

文章编号 1004-924X(2018)11-2827-11

采用 PHOG 融合特征和 多类别 Adaboost 分类器的行为识别

马世伟¹, 刘丽娜^{1,2*}, 傅琪¹, 温加睿¹

(1. 上海大学 机电工程与自动化学院, 上海 210072;

2. 山东理工大学 电气与电子工程学院, 山东 淄博 255049)

摘要:为了解决类能量图易受人体运动时间和位置移动等因素影响而难以有效描述动作细节特征的问题, 本文提出了一种基于类能量图金字塔梯度直方图(PHOG)融合特征和多类别 Adaboost 分类器的人体行为识别方法。该方法首先对经过躯体配准的运动人体目标轮廓图像构造平均运动能量图(AMEI)和增强的运动能量图(EMEI), 分别提取其分层梯度方向直方图(PHOG)特征并进行串联融合, 作为一种多层次的行为特征描述; 然后使用基于查找表的 LUT-Real Adaboost 算法设计多类别分类器, 实现图像中人体行为动作的识别。实验结果显示其在典型的人体动作数据集 DHA 上的正确识别率达 97.6%, 高于其它采用单一特征描述和 SVM 等分类器的方法。表明该方法将整体与局部特征相结合, 可以有效描述不同尺度下的动作细节特征, 增强了人体行为特征的描述能力, 提高了识别性能。

关键词:人体行为识别; 平均运动能量图; 增强运动能量图; 分层梯度方向直方图特征; 查找表型 Real Adaboost
中图分类号: TP391.4 文献标识码: A doi:10.3788/OPE.20182611.2827

Using PHOG fusion features and multi-class Adaboost classifier for human behavior recognition

MA Shi-wei¹, LIU Li-na^{1,2*}, FU Qi¹, WEN Jia-rui¹

(1. School of Mechatronic Engineering and Automation,
Shanghai University, Shanghai 210072, China;

2. School of Electrical and Electronic Engineering,
Shandong University of Technology, Zibo 255049, China)

* Corresponding author, E-mail: linaliu-126@163.com

Abstract: In order to solve the problem that energy image species (EIS) are susceptible to human movement time and position shift, i. e., it is difficult to describe the details of human behaviors, in this paper a method of human behavior recognition was present based on pyramid gradient histogram (PHOG) fusion features and a multi-class Adaboost classifier. This method first calculated the average motion energy image (AMEI) and the enhanced motion energy image (EMEI) of an object's silhouette images after human body registration, and then it extracted the PHOG features of AMEI and EMEI and series them together to form a kind of multi-level feature descriptor of human behavior.

收稿日期: 2017-12-11; 修订日期: 2018-02-11.

基金项目: 国家自然科学基金资助项目(No. 61671285); 山东省自然科学基金资助项目(No. ZR2016FP04)

Finally, a look-up table-based real Adaboost (LUT-Real Adaboost) algorithm was utilized to realize human behavior recognition by designing a multi-class classifier. Experimental results show that the correct recognition rate in typical depth-included human action datasets is 97.6% by using this method, which is higher than that of other classifiers using single feature description and support vector machine. This reveals that, by combining global and local features, the proposed method can effectively describe the detailed active features of human behavior at different scales, enhance the description ability of human behavior characteristics, and improve recognition performance.

Key words: human behavior recognition; Average Motion Energy Image(AMEI); Enhanced Motion Energy Image(EMEI); Pyramid Histogram of Oriented Gradients(PHOG); Look-Up-Table type Real Adaboost(LUT-Real Adaboost)

1 引言

基于计算机视觉的人体行为识别技术在智能监控、虚拟现实、人机交互等领域有广阔应用前景,但由于个体差异以及环境变化(如姿态、运动、相机移动、光照条件、遮挡)等因素所产生的大量变量,使得识别问题变得更加困难^[1]。如何有效提高人体行为识别技术的准确性和稳健性仍然是具有挑战性的难题。人体行为识别过程一般包含行为特征的选择提取和行为模式的分类两部分内容。

特征选取方法包括基于局部特征的方法和基于全局特征的方法。基于局部特征的方法有人体外观表示、空时兴趣点和空时体体积等,它们都需要从视频序列中提取表示局部特征的时间序列,其计算量大,且易受噪声等干扰的影响^[1]。基于全局特征的方法则将观测到的人体区域进行整体运算,以得到行为特征的描述符,其观测量主要有轮廓剪影、骨架、光流和类能量图等^[2-5]。其中,基于类能量图(Energy Image Species, EIS)的全局特征描述法不易受到轮廓噪声的干扰,已得到了广泛的关注和应用。但是,类能量图在描述人体行为细节特征方面存在不足,且易受人体运动时间和位置移动等因素的影响^[2,4]。梯度方向直方图(Histogram of Orientated Gradients, HOG)是一种描述图像形状信息的特征描述符,在图像分类和运动目标检测等领域得到了广泛应用^[6-7]。金字塔梯度直方图(Pyramid Histogram of Oriented Gradients, PHOG)则是在 HOG 特征基础上,进行多次分层并进行直方图串联,使得特征可以描述不同尺度下的形状细节^[8-9]。Liu L N,

et al^[10]提出利用增强的运动能量图和金字塔 HOG 对人体行为特征进行描述,并利用 SVM 进行人体行为识别。SHEN X X, *et al*^[11]则构建了一种新的差分历史图,并计算其轴测投影以获取行为动作的运动过程,进而提取投影的 PHOG 特征进行人体行为识别。文献[12-14]则分别将金字塔 HOG 特征用于解决人脸识别问题。其中,文献[12]首先提取 PHOG 特征对人脸表情进行描述,接着在线学习一个关于人脸表情的字典,最后用基于稀疏表示的分类方法对人脸表情进行识别;文献[13]提出了一种基于多尺度金字塔特征块提取 HOG 特征的新型人脸识别算法;文献[14]则通过多尺度分析和 HOG 特征构建人脸图像的 HOG 金字塔来实现整个人脸的特征表达,进而计算不同序列特征的相似度采用最近邻分类器进行人脸识别。

