

基于全局和局部特征的人体行为识别

刘帆*, 于凤芹

江南大学物联网工程学院, 江苏 无锡 214122

摘要 为克服方向梯度直方图(HOG)特征在人体行为识别中仅表征动作的全局梯度特征、缺乏局部细节信息、对遮挡表现不佳的问题,改进了一种基于全局特征和局部特征的方法来实现人体行为识别。该算法首先使用背景减法获得人体运动区域;方向可控滤波器能有效描述动作边缘特征,通过引入方向可控滤波器改进 HOG 特征以增强局部边缘信息,同时对加速稳健特征进行 k -means 聚类获得词袋模型;最后将融合后的行为特征输入支持向量机对行为特征进行分类识别。在数据集 KTH、UCF Sports 和 SBU Kinect Interaction 上进行仿真模拟,结果表明改进的算法识别准确率分别达到了 96.7%、94.2%和 90.8%。

关键词 图像处理; 人体行为识别; 加速稳健特征; 方向梯度直方图; 词袋模型; 支持向量机

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP57.021004

Human Action Recognition Based on Global and Local Features

Liu Fan*, Yu Fengqin

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

Abstract This paper improves a global and local feature-based method to overcome problems of the histogram of oriented gradients (HOG), such as the features only characterizing the global gradient feature of motion, lacking local detail information, and having poor performance on occlusion, in the human behavior recognition. The proposed algorithm first uses the background difference method to obtain the human motion region; then, a steerable filter can effectively describe the motion edge features to improve HOG features, therefore enhancing edge details. At the same time, k -means clustering is conducted on speeded up robust features (SURF) to obtain the bag-of-words model. Finally, the merged behavior features are input into a support vector machine (SVM) for classification and recognition. Simulation experiments perform on the KTH, UCF Sports, and SBU Kinect Interaction datasets, showing improved algorithm recognition accuracies of 96.7%, 94.2%, and 90.8%, respectively.

Key words image processing; human action recognition; speeded up robust features; histogram of oriented gradients; bag-of-words model; support vector machine

OCIS codes 100.2000; 100.3008; 100.5010

1 引言

视频中人体行为识别已经引起了计算机视觉、机器学习和模式识别领域人员的广泛关注^[1]。此外,它在智能监控、增强现实、视频标注、人机交互和体感游戏等方面都具有广泛的应用^[2]。因此,用计算机算法自动识别出视频中执行的人体行为,成为近年来计算机视觉领域的热门研究课题。

行为识别现有方法可以分为两大类:局部特征方法和全局特征方法。局部特征方法一般从视频或图像中提取局部子区域或兴趣点,如光流、尺度不变特征变换(SIFT)等。全局特征方法通常提取整个人体信息,如运动历史图(MHI)、轮廓特征等。每种特征的表征方式都有长处和不足,现在比较热门的研究方式是混合特征。

文献[3]通过决策级融合空时域特征既提升了

收稿日期: 2019-05-05; 修回日期: 2019-06-03; 录用日期: 2019-07-01

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)

* E-mail: 977216218@qq.com

识别的速度又保持了算法的精度,但是算法对遮挡因素较为敏感。文献[4]中引入运动上下文的概念来获取视频词的时空分布。这些算法虽然取得了很好的性能,但都缺少梯度特征信息。文献[5]提出使用 Harris 角点检测器在时空 3D 空间检测兴趣点,并利用这些兴趣点描述相关动作。针对文献[5]的算法缺陷,文献[6]在时空域进行 Gabor 滤波,改进了原算法。针对方向梯度直方图(HOG)特征对图像边缘描述不充分的问题,文献[7]利用改进的 HOG 特征更好地提取了建筑物的边缘信息,在多个指标上都优于传统方法。文献[8]对行为视频进行稠密采样以获得行为动作的稠密轨迹(DT),然后分别计算 DT 的运动边界直方图、光流直方图和梯度方向直方图,并将 3 个特征进行融合,该方法对红外人体行为具有较强的表达能力。文献[9]采用深度学习方法对各网络提取的特征进行加权融合,算法识别率较高、泛化能力较强,但是模型复杂。

根据以上分析,受文献[7]启发,为克服单一特征对特征描述不够充分的问题,本文提出一种新的人体行为识别特征表征方法,即将加速稳健特征(SURF)和改进的 HOG 特征进行融合来识别行为动作。为了去除无关背景带来的影响,使用背景减法将人体区域提取出来;然后分别对人体区域提取 SURF,使用 k -means 聚类算法获得视觉词典以构建直方图并融合改进的 HOG 特征;最后将融合后的特征向量输入到支持向量机(SVM)中进行训练并识别。

2 本文方法

2.1 SURF 特征提取

SURF^[10]算法主要包括两个部分:特征点定位和特征点描述。

1) 特征点定位

首先要构造 Hessian 矩阵。图像 I 中任意一个像素点 $X(x, y)$ 的 Hessian 矩阵为

$$\mathbf{H}(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix}, \quad (1)$$

