

DOI: 10.13382/j.jemi.2017.12.007

# 基于双特征融合与自适应提升机制的 图像动作识别算法\*

罗冬梅<sup>1</sup> 左金水<sup>2</sup> 余文森<sup>3</sup>

(1. 武夷学院 信息技术与实验室管理中心 武夷山 354300; 2. 浙江工商大学 管理学院 杭州 310018;  
3. 武夷学院 数学与计算机学院 武夷山 354300)

**摘要:**针对复杂环境中动作识别易受到光照变化、目标旋转、遮挡等导致目标位置不精确,目标漂移以及识别错误等问题,提出了一种基于双特征融合与自适应提升的动作识别算法。首先,基于时空上下文(spatio-temporal context, STC)机制,通过时空上下文关系与视觉系统特性来提取图像序列特征,降低光照变化、遮挡对行为动作的影响;同时,利用卷积神经网络(convolution neural network, CNN)来处理图像序列,分别获得STC特征与CNN特征;其次,引入主成分分析算子,定义双特征融合规则,对获得STC特征与CNN特征进行组合,形成一种更准确、完整的特征表示;然后,通过得到的新特征,利用自适应提升算法(adaptive boosting algorithm, ABA)进行分类训练,完成对行为动作决策判断。在Weizmann、Hollywood数据集上测试表明,相对于当前常用的动作识别方法,所提算法对各种行为动作具有更高的识别精度与鲁棒性,更能适应复杂背景和动作变化。所提算法具有较高的人体动作识别精度,在视频监控、人机交互等领域具有一定的应用价值。

**关键词:** 图像动作识别;时空上下文;双特征融合;卷积神经网络;主成分分析;自适应提升算法

**中图分类号:** TP399; TN99 **文献标识码:** A **国家标准学科分类代码:** 520.2040

## Motion recognition algorithm based on double feature fusion and adaptive boosting mechanism

Luo Dongmei<sup>1</sup> Zuo Jinshui<sup>2</sup> Yu Wensen<sup>3</sup>

(1. Information Technology and Laboratory Management Center, Wuyi University, Wuyishan 354300, China;  
2. College of Management, Zhejiang Gongshang University, Hangzhou 310018, China;  
3. College of Mathematics and Computer, Wuyi University, Wuyishan 354300, China)

**Abstract:** In order to solve the defects such as inaccurate target location, target drift and recognition error induced by influence of illumination change, target rotation, occlusion in complex environment, a motion recognition algorithm based on double feature fusion and adaptive boosting was proposed. Firstly, in order to reduce the influence of illumination variation and occlusion on behavior, spatio-temporal context was used to extract the image sequences feature based on spatiotemporal context and the visual system characteristics. At the same time, the convolution neural network was introduced to operate the image sequence's features for obtaining the STC and CNN features. Secondly, the principal component analysis operator was introduced to effectively combine the STC features and features to form a more accurate and complete feature representation. Then, by the new features, the adaptive boosting algorithm was used for classification training, the decision making of action was completed. The tests on the current popular data set show that, compared with the current commonly used behavior recognition methods, the proposed algorithm can recognize and understand all kinds of action, recognition rate is greatly improved, able to adapt for complex background and behavioral changes. This algorithm has higher accuracy and practical value in video surveillance and human-computer interaction.

**Keywords:** image motion recognition; temporal and spatial context; double feature fusion; convolution neural network; principal component analysis; adaptive lifting algorithm

收稿日期:2017-04 Received Date: 2017-04

\* 基金项目:教育部科学青年基金(13YJC630253)、福建省自然科学基金(2015J01668)、福建省中青年骨干教师教育科研项目(JB14103)资助

## 0 引言

近年来,随着计算机视觉的快速发展,基于视觉的动作识别成为当前富有挑战性的研究对象,在智能视频监控、人机交互等领域起着关键作用,越来越受到大家的关注。随着传感器设备的迅速发展,视频也成海量增加,如何快速、准确地对视频中的动作识别成为亟待解决的困难之一。动作识别主要的两个部分是特征提取和特征的识别与理解<sup>[1]</sup>。由于人体动作复杂多变,没有特定的运动轨迹和方向,可能会受到各种物体的遮挡,受环境干扰较大,使得动作特征提取难度加大,动作特征捕获的优劣对后续分类学习具有重要作用。动作的识别与理解是将提取的动作特征通过分类训练进行识别判断<sup>[2-3]</sup>。秦磊等人<sup>[4]</sup>提出了一种特征点轨迹的动作决策算法,首先利用 KLT 追踪器对局部特征追踪;然后将时空特征追踪曲线为基本单元;利用曲线相对空间、相对运动来对曲线关系建模;通过对曲线建模能更有利于提取各种动作信息;通过结合不同特征来进行分类学习,但是通过单一的局部特征描述的动作信息有限,使动作无法很好理解。Wang 等人<sup>[5]</sup>设计了一种有效的密集轨迹特征的行为识别方法。其利用光流场采样的特征点形成曲线,分别测量曲线位置矢量与其曲线中子块的 3 个信息为动作表示。3 个信息分别为光流信息(HOF)、梯度信息(HOG)和运动信息(MBH),为分类决策提供了更多参考,从而增加了动作识别率,但是密集轨迹造成较多冗余特征,计算成本太高。林水强等人<sup>[6]</sup>提出了一种姿势序列状态机动作判别技术。首先构建肢体节点体系,将肢体节点信息变换为用户结构,利用三维网格分割。然后在肢体节点体系中建立特征矢量,通过节点的空间关系作为特征表示。再通过节点运动特征构成运动路线,从而完成对动作识别,但是该方法容易受到环境干扰,稳定性不强。现有的研究中出现了基于时空上下文(STC)理论,形成了鲁棒性较高的动作表达。但是在复杂的场景中,基于 STC 识别效果有限,对遮挡、光照变化的识别率不佳。