在行为模式的分类器设计方面,常用 KNN 方法和基于不同核函数的 SVM 方法。但是, KNN 方法对特征提取的要求较高,而 SVM 则针对小样本问题,且需要对核函数进行选择。Adaboost 算法^[15]是一种自适应算法,其构成强分类器时的组合规则可根据问题自适应调整,具有从现有数据分布中筛选最优特征甚至学习特征的能力,在数据量足够的前提下,可以拟合出复杂的非线性分类超平面。Real Adaboost 算法^[16]是 Adaboost 的改进算法之一,可对若干连续置信度输出的弱分类器进行处理,输出不同类别的概率分布。基于查找表的 LUT-Real Adaboost^[17]是一种直方图分类器,通过建立概率查找表产生实数输出,能够模拟多种形式的概率分布,具有较好的抗过拟合能力,且收敛速度快,因而在人体行为识别中具有重要应用价值,但它仍然是典型的二

分类模型,不能直接处理多分类问题。

针对上述问题,本文提出了一种基于类能量图和 PHOG 融合特征的多层次人体行为特征描述,并设计出多类别的 LUT-Real Adaboost 分类器,从而实现人体行为识别。在典型的 Weizmann 和 DHA 数据集上对本文所提出的特征描述和分类器设计方法进行了实验,验证了其有效性。

2 本文方法整体框架

本文方法整体框架如图 1 所示,主要由目标轮廓图像提取和躯体配准处理、人体行为特征提取、行为识别等部分组成。

把人体行为的动作过程视作一个整体运动对象,将所有轮廓按照某种规则累计,就可以构造出类能量图。在计算类能量图之前,对目标轮廓图像进行配准处理可以有效减少运动位置和重心移动的影响。本文采用了背景减除法和质心配准方法,用于对视频序列中运动人体目标检测获得目标轮廓图像,并进行躯体配准处理。

在此基础上,求得平均运动能量图(Average Motion Energy Image, AMEI),用于统计视频序列中每个像素的灰度均值^[18]。同时,为了能够凸显人体运动过程中的动态部分,对类能量图中动作频率较大的部分进行增强并消除相对静止的区域,本方法中进一步构建了增强的运动能量图(Enhanced Motion Energy Image, EMEI)。AMEI

重点描述静态的躯体信息,而 EMEI 重点描述动态的肢体信息。然后,分别提取 AMEI 和 EMEI 各自的 PHOG 特征并直接进行串联融合,形成一种多层次的人体行为特征描述,可将整体与局部特征相结合,进一步增强行为特征的描述能力。

最后,设计多类别的 LUT-Real Adaboost 分类器,可以从高维特征中选择表达能力较好的特征,且具有较好的抗过拟合能力。通过进行分类器训练和分类器测试,实现人体行为动作的识别。多类别分类器的设计策略包括一对一投票和一对多投票。在一对多策略中,若正负样本数量差别较大,其分类效果往往不如一对一策略,因此,本文方法中采用了一对一投票策略。

3 行为特征提取

如图 1 所示,在人体行为特征描述和提取过程中,本文方法首先对经过躯体配准处理后的归一化的 AMEIs 和 EMEIs 图像分别进行四叉树结构的金字塔分解,然后分别提取他们的各层次上各子区域的金字塔模型梯度直方图 PHOG 特征,并将其直接串联构成融合的 PHOG 特征。该方法充分结合了 AMEI 和 EMEI 的优点,既描述静态的躯体信息,又描述动态的肢体信息,形成一种多层次的人体行为特征描述,可将整体与局部特征相结合,进一步增强行为特征的描述能力。下面分别给予介绍。

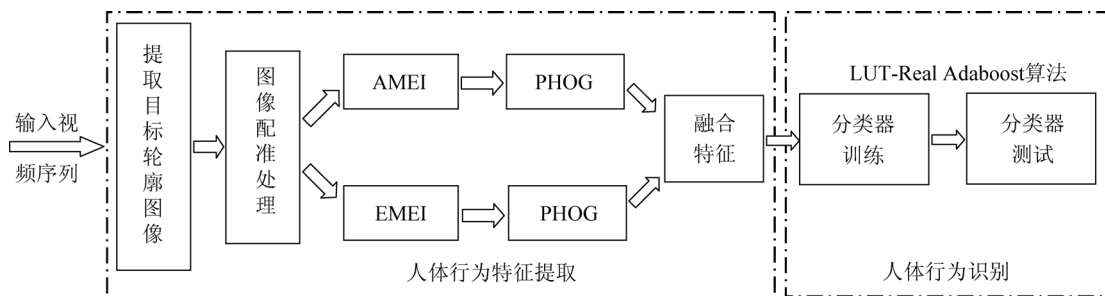


图 1 人体行为识别方法框架

Fig. 1 Framework of human behavior recognition method

3.1 目标轮廓提取和躯体配准处理

目标轮廓提取是动作识别的基础,本文利用常用的背景减除法进行视频序列中运动人体目标的检测。为了构造背景模型,假定背景静止,采用平均背景法,通过记录连续多帧灰度图像的平均

值作为背景帧图像,最后通过形态学处理得到二值化轮廓图像。视频序列中运动人体目标的位置和躯体重心会在图像中发生偏移,如果直接从轮廓图像序列构造类能量图,则容易造成拖影等问题^[4]。对此,常规处理方法都是先提取人体轮廓的

