体行为识别方法,针对其中的预处理方法和特征提取方法展开研究,提出一种基于3D点云的时空方向主成分直方图方法用于行为识别。首先,将深度图序列转化为3D点云序列,通过LHW或者LEE方法对数据库进行预处理,旨在减少不重要及次要的时空区域。其次,为了突显所有时空体内的相对时间变化,在每个时空体中使用HSTOPC描述子来提高其相关性,即在每个时空体中,对每个时间层的矩阵使用矩阵连乘的方法提取时空特征。最后,除了预处理和HSTOPC特征描述子,简单的时空分割会导致分割处信息丢失,因此采用三层时域重叠分割方法。在三个公开行为数据库中进行行为识别实验,并与其他行为识别方法进行对比,结果表明,本文算法在不同的数据库中通用性较好,且识别精度有所提高。

参 考 文 献

- [1] Dong K. Human action recognition based on Kinect [D]. Wuhan: Wuhan University of Science and Technology, 2015.
董珂. 基于 Kinect 的人体行为识别研究[D]. 武汉: 武汉科技大学, 2015.
- [2] Cai J X, Feng G C, Tang X, *et al.* Human action recognition based on poses a dictionary [J]. *Acta Optica Sinica*, 2014, 34(12): 1215002.
蔡加欣, 冯国灿, 汤鑫, 等. 基于姿势字典学习的人体行为识别[J]. *光学学报*, 2014, 34(12): 1215002.
- [3] Rahmani H, Mahmood A, Huynh D Q, *et al.* Real time human action recognition using histograms of depth gradients and random decision forests [C]. 2014 IEEE Winter Conference on Application of Computer Vision, 2014: 626-633.
- [4] Cai J X, Feng G C, Tang X, *et al.* Human action recognition based on the local contour and random forest [J]. *Acta Optica Sinica*, 2014, 34 (10): 1015006.
蔡加欣, 冯国灿, 汤鑫, 等. 基于局部轮廓和随机森林的人体行为识别[J]. *光学学报*, 2014, 34(10): 1015006.
- [5] Tang S, Wang X, Lv X, *et al.* Histogram of oriented normal vectors for object recognition with a depth sensor [C]. *Asian Conference on Computer Vision*, 2013: 525-538.
- [6] Huang C H. Human action recognition algorithm based on video [D]. Chengdu: University of Electronic Science and Technology of China, 2016.
黄成挥. 基于视频的人体行为识别算法研究[D]. 成都: 电子科技大学, 2016.
- [7] Wang J. Human action recognition based on sample learning method [J]. *Information Technology*, 2016 (7): 65-70.
王军. 基于多示例学习法的人体行为识别[J]. *信息技术*, 2016(7): 65-70.
- [8] Liu Z, Dong S D. Human action recognition based on the joint movement in the depth of video information [J]. *Computer Applications and Software*, 2017, 34 (2): 189-192.
刘智, 董世都. 利用深度视频中的关节运动信息研究人体行为识别[J]. *计算机应用与软件*, 2017, 34 (2): 189-192.
- [9] Zhang Y J, Wang H M, Fu X H, *et al.* Identification of steel plate damage position based on particle swarm support vector machine [J]. *Chinese Journal of Lasers*, 2017, 44(10): 1006006.
张燕君, 王会敏, 付兴虎, 等. 基于粒子群支持向量机的钢板损伤位置识别[J]. *中国激光*, 2017, 44 (10): 1006006.
- [10] Liu F, Sheng T S, Ma X X. Convolutional neural network based multi-band ship target recognition with feature fusion [J]. *Acta Optica Sinica*, 2017, 37(10): 1015002.
刘峰, 沈同圣, 马新星. 特征融合的卷积神经网络多波段舰船目标识别[J]. *光学学报*, 2017, 37(10): 1015002.
- [11] Bi L H, Liu Y C. Plant leaf image recognition based on improved neural network algorithm [J]. *Laser & Optoelectronics Progress*, 2017, 54(12): 121102.
毕立恒, 刘云滢. 基于改进神经网络算法的植物叶片图像识别研究[J]. *激光与光电子学进展*, 2017, 54 (12): 121102.
- [12] Rahmani H, Mahmood A, Du Q H, *et al.* HOPC: histogram of oriented principal components of 3D pointclouds for action recognition [C]. *European Conference on Computer Vision*, 2014: 742-757.
- [13] Cao L. Face recognition and human motion recognition technology and application [M]. Beijing: Electronic Industry Press, 2015.
曹林. 人脸识别与人体动作识别技术及应用[M]. 北京: 电子工业出版社, 2015.
- [14] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3D points [C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2010: 9-14.
- [15] Ji X, Cheng J, Feng W. Spatio-temporal cuboid pyramid for action recognition using depth motion sequences [C]. *Eighth International Conference on Advanced Computational Intelligence, IEEE*, 2016: 208-213.
- [16] Zhang C, Tian Y. Histogram of 3Dfacets: a depth descriptor for human action and hand gesture recognition [J]. *Computer Vision & Image Understanding*, 2015, 139: 29-39.
- [17] Wang C, Wang Y, Yuille A L. An Approach to Pose-Based Action Recognition [C]. *Computer Vision and Pattern Recognition IEEE*, 2013: 915-922.
- [18] Rahmani H, Du Q H, Mahmood A, *et al.* Discriminative human action classification using locality-constrained linear coding [J]. *Pattern Recognition Letters*, 2016, 72: 62-71.
- [19] Kong Y, Sattar B, Fu Y. Hierarchical 3D kernel descriptors for action recognition using depth

- sequences [C]. IEEE International Conference on Automatic Face and Gesture Recognition, 2015: 1-6.
- [20] Chen C, Jafari R, Kehtarnavaz N. Action Recognition from Depth Sequences Using Depth Motion Maps-based Local Binary Patterns[C]. IEEE Applications of Computer Vision, 2015:1092-1099.
- [21] Oreifej O, Liu Z. HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences [C]. Computer Vision and Pattern Recognition, IEEE, 2013: 716-723.