为了解决上述不足,并提高复杂场景中对于人体动作识别精度,提出了一种双特征融合与自适应提升的动作识别算法。其通过 STC 关系与视觉系统提取序列动作特征,得到了 STC 特征。利用具有平移不变行的卷积神经网络 CNN 技术提取动作序列 CNN 特征,有效改善其鲁棒性,使其能够对遮挡、旋转等目标实现识别。为了得到更准确、全面的特征表示,利用主成分分析(PCA)对 STC 特征与 CNN 特征有效融合。同时,根据融合新特征,利用自适应提升算法(ABA)进行分类学习,更好对动作识别与理解,准确识别动作。最后在常用的数据集进行测试。

## 1 CTC 特征

对于视觉跟踪,局部上下文具有某个特征和其周围背景区域。在连续帧间特征的局部背景具有很强的时空联系。如果一个特征被阻挡,导致其外形产生了较大的变化。但是,由于只有小部分的上下文区域被阻挡,前后的上下文区域具有相似性,因此,该特征的局部上下文变化较少。所以,可通过当前帧的局部上下文预测下一帧的特征关系。

STC 主要是基于 Bayes 理论提取对象与其局部上下文的时空信息,获取对象与其附近低级特征的统计性质<sup>[7]</sup>。利用时空信息与视觉系统获取对象的置信图,置信最高的点为对象的坐标,找到似然概率最高特征。将对象追踪看作为测量对象的置信图  $m(x)$ ,表示为<sup>[8]</sup>:

$$m(x) = p(x|o) \quad (1)$$

式中:  $x$  为特征位置,  $o$  为跟踪对象。令当前特征中心点是  $x^o$ ,则上下文特征表示如下:

$$X^o = \{c(z) = (I(z), z) | z \in \Omega_c(x^o)\} \quad (2)$$

式中:  $I(z)$  为在  $z$  处的灰度值,  $\Omega_c(x^o)$  为特征中心的  $x^o$  及附近的上下文区域。

通过将上下文信息与跟踪过程相结合可得到式(1)中的置信图进一步表达式:

$$m(x) = \sum_{c(z) \in X^o} p(x, c(z) | o) = \sum_{c(z) \in X^o} p(x | c(z) | o) p(c(z) | o) \quad (3)$$

式中:  $p(c(z) | o)$  为上下文先验概率,表示外观特征;  $p(x | c(z), o)$  为条件概率,表示特征位置和空间联系。

在 STC 中,主要通过先验概率,空间模型和时空模型构成。首先,通过局部上下文构建先验概率:

$$p(c(z) | o) = I(z) w_\sigma(z - x^o) \quad (4)$$

其中,  $w_\sigma$  为高斯加权算子。一般来说,离  $x^o$  越近,其追踪特征越重要,权重也越大。

对于空间模型中,通过在  $p(x | c(z), o)$  嵌入径向非对称算子进行表示:

$$p(x | c(z), o) = h^{sc}(x - z) \quad (5)$$

$p(x | c(z), o)$  反应了特征位置  $x$  和局部背景  $z$  的距离与方向。因此,其表示特征与局部区域的空间关系。

根据计算的式(4)和(5)可演变得到置信图  $m(x)$ :

$$m(x) = \sum_{c(z) \in X^o} p(x | c(z) | o) p(c(z) | o) = \sum_{c(z) \in X^o} h^{sc}(x - z) I(z) w_\sigma(z - x^o) = h^{sc}(x) \otimes (I(x) w_\sigma(x - x^o)) \quad (6)$$

其中,  $\otimes$  为卷积运算。

通过 STC 特征描述,其利用前景运动特征和形态特

征可很好地对运动目标进行表征,有效反应各种复杂的运动属性。包含物体类别、速度、方向信息。将 *STC* 得到的置信图  $m(x)$  扩展为一个矩阵  $M$ ,表示为:

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1N} \\ M_{21} & M_{22} & \cdots & M_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ M_{M1} & M_{M2} & \cdots & M_{MN} \end{bmatrix} \quad (7)$$