外接矩形框,即感兴趣区域(Region of Interest, ROI),然后将框内图像缩放为标准大小,再对标准图像序列生成类能量图。

但是这种方法在缩放阶段需要对每幅图像运用插值等算法进行标准化处理,且对于动作幅度或身体形变较大的动作,如弯腰、伸展等,需要预先估算不同动作的大致形状,在每帧图片中提取出一个足够大的区域,因而难以适应动作类型较多与图像分辨率变化的情况。为此,本文采取的解决办法是先进行基于质心配准的图像配准处理,再合成类能量图,最后再提取其轮廓 ROI,并一次性标准化图像,以减少计算量和提高配准效果。具体过程为:

(1)采用背景减除法获取视频序列中每帧图像的目标轮廓图像,计算原始图像区域中的人体前景轮廓的质心;

(2)将其质心移动到整幅画面的中心位置进行质心配准;

(3)利用配准后的图像序列构造类能量图;

(4)提取类能量图中人体轮廓区域的最小外接矩形作为 ROI,缩放为标准图像,进行后续的特征提取操作。

其中,为了解决 ROI 区域大小不同造成图像尺寸不一的问题,需要对提取的 ROI 区域进行标准化处理。为了处理方便,本文实验中将其统一缩放为 64×64 的标准图像。

3.2 构造 AMEI 和 EMEI

类能量图是全局动作描述方法,通过将一个动作周期内的轮廓图像进行组合,构成动作的模板。平均运动能量图 AMEI 可用于统计视频序列中每个像素的灰度均值,其计算公式为^[18]:

$$E_{AMEI}(x, y) = \frac{1}{N} \sum_{t=1}^N I(x, y, t), \quad (1)$$

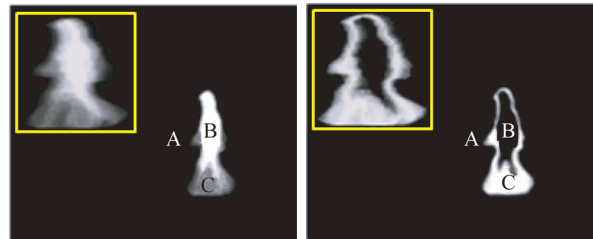
其中: N 为视频序列的长度, $I(x, y, t)$ 表示第 t 帧二值图像 I 中 (x, y) 处的像素值。一个人体行走动作的 AMEI 如图 2(a)所示,其中有 3 种不同亮度的区域:区域 A 为背景区域,需通过 ROI 去除;区域 B 亮度最高,代表了人体运动过程中相对静止的躯干区域;区域 C 灰度变化明显,代表了人体运动过程中肢体摆动频率较高、肢体摆幅较大的部分。

本文在 AMEI 的基础上进一步提取增强的

运动能量图 EMEI,以凸显运动过程中的动态部分,对类能量图中动作频率较大的部分进行增强,并消除相对静止的区域。EMEI 的计算公式为:

$$E_{EMEI}(x, y) = \frac{1}{N} \sum_{t=1}^N \| I(x, y, t) - E_{AMEI}(x, y) \|. \quad (2)$$

一个人体行走动作的 EMEI 如图 2(b)所示,可以看出 EMEI 与 AMEI 相比在区域 B 和区域 C 有非常明显的差别:较亮区域 C 为移动较多区域,其白色轮廓线内部的黑色区域 B 为运动过程中躯干相对静止的部分。可见 EMEI 重点突出了肢体在一个运动周期内的变化范围。在图 2 中,左上角 ROI 区域内为标准化的图像。

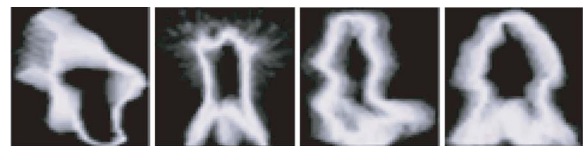


(a)平均运动能量图 (a)AMEI (b)增强的运动能量图 (b)EMEI

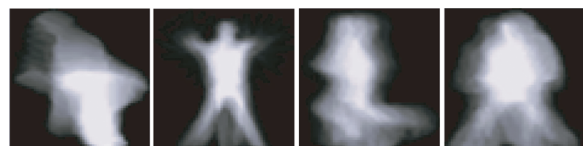
图 2 人体行走动作的 AMEI 和 EMEI

Fig. 2 AMEI and EMEI of the walk behavior

图 3 展示了 Weizmann 数据集中多个动作的 AMEIs 和 EMEIs,从左到右依次为弯腰、开合跳、跑步和侧步走。



(a)平均运动能量图 (a)AMEIs



(b)增强的运动能量图 (b)EMEIs

图 3 Weizmann 数据集中多个动作的 AMEIs 和 EMEIs

Fig. 3 AMEIs and EMEIs of behaviors in Weizmann datasets

由图可见,AMEIs 和 EMEIs 分别能够很好地描述原始图像序列的运动特征,前者重点描述静态的躯体信息,后者重点描述动态的肢体信息。分析 AMEI 和 EMEI 的计算过程可知,二者对时间长度、动作周期个数不敏感,具有很好的时间不变性。因此,本文方法结合了 AMEI 和 EMEI 进行动作描述,不但能克服传统类能量图描述法易受运动时间影响的缺点的,还能够清晰地反映人体动作的静态与动态信息。

3.3 PHOG 特征提取及融合

HOG 特征是描述图像形状信息的特征描述符,PHOG 是在 HOG 特征的基础上进行多次分层,使得特征可以从多个尺度来描述图像的纹理和形状信息,因而是一种更为高效的边缘梯度描述方法。类能量图中 HOG 特征的提取过程为:

(1) 计算类能量图中各个像素点的梯度方向和梯度幅值;

