

# 一种基于线性序列差异分析降维的人体行为识别方法

鹿天然, 于凤芹, 陈 莹

(江南大学 物联网工程学院, 江苏 无锡 214122)

**摘 要:** 在视频数据处理过程中容易出现维数灾难的问题。为此, 提出一种线性序列差异分析方法, 对视频数据降维来进行人体行为识别。运用 ViBe 算法对视频帧进行背景减除操作获取行为区域, 在该区域内提取稠密轨迹特征从而去除背景数据的干扰。使用 Fisher Vector 对特征编码后进行线性序列差异分析, 采用动态线性规整算法计算序列类别间相似度, 得到最小化类内残差和最大化类间残差的线性变换, 将特征从高维空间投影至低维空间, 降低特征维数。利用降维后的特征训练支持向量机, 实现人体行为识别。在 KTH 数据集和 UCF101 数据集上进行数据仿真, 结果表明, 与主成分分析算法、线性判别分析法等相比, 该方法可有效提高识别准确率。

**关键词:** 人体行为识别; 背景减除; 稠密轨迹; 线性序列差异分析; 降维

**中文引用格式:** 鹿天然, 于凤芹, 陈莹. 一种基于线性序列差异分析降维的人体行为识别方法[J]. 计算机工程, 2019, 45(3): 237-241, 249.

**英文引用格式:** LU Tianran, YU Fengqin, CHEN Ying. A human action recognition method based on LSDA dimension reduction[J]. Computer Engineering, 2019, 45(3): 237-241, 249.

## A Human Action Recognition Method Based on LSDA Dimension Reduction

LU Tianran, YU Fengqin, CHEN Ying

(School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China)

**【Abstract】** Aiming at the problem that dimensionality disaster easily occurs in the processing of dealing with video data, a dimension reduction method called Linear Sequence Discriminant Analysis (LSDA) is proposed for human action recognition. ViBe algorithm is used to subtract the backgrounds of video frames to get action areas, and dense trajectories are extracted in these areas to suppress the noise caused by camera movements. Fisher Vector is used to encode the features and linear sequence discriminant analysis is conducted on them, the sequence class separability is measured by dynamic time warping distance. In order to reduce the data dimension, a linear discriminative projection of the feature vectors in sequences is mapped to a lower-dimensional subspace by maximizing the between-class separability and minimizing the within-class separability. Support Vector Machine (SVM) is learned from the reduced dimension features, and then get the results of human action recognition. Simulation results on KTH datasets and UCF101 datasets show that compared with Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and other dimension reduction methods, the proposed method can effectively improve the recognition accuracy.

**【Key words】** human action recognition; background subtraction; dense trajectories; Linear Sequence Discriminant Analysis (LSDA); dimension reduction

DOI: 10.19678/j.issn.1000-3428.0050160

## 0 概述

人体行为识别作为模式识别和机器视觉领域的研究热点, 不仅在智能监控<sup>[1]</sup>、人机交互和运动分析方面被广泛应用, 而且在公共场所人员身份鉴别和异常行为监控等方面具有重要的应用前景<sup>[2-3]</sup>。由于人体行为动作的复杂性, 描述目标在视频中的时空运动变化时, 待处理的数据量往往十

分庞大, 需要大量的计算时间。如何在不同的光照、视角和背景下对视频中的人体行为进行快速、精确的分类识别, 是目前亟待解决的难题。

国内外研究者采用降维方法处理视频数据来进行人体行为识别。典型代表有主成分分析 (Principal Component Analysis, PCA) 算法<sup>[4]</sup> 和线性判别分析 (Linear Discriminant Analysis, LDA) 算法<sup>[5]</sup>。文献[6]将 PCA 方法用于人体行为识别, 以非监督方