为避免光照影响,对图像去均值化;并通过 Hamming 窗来降低边缘对快速傅里叶变换 (FFT) 带来的频率干扰。

## 2 CNN 特征

CNN 为深度学习在图像领域的应用,CNN 利用权值共享降低参数扩张,并向前运算中利用卷积函数对信号完成卷积运算<sup>[9]</sup>。然后,通过非线性函数对当前层的结果输出。CNN 某个层的输出可作为信号的另一种表达,并可作为其通过网络学习产生的特征。CNN 消除了繁琐的预处理,因而得到了广泛的应用。CNN 的主要由 2 层构成,其一为卷积层 (convolution layer, CL),另一个为下采样层 (down-sampling layer, DSL),通过利用 sigmoid 算子作为 DSL 的激活算子,使 DSL 保持平移不变性。CNN 的每一个 CL 包含一个用来求局部平均与二次提取的计算层,通过两次特征提取可有效降低了特征分辨率。对于一个神经网络,其结构如图 1 所示。

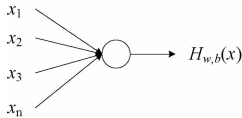


图 1 神经网络  
Fig. 1 Neural Network

相应的公式表示为:

$$h_{w,b}(x) = f(\mathbf{W}^T x) = f\left(\sum_{i=1}^3 W_i x_i + b\right) \quad (8)$$

图 1 表示的也称之为 Logisti 回归模型。根据多个网络节点形成的含有多层关系就得到了神经网络模型,表示如图 2 所示。

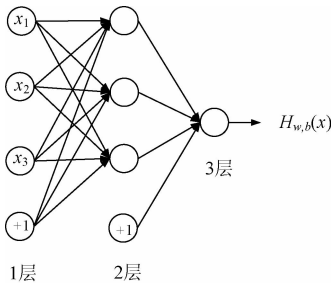


图 2 多层神经网络  
Fig. 2 Multi-layer neural network

在 CNN 中,对于每一个 CL,上一级的信息映射可通过卷积核完成,并利用一个激活算子,得到后续信息映射。通过多个输出信息映射可得到 CNN 的映射值<sup>[10]</sup>:

$$X_j^{\ell-1} = f\left(\sum_{i \in M_j} X_i^{\ell-1} * K_{ij}^{\ell} + b_j^{\ell}\right) \quad (9)$$

其中:  $M_j$  为输入映射的集,  $K_{ij}^{\ell}$  为卷积核,  $j, i$  分别为输入与输出特征,  $b$  为偏置参数。设每个 CL 会对应一个 DSL,对于第  $i$  个 CL 对应的权重,先测量所有网络的灵敏度  $\delta$ ,定义为:

$$\delta = \frac{\partial E}{\partial u} \quad (10)$$

式中:  $E$  为总误差,  $u$  为净激活。  $E$  为一个多分类学习的总误差,定义如下:

$$E(w, \beta, k, b) = \frac{1}{2} \sum_{n=1}^N |t_n - y_n|^2 \quad (11)$$

式中:  $k$  为核参数,  $\beta$  为下采样权值,  $w$  为全连接权值,  $y_n$  为第  $n$  个数据在前向传播预测结果,  $t_n$  为第  $n$  个数据的类别真值。通过  $\delta$  获取误差较少的方向。

因下采样的存在,在 DSL 中的  $\delta$  小于 CL 中的  $\delta$ ,所以,需将 DSL 中第  $l+1$  的  $\delta$  上采样至 CL 中  $l$  的  $\delta$ :

$$\delta_j^l = \frac{\partial E}{\partial u_j^l} = \beta_j^{l+1} (f'(u_j^l) \cdot \text{UP}(\delta_j^{l+1})) \quad (12)$$

式中:  $f$  为激活函数,  $u_j^l$  为 CL $l$  的第  $j$  个通道的净激活,  $\text{UP}()$  为上采样运算,  $\cdot$  为每个节点相乘。为获得 DSL 的  $\delta$ ,需要寻找当前卷积核相乘。因此,DSL 中  $l$  第  $j$  个通道的  $\delta$  可表示为:

$$\delta_j^l = f'(u_j^l) \cdot \text{conv2}(\delta_j^{l+1}, \text{rot180}(k_j^{l+1}), \text{full}') \quad (13)$$

式中:  $\text{rot}(180)$  为对卷积核旋转  $180^\circ$ ,  $\text{conv2}$  为运算函数。然后,  $E$  对偏移量  $b$  的偏导和上一个 CL 的相同,只需对  $\delta$  中的所有节点  $\delta$  加起来,表示为:

$$\frac{\partial E}{\partial b_j^l} = \sum_{u,v} (\delta_j^l)_{u,v} \quad (14)$$

CNN 一般采用 CL 与 DSL 交替设置,即一个 CL 连接一个 DSL,CL 后接 DSL,因此,通过 CL 提取特征,再通过组合得到更抽象的特征,最后得到对象的描述特征。CNN 的框图表示如图 3 所示。