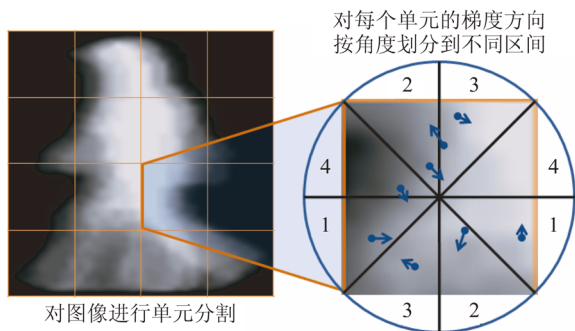
(2) 将图像按横向和纵向划分为若干个单元,分别用 c_H 和 c_W 表示,则总单元数为 $c_H \times c_W$ 。本文实验中的划分模式如图 4(a) 所示,每个类能量图被划分为 16 个单元;

(3) 将每个单元按梯度方向等角度划分为 bins 个区间,取梯度方向在各区间上的投影,则每个单元得到一个长度为 bins 的直方图向量。例如图 4(a) 所示,每个单元划分为 4 个 bins;

(4) 将相邻单元组合成块,块与块之间可以存在单元的重叠。分别用 B_H 和 B_W 表示块的高度和宽度,本文采用的块组合模式如图 4(b) 所示;

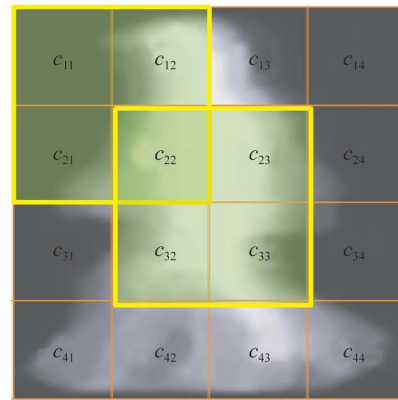
(5) 将所有块中的直方图向量进行串联就构成了整个 HOG 特征向量,其维数为:

$$\dim_{HOG} = bins \times B_H \times B_W \times (c_H - B_H + 1) \times (c_H - B_H + 1). \quad (3)$$



(a) 划分单元与梯度方向区间

(a)Partitions of cell and gradient direction



(b) 将相邻的单元组合成块

(b)Combine adjacent cells into blocks

图 4 HOG 特征划分示意图

Fig. 4 Schematic diagram of HOG feature division

PHOG 特征提取时根据二叉树结构对图像进行多层分割。在二叉树中,每一层在上一层分割得到单元的基础上,再次分割出 4 个大小相等的子单元。分别在不同分割层次的图片上提取 HOG 特征,最后将每一层的 HOG 串联成为一个维度更高的直方图,作为图像在多尺度下的 PHOG 特征。分别提取 AMEIs 和 EMEIs 各自的 PHOG 特征,并直接进行串联融合,就形成了一种多层次的人体行为特征描述。以一个人体行走动作的 AMEI 为例,其 PHOG 特征提取过程示意图如图 5 所示。

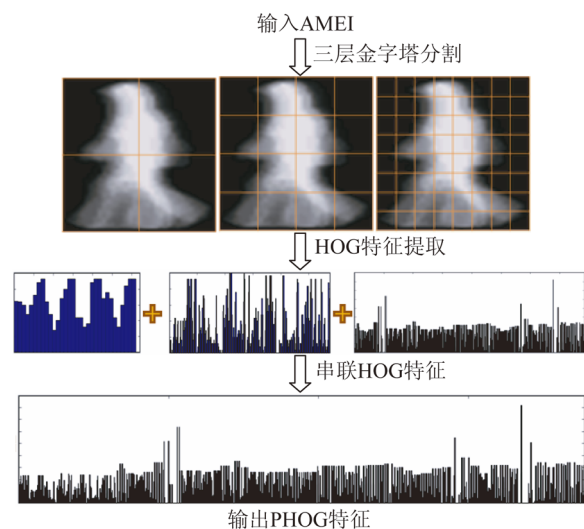


图 5 AMEI 的 PHOG 特征提取过程

Fig. 5 Process of PHOG feature extraction of AMEI

4 行为识别策略

如图 1 框架所示,本文采用具有连续概率输出的查找表作为 Real Adaboost 的弱分类器(LUT-Real Adaboost),设计多类别分类器。其中,Real Adaboost 是一种通过前向分布算法进行学习的二分类机器学习算法,其损失函数为指数函数,算法流程如表 1 所示,可将样本空间 χ 映射到实数空间 \mathbf{R} ,通过将不同弱分类器的置信度求和,得出未知样本的联合置信度。LUT 型弱分类器的设计过程如表 2 所示,它通过对样本数据进行维度选择与样本空间投影,根据样本的分布情况建立概率密度查找表,从而计算在当前维度下样本为正样本的概率。

表 1 Real Adaboost 算法流程

Tab.1 Algorithm flow of Real Adaboost

输入:训练样本 $\mathbf{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$, 其中 $(x_i, y_i) \in \chi \times \{1, -1\}$; 弱分类器池 L ; 弱分类器数目 M ;

输出:强分类器 $H(x)$;

(1) 初始化训练集各样本权重: $\omega_i = 1/N, i = 1, 2, \dots, N$,

$\mathbf{W}_1 = \{\omega_1, \omega_2, \dots, \omega_N\}$

(2) for $m = 1, \dots, M$

a) 训练弱分类器 $h_m = L(\mathbf{D}, \mathbf{W}_m)$, 使得弱分类器输出:

$$p_m(x) = \hat{P}_m(y=1|x) \in [0, 1];$$

弱分类器预测正样本的置信度为:

$$h_m(x) = \frac{1}{2} \log(p_m(x)/(1-p_m(x))) \in \mathbf{R};$$

b) 更新样本分布权重参数:

$$\mathbf{W}_{m+1}(i) = \mathbf{W}_m(i) \exp(-y_i h_m(x_i)),$$

并归一化,使得 $\sum_{i=1}^N \mathbf{W}_{m+1}(i) := 1 \sum_{i=1}^N \mathbf{W}_{m-1}(i)$

End for

(3) 输出强分类器:

$$H(x) = \text{sign}\left(\sum_{m=1}^M h_m(x) - b\right);$$

其中: b 为阈值,其默认值为 0; 强分类器的置信度为:

$$\text{conf}_H(x) = \left| \sum_{m=1}^M h_m(x) - b \right|.$$

但是,由于 SVM 和 Real Adaboost 等本质上都是二分类模型,不能直接处理类别较多的分类问题。因此需要对二分类模型进行推广,得到多类别的分类器。其设计策略一般包括一对一投票