式中: σ 为尺度; $L_{xy}(X, \sigma)$ 为高斯二阶微分 $\frac{\partial^2 g(\sigma)}{\partial x \partial y}$ 在点像素点 X 处与图像 I 的卷积; $L_{xx}(X, \sigma)$ 和 $L_{yy}(X, \sigma)$ 含义相似。 $g(\sigma)$ 为高斯函数,表达式为

$$g(\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (2)$$

使用 D_{xx} 、 D_{xy} 和 D_{yy} 表示盒子滤波器与图像在

不同方向上的卷积结果,则简化后的 Hessian 矩阵的行列式可表示为

$$\det(\mathbf{H}_{\text{approx}}) = D_{xx}D_{yy} - (0.9D_{xy})^2, \quad (3)$$

式中:0.9 为经验值,用来平衡使用盒子滤波器近似所带来的误差。

将所有经过 Hessian 矩阵处理的像素点与其 $3 \times 3 \times 3$ 的立体邻域内的 26 个点进行非极大值比较,当 Hessian 矩阵的行列式取得局部极大值或极小值时,将该极值点作为候选特征点。然后,在尺度空间和图像空间中进行线性插值运算来获得最后稳定的特征点。

2) 特征点描述

在兴趣点附近建立一个 $20s$ (s 为特征点的尺度值)大小的方形区域,将其旋转到主方向,然后再将方形区域分成 $4 \times 4 = 16$ 个子区域,对每个子区域统计 $5 \times 5 = 25$ 个像素的 haar 小波特征。该 haar 小波特征为水平方向 haar 小波响应值之和,水平方向 haar 小波响应值的绝对值之和,垂直方向 haar 小波响应值之和,垂直方向 haar 小波响应值的绝对值之和。这样每个子区域就有 4 个值,所以每个特征点就是 $16 \times 4 = 64$ 维的向量。

2.2 构建视觉词典

为了生成视觉字典,本文使用最常用的 k -means 聚类算法(k 代表字典大小)。欧几里得距离的表达式为

$$E_{ss} = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2, \quad (4)$$

式中: E_{ss} 是所有对象的聚类误差; p 为空间中的点; m_i 为簇 C_i 的平均值。本文通过欧几里得距离把所有样本划分到与之距离最近的聚类中心所代表的类中,形成 k 个聚类,然后对这 k 个聚类再次进行计算得到新的聚类中心,并按照新的聚类中心重新划分样本类别,一直迭代下去,直到质心的位置变化不大,趋于稳定。

将每一个聚类的中心作为视觉词典中的单词,然后计算每个特征到这 k 个视觉单词的距离,并将其映射到距离最近的视觉单词中构建直方图。

2.3 改进的 HOG 特征提取

可控滤波器由 Freeman 等^[11]提出,它是一种特殊的滤波器,可以任意旋转,在图像纹理分析方面表现较好。可控滤波器由一组基滤波器构成,即

$$f_{\theta}(x, y) = \sum_{i'=1}^M k_{i'}(\theta) f_{\theta_{i'}}(x, y), \quad (5)$$

式中: $f_{\theta}(x, y)$ 为 θ 方向上的滤波器; $k_{i'}(\theta)$ 为插值

函数, $f_{\theta_i}(x, y)$ 为基函数, 插值函数和基函数成对存在; M 为基函数的个数。

高斯函数二阶导数近似于边界, 可以用一个圆对称的窗函数和一个多项式的乘积来表示。可控滤波器由高斯函数二阶导数构成, 可表示为

$$G_2^\theta = k_1(\theta)G_2^{0^\circ} + k_2(\theta)G_2^{60^\circ} + [1 - k_1(\theta)]G_2^{120^\circ}, \quad (6)$$

式中: $k_1(\theta)$ 、 $k_2(\theta)$ 为插值函数; $G_2^{0^\circ}$ 、 $G_2^{60^\circ}$ 和 $G_2^{120^\circ}$ 表示高斯函数在角度为 0° 、 60° 和 120° 时的二阶导数。插值函数和基函数为

$$\begin{cases} k_1(\theta) = \cos^2\theta \\ k_2(\theta) = -\sin 2\theta \end{cases}, \quad (7)$$

$$\begin{cases} G_2^{0^\circ}(x, y) = 0.9213(2x^2 - 1)\exp[-(x^2 + y^2)] \\ G_2^{60^\circ}(x, y) = 1.843xy\exp[-(x^2 + y^2)] \\ G_2^{120^\circ}(x, y) = 0.9213(2y^2 - 1)\exp[-(x^2 + y^2)] \end{cases}. \quad (8)$$

HOG 特征最早是由 Dalal 等^[12] 提出, 被应用于行人检测。HOG 方法思想是计算稠密网格中归一化的局部方向梯度直方图。改进的 HOG 特征提取过程如下。

1) 首先为了消减阴影和光照因素带来的影响, 需要对图像进行归一化处理。

2) 求取 0° 和 90° 方向的梯度。假设图像 I 中任一像素点的坐标为 (x, y) , 则该点的 0° 和 90° 方向的梯度 $G_x(x, y)$ 、 $G_y(x, y)$ 为

$$\begin{cases} G_x(x, y) = I \times f_{0^\circ} \\ G_y(x, y) = I \times f_{90^\circ} \end{cases}, \quad (9)$$

像素点处的梯度幅值和梯度方向可分别表示为

$$\begin{cases} G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ \theta(x, y) = \arctan[G_y(x, y)/G_x(x, y)] \end{cases}. \quad (10)$$

3) 为了保持对图像中人的姿态以及外观的弱

敏感性, 将图像分成若干个单元格 (Cell) 并将每个单元格分为 9 个方向块, 然后为每个单元格构建梯度方向直方图。

4) 把步骤 3) 得到的单元格组合成大的区间 (Block), 对区间内的直方图进行 L_2 范数归一化处理, 直方图归一化处理的方程为

$$L_2(\mathbf{v}) = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + e^2}}, \quad (11)$$

式中: e 为常数参数, 它的值非常小; \mathbf{v} 表示区间内的直方图向量。

5) 连接所有区间内的直方图向量, 最终得到整张图像的 HOG 特征。

6) 另外, HOG 特征的维数计算方程为

$$D_{\text{HOG}} = D_{\text{Cell}} \times \frac{S_{\text{Block}}}{S_{\text{Cell}}} \times \left(\frac{h - S'_{\text{Block}}}{S'_{\text{Block}}} + 1 \right) \times \left(\frac{w - S'_{\text{Block}}}{S'_{\text{Block}}} + 1 \right), \quad (12)$$

式中: D_{Cell} 表示 Cell 的维数; S_{Block} 为 Block 的大小; S_{Cell} 为 Cell 的大小; S'_{Block} 表示 Block 的步长; h 和 w 分别表示一幅图像的高和宽。

改进的 HOG 特征对光照以及几何变化具有良好的鲁棒性, 可以有效提取边缘纹理信息。

本文使用 SVM 对数据进行分类, SVM 在高维空间中使用超平面来划分具有最大边距的数据, 它的数学表达式为

$$\min \frac{\|\boldsymbol{\omega}\|^2}{2} + C \sum_j^n \xi_j, \text{ s.t. } y_j(\boldsymbol{\omega}x_j + b) \geq 1 - \xi_j, j \in \{1, \dots, n\}, \quad (13)$$

式中: $\boldsymbol{\omega}$ 为超平面的法向量; C 为惩罚因子; n 是样本点的数量; ξ 为松弛变量; b 为偏置。

2.4 算法实现步骤

基于全局特征和局部特征的人体行为识别算法的原理流程如图 1 所示。

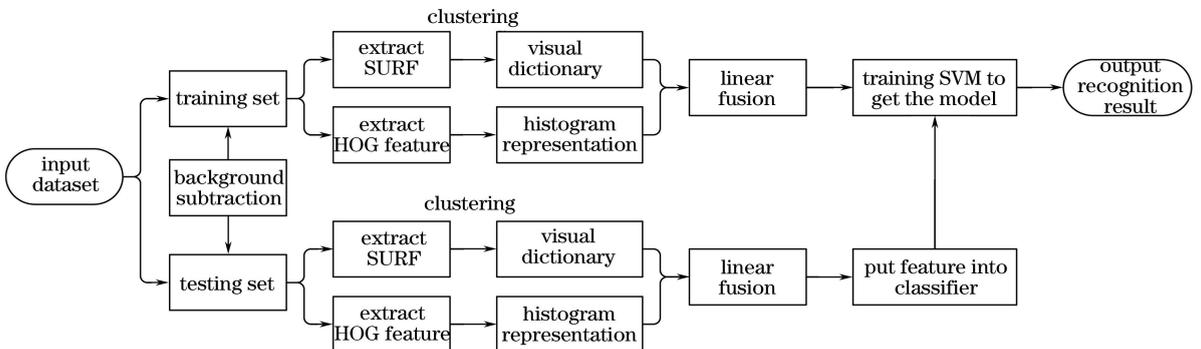


图 1 算法流程图

Fig. 1 Flow chart of algorithm

算法具体实现步骤如下。

1) 在输入的视频帧上应用背景减法, 仅从视频的每个帧中提取人类所在的区域, 称该区域为动作区域。

2) 对运动区域进行兴趣点检测, 用 SURF 对其进行描述, 采用词袋(BOW)模型对特征进行直方图构建。每个特征点的维数是 64。

3) 对每个提取出的动作区域分别在水平和垂直方向进行 HOG 特征提取。HOG 特征的维数是 D_{HOG} , 本文为 3780。

4) 将 BOW 模型构建好的直方图特征和 HOG 特征融合以编码视频信息。将最终的行为特征输入训练好的 SVM 分类器来实现动作识别。

3 仿真实验结果与分析

仿真所用软件环境为 Windows10 64 位, Intel Core i5-760 CPU 处理器, 8G 内存, