多聚点子空间下的时空信息融合及其在行为识别中的应用

杨天金[1] 侯振杰[1,2] 李 兴[1] 梁久祯[1] 宦 娟[1] 郑纪翔[1]

摘 要 基于深度序列的人体行为识别, 一般通过提取特征图来提高识别精度, 但这类特征图通常存在时序信息缺失的问题. 针对上述问题, 本文提出了一种新的深度图序列表示方式, 即深度时空图 (Depth space time maps, DSTM). DSTM 降低了特征图的冗余度, 弥补了时序信息缺失的问题. 本文通过融合空间信息占优的深度运动图 (Depth motion maps, DMM) 与时序信息占优的 DSTM, 进行高精度的人体行为研究, 并提出了多聚点子空间学习 (Multi-center subspace learning, MCSL) 的多模态数据融合算法. 该算法为各类数据构建多个投影聚点, 以此增大样本的类间距离, 降低了投影目标区域维度. 本文在 MSR-Action3D 数据集和 UTD-MHAD 数据集上进行人体行为识别. 最后实验结果表明, 本文方法相较于现有人体行为识别方法有着较高的识别率.

关键词 行为识别, 信息融合, 深度时空图, 多聚点子空间学习

引用格式 杨天金, 侯振杰, 李兴, 梁久祯, 宦娟, 郑纪翔. 多聚点子空间下的时空信息融合及其在行为识别中的应用. 自动化学报, 2022, **48**(11): 2823−2835

DOI 10.16383/j.aas.c190327

Recognizing Action Using Multi-center Subspace Learning-based Spatial-temporal Information Fusion

YANG Tian-Jin[1] HOU Zhen-Jie[1,2] LI Xing[1] LIANG Jiu-Zhen[1] HUAN Juan[1] ZHENG Ji-Xiang[1]

Abstract Human action recognitions from depth map sequences improve the recognition accuracy by extracting feature maps. A new representation of depth map sequences called depth space time map (DSTM) is proposed in this paper for overcoming the lack of temporal information in the feature maps. DSTM reduces the redundancy of action features. We conduct high-precision human action recognitions by fusing depth motion maps (DMM) and DSTM based on a new multi-modal data fusion algorithm called multi-center subspace learning (MCSL). The algorithm constructs multiple projection centers for each class data to expand the samples inter-class distance and reduce the projection target area dimension. Experiments conducted on MSR-Action3D and UTD-MHAD depth database show the effectiveness of the proposed method.

Key words Action recognition, information fusion, depth space time maps (DSTM) , multi-center subspace learning (MCSL)

Citation Yang Tian-Jin, Hou Zhen-Jie, Li Xing, Liang Jiu-Zhen, Huan Juan, Zheng Ji-Xiang. Recognizing action using multi-center subspace learning-based spatial-temporal information fusion. *Acta Automatica Sinica*, 2022, **48**(11): 2823−2835

人体行为识别是计算机视觉领域和模式识别领域的一个重要的分支, 应用范围十分广泛, 在智能监控、虚拟现实等应用中表现十分优秀[1−5]. 传统的人体行为识别使用的是彩色摄像机[6] 生成的 RGB 图像序列, 而 RGB 图像受光照、背景、摄像器材的影响很大, 识别稳定性较差.

随着技术的发展, 特别是微软 Kinect 体感设备的推出, 基于图像序列的人体行为识别研究得到了进一步的发展. 相比于彩色图像序列, 深度图序列更有优势. 不仅可以忽略光照和背景带来的影响, 还可以提供深度信息, 深度信息表示为在可视范围内目标与深度摄像机的距离. 深度图序列相较于彩色图序列, 提供了丰富的人体 3D 信息, 胡建芳等[7] 详细描述了 RGB-D 行为识别研究进展和展望. 至今已经探索了多种基于深度图序列的表示方法, 以

收稿日期 2019-04-30 录用日期 2019-11-01
Manuscript received April 30, 2019; accepted November 1, 2019
国家自然科学基金 (61803050, 61063021), 江苏省物联网移动互联技术工程重点实验室开放课题基金 (JSWLW-2017-013), 浙江省公益技术研究社会发展项目 (2017C33223) 资助
Supported by National Natural Science Foundation of China (61803050, 61063021), Jiangsu Province Networking and Mobile Internet Technology Engineering Key Laboratory Open Research Fund Project (JSWLW-2017-013), and Zhejiang Public Welfare Technology Research Social Development Project (2017C33223)
本文责任编委 桑农
Recommended by Associate Editor SANG Nong
1. 常州大学计算机与人工智能学院 阿里云大数据学院 软件学院 常州 213164 2. 江苏省物联网移动互联技术工程重点实验室 淮安 223003
1. School of Computer Science and Artificial Intelligence Aliyun School of Big Data School of Software, Changzhou 213164 2. Jiangsu Key Laboratory of Internet of Things Mobile Internet Technology Engineering, Huai′an 223003

Bobick 等[8] 的运动能量图 (Motion energy images, MEI)、运动历史图 (Motion history images, MHI) 作为时空模板的人体行为识别的特征提取方法, 提高了识别的稳健性; 苏本跃等[9] 采用函数型数据分析的行为识别方法; Anderson 等[10] 基于 3 维 Zernike 的图像数据尝试行为分类, 并且该分类对于具有低阶矩的行为是有效的; Wu 等[11] 基于 3 维特征和隐马尔科夫模型对人体行为动作进行分类并加以识别; Wang 等[12] 从深度视频中提取随机占用模式 (Random occupancy pattern, ROP) 特征, 并用稀疏编码技术进行重新编码; Zhang 等[13] 使用梯度信息和稀疏表达将深度和骨骼相结合, 用于提高识别率; Zhang 等[14] 从深度序列中提取的动作运动历史图像 (Sub-action motion history image, SMHI) 和静态历史图像 (Static history image, SHI); Liu 等[15] 利用深度序列和相应的骨架联合信息, 采用深度学习的方法进行动作识别; Xu 等[16] 提出了深度图和骨骼融合的人体行为识别; Wang 等[17−19] 采用卷积神经网络进行人体行为识别; Yang 等[20] 提出了深度运动图 (Depth motion maps, DMM), 将深度帧投影到笛卡尔直角坐标平面上, 生成主视图、俯视图、侧视图, 得到三个 2 维地图, 在此基础上差分堆叠整个深度序列动作能量图生成 DMM. DMM 虽然展现出人体行为丰富的空间信息, 但是无法记录人体行为的时序信息. 针对现有深度序列特征图时序信息缺失的问题, 本文提出了一种新的深度序列表征方式, 即深度时空图 (Depth space time maps, DSTM).

DMM 侧重于表征人体行为的空间信息, 而 DSTM 侧重于表征人体行为的时序信息. 通过融合空间信息与时序信息进行人体行为识别, 可以提高人体行为识别的鲁棒性, 其中融合算法的可靠性直接影响了识别的精确度. 在一些实际应用中, 多模态数据虽然通过不同方式收集, 但表达的是相同语义. 通过分析多模态的数据, 提取与融合有效特征, 解决快速增长的数据量问题. 常见的融合方法有子空间学习, 例如 Li 等[21] 将典型性相关分析 (Canonical correlation analysis, CCA) 应用于基于非对应区域匹配的人脸识别, 使用 CCA 来学习一个公共空间, 测量两个非对应面部区域是否属于同一面部的可能性; Haghighat 等[22] 改进 CCA 提出的 判别相关分析 (Discriminant correlation analysis, DCA); Rosipal 等[23] 将偏最小二乘法 (Partial least squares, PLS) 用于执行多模态人脸识别; Liu 等[24] 的字典学习 (Dictionary learning method) 广泛应用于多视图的人脸识别; Zhuang 等[25] 使用基于图的学习方法 (Graph-based learning method) 进行多模态的融合; Sharma 等[26] 将线性判别分析 (Linear discriminant analysis, LDA) 和边际 Fisher 分析 (Marginal Fisher analysis, MFA) 扩展到它们的多视图对应物, 即广义多视图 LDA (Generalized multi-view LDA, GMLDA) 和广义多视图 MFA (Generalized multi-view MFA, GMMFA), 并将它们应用于跨媒体检索问题; Wang 等[27] 对子空间学习进行改进, 同样将它们应用于跨媒体的检索问题. 本文提出多聚点子空间学习算法以用于融合空间信息与时序信息进行人体行为识别.

1 相关工作

1.1 深度序列特征图

1.1.1 运动能量图和运动历史图

Bobick 等[8] 通过对彩色序列中相邻帧进行图片差分, 获得人体行为的区域, 在此基础上进行二值化后生成二值的图像序列 $D(x, y, t)$, 进一步获得二值特征图 MEI, 计算式为

$$E_\tau(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t-i) \tag{1}$$

其中, $E_\tau(x, y, t)$ 为视频序列中 t 帧处的能量, 由 τ 帧序列生成的 MEI.

同时, Bobick 等[8] 在 MEI 的基础上, 为了表示出行为的时序性, 提出了 MHI. 在 MHI 中像素亮度是该点处运动的时间历史函数. MHI 通过简单的替换和衰减运算获得, 计算式为

$$H_\sigma(x, y, t) = \begin{cases} \sigma, & \text{若 } D(x, y, t) = 1 \\ \max\left(0, H_\sigma(x, y, t-1) - 1\right), & \text{否则} \end{cases} \tag{2}$$

其中, $H_\sigma(x, y, t)$ 的初始像素亮度为 σ, $D(x, y, t)$ 为整个图像序列.

1.1.2 深度运动图

Yang 等[20] 提出将深度序列中的深度帧投影到笛卡尔直角坐标平面, 获取 3D 结构和形状信息. 在整个过程中提出了深度运动图 (DMM) 描述行为, 每个深度帧在投影后获得主视图、侧视图和俯视图三个 2 维投影图, 表示为 map_v. 假设一个有 N 帧的深度图序列, DMM_v 特征计算式为

$$DMM_v = \sum_{i=2}^{N} \left(\left|map_v^i - map_v^{i-1}\right| v \in \{f, s, t\}\right) \tag{3}$$

其中, i 表示帧索引, map_v^i 表示第 i 帧深度帧在 v 方向上的投影, f 表示主视图, s 表示侧视图, t 表示俯视图.

1.2　典型性相关分析

子空间学习的本质是庞大的数据集样本背后最质朴的特征选择与降维. 子空间学习的基础是 Harold Hotelling 提出的典型性相关分析 (CCA)[15], CCA 的主要思想是在两组随机变量中选取若干个有代表性的综合指标(变量的线性组合), 这些指标的相关关系来表示原来的两组变量的相关关系. 假设有两组数据样本 X 和 Y, 其中 X 为 $x_1 \times m$ 的样本矩阵, Y 为 $x_2 \times m$ 的样本矩阵, 对 X, Y 做标准化后 CCA 的计算式为

$$\arg\max(a,\, b) = \frac{cov\,(X',\, Y')}{\sqrt{D\,(X')\,D\,(Y')}} \tag{4}$$

其中, a, b 分别为 X, Y 的投影矩阵, $X' = a^{\mathrm{T}}X$, $Y' = a^{\mathrm{T}}Y$, cov 为协方差, $cov\,(X', Y')$ 协方差和方差的计算式为

$$\begin{aligned} cov\,(X',\, Y') = cov\left(a^{\mathrm{T}}X,\, b^{\mathrm{T}}Y\right) = \\ \mathrm{E}\left(\langle a^{\mathrm{T}}X,\, b^{\mathrm{T}}\rangle\right) = \\ a^{\mathrm{T}}\mathrm{E}\left(XX^{\mathrm{T}}\right)b \end{aligned} \tag{5}$$

$$D(X) = cov(X,\, X) = \mathrm{E}\left(XX^{\mathrm{T}}\right) \tag{6}$$

CCA 的优化目标计算式为

$$\arg\max(a,\, b) = \frac{a^{\mathrm{T}}cov(X,\, Y)b}{\sqrt{a^{\mathrm{T}}cov(X,\, X)a}\sqrt{b^{\mathrm{T}}cov(Y,\, Y)b}} \tag{7}$$

以 CCA 为基础的子空间学习将大规模的数据样本进行优化, 但它的计算复杂度很高, 无法消除阶级间的相关性并无法限制类内的相关性.

2　深度时空图

针对 DMM 时序信息的缺失的问题, 本文提出一种深度图序列表示算法 DSTM. DSTM 反映的是人体 3D 时空行为在空间直角坐标轴上的分布随着时间变化的情况, 人体所在空间直角坐标系三个轴分别为宽度轴 (w) 代表宽度方向、高度轴 (h) 代表高度方向、深度轴 (d) 代表深度方向, 图 1 为 DSTM 的流程图.

如图 1 所示, 首先将深度帧投影在三个笛卡尔正交面上, 获得主视图、侧视图和俯视图三个 2 维投影图, 表示为 map_v, $v \in \{f, s, t\}$. 然后根据每个 2 维投影图得到两个轴的行为分布情况. 任选两个 2 维投影图即可得到宽度轴、高度轴、深度轴的行为分布情况.

对 a 轴上的投影列表为

$$\begin{aligned} sum_a(i) = \sum_{x=1}^{W} map_v(x,\, i) \text{ 或} \\ sum_a(i) = \sum_{y=1}^{H} map_v(i,\, y) \end{aligned} \tag{8}$$

其中, $a \in \{w, h, d\}$, W, H 分别表示 2 维投影图的宽度和高度. sum_a 表示 2 维投影图序列在 a 轴上投影列表. 对 2 维投影图序列在 a 轴上的投影列表进行二值化, 即

$$list_a(i) = \begin{cases} 1, & sum_a(i) > \varepsilon \\ 0, & \text{其他} \end{cases} \tag{9}$$

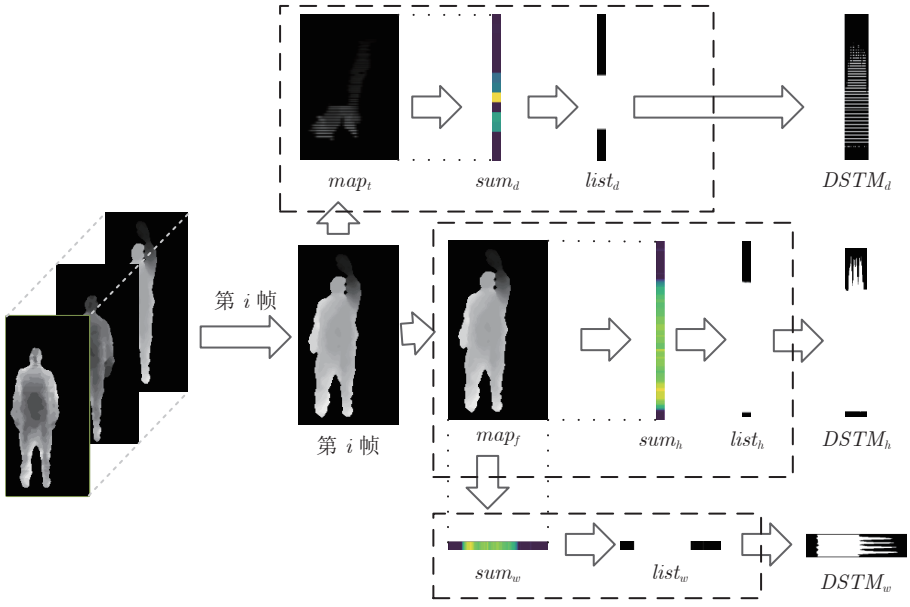


图 1　DSTM 流程图

Fig. 1　DSTM flowchart

其中, $list_a$ 表示对 2 维投影图序列在 a 轴上的投影列表进行二值化, $a \in \{w, h, d\}$, ε 表示二值化的阈值. 假设有 N 帧投影, $DSTM$ 的计算式为

$$DSTM_a(t) = list_a^t \tag{10}$$

其中, $list_a^t$ 表示第 t 帧 2 维投影图序列在 a 轴上投影列表进行二值化. $DSTM_a(t)$ 表示 $DSTM_a$ 的第 t 行.

最后对 $DSTM$ 进行感兴趣区域 (Region of interest, ROI) 处理, 根据感兴趣区域的主旨, 对图片进行裁剪、大小归一化处理.

3 多聚点子空间学习

子空间学习存在着计算复杂度高, 无法消除阶级间相关性的缺陷, 本文提出了多聚点子空间学习的方法, 在约束平衡模态间样本关系的同时, 通过构建同类别各样本的多个投影聚点, 疏远不同类别样本的类间距离, 降低了投影目标区域维度. 多聚点子空间学习算法的思想可表示为

$$\min_{U_1,\cdots,U_M} \sum_{p=1}^{M} \left\|X_p^{\mathrm{T}}U_p - Y\right\|_{\mathrm{F}}^2 + \lambda_1 \sum_{p=1}^{M} \|U_p\|_{21} + \lambda_2 \Omega(U_1, \cdots, U_M) + \lambda_3 \sum_{p=1}^{M}\sum_{c=1}^{L-1} \left\|X_p^{\mathrm{T}}U_p - G_c\right\|_{\mathrm{F}}^2 \tag{11}$$

其中, X_p 表示未经投影各模态样本, 即原空间样本; U_p, $p = 1, \cdots, M$ 表示各模态样本的投影矩阵; $X_p^{\mathrm{T}}U_p$ 表示经投影后各模态样本, 即子空间样本; L 表示类别总数; Y 为子空间内目标投影矩阵, 由各类别样本目标投影聚点 y_i 组成; G_c 为多个各模态同一类别样本新建目标投影点矩阵; λ_1, λ_2, λ_3 为各项超参.

3.1 聚点与子空间学习

本文将传统子空间学习称为单聚点子空间学习. 多聚点子空间学习与单聚点子空间学习的主要区别是聚点个数的不同, 具体定义如下:

1) 单聚点子空间学习. 通过学习每种模态数据的投影矩阵, 将不同类别数据投影到公共子空间. 投影矩阵的学习通常是最小化投影后样本与各类数据唯一主聚点的距离得到, 计算式为

$$\min_{U_1,\cdots,U_M} \sum_{p=1}^{M} \left\|X_p^{\mathrm{T}}U_p - Y\right\|_{\mathrm{F}}^2 + \lambda_1 \sum_{p=1}^{M} \|U_p\|_{21} \tag{12}$$

其中, Y 为子空间内目标投影矩阵, 由各类别样本目标投影聚点 y_i 组成, 可表示为 $Y = [y_1, y_2, \cdots, y_N]^{\mathrm{T}}$, 其中, $y_i = (v_1, v_2, \cdots, v_j, \cdots, v_L)$, $j = 1, \cdots, L$, $v_j = \begin{cases} 1, & x_i \in \text{第 } j \text{ 类}, \ x_i \text{ 为样本} \\ 0, & \text{其他} \end{cases}$

图 2 为单聚点子空间学习. 通过最小化子空间样本与各类别投影聚点之间距离来减少样本的类内距离.

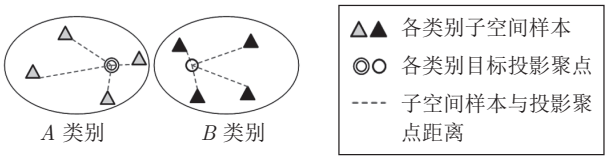


图 2 单聚点子空间学习

Fig. 2 Subspace learning

2) 多聚点子空间学习. 多聚点子空间学习是对单聚点子空间学习的优化, 都是通过学习每种模态数据的投影矩阵, 将不同类别数据投影到公共子空间. 不同的是, 投影矩阵的学习由同时最小化投影后样本与各类数据唯一主聚点以及与多个副聚点的总距离得到, 计算式为

$$\min_{U_1,\cdots,U_M} \sum_{p=1}^{M} \left\|X_p^{\mathrm{T}}U_p - Y\right\|_{\mathrm{F}}^2 + \lambda_1 \sum_{p=1}^{M} \|U_p\|_{21} + \lambda_3 \sum_{p=1}^{M}\sum_{c=1}^{L-1} \left\|X_p^{\mathrm{T}}U_p - G_c\right\|_{\mathrm{F}}^2 \tag{13}$$

其中, G_c 为各类别样本的第 c 个副投影聚点集合矩阵. 副投影聚点为其他类别投影聚点关于当前类别目标投影聚点的对称聚点. G_c 的构建算法步骤如下.

算法 1. Gc 的构建算法

输入. 子空间样本: $Y = \{y_i\}$, $i = 1, \cdots, L$; 类别数: H.

输出. 多聚点子空间内目标投影矩阵: G_c.

```
A ⇐ Y
for all c ⇐ {1, ···, L − 1} do
    for all j ⇐ {1, ···, L} do
        if c == 0 then
            B^0 ⇐ A^{j−1}
        else
            B^j ⇐ A^{j−1}
        end if
end for
        A ⇐ B
        G_c ⇐ 2Y_j − A
end for
```

注. B^j 为矩阵 B 中第 j 列.

图 3 为多聚点子空间学习. 通过为各类别样本

构建多个投影聚点并使用模态内、模态间数据相似度关系, 使得子空间样本向多个投影目标点附近的超平面聚拢, 有效增大了子空间样本之间的距离, 降低了投影目标区域的维度, 使投影目标区域从 n 维的超球体变为 $n-1$ 维的超平面, 同类别的子空间样本更为紧凑, 从而有效地提高了算法的特征优化效果. 因此结合使用数据模态内、模态间相似度关系的多聚点子空间学习可表示为

$$\min_{U_1,\cdots,U_M}\sum_{p=1}^{M}\left\|X_p^{\mathrm{T}}U_p-Y\right\|_{\mathrm{F}}^2+\lambda_1\sum_{p=1}^{M}\|U_p\|_{21}+\lambda_2\Omega\left(U_1,\cdots,U_M\right)+\lambda_3\sum_{p=1}^{M}\sum_{c=1}^{\mathrm{L}-1}\left\|X_p^{\mathrm{T}}U_p-G_c\right\|_{\mathrm{F}}^2 \tag{14}$$

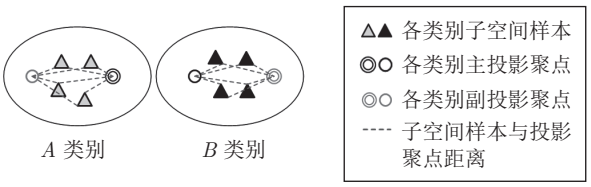


图 3　多聚点子空间学习

Fig. 3　Multi-center subspace learning

3.2　超参计算

本文以式 (14) 第 1 项为基准确定式中各项超参, 设定子空间样本与目标投影聚点之间约束程度等同于同类别子空间样本之间约束程度. 第 1 项中子空间样本与目标投影点之间约束共有 F_1 个, F_1 计算式为

$$F_1=M\times N \tag{15}$$

其中, M 为模态数, N 为样本数.

式 (14) 第 3 项中子空间样本之间约束共有 F_2 个, 其中同一模态子空间样本相似度的约束共有 F_a 个, 不同模态同一类别的子空间样本之间的相似度的约束共有 F_b 个, F_2, F_a, F_b 计算式为

$$F_a=\frac{M\times N\times N}{2} \tag{16}$$

$$F_b=\sum_{i=1}^{L}\frac{N_i\times M\times(N_i\times M+1)}{2} \tag{17}$$

$$F_2=F_a+F_b \tag{18}$$

其中, L 为样本类别数; N_i 为各类样本数, 并且 $N=\sum_{i=1}^{L}N_i$.

式 (14) 第 4 项中子空间样本与目标投影聚点之间约束共有 F_3 个, F_3 计算式为

$$F_3=F_1\times(L-1)=M\times N\times(L-1) \tag{19}$$

在子空间样本与目标投影聚点之间约束程度等同于同类别子空间样本之间约束. 根据 F_1, F_2, F_3 比例关系, 可以确定式 (14) 的第 3 项和第 4 项超参的计算式为

$$\lambda_2=\frac{F_1}{F_2}=\frac{2\times M\times N}{M\times N\times N+\sum\limits_{i=1}^{L}\left[N_i\times M\times(N_i\times M+1)\right]}=\frac{2}{N}+\frac{2\sum\limits_{i=1}^{L}N_i}{M\sum\limits_{i=1}^{L}N_i^2+\sum\limits_{i=1}^{L}N_i} \tag{20}$$

$$\lambda_3=\frac{F_1}{F_3}=\frac{M\times N}{M\times N\times(L-1)}=\frac{1}{L-1} \tag{21}$$

最后本文通过实验, 以最终识别率为依据, 确定 λ_1.

3.3　公式优化与投影矩阵求取

对于式 (16) 中的几项可进行优化, 式 (16) 中的第 2 项是对各模态的数据样本投影矩阵的约束项, 防止算法过拟合. 第 2 项中含有 $l_{2,1}$ 范数, 它是非平滑且不能得到的一个闭式解[28]. 对于投影矩阵, 其 $l_{2,1}$ 范数定义为

$$\sum_{p=1}^{M}\|U_p\|_{21}=\sum_{p=1}^{M}\left(\sum_{i=1}^{m}\sqrt{\sum_{j=1}^{n}u_{ij}^2}\right)=\sum_{p=1}^{M}\mathrm{tr}\left(U_p^{\mathrm{T}}R_pU_p\right) \tag{22}$$

其中, $R_p=[r_{ij}]$ 是一个对角阵, $r_{ij}=\dfrac{1}{2||u_p||_2}$, u_p 表示投影矩阵 U 的第 i 个行向量, 为了避免 $||u_p||_2$ 的值为 0, 根据文献 [29] 对于 $l_{2,1}$ 的分析, 引入一个不为 0 的无穷小数 ε, r_{ij} 重新定义为

$$r_{ij}=\frac{1}{2\sqrt{||u_p||_2^2+\varepsilon}} \tag{23}$$

式 (14) 中第 3 项是不同模态同一类别的子空间样本之间的约束. 第 3 项可以通过如下方式进行推导

$$\begin{aligned}\Omega\left(U_1,\cdots,U_M\right)&=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}\left\|f_i-f_j\right\|^2=\\&\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}f_i^2-\sum_{i=1}^{N}\sum_{j=1}^{N}W_{ij}f_if_j=\\&FDF^{\mathrm{T}}-FWF^{\mathrm{T}}=\\&\mathrm{tr}\left(FLF^{\mathrm{T}}\right)=\\&\sum_{p=1}^{M}\sum_{q=1}^{M}\mathrm{tr}\left(U_p^{\mathrm{T}}X_p^bL_{pq}\left(X_q^b\right)^{\mathrm{T}}U_q\right)\end{aligned} \tag{24}$$

其中, N' 是所有模态的样本总数, p, q 为两个不同

的模态, L 是拉普拉斯矩阵并且 $F=(F_1^{\mathrm{T}},\cdots,F_M^{\mathrm{T}})=(U_1^{\mathrm{T}}X_1^b,\cdots,U_M^{\mathrm{T}}X_M^b)$, W 为模态相似度矩阵, 其定义为

$$W_{ij}^{pq}=\begin{cases}1, & x_i^p \text{ 与 } x_j^q \text{ 是同一类别}\\ 0, & \text{其他}\end{cases} \tag{25}$$

式 (14) 通过优化后可以重新表达为

$$\begin{aligned}\min_{U_1,\cdots,U_m}&\sum_{p=1}^{M}\left\|X_p^{\mathrm{T}}U_p-Y\right\|_{\mathrm{F}}^2+\lambda_1\sum_{p=1}^{M}\mathrm{tr}\left(U_p^{\mathrm{T}}\boldsymbol{R}_pU_p\right)+\\ &\lambda_2\sum_{p=1}^{M}\sum_{q=1}^{M}\mathrm{tr}\left(U_p^{\mathrm{T}}X_p^bL_{\mathrm{pq}}\left(X_q^b\right)^{\mathrm{T}}U_q\right)+\\ &\lambda_3\sum_{p=1}^{M}\sum_{c=1}^{L-1}\left\|X_p^{\mathrm{T}}U_p-G_c\right\|_{\mathrm{F}}^2\end{aligned} \tag{26}$$

本节通过下述算法步骤求解线性系统问题来计算式 (26)的最优解.

算法 2. 计算子空间学习的最优解

输入. 原空间样本: $X_p,\ p=1,\cdots,M$;

子空间样本: $Y=\{y_i\},\ i=1,\cdots,L$.

输出. 子空间内目标投影矩阵: $U_p,\ p=1,\cdots,M$.

计算 L 的拉普拉斯矩阵

设置 $t=0$, 初始化 U_p

repeat

1) 通过求解方程 (26) 中的线性系统问题, U_p^t 更新如下:

$$\begin{aligned}U_p^{t+1}=&\left(X_pX_p^{\mathrm{T}}+\lambda_{\mathrm{s}}X_pX_p^{\mathrm{T}}+\lambda_1R_p+\right.\\ &\lambda_2X_pL_{pp}\left(X_p\right)^{\mathrm{T}}\Big)^{-1}\Big(X_pY+\lambda_{\mathrm{s}}\sum_{c=1}^{L}X_pG_c-\\ &\left.\lambda_2\sum_{p\neq q}X_pL_{pq}\left(\mathrm{X}_q\right)^{\mathrm{T}}U_q^t\right)\end{aligned} \tag{27}$$

2) $t=t+1$

until convergence

通过算法 2 进行求解, 先计算出拉普拉斯矩阵, 然后求解出 U_p^1 并代入式 (27) 进行重复求解, 直至收敛.

4 实验结果与分析

4.1 数据库

文献 [30] 对数据集进行了详细的研究, 本文采用的是由 Kinect 摄像头采集的 MSR-Action3D[31] 数据库和 UTD-MHAD[32] 数据库.

MSR-Action3D (MSR) 数据库由 10 个人 20 个动作重复 2 ~ 3 次, 共计 557 个深度图序列, 涉及人的全身动作. 详情如表 1 所示.

UTD-MHAD (UTD) 数据库由 8 个人 (4 男 4 女) 27 个动作重复 4 次, 共计 861 个深度图序列. 详情如表 2 所示.

表 1 MSR 数据库中的人体行为

Table 1 Human actions in MSR

动作	样本数	动作	样本数
高挥手 (A01)	27	双手挥 (A11)	30
水平挥手 (A02)	26	侧边拳击 (A12)	30
锤 (A03)	27	弯曲 (A13)	27
手抓 (A04)	25	向前踢 (A14)	29
打拳 (A05	26	侧踢 (A15)	20
高抛 (A06)	26	慢跑 (A16)	30
画叉 (A07)	27	网球挥拍 (A17)	30
画勾 (A08)	30	发网球 (A18)	30
画圆 (A09)	30	高尔夫挥杆 (A19)	30
拍手 (A10)	30	捡起扔 (A20)	27

表 2 UTD 数据库中的人体行为

Table 2 Human actions in UTD

动作	样本数	动作	样本数
向左滑动 (B01)	32	挥网球 (B15)	32
向右滑动 (B02)	32	手臂卷曲 (B16)	32
挥手 (B03)	32	网球发球 (B17)	32
鼓掌 (B04)	32	推 (B18)	32
扔 (B05)	32	敲 (B19)	32
双手交叉 (B06)	32	抓 (B20)	32
拍篮球 (B07)	32	捡起扔 (B21)	32
画叉 (B08)	31	慢跑 (B22)	31
画圆 (B09)	32	走 (B23)	32
持续画圆 (B10)	32	坐下 (B24)	32
画三角 (B11)	32	站起来 (B25)	32
打保龄球 (B12)	32	弓步 (B26)	32
冲拳 (B13)	32	蹲 (B27)	32
挥羽毛球 (B14)	32		

为了验证时序信息在人体行为中的重要性, 本文将与原深度图序列顺序相反的行为称为反序行为. 本文中的反序行为是通过将正序行为的深度图序列进行反序排列操作得到新数据库 D1, D2, 其中 D1 为 MSR 数据库及 MSR 反序数据库, D2 为 UTD 数据库及 UTD 反序数据库. D1 正反高抛动作如图 4 所示.

4.2 实验设置

本文采用 10×10 像素的图像单元分割图像, 每 2×2 个图像单元构成一个图像块, 以 10 像素为步长滑动图像块来提取图像的方向梯度直方图 (Histogram of oriented gradient, HOG)[26] 特征. 采

(a) MSR 数据集中正序高抛动作
(a) Positive sequence high throw in MSR Database

(b) MSR 数据集中反序高抛动作
(b) Negative sequence high throw in MSR Database

图 4　正反高抛动作
Fig. 4　Positive and negative high throwing action

用采样半径为 2, 采样点数为 8 的参数设置来提取图像局部二值模式 (Local binary patterns, LBP)[33] 特征. 尺寸归一化后 DMM_f 大小为 320×240, DMM_s 大小为 500×240, DMM_t 大小为 320×500, 所以 DMM-HOG 的特征数量为 120 924. DMM-LBP 的特征数量为 276 800. 同样尺寸归一化后 $DSTM_w$ 大小为 320×60, $DSTM_h$ 大小为 240×60, $DSTM_d$ 大小为 500×60, 所以 DMM-HOG 的特征数量为 18 540. DMM-LBP 的特征数量为 63 600.

实验中分为两个设置. 设置 1 在 MSR 数据库上将 20 个行为分为 3 组 (AS1、AS2、AS3)[33], 行为分布情况如表 1, 其中 AS1 和 AS2 组内相似度较高, AS3 组内相似度较低. 如表 3 所示.

表 3　MSR-Action3D 数据分组
Table 3　MSR-Action3D data grouping

AS1	AS2	AS3
A02	A01	A06
A03	A04	A14
A05	A07	A15
A06	A08	A16
A10	A09	A17
A13	A11	A18
A18	A14	A19
A20	A12	A20

设置 2 在 MSR 数据库和 UTD 数据库上选取全部的动作.

在设置 1 和设置 2 中可采用 4 种测试方法. 测试 1: 1/3 作为训练数据, 2/3 作为测试数据; 测试 2[12]: 1/2 作为训练数据, 1/2 作为测试数据; 测试 3: 2/3 作为训练数据, 1/3 作为测试数据; 测试 4: 采用 5 折交叉验证

4.3　参数设置

在本文提出的人体识别的模型中, 首先要确定参数 λ_1, λ_2, λ_3 的值. 在进行子空间学习的时候, 参数对于结果有着巨大的影响, 需要优先估计最优的参数. 通过选择不同的参数, 并以识别率作为评判标准. 识别率 = 预测正确测试样本数/总测试样本数. 通过采用设置 1 测试 1 的方法和 HOG 特征进行实验. 根据式 (20) 和式 (21) 分别可以得到 $\lambda_2 = 1/13\,847$, $\lambda_3 = 1/19$. 根据图 5 可知, 当 $\lambda_1 = 20$ 时, 本文算法具有较高的人体识别性能.

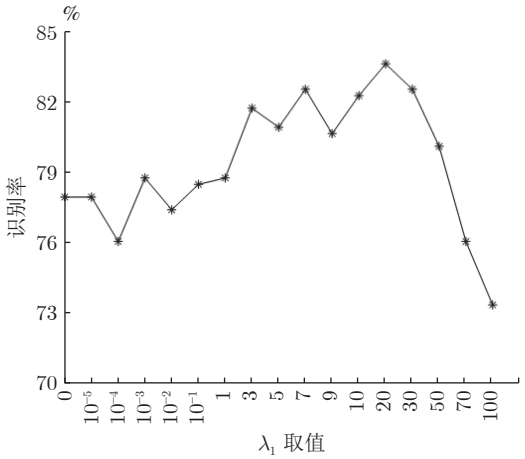


图 5　参数选择
Fig. 5　The parameter of selection

4.4　实验结果分析

4.4.1　分类器选择

对同一种特征图而言, 采用不同的分类器识别效果会有较大的差异. 为了选择对特征图识别效果较好的分类器, 本实验通过比较 DSTM 在不同的分类器的识别效果, 最终以识别率作为标准, 采用设置 1 测试 3 的方法, 如图 6 所示.

从图 6 中可以发现 HOG 特征采用了不同的分类器, 得到的识别率差异较大, 不同特征图采用同一分类器, 与同一特征图采用不同分类器, 支持向量机 (Support vector machine, SVM) 的识别效果较好, 下面实验均采用 SVM 作为分类器.

4.4.2　特征选择

为了筛出空间信息和时序信息的特征图, 采用设置 1, 在 MSR 数据库上使用测试 1、测试 2、测试 4 的方法进行实验, 并且对 3 组实验结果设置了平均值; 采用设置 2, 在 UTD 数据库上使用测试 1、测试 2、测试 3 的方法进行实验. 通过个体识别率和平均识别率来筛出空间信息和时序信息的特征图.

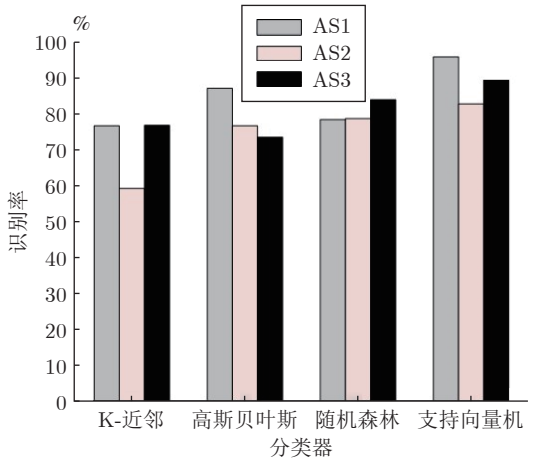


图 6 DSTM 在不同分类器识别效果

Fig. 6 DSTM recognition of different classifiers

表 4 和表 5 使用 HOG 和 LBP 两个特征图序列. 由表 4 中的单个识别率或平均识别率以及表 5 中所有动作的识别率可以得出结论: 在同一特征图中, HOG 特征较 LBP 特征有着更高的识别率. LBP 特征反映的是像素周围区域的纹理信息; HOG 特征能捕获轮廓、弱化光照, 对于深度图有着更佳的表现, 有着更好的识别效果. 就本文实验而言. HOG特征更适合于本实验.

在表 4 和表 5 中选择同为 HOG 特征的特征图, 从表中的识别率可以得出, DMM 和 DSTM 与 MEI 和 MHI 相比有更高的识别率. 主要原因是 MEI 将深度帧二值化后进行叠加, 掩盖了时序图中每张图的轮廓信息, 丢失了时序图自身的深度信息, 但反映出一定的轮廓信息, 保留了一定的空间信息; MHI 虽然通过图像的亮度衰减, 增加了一部分时序信息, 但由于人为干预图像的亮度, 导致了图像自身的深度信息的丢失.

使用 DSTM 和 DMM 的优势主要有以下几点: 1) DMM 是将深度帧投影到笛卡尔直角坐标平面上, 生成主视图、俯视图、侧视图三个 2 维地图, 在此基础上差分堆叠整个深度序列动作能量图. 相较于 MEI, DMM 充分地使用了时序图的深度信息, 丰富了特征中的空间信息, 很大程度上保留了轮廓信息, 并且从三个方向上可以很明显地看出行为动作, 充分展现了空间信息. 2) DSTM 是将深度帧投影到笛卡尔直角坐标平面上, 生成主视图、俯视图、侧视图三个 2 维地图, 提取任意两个 2 维地图投影到 3 个正交轴上获取三轴坐标投影, 将获得的坐标投影二值化后按时间顺序进行拼接. DSTM 将深度帧的时序信息很好地保留了下来, 相较于 MHI 有了很大程度上的改善. DSTM 较好地保存了时序信息.

时序信息在行为识别中有着重要的作用. 对比 DMM, DSTM 蕴含着重要的时序信息. 本文在 D1 和 D2 数据库上采用设置 2, 使用测试 1 的方法

通过对比表 6 的识别率和表 7 的时间复杂度, 在 D1 与 D2 数据库的实验证明, DMM 由于未含有时序信息, 与 DSTM 识别率差异较大. 另外 DMM 相较于 DSTM 时间复杂度较高, DSTM 的时序信息在行为识别中起着重要的作用.