**基金项目:** 国家自然科学基金 (61573168); 中央高校基本科研业务费专项资金 (JUSRP51733B)。

**作者简介:** 鹿天然 (1993—), 女, 硕士研究生, 主研方向为模式识别、智能系统; 于凤芹、陈 莹, 教授、博士。

**收稿日期:** 2018-01-18    **修回日期:** 2018-02-20    **E-mail:** ltr\_lavender@163.com

式、最大化方差将高维数据映射到低维空间。文献[7]利用 LDA 方法以监督方式得到将高维数据映射到低维空间的投影矩阵。这 2 种方法均有效地降低了数据维度,但都是以相同方式处理不同长度的特征向量,忽略了视频序列间的时间相关性。针对这一问题,文献[8]提出基于流形学习的降维方法,其中应用最为广泛的是局部保持投影(Locality Preserving Projections, LPP)算法,其假设低维数据来源于高维空间的一个潜在流形上。文献[9]将 LPP 运用于人体行为识别中,但由于 LPP 是非监督算法,没有考虑类别标签,分类效果不佳。稀疏表示(Sparse Representation, SR)<sup>[10]</sup>近年来在模式识别领域被广泛应用,它可以对高维数据进行有效表示,通常与其他降维方法相结合在低维空间中表示数据并实现有效分类。然而,该方法采用的 L1 范数计算过程复杂,处理的数据量大,文献[11]将其应用于行为识别,虽取得了较高的识别准确率,但实时效果不佳。

针对稠密轨迹行为识别方法不能很好地区分行为和背景的问题,本文利用 ViBe 算法<sup>[12]</sup>进行背景减除获取行为区域,去除背景冗余数据的干扰。在行为区域提取稠密轨迹特征并对编码后的特征进行线性序列差异分析(Linear Sequence Discriminant Analysis, LSDA),以降低数据维度。由于不同的序列有不同的时续长度,本文采用动态时间规整(Dynamic Time Warping, DTW)算法<sup>[13]</sup>计算序列间的距离,从而衡量序列类别间差异性的,并将序列的平均方差作为序列类别内差异性的衡量准则。

## 1 基本原理

### 1.1 基于 ViBe 算法的背景减除方法

人体行为复杂、背景更新速度慢会导致行为对象部分漏检,影响识别结果。ViBe 算法构建背景速度快,运算效率高,其具体思想是:为每个像素点存储一个样本集,样本集中的采样值为该像素点过去的像素值和其邻居结点的像素值,然后将每一个新的像素值和样本集进行比较来判断是否为背景点。算法主要包括以下 3 个步骤。

#### 1) 背景模型初始化

初始背景模型( $t=0$ )中像素点 $(x, y)$ 处的像素值可以表示为:

$$B^0(x, y) = f^0(x_N, y_N) \quad (1)$$

其中,  $f^0(x_N, y_N)$  表示当前帧像素点 $(x_N, y_N)$ 处的像素值,  $(x_N, y_N)$  为在像素点 $(x, y)$ 周围随机选取的像素点。 $M$  为选取次数,即可以构建的背景模型为  $M$  个,本文中取  $M=20$ 。

#### 2) 前景提取

当前帧的像素点 $(x, y)$ 对应的  $M$  个背景模型在当前时刻  $t$  的像素值为  $B_i^t(x, y)$ ,  $i=1, 2, \dots$ ,

$M$ , 记:

$$b_i(x, y) = \begin{cases} 1, & |f^t(x, y) - B_i^t(x, y)| \leq T \\ 0, & |f^t(x, y) - B_i^t(x, y)| > T \end{cases} \quad (2)$$

其中,参数  $T$  表示前景与背景像素值差异的下限,本文中取 20。

利用  $M$  个背景模型对像素点 $(x, y)$ 的属性进行投票来判断其所属类别,判断准则为:

$$p(x, y) = \begin{cases} \text{背景}, & \sum_{i=1}^M b_i(x, y) \geq \eta \\ \text{前景}, & \sum_{i=1}^M b_i(x, y) < \eta \end{cases} \quad (3)$$

其中,参数  $\eta$  表示设置的经验值,本文中取 3。

### 3) 背景模型更新

从时域和空域两方面考虑 ViBe 算法在背景更新时的随机性。首先在时域上随机抽取一个模型,记为  $B_i^t(x, y)$ 。当前帧图像  $f^t(x, y)$  经过前景提取之后得到一些背景像素点,在更新背景模型  $B_i^t(x, y)$  时,用当前帧图像上对应位置的像素值进行替换。在空域上,对于  $B_i^t(x, y)$  中的任一像素点 $(x, y)$ ,从其 8 邻域中随机抽取一个像素点 $(x_N, y_N)$ ,用当前帧图像上像素点 $(x, y)$ 处的像素值  $f^t(x, y)$  来替换该点处的像素值。