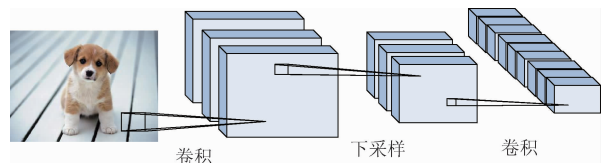


图 3 卷积神经网络  
Fig. 3 Convolution neural network

## 3 本文图像动作识别算法

针对单一特征对动作特征描述能力不足,特别是在

复杂场景下,容易受到光照变化、目标旋转、遮挡等环境的影响,使其难以准确提取鲁棒动作特征,本文定义了双特征融合机制,为了降低光照变化、遮挡对行为动作的影响,引入 STC 技术,通过时空上下文关系与视觉系统特性提取图像序列 STC 特征。同时,为了避免复杂图像的预处理和增加特征提取鲁棒性,利用具有平移不变性的 CNN 对图像序列操作,得到 CNN 特征。为了对提取的两个特征融合,引入 PCA 融合技术,PCA 通过线性变换形成一组互不相关的主特征,在完成特征融合的同时也能有效降低冗余信息。根据得到的融合双特征,利用 ABA 算法进行分类训练,完成动作的判断。整个动作识别过程如图 4 所示。

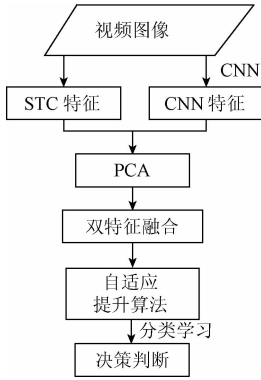


图 4 本文算法流程

Fig. 4 Flow chart of the algorithm proposed in this paper

### 3.1 基于 PCA 特征融合

对于图像序列提取的 STC 特征与 CNN 特征,其能某点或者场景下很好捕抓动作序列的局部特征,但不能完整、准确地对动作特征表达<sup>[11]</sup>。为了获得更准确、全面的动作特征,本文引入一种基于 PCA 的特征融合算法,将提取的 STC 特征与 CNN 特征组合,形成一个更加有效的特征表示。

在 PCA 算法中,当利用新的分量对原始信号重构时,在最小均方误差下逼近效果最佳,表示为<sup>[12]</sup>:

$$\boldsymbol{\varepsilon}^2(m) = E\left\{\left\|\mathbf{x} - \sum_{i=1}^m (\mathbf{u}_i^T \mathbf{x}) \mathbf{u}_i\right\|^2\right\} \quad (15)$$

式中:  $\boldsymbol{\varepsilon}$  为最小均方差,  $\mathbf{x} = \{x_1, x_2, x_n\}^T$ , 均值为  $E$ , 协方差  $C_x = E\{\mathbf{x}\mathbf{x}^T\}$ 。为了寻求矩阵  $\mathbf{U}^T$ , 对  $\mathbf{x} = \{x_1, x_2, x_n\}^T$  正交处理, 得到新的矢量  $\mathbf{y} = \{y_1, y_2, y_n\}^T$  无相关性, 表示为:

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} = \{u_1, u_2, u_n\}^T \mathbf{x} \quad (16)$$

选择  $\mathbf{x}$  在  $\mathbf{y}$  的展开式的  $m$  项在最小均方差下线性估计  $\hat{\mathbf{x}}$ , 表示为:

$$\hat{\mathbf{x}} = \sum_{i=1}^m \mathbf{y}_i \mathbf{u}_i \quad (17)$$

因此,均方误差估计可表示为:

$$\boldsymbol{\varepsilon}^2(m) = E\left\{\left\|(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\right\|\right\} = \sum_{i=m+1}^n E[\mathbf{y}_i \mathbf{y}_i^T] = \sum_{i=m+1}^n E \mathbf{u}_i^T R_x \mathbf{u}_i \quad (18)$$

在此约束条件下,使  $\boldsymbol{\varepsilon}^2(m)$  最小的条件满足:

$$J = \sum_{i=m+1}^n \mathbf{u}_i^T R_x \mathbf{u}_i - \sum_{i=m+1}^n \lambda_i (\mathbf{u}_i^T \mathbf{u}_i - 1) \quad (19)$$

PCA 为一种线性变换,通过对初始状态具有一定相关性的信息经过线性变换形成一组新的互不相关的主特征,有效降低冗余信息<sup>[13]</sup>。图 5 所示为通过 PCA 方法将 STC 特征与 CNN 特征融合的结果,图 5(a) 为图像序列,图 5(b)、(c) 分别为通过 STC 与 CNN 方法的特征表示,图 5(d) 为通过 PCA 对两种特征结合所形成的新的特征表示。从图中看出,STC 与 CNN 方法得到的某些局部特征,经过 PCA 特征融合后更有效、全面反映了动作特征,为分类学习提供更有力的决策依据。

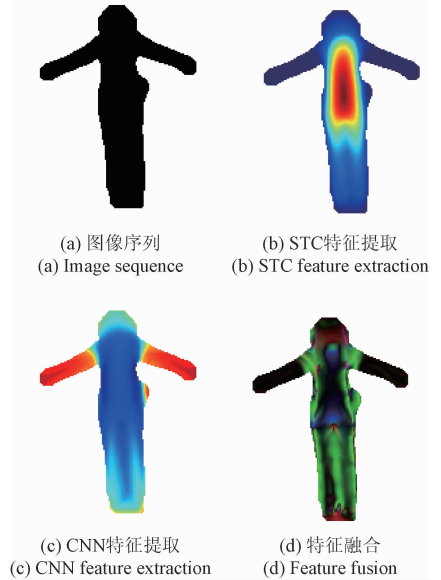


图 5 基于 PCA 的特征融合

Fig. 5 Feature fusion based on PCA

### 3.2 基于自适应提升分类训练

ABA 算法是一种迭代算法,在最终分类误差率小于预定阈值  $\tau$  时,算法停止工作<sup>[14]</sup>。ABA 主要是针对同一个训练数据集迭代地弱分类训练,在每次迭代过程都会对总体数据集的权值的调整,同时利用误差来分配弱分类的权值。初始时,每一个训练数据的权值都是一致的。通过每次迭代后,对数据权值分配,分配的依据主要是:对正确种类降低权重,对错误种类增加权重。因此,经过多次迭代后,错误分类的数据被搜索出来,并获得了一个用于下一个弱分类器的新数据分布。通过预设的  $T$  个迭代,产生了  $T$  个弱分类。利用每个弱分组按权值生成最终的强分组。ABA 步骤如下:

1) 输入训练数据集  $S = [(x_1, y_1), (x_i, y_i), \dots, (x_M, y_M)]$ ,  $i \in \{1, M\}$ ,  $y_i \in \{-1, 1\}$ , 迭代数量  $N$ 。

2) 初始化数据权值分布  $W_1(i) = \frac{1}{m}$ , 初始分类器  $L$ 。

3) 通过  $W$  与弱分组  $r_i = L(S, W)$ , 对  $r_i$  的误差  $\varepsilon_i$  测算, 表示为:

$$\varepsilon_i = P_{S_i}(r_i(x_i) \neq y_i) \quad (20)$$

如果  $\varepsilon_i \geq 0.5$ , 那么迭代停止, 否则继续执行。

4) 计算  $r_i$  的权值, 表示为:

$$\omega_i = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_i}{\varepsilon_i} \right) \quad (21)$$

5) 更新数据集的权值分布, 表示为:

$$W_{i+1}(i) = \frac{W_i(i) \exp(-\omega_i y_i r_i(x_i))}{Z_i} \quad (22)$$

式中:  $Z_i$  为归一化因数。

6) 通过迭代操作, 获得了强分组  $Q(X)$ :

$$Q(X) = \text{sign} \left( \sum_{i=1}^r \omega_i r_i(X) \right) \quad (23)$$

为了将 ABA 算法的二分类演变为多分组, 在初始化阶段设定不同数据类别标签  $\zeta$ , 因此, 权值更新可表示为:

$$W_{i+1}(i) = \frac{W_i(i, \zeta) \exp(-\omega_i y_i [\zeta] r_i(x_i, \zeta))}{Z_i} \quad (24)$$

因此, 强分类器  $Q(X, \zeta)$  可表示为:

$$Q(X, \zeta) = \text{sign} \left( \sum_{i=1}^r \omega_i r_i(X, \zeta) \right) \quad (25)$$

## 4 实验与分析

为了对算法性能验证, 在公开的 Hollywood 与 Weizmann 数据集上完成测验。实验环境为 Core I3, 3.30 GHz 的 CPU, 4 GB 运行内存, Win8 操作系统。Hollywood 是由 69 个 Hollywood 中题材组成, 其中的情绪、衣服、动作, 运动、光照、阴影等与实际生活较相似, 因此, 对于动作识别提供了良好的模拟环境<sup>[15]</sup>。Hollywood 数据集由 10 个场景的 12 个动作类别的 3 669 个序列组成, 该 12 种类别分别为 AnswerPhone、Drivecar、Eat、Fightperson、Getoutcar、HandShake、HugPeron、Kiss、Run、SitDown、SitUp、StandUp。Weizmann 数据集<sup>[16]</sup>由含有 Bend、Jack、Jump、P-jump、Run、Side、Skip、Walk、Wavl、Wave2 的 10 种动作类别组成。

为了显示算法的优越性, 选择当前流行的动作识别算法作为对比, 文献[4-6]算法, 分别记为 A、B、C 算法。设预定阈值  $\tau = 0.05$ , 迭代次数为 50。为了对算法性能定量评价, 通过当前常用的查准率、查全率和识别率<sup>[17]</sup>进行定量评估。

为了评价算法在每个动作类别上的识别率, 测量在 Weizmann 数据集中混淆矩阵, 如表 1 所示。表中第  $i$  行第  $j$  为第  $i$  行识别为第  $j$  种动作的概率, 相同行累加等于 1。表 1 为 Weizmann 测试集中各个动作的混淆矩阵。从表 1 中得出, Run 和 Side 的识别率较低, Run 易被判别为 Side 与 Jump, Side 易被认为 Jump。主要是这几种类别特征相似性高, 区分难度较大, 而 Jump 与 Wavl 的正确率最高。

表 1 Weizmann 数据集的混淆矩阵

Table 1 Confusion matrix of Weizmann data set

	Bend	Jack	Jump	P-jump	Run	Side	Skip	Walk	Wavl	Wave2
Bend	0.94				0.02			0.04		
Jack		0.93	0.02				0.02			0.03
Jump			1.00							
P-jump		0.03		0.94				0.03		
Run		0.02	0.03		0.88	0.05			0.02	
Side		0.02	0.04		0.01	0.91		0.02		
Skip			0.02			0.01	0.95		0.02	
Walk		0.03		0.03				0.93		0.02
Wavl									1.00	
Wave2		0.01	0.03							0.94

表 2 所示为 Hollywood 数据集中各个动作的的混淆矩阵。从表 2 中得出, Getoutcar 和 Kiss 的识别率较低, Getoutcar 易被识别为 Eat 与 HandShake 动作, Kiss 易被认

为 SitUp 动作。此外, Sit-Up 动作的识别率也相对较低, 容易被判断为 St-Up 动作, 主要是这几种动的特征相似性较高。

表 2 Hollywood 数据集的混淆矩阵  
Table 2 Confusion matrix of Hollywood data set

	An-Ph	Drivecar	Eat	Fi-per	Get-car	Ha-Sh	Hu-Pe	Kiss	Run	Sit-Do	Sit-Up	St-Up
An-Ph	0.91		0.01	0.02	0.01				0.03	0.02		
Drivecar	0.02	0.92		0.01			0.02			0.03		
Eat		0.03	0.92	0.01		0.03		0.01				
Fi-per		0.02		0.93		0.02		0.02			0.01	
Get-car	0.02	0.01	0.04		0.86	0.05		0.01			0.01	
Ha-Sh		0.01		0.02	0.02	0.92		0.01		0.02		
Hu-Pe	0.02		0.01			0.02	0.93		0.02			
Kiss		0.02		0.02	0.02		0.02	0.85		0.02		0.05
Run	0.02		0.01		0.02		0.02		0.90		0.03	
Sit-Do			0.02				0.01		0.02	0.94		0.01
Sit-Up	0.03				0.02			0.02			0.88	0.05
St-Up		0.02	0.02			0.02		0.02				0.92

分别在 Hollywood 和 Weizmann 中测验,两组测验的识别率如图 6 所示。从图 6 中得出,本文的方法取得了较高的平均识别率,主要是本文算法利用了双特征来表征动作特征,利用两种特征的互补信息,提高了动作特征在分类学习了理解力。在 Weizmann 数据集的效果优于 Hollywood,说明了在越复杂的环境下,算法的识别率会受到影响。

表 3 所示为利用不同方法在 Weizmann 中得到结果。依据表 3 中得知,在这 10 个动作判别效果上,本文算法均取得了一定的优势。在 Jump、Wavl、Walk、Wave2 动作上识别率超过了 95%。而 A 方法、B 方法、C 方法对不同的动作的识别率波动较大,并且与提出的方法具有一定的差距。主要是本文方法分别利用了两种特征对动作特征提取,并通过 PCA 将提取的两种特征融合,通过优势互补,形成的一种新的特征。能够较好消除单一特征对

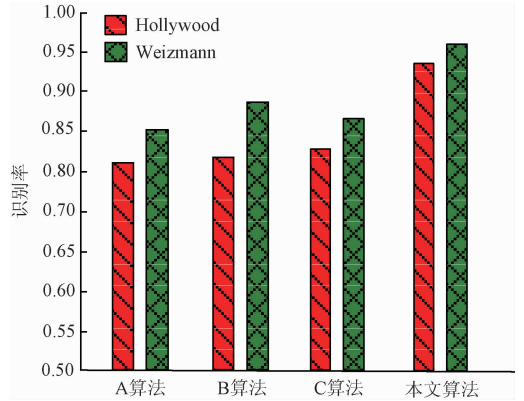


图 6 Hollywood 与 Weizmann 识别率测试

Fig. 6 Recognition rate test of Hollywood and Weizmann

动作描述不足,并通过自适应提升算法对特征进行分类训练,从而提高了复杂环境中动作的识别率。

表 3 Weizmann 中识别率比较  
Table 3 Comparison of recognition rates in Weizmann (%)

	Bend	Jack	Jump	P-jump	Run	Side	Skip	Walk	Wavl	Wave2
A 方法	79.36	78.54	83.14	79.45	83.74	77.84	81.25	82.63	77.54	82.95
B 方法	84.52	83.15	82.17	83.65	85.64	83.98	88.45	86.85	86.84	85.71
C 方法	86.21	86.52	86.31	87.06	88.94	86.52	89.54	86.15	88.20	84.58
本文方法	93.51	92.64	96.46	87.62	86.88	89.12	90.56	95.55	97.12	97.68