4.4.3 特征选择实验结果

本文选取的深度运动图代表的空间信息与深度时空图代表的特征图使用多聚点子空间学习的算法 (简称本文方法). 为了表征本文方法对于单一特征有着更高的识别率以及本文方法对于融合方法同样有着更高的识别率, 将本文方法与当前主流单一算法和融合算法进行比较. 在 MSD-Action3D 上采用设置 2 测试 2、设置 2 测试 4 的方法; 在 UTD-MHAD 上采用设置 2 测试 4 的方法.

表 8 均采用文献 [12] 方法中的实验设置, 其中文献 [34−40] 方法使用了深度学习的模型框架. 识

表 4 MSR 数据库上不同特征的识别率 (%)

Table 4 Different of feature action recognition on MSR (%)

方法	测试 1				测试 2				测试 3			
	AS1	AS2	AS3	均值	AS1	AS2	AS3	均值	AS1	AS2	AS3	均值
MEI-HOG	69.79	77.63	79.72	75.71	84.00	89.58	93.24	88.94	86.95	86.95	95.45	89.78
MEI-LBP	57.05	56.58	64.19	59.27	66.66	69.79	78.37	71.61	69.56	73.91	77.27	73.58
DSTM-HOG	83.22	71.71	87.83	80.92	94.66	84.37	88.23	89.80	91.30	82.61	95.95	89.95
DSTM-LBP	84.56	71.71	87.83	81.37	88.00	82.29	95.94	88.74	86.96	82.61	95.45	88.34
MHI-HOG	69.79	72.36	70.95	71.03	88.00	84.37	89.19	87.19	95.65	82.60	95.45	91.23
MHI-LBP	51.67	60.52	54.05	55.41	73.33	70.83	78.37	74.18	82.60	65.21	72.72	73.51
DMM-HOG	88.00	87.78	87.16	87.65	94.66	87.78	100.00	94.15	100.00	88.23	95.45	94.56
DMM-LBP	89.52	87.78	93.20	90.17	93.11	85.19	100.00	92.77	94.03	88.98	92.38	91.80

表 5　UTD 数据库上不同特征的识别率 (%)
Table 5　Different of feature action recognition on UTD (%)

方法	测试 1	测试 2	测试 3
MEI-HOG	69.51	65.42	68.20
MEI-LBP	45.12	51.97	52.61
DSTM-HOG	71.08	80.28	89.54
DSTM-LBP	68.81	80.97	86.06
MHI-HOG	56.44	66.58	73.14
MHI-LBP	49.82	53.82	57.40
DMM-HOG	78.39	75.40	87.94
DMM-LBP	68.98	74.94	86.75

表 6　DMM 和 DSTM 对比实验结果 (%)
Table 6　Experimental results of DMM and DSTM (%)

方法	D1	D2
DSTM	62.83	81.53
DMM	32.17	63.93

表 7　DMM 和 DSTM 平均处理时间 (s)
Table 7　Average processing time of DMM and DSTM (s)

方法	D1	D2
DSTM	2.1059	3.4376
DMM	5.6014	8.6583

表 8　MSR-Action3D[1] 上的实验结果
Table 8　Experimental results on MSR-Action3D[1]

方法	识别率 (%)
文献 [12]	86.50
文献 [34]	91.45
文献 [35]	90.01
文献 [36]	89.40
文献 [37]	77.47
文献 [38]	81.7
文献 [39]	90.01
文献 [40]	89.48
本文学习方法	90.32

注: MSR-Action3D[1] 采用设置 2 测试 2.

别率最高为 91.45. 本文的识别率达到了 90.32%, 接近文献 [34] 中的最优结果, 主要原因是: 本文提出的 DSTM 算法可以将深度帧的时序信息很好地保留下来, 获得的特征信息更加丰富和完善. 多聚点子空间的方法构建了多个投影聚点并使用了模态内、模态间数据相似度关系, 使得子空间样本向多个投影目标点附近的超平面聚拢, 有效增大了子空间样本之间的距离, 所以在行为识别中表现出了较为优越的性能. 表 9 和表 10 在多聚点子空间学习加单个特征图的识别率有一定的提升, 但相较于融合 DSTM 特征和 DMM 特征图略有不足. 本文在采用不同的融合方法时, 识别率也有一定提升. 本文方法的识别率在 MSR 数据库达到 98.21% 和 UTD 数据库达到 98.84%. 为了更深层次的了解本文方法的识别效果, 本文给出了本文方法的每个动作识别效果的混淆矩阵.

表 9　MSR-Action3D[2] 上的实验结果
Table 9　Experimental results on MSR-Action3D[2]

方法	识别率 (%)
MHI-LBP	68.75
MEI-LBP	71.43
DCA[22]	94.64
DSTM-LBP	87.50
DSTM-HOG	89.28
MCSL+DMM	89.28
MCSL+DSTM	91.96
CCA[21]	83.05
子空间学习	92.85
本文学习方法	98.21

注: MSR-Action3D[2] 采用设置 2 测试 4; MCSL 为多聚点子空间学习.

表 10　UTD-MHAD 在设置 2 测试 4 上的实验结果
Table 10　Experimental results on UTD-MHAD

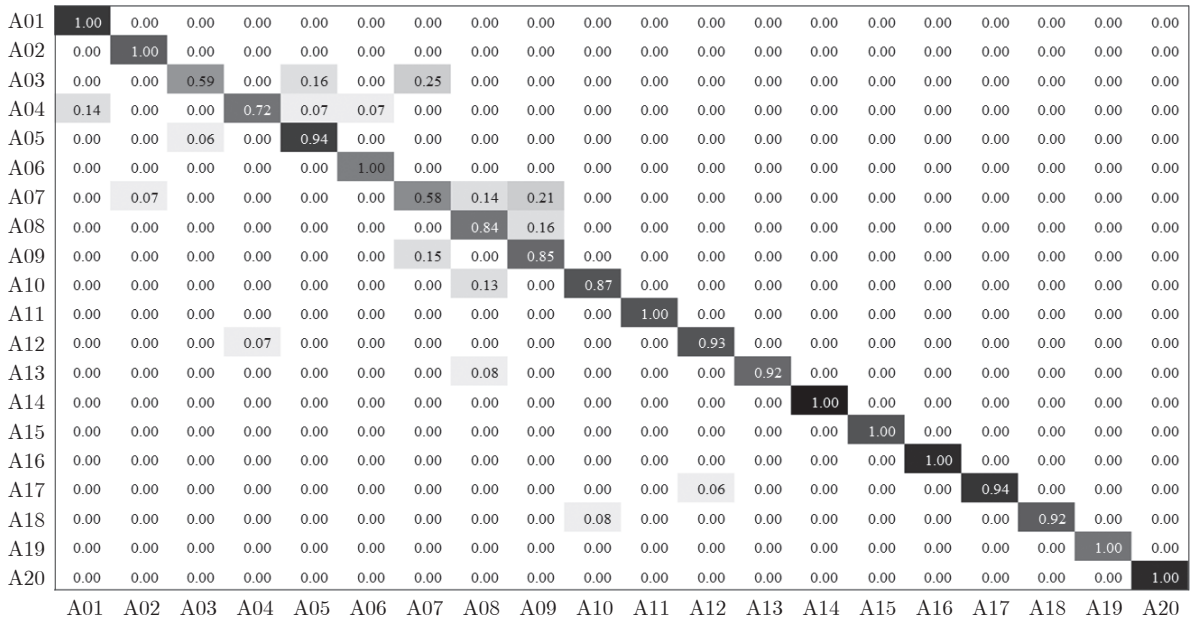
方法	识别率 (%)
MHI-LBP	62.40
MEI-LBP	57.80
DCA[22]	92.48
DSTM-LBP	89.59
DSTM-HOG	91.90
MCSL+DMM	93.64
MCSL+DSTM	95.37
CCA[21]	87.28
子空间学习	93.64
本文学习方法	98.84

本文通过融合 DMM 的空间信息和 DSTM 的时序信息的两种特征图后, 得到空间时序特征. 多聚点子空间学习是通过为各类别样本构建了多个投影聚点. 图 7(a) 和图 7(b) 为 MSR 的混淆矩阵. 其中, MSR-Action3D[1] 采用设置 2 测试 2; MSR-Action3D[2] 采用设置 2 测试 4. 从中可以看出整体识别率, 图中显示本文方法将画叉识别成画圈, 将发网球识别成了画勾. 两类动作差异性小, 因此较容易

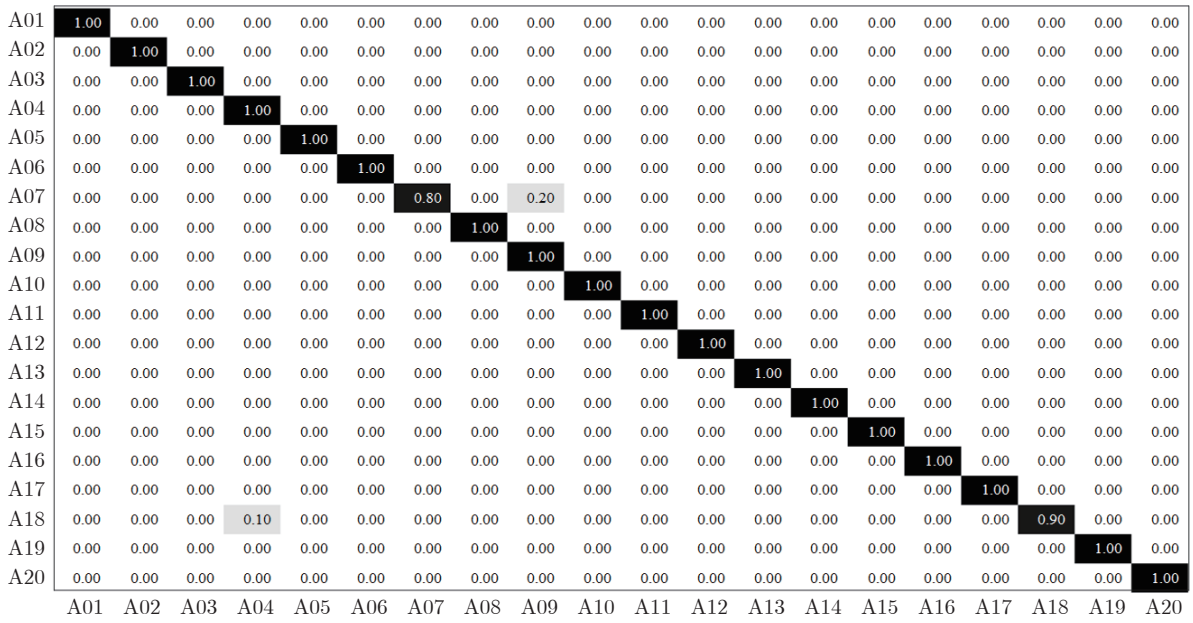
出错. 图 7 (c) 为 UTD 的混淆矩阵, 图中显示本文方法将慢跑变成走路. 出现错误原因是动作行为轨迹相似性较大.

5 结束语

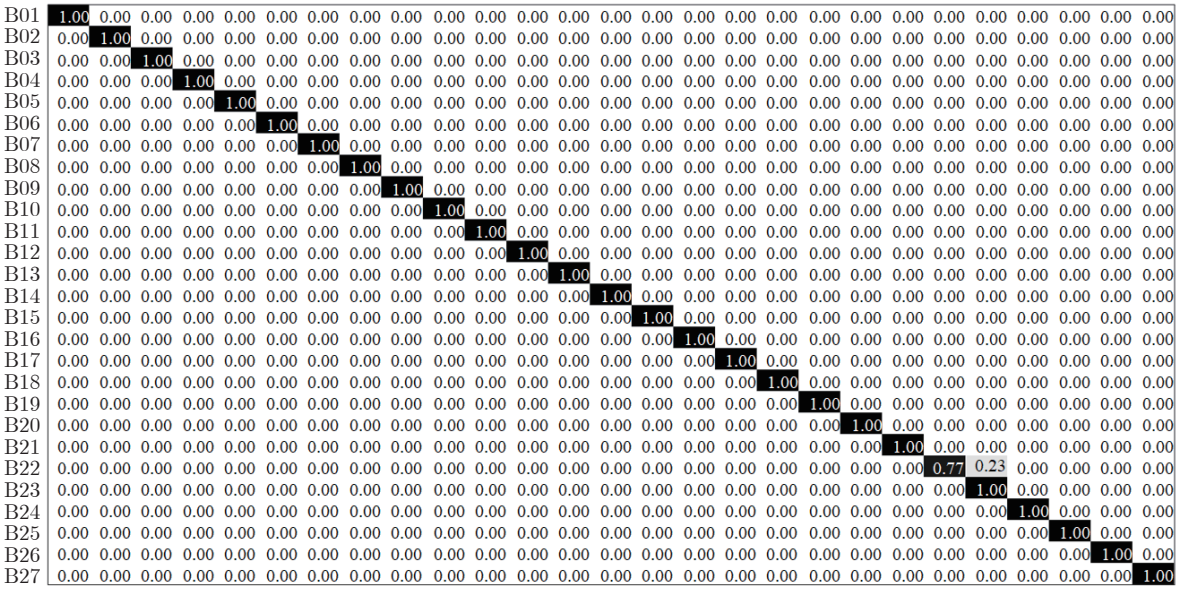
针对现有的深度图序列特征图冗余过多、时序和空间信息缺失等问题, 本文提出一种新的深度序列表示方式 DSTM 和多聚点子空间学习, 并在此基础上进行了人体行为识别研究. 深度帧投影二值化后按时间顺序进行拼接生成 DSTM, 对每张 DSTM 提取 HOG 特征以获得时序信息. 对 DMM 提取 HOG 特征以获得空间信息. 多聚点子空间学习, 在约束平衡模态间样本关系的同时, 构建同类别各样本的多个副投影聚点, 疏远不同类别样本的类间距离, 降低了投影目标区域维度, 最后送入分类器进行人体行为识别. 本实验表明本文提出的 DSTM 和多聚点子空间学习的方法能够减少深度序列的冗余, 保留丰富的空间信息和良好的时序信



(a) MSR-Action3D[1] 混淆矩阵
(a) Confusion matrix of MSR-Action3D[1]



(b) MSR-Action3D[2] 混淆矩阵
(b) Confusion matrix of MSR-Action3D[2]



(c) UTD 混淆矩阵

(c) Confusion matrix of UTD

图 7　混淆矩阵

Fig. 7　Confusion matrix

息, 有效地提高行为识别的准确性.

References

1 Yousefi S, Narui H, Dayal S, Ermon S, Valaee S. A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine*, 2017, **55**(10): 98−104

2 Ben Mabrouk A, Zagrouba E. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 2018, **91**: 480−491

3 Fang C C, Mou T C, Sun S W, Chang P C. Machine-learning based fitness behavior recognition from camera and sensor modalities. In: Proceedings of the 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). Taichung, China: IEEE, 2018. 249−250

4 Chen C, Liu K, Jafari R, Kehtarnavaz N. Home-based senior fitness test measurement system using collaborative inertial and depth sensors. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Chicago, USA: IEEE, 2014. 4135−4138

5 Laver K E, Lange B, George S, Deutsch J E, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. *Cochrane Database of Systematic Reviews*, 2017, **11**(11): Article No. CD008349

6 Sun J, Wu X, Yan S C, Cheong L F, Chua T S, Li J T. Hierarchical spatio-temporal context modeling for action recognition. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 2004−2011

7 Hu Jian-Fang, Wang Xiong-Hui, Zheng Wei-Shi, Lai Jian-Huang. RGB-D action recognition: Recent advances and future perspectives. *Acta Automatica Sinica*, 2019, **45**(5): 829−840
(胡建芳, 王熊辉, 郑伟诗, 赖剑煌. RGB-D行为识别研究进展及展望. 自动化学报, 2019, **45**(5): 829−840)

8 Bobick A F, Davis J W. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(3): 257−267

9 Su Ben-Yue, Jiang Jing, Tang Qing-Feng, Sheng Min. Human dynamic action recognition based on functional data analysis. *Acta Automatica Sinica*, 2017, **43**(5): 866−876
(苏本跃, 蒋京, 汤庆丰, 盛敏. 基于函数型数据分析方法的人体动态行为识别. 自动化学报, 2017, **43**(5): 866−876)

10 Anderson D, Luke R H, Keller J M, Skubic M, Rantz M J, Aud M A. Modeling human activity from voxel person using fuzzy logic. *IEEE Transactions on Fuzzy Systems*, 2009, **17**(1): 39−49

11 Wu Y X, Jia Z, Ming Y, Sun J J, Cao L J. Human behavior recognition based on 3D features and hidden Markov models. *Signal, Image and Video Processing*, 2016, **10**(3): 495−502

12 Wang J, Liu Z C, Chorowski J, Chen Z Y, Wu Y. Robust 3D action recognition with random occupancy patterns. In: Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012. 872−885

13 Zhang H L, Zhong P, He J L, Xia C X. Combining depth-skeleton feature with sparse coding for action recognition. *Neurocomputing*, 2017, **230**: 417−426

14 Zhang S C, Chen E Q, Qi C, Liang C W. Action recognition based on sub-action motion history image and static history image. *MATEC Web of Conferences*, 2016, **56**: Article No. 02006

15 Liu Z, Zhang C Y, Tian Y L. 3D-based deep convolutional neural network for action recognition with depth sequences. *Image and Vision Computing*, 2016, **55**: 93−100

16 Xu Y, Hou Z J, Liang J Z, Chen C, Jia L, Song Y. Action recognition using weighted fusion of depth images and skeleton′s key frames. *Multimedia Tools and Applications*, 2019, **78**(17): 25063−25078

17 Wang P C, Li W Q, Li C K, Hou Y H. Action recognition based on joint trajectory maps with convolutional neural networks. *Knowledge-Based Systems*, 2018, **158**: 43−53

18 Kamel A, Sheng B, Yang P, Li P, Shen R M, Feng D D. Deep

convolutional neural networks for human action recognition using depth maps and postures. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019, **49**(9): 1806−1819

19 Li C K, Hou Y H, Wang P C, Li W Q. Joint distance maps based action recognition with convolutional neural networks. *IEEE Signal Processing Letters*, 2017, **24**(5): 624−628

20 Yang X D, Zhang C Y, Tian Y L. Recognizing actions using depth motion maps-based histograms of oriented gradients. In: Proceedings of the 20th ACM International Conference on Multimedia. Nara, Japan: ACM, 2012. 1057−1060

21 Li A N, Shan S G, Chen X L, Gao W. Face recognition based on non-corresponding region matching. In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1060−1067

22 Haghighat M, Abdel-Mottaleb M, Alhalabi W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on Information Forensics and Security*, 2016, **11**(9): 1984−1996

23 Rosipal R, Krämer N. Overview and recent advances in partial least squares. In: Proceedings of the 2006 International Statistical and Optimization Perspectives Workshop "Subspace, Latent Structure and Feature Selection". Bohinj, Slovenia: Springer, 2006. 34−51

24 Liu H P, Sun F C. Material identification using tactile perception: A semantics-regularized dictionary learning method. *IEEE/ASME Transactions on Mechatronics*, 2018, **23**(3): 1050−1058

25 Zhuang Y T, Yang Y, Wu F. Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval. *IEEE Transactions on Multimedia*, 2008, **10**(2): 221−229

26 Sharma A, Kumar A, Daume H, Jacobs D W. Generalized multiview analysis: A discriminative latent space. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2160−2167

27 Wang K Y, He R, Wang L, Wang W, Tan T N. Joint feature selection and subspace learning for cross-modal retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(10): 2010−2023

28 Nie F, Huang H, Cai X, Ding C. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver British, Canada: Curran Associates Inc., 2010. 1813−1821

29 He R, Tan T N, Wang L, Zheng W S. $l_{2,1}$ regularized correntropy for robust feature selection. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2504−2511

30 Zhu Hong-Lei, Zhu Chang-Sheng, Xu Zhi-Gang. Research advances on human activity recognition datasets. *Acta Automatica Sinica*, 2018, **44**(6): 978−1004
(朱红蕾, 朱昶胜, 徐志刚. 人体行为识别数据集研究进展. 自动化学报, 2018, **44**(6): 978−1004)

31 Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, 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). Colorado Springs, USA: IEEE, 2011. 1297−1304

32 Chen C, Jafari R, Kehtarnavaz N. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP). Quebec City, Canada: IEEE, 2015. 168−172

33 Chen C, Jafari R, Kehtarnavaz N. Action recognition from depth sequences using depth motion maps-based local binary patterns. In: Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE, 2015. 1092−1099

34 Koniusz P, Cherian A, Porikli F. Tensor representations via kernel linearization for action recognition from 3D skeletons. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 37−53

35 Ben Tanfous A, Drira H, Ben Amor B. Coding Kendall′s shape trajectories for 3D action recognition. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2840−2849

36 Vemulapalli R, Chellappa R. Rolling rotations for recognizing human actions from 3D skeletal data. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 4471−4479

37 Wang L, Huynh D Q, Koniusz P. A comparative review of recent kinect-based action recognition algorithms. *IEEE Transactions on Image Processing*, 2019, **29**: 15−28

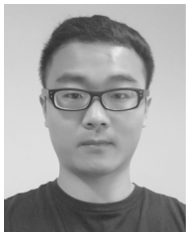
38 Rahmani H, Mian A. 3D action recognition from novel viewpoints. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 1506−1515

39 Ben Tanfous A, Drira H, Ben Amor B. Sparse coding of shape trajectories for facial expression and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(10): 2594−2607

40 Ben Amor B, Su J Y, Srivastava A. Action recognition using rate-invariant analysis of skeletal shape trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(1): 1−13

杨天金 常州大学信息科学与工程学院硕士研究生. 主要研究方向为行为识别, 机器学习.
E-mail: yangtianjin128@163.com
(**YANG Tian-Jin** Master student at the School of Information Science and Engineering, Changzhou University. His research interest covers behavior recognition and machine learning.)

侯振杰 常州大学信息科学与工程学院教授. 2015 年获内蒙古农业大学机械专业博士学位. 主要研究方向为行业识别, 机器学习. 本文通信作者.
E-mail: houzj@cczu.edu.cn
(**HOU Zhen-Jie** Professor at the School of Information Science and Engineering, Changzhou University. He received his Ph.D. degree in mechanical engineering from Inner Mongolia Agricultural University in 2015. His research interest covers behavior recognition and machine learning. Corresponding author of this paper.)



李 兴 常州大学信息科学与工程学院硕士研究生. 主要研究方向为行为识别, 机器学习.
E-mail: lixing03201012@163.com
(**LI Xing** Master student at the School of Information Science and Engineering, Changzhou University. His research interest covers behavior recognition and machine learning.)

梁久祯 常州大学信息科学与工程学院教授. 2001 年获北京航空航天大学计算机软件与理论工学博士学位. 主要研究方向为行为识别, 机器学习.
E-mail: jzliang@cczu.edu.cn
(**LIANG Jiu-Zhen** Professor at the School Information Science and Engineering, Changzhou University. He received his Ph.D. degree in computer software and theory engineering from Beijing University of Aeronautics and Astronautics in 2001. His research interest covers behavior recognition and machine learning.)

宦 娟 常州大学信息科学与工程学院副教授. 2019 年获江苏大学农业电气化与自动化专业博士学位. 主要研究方向为信息智能处理.
E-mail: huanjuan@cczu.edu.cn
(**HUAN Juan** Associate professor at the School of Information Science and Engineering, Changzhou University. She received her Ph.D. degree in agricultural electriflcation automation from Jiangsu University in 2019. Her main research interest is information intelligence processing.)

郑纪翔 2020 年于常州大学信息科学与工程学院获得学士学位. 主要研究方向为行为识别, 机器学习.
E-mail: zjx991031@163.com
(**ZHENG Ji-Xiang** Received his bachelor degree from the School of Information Science and Engineering, Changzhou University in 2020. His research interest covers behavior recognition and machine learning.)