### 1.2 基于 DTW 算法的序列间相似度衡量方法

DTW 算法可以处理不等长序列间距离的衡量问题。设时间序列  $S^1(t) = \{s_1^1, s_2^1, \dots, s_m^1\}$ ,  $S^2(t) = \{s_1^2, s_2^2, \dots, s_n^2\}$ , 其长度分别为  $m$  和  $n$ 。根据它们的时间位置排序,构造  $m \times n$  矩阵  $A_{m \times n}$ ,  $A_{m \times n}$  中的元素  $a_{ij} = d(s_i^1, s_j^2) = \sqrt{(s_i^1 - s_j^2)^2}$ , 在矩阵  $A_{m \times n}$  中,把一组相邻的矩阵元素集成规整路径,记为  $W = w_1, w_2, \dots, w_K$ ,  $W$  的第  $k$  个元素  $w_k = (a_{ij})_k$ , 则  $DTW(S^1, S^2) = \min(\frac{1}{K} \sqrt{\sum_{i=1}^K w_i})$ 。DTW 算法运用动态规划思想寻找一条具有最小规整代价的最佳路径,即:

$$\begin{cases} D(1, 1) = a_{11} \\ D(i, j) = a_{ij} + \min(D(i-1, j-1), D(i, j-1), D(i-1, j)) \end{cases} \quad (4)$$

其中,  $i=2, 3, \dots, m$ ,  $j=2, 3, \dots, n$ 。  $D(m, n)$  表示  $A_{m \times n}$  中规整路径的最小累加值。

对于 2 个序列样本  $U_n$  和  $U_{n'}$ , 它们之间的 DTW 距离可以利用下式计算:

$$\min_{\Pi_n, \Pi_{n'}} \text{tr}((U_n \Pi_n - U_{n'} \Pi_{n'}) \cdot (U_n \Pi_n - U_{n'} \Pi_{n'})^T) \quad (5)$$

其中,  $\Pi_n \in \{0, 1\}^{P_n \times T_{nn'}}$ ,  $\Pi_{n'} \in \{0, 1\}^{P_{n'} \times T_{nn'}}$  为最佳规整路径,  $P_n$  和  $P_{n'}$  分别为 2 个序列的长度,  $T_{nn'}$  为规整路径的长度。

### 1.3 基于 LSDA 的降维原理

LSDA 算法<sup>[14]</sup>是为了处理序列数据的降维问题,寻找最佳线性判别变换来最大化序列类别间可

分离度并最小化类内可分离度,同时利用有监督信息来解决时间依赖问题的一种有效方法。降维是为了找到一个线性变换  $W \in \mathbb{R}^{d \times d'}$ ,将维度为  $d$  的原始向量  $u_p \in \mathbb{R}^d$  降至维度为  $d'$ :  $W^T u_p \in \mathbb{R}^{d'}$ 。因此,初始序列  $U = [u_1, u_2, \dots, u_p] \in \mathbb{R}^{d \times p}$  变为  $[W^T u_1, W^T u_2, \dots, W^T u_p] \in \mathbb{R}^{d' \times p}$ 。

经过投影后,来自不同序列类别的特征尽可能不同,同时相同类别的特征尽可能相同。类别  $i$  的平均序列表示为  $M^i = [m_1^i, m_2^i, \dots, m_{L^i}^i] \in \mathbb{R}^{d \times L^i}$ ,其中  $L^i$  为  $M^i$  的长度,类别  $i$  的方差为  $\Gamma^i$ 。所有序列类别间的 DTW 距离总和和视为衡量序列类别间可分离性的标准,序列类内差异定义为平均方差  $\Gamma_w = \sum_{i=1}^C p^i \Gamma^i$ ,其中,  $p^i$  为序列类别  $i$  的先验概率。类别  $i$  变换后的平均序列为:

$$W^T M^i = [W^T m_1^i, W^T m_2^i, \dots, W^T m_{L^i}^i] \quad (6)$$

$W^T \Gamma_w W$  为类别  $i$  变换后的序列类内残差,序列类内差异包含在  $W$  中。计算 DTW:

$$\begin{aligned} \max_{\Pi} \min_{\Pi'} (\text{tr}(W^T \Gamma_w W)^{-1} \times (\sum_i \sum_j \text{tr}((W^T M^i \Pi^i - \\ W^T M^j \Pi^j) \cdot W^T M^i \Pi^i - W^T M^j \Pi^j)^T)) \Leftrightarrow \\ \max_{\Pi} \min_{\Pi'} (\text{tr}(W^T \Gamma_w W)^{-1} \text{tr}(W^T B_s(\Pi) W)) \end{aligned} \quad (7)$$

其中,  $\Pi^i \in \{0, 1\}^{L^i \times T_{ij}}$ ,  $\Pi^j \in \{0, 1\}^{L^j \times T_{ij}}$ ,  $T_{ij}$  为对齐 2 个平均序列  $M^i$  和  $M^j$  所需要的步长,  $\Pi = \{(\Pi^i, \Pi^j), 1 \leq i \leq j \leq C\}$ ,  $B_s(\Pi)$  定义为  $B_s(\Pi) = \sum_i \sum_j B(\Pi^i, \Pi^j)$ ,  $B(\Pi^i, \Pi^j) = (M^i \Pi^i - M^j \Pi^j)(M^i \Pi^i - M^j \Pi^j)^T$ 。

用  $(\Pi^{i*}, \Pi^{j*})$  表示由 DTW 在初始空间中找到的最佳规整路径,则 LDA 可由下式计算:

$$\max_{\Pi^*} \text{tr}((W^T \Gamma_w W)^{-1} W^T B_s(\Pi^*) W) \quad (8)$$

其中,  $\Pi^* = \{(\Pi^{i*}, \Pi^{j*}), 1 \leq i \leq j \leq C\}$ ,  $B_s(\Pi^*)$  表示序列间差异矩阵。

## 2 LSDA 算法实现步骤

本文运用 LSDA 降维的人体行为识别算法,具体步骤如下:

**步骤 1** 对初始视频帧运用 ViBe 算法进行背景减除操作,获取人体行为区域,去除背景数据的干扰。

**步骤 2** 在行为区域内对视频进行密集采样,提取稠密轨迹<sup>[15]</sup>构建特征描述子,采集方向梯度直方图(Histogram of Oriented Gradient, HOG)和轨迹形状作为形状描述符,利用光流场方向直方图(Histograms of Oriented Optical Flow, HOF)和运动边界直方图(Motion Boundary Histograms, MBH)作为运动描述符,最终得到一个 426 维的描述符。采用 PCA 对获取的特征进行降维处理,减少计算量和数据冗余,然后采用高斯混合模型对特征进行建模,

高斯聚类个数  $K = 256$ 。

**步骤 3** 使用 Fisher Vector 对特征进行编码,编码后得到的特征维数为 109 056 维。

**步骤 4** 对降维后的特征进行线性序列差异分析,利用 DTW 算法衡量类别间的可分离性。

**步骤 5** 计算序列类别的平均方差,并将其作为类内差异,将原始数据进行线性变换,从高维空间投影到低维空间,使类间残差尽可能大而类内残差尽可能小,从而降低数据维度,降维后的数据维度为 38 170 维。

**步骤 6** 利用降维后的特征向量训练 SVM 分类器实现最终的人体行为识别。

具体流程如图 1 所示。

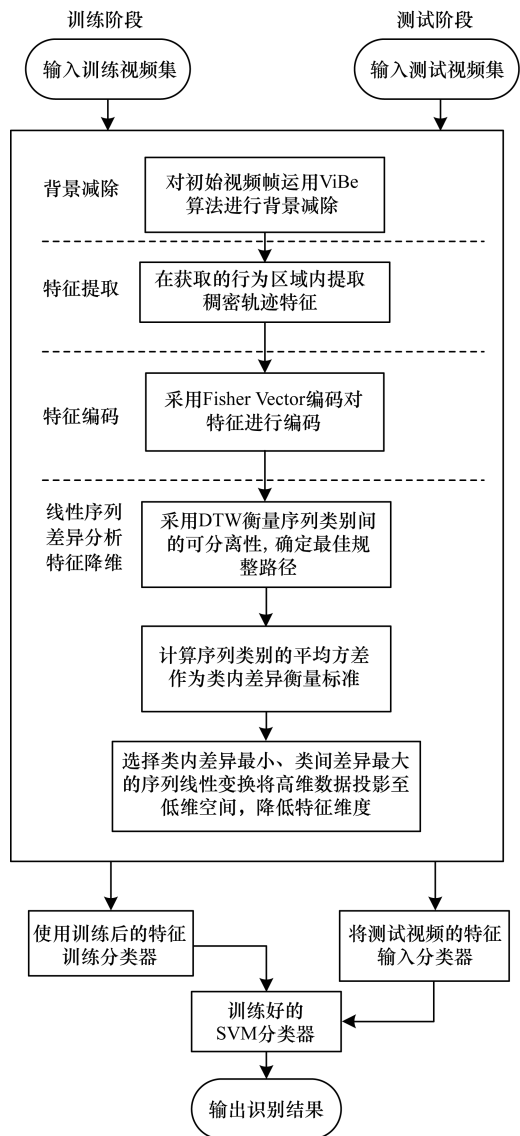


图 1 运用 LSDA 降维的人体行为识别方法流程

## 3 仿真结果与分析

实验所用软件环境为 OpenCV 和 Matlab R2016a, 其中,采用 OpenCV 提取稠密轨迹以及相关特征,采

用 Matlab 实现 ViBe 算法背景减除、线性序列差异分析以及分类预测。本文采用在动作识别领域广为应用的 KTH 数据集<sup>[16]</sup>和 UCF101 数据集<sup>[17]</sup>进行算法效果的检验。KTH 数据集包含室外、室内、室外尺度变化、室外着装变化 4 种场景下的 6 种人体动作,每个动作由 25 个人分别完成,在大多数场景下背景是同质且静止的。UCF101 拥有 13 320 个视频,101 个分类,每一类动作由 25 个人完成,是目前动作类别数、样本数最多的数据库之一。

### 3.1 基于 ViBe 算法的背景减除操作

本文首先对初始视频帧运用 ViBe 算法进行背景减除操作,获取行为区域并去除背景数据的干扰。分别对 KTH 数据集中的“走路”视频和“拳击”视频以及 UCF101 数据集中的“骑车”视频和“弹钢琴”视频进行背景减除操作,获取的效果图与原视频对比如图 2、图 3 所示。

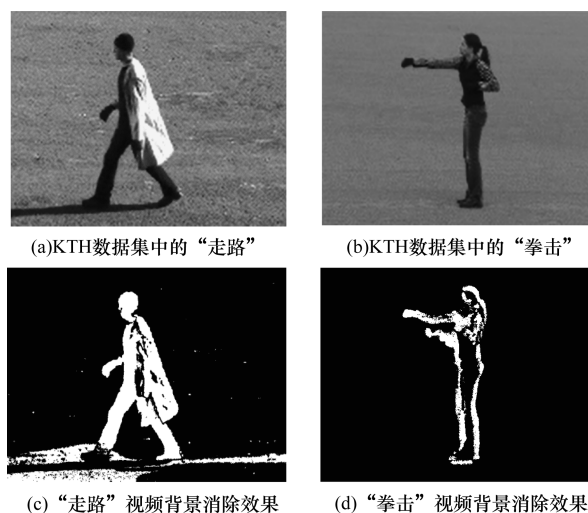


图 2 在 KTH 数据集上 ViBe 算法背景减除前后效果对比

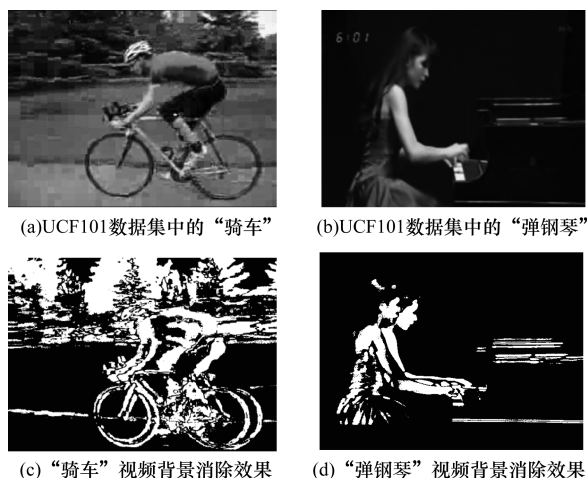


图 3 在 UCF101 数据集上 ViBe 算法背景减除前后效果对比

从图 2、图 3 可以看出,运用 ViBe 算法进行背景减除操作后,可以去除一些类似于相机移动等背景

变化的影响。在图 2 中,由于 KTH 数据集背景简单且静止,因此运用 ViBe 算法进行背景减除后效果较好,可基本获取人体行为区域。而 UCF101 数据集背景较为复杂且在变化,其效果不如 KTH 数据集上取得的效果。

分别对 KTH 数据集中的“走路”视频以及 UCF101 数据集中的“骑车”视频进行背景减除操作后提取稠密轨迹,并与未进行背景减除操作的稠密轨迹提取过程进行对比,获取效果如图 4 所示。

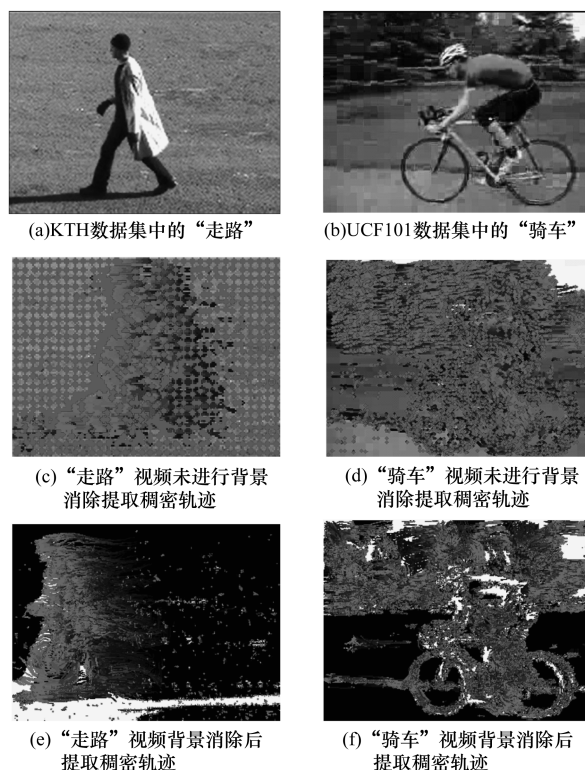


图 4 背景减除操作前后稠密轨迹提取效果对比

从图 4 的对比效果可以看出,经过背景减除操作后在获取的运动主体区域内提取稠密轨迹,可有效去除背景冗余信息的影响,降低特征维数的同时减少计算量。由于 KTH 数据集的背景静止、单一,而 UCF101 数据集的背景复杂且随人体运动,本文的背景减除操作在 KTH 数据集上效果较好。

在 KTH 数据集和 UCF101 数据集上进行背景减除操作,特征所需的存储空间以及以稠密轨迹为特征获得的识别准确率( $\eta$ )结果如表 1 所示。

表 1 背景减除操作前后存储空间和识别准确率对比

数据集	特征所需存储空间/GB		节省的存储空间占比/%	$\eta$ /%
	背景减除前	背景减除后		
KTH	17.40	7.11	59.15	95.40
UCF101	679.00	530.00	21.93	86.00

从表 1 可以看出,对视频进行背景减除操作可以去除背景数据的干扰,节省存储空间。在 KTH 数

数据集上可以节省 59.15% 的存储空间,在 UCF101 数据集上可以节省 21.93% 的存储空间。这是因为 KTH 数据集大多数视频背景是同质化且静止的,较为单一,进行背景减除操作效果明显。UCF101 数据集的背景较为复杂,且背景大多随人体运动,因此效果相对不明显。