为了更好地描述算法的性能,对算法的查准率与查全率评估,得到了查准-查全率曲线,图 7 所示为在 Hollywood 数据集中利用 A 方法、B 方法、C 方法与本文方法得到的查准-查全曲线。从图 7 看到,本文方法的曲线相对比较平稳,且在同样的查全率下,本文方法的查准率相对最好。

根据上述实验结果可得出,在 Hollywood 和 Weizmann 数据集中,Weizmann 的识别率高于 Hollywood

数据集,主要是 Weizmann 中场景简单,干扰因素较少。而 Hollywood 中的场景复杂,受到光照、阴影等的影响,得到的正确率相对较低。与当前流行的其他方法相比较,本文算法具有更高的通查全率-查准率与识别率。相对于其他单一的动作特征,本文方法通过 STC 和 CNN 分别提取图像序列的动作特征,将两者动作特征通过 PCA 技术融合形成一个新的特征,且定义了自适应提升算法,对动作特征进行分类训练。而文献[4-6]3 种算法的动作



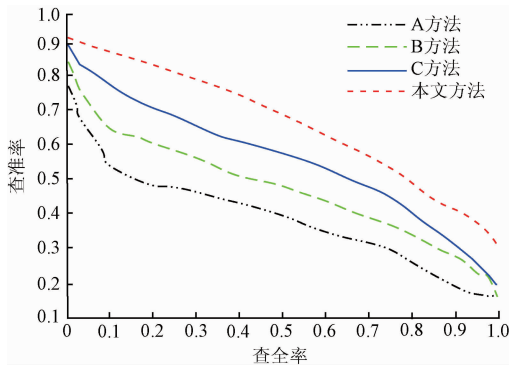


图7 查准-查全曲线

Fig.7 The curve of precision ratio-recall

识别准确度不高,鲁棒性较弱。主要是这些算法只通过单一的特征表征动作信息,在很多情况下受到各种环境影响,难以准确、完整表示动作信息,因此降低了动作识别精度。

## 5 结论

针对复杂场景中动作容易受到各种环境的干扰,设计了一种双特征融合与自适应提升的动作识别方案。为了降低光照变化、遮挡对动作的干扰,通过STC关系与视觉系统特性提取图像序列特征,得到了STC特征。同时,为了避免复杂图像的预处理和增加特征提取鲁棒性,利用具有平移不变性的CNN技术对图像序列操作,提取特征,获得CNN特征。为了更好地融合得到的两种特征,利用PCA算子对获得STC特征与CNN特征有效组合,形成一种更准确、完整的特征表示。根据新特征,利用ABA进行分类训练,完成对行为动作决策判断。实验表明,在Hollywood中的识别率为94.56%,在Weizmann中的识别率为96.13%。证明了本文方法能够有效对各种动作完成识别和理解,可降低光照变化、遮挡的干扰,在复杂场景中仍有较好表现,稳定性更佳。

## 参考文献

[1] 张飞燕,李俊峰,沈军民. 基于梯度和光流统计特性的人体行为识别[J]. 光电子·激光, 2015, 26(8): 1593-1601.  
ZHANG F Y, LI J F, SHEN J M. Human behavior recognition based on the statistical characteristics of gradient and optical flow[J]. Journal of Optoelectronics Laser, 2015, 26(8): 1593-1601.

[2] LI X X, CHEN SH X, ZHANG H SH. Towards reducing the impacts of unwanted movements on identification of motion intentions [J]. Electromyography Kinesiol, 2016, 28(5): 90-98.

[3] 于宗泽,方勇,李兆元. 基于压缩稀疏融合的动作识别方法[J]. 电子测量技术, 2016, 39(11): 155-159.  
YU Z Z, FANG Y, LI ZH Y. Motion recognition method based on compressed sparse fusion [J]. Electronic Measurement Technology, 2016, 39(11): 155-159.

[4] 秦磊,胡琼,黄庆明. 基于特征点轨迹的动作识别[J]. 计算机学报, 2014, 37(6): 1281-1288.  
QIN L, HU Q, HUANG Q M. Action recognition using trajectories of spatio-temporal feature points [J]. Chinese Journal of Computers, 2014, 37(6): 1281-1288.

[5] WANG H, KLASER A. Action recognition by dense trajectories [J]. Computer Vision and Pattern Recognition, 2013, 62(11): 3169-3176.

[6] 林水强,吴亚东,余芳. 姿势序列有限状态机动作识别方法[J]. 计算机辅助设计与图形学学报, 2014, 26(9): 1403-1411.  
LIN SH Q, WU Y D, YU F. Posture sequence finite-state machine method for motion recognition [J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(9): 1403-1411.

[7] 刘万军,董帅含,曲海成. 时空上下文抗遮挡视觉跟踪[J]. 中国图象图形学报, 2016, 21(8): 1057-1062.  
LIU W J, DONG SH H, QU H C. Anti-occlusion visual tracking algorithm based on spatial Temporal context learning [J]. Journal of Image and Graphics, 2016, 21(8): 1057-1062.