和一对多最大响应投票,在一对多策略中,若正负样本数量差别较大,其分类效果往往不如一对一策略,因此,本文采用如下的一对一投票策略。从训练样本中取出一类作为正样本,再从剩下样本中选出一类作为负样本,构成一个训练集。

表 2 LUT 型弱分类器的设计过程

Tab.2 Design process of LUT type weak classifier

输入:训练样本空间 χ , 第 m 次迭代的样本权重分布: $\mathbf{W}_m = \{\mathbf{W}_1, \dots, \mathbf{W}_N\}$;

输出:第 m 次迭代的弱分类器 $h_m(x)$;

(1) 对 L 中的每个弱分类器 h 作如下操作:

a. 将样本空间 χ 划分为 n 个不相交的子空间 $\{X_j\}, j = 1, \dots, n$;

b. 在样本分布 \mathbf{W}_m 下,计算:

$$\mathbf{W}_i^l = P(x_i \in X_j, y_i = l) = \sum_{i: x_i \in X_j \wedge y_i = l} \mathbf{W}_m(i), l = \pm 1.$$

c. 设置弱分类器在每个子空间 X_j 上的输出:

$$\forall x \in X_j, h(x) = \frac{1}{2} \ln\left(\frac{\mathbf{W}_{l_1}^j + \epsilon}{\mathbf{W}_{l_2}^j + \epsilon}\right).$$

其中:为了避免分母为 0,设置平滑因子 ϵ 为一较小正数。

d. 计算评估系数:

$$Z = 2 \sum_j \sqrt{\mathbf{W}_{l_1}^j \mathbf{W}_{l_2}^j};$$

(2) 选择 h_m 使得 Z 最小化,即 $Z_m = \min_{h \in L} Z$, 则相应的弱分类器为: $h_m = \underset{h \in L}{\text{argmin}} Z$.

设待训练样本一共有 N 类,则一共可以构造 $N(N-1)/2$ 个独立的训练样本集,通过训练可得 $N(N-1)/2$ 个分类器。每一类对应一个投票得分 vote ,对于由第 i 和第 j 类样本得到的分类器,输入一个测试样本,如果被判定类标签属于第 i 类,则该类投票得分 $\text{vote}(i) = \text{vote}(i) + 1$ 。测试样本被分别输入上述分类器,从投票得分中选取最大值,则该得分对应的类标号就是测试样本被判定的类别。

利用上述的多类别 LUT-Real Adaboost 分类器,可以从高维特征中选择表达能力较好的特征,且具有较好的抗过拟合能力。最后,通过进行分类器训练和分类器测试,实现人体行为动作的识别。

5 实验及结果分析

为了验证本文方法的有效性,在两个经典的人体运动数据库 Weizmann 和 DHA^[10] 上做了仿

真实实验。为了便于与其他同类文献进行对比,本文采用留一法(将原始数据的每一个样本作为测试集,其余所有样本作为训练集)进行交叉验证,并取每次测试准确率的均值作为最终标准,对所提算法的性能进行评估。

5.1 参数设置

在行为识别实验中,需要设置的模型参数包括 PHOG 特征参数和分类器参数。其中,PHOG 特征参数包括 HOG 特征的相关参数和金字塔的结构层次,二者决定了 PHOG 特征向量的维数;分类器参数包括弱分类器数目和样本空间 χ 的划分值 n 。

(1) PHOG 特征参数的设置

本文先使用训练速度较快的线性核 SVM 分类器在规模较小的 Weizmann 数据集上对特征参数进行初步筛选实验,然后依据实验结果确定 DHA 数据集的特征参数。这样一方面能够验证本文特征是否有效,另一方面可排除一些效果较差的参数以简化实验。以四叉树金字塔结构分层的 PHOG 参数设置和特征维数见表 3 所示:金字塔采用三层结构,各层的 HOG 划分如表中左侧所示;抽取金字塔第三层特征作为 HOG 特征,并将每个块按照 $0^\circ \sim 180^\circ$ 划分为 $bins$ 个区间;不同 $bins$ 取值对应的 HOG 和 PHOG 特征维数如表中右侧所示。

表 3 PHOG 特征参数设置

Tab. 3 Parameter settings for PHOG features

金字塔层数	PHOG 特征参数			特征维数	
	单元	块中单元数	Bins 数	HOG	PHOG
1	4	2×2	4	1 568	1 888
2	16	3×3	8	3 126	3 776
3	64	2×2	9	3 528	4 248

筛选实验结果显示,在 Weizmann 数据集上,所有的 PHOG 特征参数在 SVM 分类器下的正确识别率均达到了 100%,表明了采用 PHOG 特征描述人体行为是有效的。但是在 DHA 数据集上,不同的 PHOG 特征参数表现出了不同的识别性能,其中 $bins$ 为 4 和 8 时的识别率分别为 95.5% 和 95.8%。因此,本文实验中选择 $bins$ 为 8 的 PHOG 特征。

(2) 分类器参数的设置

采用多类别 LUT-Real Adaboost 分类器进行人体行为动作的识别实验,需要设置分类器参数。由表 3 可知, $bins$ 为 8 时的 PHOG 特征维数为 3 776。由于 DHA 数据集包含 17 类动作,则基于 LUT-Real Adaboost 的多类别分类器将包含 $17 \times 16/2 = 136$ 个二分类器,若将弱分类器数目设置为 8 000,则在学习过程中需要从 3 776 维特征中进行重复选择,每次训练的特征选择次数多达 4 108 288 000 次,耗时较长且存在过拟合现象。

为了解决此问题,本文实验中将 DHA 数据集中每个类别的样本按照 2:1 的比例随机分为训练集与测试集,不作分类器数量上的交叉验证,而采用数量较多的弱分类器进行训练,通过迭代次数—识别率曲线来分析分类器的抗过拟合能力,并选取最优弱分类器的个数。图 6 所示为 LUT-Real Adaboost 分类器在 DHA 数据集上的错误识别率变化曲线。由图可见,随着分类器数目增加,错误识别率快速下降,当分类器数目达 5 284 时,错误识别率最低为 2.4%,随着分类器数目的进一步增加,错误识别率趋于稳定,表明算法具有良好的抗过拟合能力。因此,下面的实验中设置弱分类器的数目为 6 000,将样本空间 χ 划分为 64 个子空间。

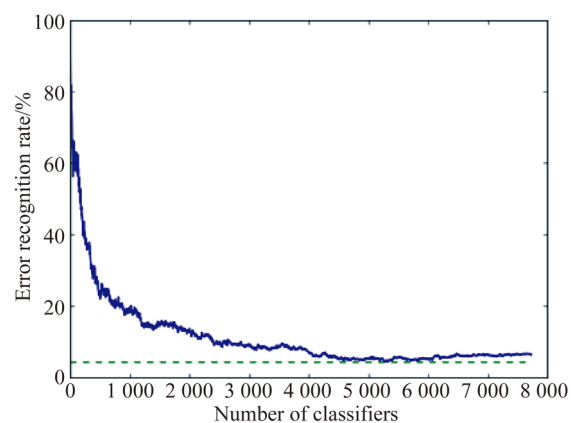


图 6 分类器在 DHA 数据集上的错误识别率变化曲线

Fig. 6 Error rate change curve of classifier on DHA datasets

5.2 不同数据集上的性能对比

(1) Weizmann 数据集实验

在该数据集上,将本文提出的融合特征和分

类器设计方法进行实验得到的正确识别率,与其它类能量图加单一特征方法如 HOG、3DHOG、归一化 R 变换等^[10, 20-23]以及线性核 SVM 和 Adaboost 等分类器组合得到的结果进行了对比,如表 4 所示。由表可见,基于 AMEI 和 EMEI 的 PHOG 融合特征的正确识别率达到 100%,性能优于其它基于单一特征的方法,且多类别 LUT-Real Adaboost 分类器的效果比 Adaboost、KNN 和 K-means 要好。

表 4 Weizmann 数据集上的性能比较

Tab. 4 Performance comparison on Weizmann datasets

类能量图+特征提取	分类器	正确识别率/%
3DHOG+PCA ^[20]	SVM	90
三维频域滤波+PCA ^[21]	SVM	97
EGEI+2DLPP ^[22]	KNN	91.2
STIG+PHOG+PHOW ^[23]	Adaboost	94.75
MEI+归一化 R ^[24]	K-means	92.47
AMEI+PHOG ^[10]	SVM	96.3
MEI+PHOG ^[10]	SVM	95.1
EMEI+PHOG	SVM	93.8
AMEI+EMEI+HOG	SVM	100
AMEI+EMEI+PHOG	SVM	100
AMEI+EMEI+PHOG LUT-Real Adaboost		100

(2)DHA 数据集实验

在该数据集上,将本文算法分别与文献[8]提出的 DDHMI、DLRDMHI 和文献[25]提出的 DDHMI+ HBM+ GIST,以及文献[10]提出的 AMEI、MEI、MHI 和 MEI 方法进行了对比实验,性能比较结果如表 5 所示。由表可见,本文提出基于 AMEI 和 EMEI 的 PHOG 融合特征的识别性能要优于其它方法,且使用多类别 LUT-Real Adaboost 分类器的识别率达到 97.6%,高于其它使用 SVM 分类器的结果。

综合分析上述结果,与其他采用类能量图进行全局动作描述的行为识别方法相比,本文算法采用 AMEI 和 EMEI 两种类能量图的 PHOG 融合特征来描述人体全局动作,将两种类能量图的优点充分结合,形成了多层次的人体行为特征描述符,增强了动作的描述能力,使得识别率有较大地提升。

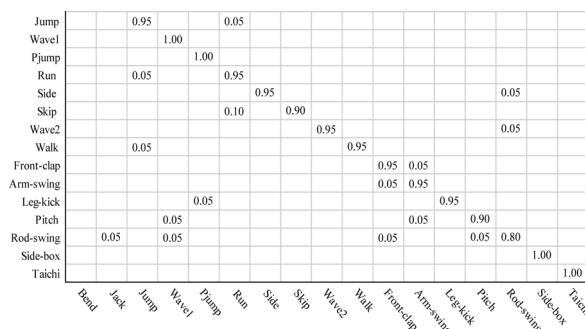
表 5 DHA 数据集上的性能比较

Tab. 5 Performance comparison on DHA datasets

类能量图+特征提取	分类器	正确识别率/%
DDHMI + PHOG ^[8]	SVM	92.4
DLRDMHI+ PHOG ^[8]	SVM	92.7
DDHMI+HBM+ GIST ^[25]	SVM	89.0
AMEI+PHOG ^[10]	SVM	94.2
MEI+PHOG ^[10]	SVM	92.9
MEI+PHOG ^[10]	SVM	89.0
MHI+PHOG ^[10]	SVM	86.7
AMEI+EMEI+ PHOG	SVM	95.8
AMEI+EMEI+ PHOG LUT-Real Adaboost		97.6

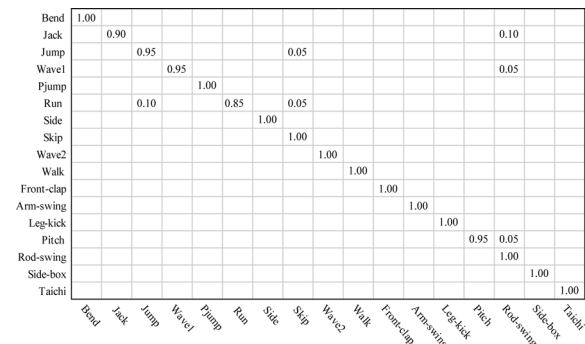
5.3 混淆矩阵分析

相比 Weizmann 数据集,具有挑战性的 DHA 数据集中包含更为丰富的动作类型(共 17 种),造成有误差分类情况。为了进一步分析本文方法在 DHA 数据集中对于不同动作的分类情况,分别使用 SVM 和多类别 LUT-Real Adaboost 分类器,利用混淆矩阵进行了相关分析,具体对比图 7 (a)和 7(b)可见:



(a)使用 SVM 的混淆矩阵

(a)Confusion matrix for SVM



(b)使用 LUT-Real Adaboost 的混淆矩阵

(b)Confusion matrix for LUT-Real Adaboost

图 7 DHA 数据集上的混淆矩阵

Fig. 7 Confusion matrices of DHA datasets

(1)使用 SVM 仅有 6 类动作得到了正确识别,识别率最低的是 rod-swing 动作,而使用 LUT-Real Adaboost 则有 12 类动作得到了正确识别,包括 rod-swing 动作;

(2) run、side、skip、jump、walk 等动作有一定相似性,使用 SVM 容易互相混淆,而使用 LUT-Real Adaboost 通过大量弱分类器进行学习可以达到较好的识别效果;

(3) front-clap 与 arm-swing 均为正面的手部动作,轮廓图之间差异不明显,使用 SVM 会相互混淆,而使用 LUT-Real Adaboost 可完全区别。

6 结 论

本文提出的采用 AMEI 和 EMEI 类能量图 PHOG 融合特征和多类别 LUT-Real Adaboost 分类器的图像中人体行为动作识别方法,具有以下特点:在构造 AMEI 和 EMEI 之前,采用背景减除和质心配准方案对目标轮廓图像进行躯体配

准处理,可有效减少运动位置和重心移动的影响,克服拖影现象;AMEI 重点描述静态的躯体信息,EMEI 重点描述动态的肢体信息,分别提取各自的 PHOG 特征并进行串联融合,可将整体与局部特征相结合,形成一种多层次的人体行为特征,增强了人体动作特征描述能力;采用一对一投票策略设计的多类别 LUT-Real Adaboost 分类器,可以从高维特征中选择表达能力较好的特征,且具有较好的抗过拟合能力。

在典型的人体动作数据库 Weizmann 和 DHA 上的仿真实验和交叉验证结果表明,本文方法的正确识别率高于其它采用单一特征描述和 SVM 等分类器的方法,验证了其有效性。同时,采用混淆矩阵法进行分析表明,对于一些相似度较高且存在运动方向和前后位置区别的动作,如 run、jack、jump 等容易出现错分,可能原因是这些动作都是腿部运动存在深度信息,而类能量图像不能很好的描述深度信息,即使采用 PHOG 融合特征亦难以区分这几类动作。因此,下一步将针对深度图像进行建模,进一步提升算法性能。

参考文献:

- [1] GUO G D, LAI A. A survey on still image based human action recognition [J]. *Pattern Recognition*, 2014, 47(10), 3343-3361.
- [2] 刘智,黄江涛,冯欣. 构建多尺度深度卷积神经网络行为识别模型 [J]. *光学精密工程*, 2017, 25(3):799-805.
LIU ZH, HUANG J T, FENG X. Action recognition model construction based on multi-scale deep convolution neural network [J]. *Opt. Precision Eng.*, 2017, 25(3):799-805. (in Chinese)
- [3] 张国梁,贾松敏,张祥银,等. 采用自适应变异粒子群优化 SVM 的行为识别 [J]. *光学精密工程*, 2017, 25(6):1669-1678.
ZHANG G L, JIA S M, ZHANG X Y, et al.. Action recognition based on adaptive mutation particle swarm optimization for SVM [J]. *Opt. Precision Eng.*, 2017, 25(6):1669-1678. (in Chinese)
- [4] 裴晓敏,范慧杰,唐延东. 时空特征融合深度学习网络人体行为识别方法 [J]. *红外与激光工程*, 2018, 47(2): 0203007.
PEI X M, FAN H J, TANG Y D. Action recognition method of spatio-temporal feature fusion deep learning network [J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203007. (in Chinese)
- [5] 李庆辉,李艾华,崔智高,等. 结合限制密集轨迹与时空共生特征的行为识别 [J]. *光学精密工程*, 2018, 26(1):230-237.
LI Q H, LI A H, CUI ZH G, et al.. Action recognition via restricted dense trajectories and spatio-temporal co-occurrence feature [J]. *Opt. Precision Eng.*, 2018, 26(1):230-237. (in Chinese)
- [6] SAIKA S, TAKAHASHI S, TAKEUCHI M, et al.. Accuracy improvement in human detection using HOG features on train-mounted camera [C]. *IEEE Global Conference on Consumer Electronics, IEEE*, 2016:1-2.
- [7] GAO H L, CHEN W J. Image Classification Based on the Fusion of Complementary Features [J]. *Journal of Beijing Institute of Technology*, 2017, 26(2):197-205.
- [8] 申晓霞,张桦,高赞,等. 基于 Kinect 和金字塔特征的行为识别算法 [J]. *光电子激光*, 2014, 25(2): 357-363.
SHEN X X, ZHANG H, GAO Z, et al.. Human behavior recognition based on Kinect and pyramid features [J]. *Journal of Optoelectronics · Laser*,