### 3.2 基于 LSDA 的降维操作

在 KTH 数据集上运用本文方法对特征降维来进行人体行为识别,各动作之间识别准确率的混淆矩阵如表 2 所示。

表 2 本文方法在 KTH 数据集上的分类混淆矩阵

类别	拳击	鼓掌	挥手	走路	慢跑	奔跑
拳击	1.00	0.00	0.00	0.00	0.00	0.00
鼓掌	0.00	0.97	0.03	0.00	0.00	0.00
挥手	0.00	0.03	0.97	0.00	0.00	0.00
走路	0.00	0.00	0.00	0.96	0.04	0.00
慢跑	0.00	0.00	0.00	0.04	0.87	0.09
奔跑	0.00	0.00	0.00	0.00	0.09	0.91

由表 2 可以看出,拳击识别率为 100%。挥手和鼓掌之间有一定的混淆,鼓掌识别率为 97%,其中有 3% 被误识为挥手。走路、慢跑和奔跑由于相似度较大,混淆度稍大,慢跑有 9% 的误差落在奔跑上,有 4% 的误差落在走路上,另外有 9% 的奔跑被误识为慢跑。

本文采用不同的降维方法来降低以稠密轨迹特征作为视频特征的数据维度,在此基础上进行人体行为识别,并对结果进行对比。在 UCF101 数据集上进行实验,得到降维后的识别率以及相关特征数据维度的对比结果如表 3 所示。

表 3 在 UCF101 数据集上 5 种池化方法的效果对比

方法	识别率/%	特征维度
PCA	81.2	54 528
LDA	82.9	44 614
LPP	87.8	54 528
SR	88.6	40 896
LSDA	89.3	38 170

由表 3 可以看出,经过 PCA 和 LPP 降维处理后,特征维度均降为原特征维度的一半,但 LPP 降维后的识别率较 PCA 稍高。LDA 虽然对特征维度的降低效果较好,但识别率与 PCA 类似。本文提出的 LSDA 方法可在获取最佳降维效果的基础上保持较高的识别准确率,其特征维度比 LDA 降低 14.4%,同时识别率较 LPP 高 1.5%。

为进一步验证本文方法的有效性,采用识别准确率( $\eta$ )为评价指标,分别与文献[6-7,9,12]方法在 KTH 数据集和 UCF101 数据集上的实验结果进行对比,结果如表 4 所示。

表 4 不同方法的识别准确率对比 %

方法	KTH	UCF101
文献[6]方法	94.5	84.0
文献[7]方法	94.7	84.2
文献[9]方法	94.9	88.3
文献[12]方法	95.4	88.9
本文方法	95.6	89.3

由表 4 可以看出,由于 KTH 数据集中的视频背景静止且简单,上述 5 种方法均可取得较高的识别准确率。本文方法在 KTH 数据集上取得 95.6% 的识别准确率,与文献[6-7,9,12]相比分别提升 1.1%、0.9%、0.7% 和 0.2%,在一定程度上提高了人体行为的识别准确率。在 UCF101 数据集上,本文方法取得 89.3% 的识别准确率,与文献[6-7,9,12]相比分别提升 5.3%、5.1%、1.0% 和 0.4%,尤其是与文献[6]的 PCA 方法和文献[7]的 LDA 方法相比,本文方法的准确率提升较为明显,实现了良好的识别效果。

## 4 结束语

本文应用 LSDA 的降维方法进行人体行为识别。利用 ViBe 算法对视频帧进行背景减除操作,得到人体行为区域并提取该区域的稠密轨迹特征。采用 Fisher Vector 对特征进行编码及线性序列差异分析,将序列中的特征向量线性映射到一个低维子空间,降低特征维数。利用降维后的特征训练 SVM 实现人体行为识别。仿真结果表明,该方法在 KTH 数据集和 UCF101 数据集上均取得较好的识别效果,可提高人体行为识别准确率。但是,该方法在应用场景方面还有一定的局限性,下一步将降低算法复杂度并拓展其应用场景。

### 参考文献

- [1] 黄凯奇,陈晓棠,康运锋,等. 智能视频监控技术综述[J]. 计算机学报,2015,20(6):1093-1118.
- [2] 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.
- [3] 李瑞峰,王亮亮,王珂. 人体动作行为识别研究综述[J]. 模式识别与人工智能,2014,27(1):35-48.
- [4] TURK M, PENTLAND A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience,1991,3(1):71-86.
- [5] BELHUMEUR P N, HESPAHNA J P, KRIEGMAN D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[C]//Proceedings of the 4th European Conference on Computer Vision. Berlin, Germany: Springer,1996:45-58.