[8] 徐建强,陆耀. 一种基于加权时空上下文的鲁棒视觉跟踪算法[J]. 自动化学报, 2015, 41(11): 1901-1912.  
XU J Q, LU Y. Robust visual tracking via weighted spatio-temporal context learning [J]. Acta Automatica Sinica, 2015, 41(11): 1901-1912.

[9] 余永维,杜柳青,曾翠兰. 基于深度学习特征匹配的铸件微小缺陷自动定位方法[J]. 仪器仪表学报, 2016, 37(6): 1364-1370.  
YU Y W, DU L Q, ZENG C L. Automatic location method of small defects in castings based on depth learning feature matching [J]. Chinese Journal of Instrument and Meters, 2016, 37(6): 1364-1370.

[10] 曲景影,孙显,高鑫. 基于CNN模型的高分辨率遥感图像目标识别[J]. 国外电子测量技术, 2016, 12(8): 45-50.  
QU J Y, SUN X, GAO X. Target recognition of high resolution remote sensing images based on CNN model [J]. Foreign Electronic Measurement Technology, 2016, 12(8): 45-50.

[11] 霍雷刚,冯象初. 基于主成分分析和字典学习的高光谱遥感图像去噪方法[J]. 电子与信息学报, 2014,

36(11):2723-2729.

HUO L G, FENG X CH. Denoising of hyperspectral remote sensing image based on principal component analysis and dictionary learning [J]. Journal of Electronics & Information Technology, 2014, 36(11): 2723-2729.

- [12] 杨文文, 毛建旭. 基于分块 LBP 和分块 PCA 的指静脉识别方法[J]. 电子测量与仪器学报, 2016, 30(7): 1000-1007.

YANG W W, MAO J X. Finger vein recognition method based on block LBP and block PCA [J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(7): 1000-1007.

- [13] LIN ZH H, WANG L L, GAO J G. Interpretable functional principal component analysis [J]. Biometrics, 2016, 72(3): 846-854.

- [14] 付忠良. 多标签代价敏感分类集成学习算法[J]. 自动化学报, 2014, 40(6): 1075-1085.

FU ZH L. Ensemble learning algorithm for multi label cost sensitive classification [J]. Chinese Journal of Automation, 2014, 40(6): 1075-1085.

- [15] 徐勤军, 吴镇扬. 视频序列中的行为识别研究进展[J]. 电子测量与仪器学报, 2014, 28(4): 343-351.

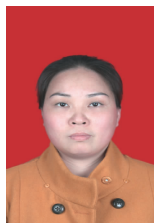
XU Q J, WU ZH Y. Research progress on activity recognition in video [J]. Journal of Electronic Measurement and Instrumentation, 2014, 28(4): 343-351.

- [16] 田国会, 尹建芹, 韩旭. 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.

TIAN G H, YIN J Q, HAN X. A novel human activity recognition method using joint points information [J]. Robot, 2014, 36(3): 285-292.

- [17] RAO L K, RAO D V. Local quantized extreme patterns for content based natural and texture image Retrieval [J]. Human-Centric Computing and Information Sciences, 2015, 26(5): 1-24.

## 作者简介



罗冬梅, 1975 年出生, 1999 年于湖北师范学院获得学士学位, 2008 年于福州大学获得硕士学位。现为武夷学院讲师, 主要研究方向为图像处理、人工智能与模式识别、数据挖掘。

E-mail: luodnGmei1975wy@sohu.com

**Luo Dongmei** was born in 1975, received B. Sc. from Hubei Normal University in 1999, and M. Sc. from Fuzhou University in 2008, respectively. Now she is a lecturer in Wuyi University. Her main research interests include image processing, artificial intelligence and pattern recognition, and data mining.



左金水, 1977 年出生, 2000 年于西安外国语大学获得学士学位, 分别在 2005、2010 年于日本筑波大学获得硕士学位和博士学位。现为浙江工商大学讲师, 主要研究方向为电子商务、数据挖掘。

**Zuo Jinshui** was born in 1977, received B. Sc. from Xi'an Foreign Language University in 2000, M. Sc. and Ph. D. from Tsukuba University, Japan, in 2005 and 2010, respectively. Now he is a lecturer in Zhejiang Gongshang University. His main research interests include image processing and data mining.



余文森, 1973 年出生, 1996 年于福建师范大学获得学士学位, 2004 年于江西师范大学获得硕士学位, 2013 年于四川大学获得博士学位。现为武夷学院副教授, 主要研究方向为图像处理、数据挖掘、机器学习。

**Yu Wensen** was born in 1973, received B. Sc. from Fujian Normal University in 1996, M. Sc. from Jiangxi Normal University in 2004, and Ph. D. from Sichuan University in 2013, respectively. Now he is an associate professor in Wuyi University. His main research interests include image processing, data mining, and machine learning