- 2014,25(2):357-363. (in Chinese)
- [9] 周英姿, 王正勇, 卿鄰波, 等. 基于局部块模型的复杂场景行为识别算法 [J]. 液晶与显示, 2017, 32(9):748-754.
ZHOU Y Z, WANG ZH Y, QING L B, *et al.*. Action recognition algorithm under complex scenes based on local part mode [J]. *Journal of Liquid Crystals and Display*, 2017, 32(9):748-754. (in Chinese)
- [10] LIU L N, WEN J R, MA S W, *et al.*. Human Behavior Recognition Method Based on Improved Energy Image Species and Pyramid HOG Feature [M]. Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems. Asia Sim 2016/ SCS Autumn Sim 2016, Part IV, CCIS 646, Singapore, 2016:216-2224.
- [11] SHEN X X, ZHANG H, GAO Z, *et al.*. Human behavior recognition based on axonometric projections and PHOG feature [J]. *Journal of Computational Information Systems*, 2014, 10(8):3455-3463.
- [12] WANG H, GAO J, TONG L, *et al.*. Facial recognition based on PHOG feature and sparse representation [C]. *Proceedings of the 35th Chinese Control Conference*. Chengdu, China, IEEE, 2016:3869-3874.
- [13] 张昊. 基于多尺度金字塔特征块提取 HOG 特征的新型人脸识别算法[D]. 长春: 吉林大学计算机科学与技术学院, 2017.
ZHANG H. A novel face recognition method using HOG features deriving from multi-layer pyramid feature blocks [D]. Changchun: School of Computer Science and Technology, Jilin University, 2017. (in Chinese)
- [14] 杨冰, 王小华, 杨鑫, 等. 基于 HOG 金字塔人脸识别方法 [J]. 浙江大学学报(工学版), 2014, 48(9):1564-1569.
YANG B, WANG X H, YANG X, *et al.*. Face recognition method based on HOG pyramid [J]. *Journal of Zhejiang University (Engineering Science)*, 2014, 48(9):1564-1569. (in Chinese)
- [15] 徐超, 高敏, 杨锁昌, 等. 视觉注意机制下的粒子窗快速目标检测 [J]. 光学 精密工程, 2015, 23(11): 3227-3237.
XU CH, GAO M, YANG S CH, *et al.*. Visual attention mechanism-aided fast target detection by particle window [J]. *Opt. Precision Eng.*, 2015, 23(11): 3227-3237. (in Chinese)
- [16] HUANG W B, WANG K, YAN Y. Automatic detection method of blood vessel for color retina fundus images [J]. *Opt. Precision Eng.*, 2017, 25(5): 1378-1386.
- [17] CHEN C Y, ZHANG P Z, LUO L M. Face detection using real Adaboost on granular features [J]. *Caai Transactions on Intelligent Systems*, 2009.
- [18] HAN J, BHANU B. Individual recognition using gait energy image [J]. *IEEE Transation on Pattern Analysis & Machine Intelligence*, 2006, 28(2):316-322.
- [19] LIN Y C, HU M C, CHENG W H, *et al.*. Human action recognition and retrieval using sole depth information [C]. *The 20th ACM International conference on Multimedia(MM'12)*, Nara, Japan, ACM, 2012:1053-1056.
- [20] KLASER A, MARSZALEK M. A spatio-temporal descriptor based on 3d-gradients [C]. *The 19th British Machine Vision Conference(BMVC 2008)*, Leeds, United Kingdom, British Machine Vision Association, BMVA, 2008, 9:1-4.
- [21] SOLMAZ B, ASSARI S M, SHAH M. Classifying web videos using a global video descriptor [J]. *Machine Vision and Applications*, 2013, 24(7): 1473-1485.
- [22] 林春丽, 王科俊, 李玥, 等. 基于增强能量图和二维保局映射的行为分类算法 [J]. 计算机应用, 2011, 31(3): 721-723.
YANG CH L, WANG K J, LI Y, *et al.*. Behavior classification algorithm based on enhanced gait energy image and two-dimensional locality preserving projection [J]. *Journal of Computer Applications*, 2011, 31(3): 721-723. (in Chinese)
- [23] 杨丽召. 基于多特征融合的行为识别算法研究 [D]. 成都: 电子科技大学计算机科学与工程学院, 2013.
YANG L ZH. A research of Behavior Recognition Algorithms Based on Multi-features Fusion [D]. Chengdou: School of Computer Science and Engineering, University of Electronic Science and Technology of China, 2013. (in Chinese)
- [24] 欧阳寒, 范勇, 高琳, 等. 基于归一化 R 变换分层模型的人体行为识别 [J]. 计算机工程与设计, 2013, 34(6): 2170-2174.
OUYANG H, FAN Y, GAO L, *et al.*. Hierarchical human action recognition based on normalized R-transform [J]. *Computer Engineering and*

Design, 2013, 34(6): 2170-2174. (in Chinese)

- [25] 申晓霞, 张桦, 高赞, 等. 一种鲁棒的基于深度数据的行为识别算法 [J]. *光电子激光*, 2013(8): 1613-1618.

SHEN X X, ZHANG H, GAO Z, *et al.*. A robust behavior recognition algorithm based on sole depth information [J]. *Journal of Optoelectronics • Laser*, 2013(8): 1613-1618. (in Chinese)

导师简介:



马世伟(1965—),男,甘肃嘉峪关人,博士,教授,博士生导师,1986年、1991年于兰州大学分别获得学士、硕士学位,2000年于上海大学获得博士学位,主要从事信号处理、图像处理和模式识别等的研究。E-mail: masw@shu.edu.cn

作者简介:



刘丽娜(1981—),女,山东邹平人,博士研究生,讲师,2003年于青岛大学获得学士学位,2006年于山东大学获得硕士学位,主要从事模式识别与信息处理自动化、图像处理等的研究。E-mail: linaliu-126@163.com