(下转第 249 页)

- [9] OCHS P, BROX T. Higher order motion models and spectral clustering [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2012: 614-621.
- [10] WANG W, SHEN J, PORIKLI F. Saliency-aware geodesic video object segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2015: 3395-3402.
- [11] LEE Y J, KIM J, GRAUMAN K. Key-segments for video object segmentation [C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA; IEEE Press, 2011: 1995-2002.
- [12] MA T, LATECKI L J. Maximum weight cliques with mutex constraints for video object segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2012: 670-677.
- [13] PAPAOGLOU A, FERRARI V. Fast object segmentation in unconstrained video [C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA; IEEE Press, 2013: 1777-1784.
- [14] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.
- [15] ZHANG D, JAVED O, SHAH M. Video object segmentation through spatially accurate and temporally dense extraction of primary object regions [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2013: 628-635.
- [16] BAMPIS C G, MARAGOS P, BOVIK A C. Graph-driven diffusion and random walk schemes for image segmentation [J]. IEEE Transactions on Image Processing, 2017, 26(1): 35-50.
- [17] BANJAC G, GOULART P, STELLATO B, et al. Infeasibility detection in the alternating direction method of multipliers for convex optimization [J]. IEEE Transactions on Power Electronics, 2017, 32(5): 4007-4020.
- [18] GRANT M, BOYD S, YE Y. CVX: Matlab software for disciplined convex programming [M]. Berlin, Germany: Springer, 2008: 109-120.
- [19] KANDA N, LU X, KAWAI H. Maximum-a-posteriori-based decoding for end-to-end acoustic models [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 1023-1034.

编辑 樊丽娜

(上接第241页)

- [6] HUANG S, YE J, WANG T, et al. Extracting refined low-rank features of robust pca for human action recognition [J]. Arabian Journal for Science and Engineering, 2015, 40(5): 1427-1441.
- [7] 王森, 孙季丰, 余家林. 基于特征层融合和随机投影的行为识别算法 [J]. 科学技术与工程, 2017, 17(13): 210-215.
- [8] HE X, NIYOGI P. Locality preserving projections [J]. Advances in Neural Information Processing Systems, 2002, 16(1): 186-197.
- [9] 王鑫, 沃波海, 管秋, 等. 基于流形学习的人体动作识别 [J]. 中国图象图形学报, 2014, 19(6): 914-923.
- [10] WRIGHT J, MA Y, MAIRAL J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044.
- [11] 张瑞杰, 魏福山. 结合 Fisher 判别分析和稀疏编码的图像场景分类 [J]. 计算机辅助设计与图形学学报, 2015, 27(5): 808-814.
- [12] 肖玉玲. 结合 HOG/HOF 级联特征和多层分类器的人体行为识别 [J]. 计算机工程与设计, 2017, 38(9): 2567-2572.
- [13] KEOGH E, RATANAMAHATANA C A. Exact indexing of dynamic time warping [J]. Knowledge and Information Systems, 2005, 7(3): 358-386.
- [14] SU B, DING X, WANG H, et al. Discriminative dimensionality reduction for multi-dimensional sequences [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(1): 77-91.
- [15] WANG H, SCHMID C. Action recognition with improved trajectories [C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA; IEEE Press, 2014: 3551-3558.
- [16] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach [C]//Proceedings of International Conference on Pattern Recognition. Washington D. C., USA; IEEE Press, 2004: 32-36.
- [17] SOOMRO K, ZAMIR A R, SHAH M. UCF101: a dataset of 101 human actions classes from videos in the wild [EB/OL]. [2017-12-25]. <https://arxiv.org/pdf/1212.0402.pdf>.

编辑 樊丽娜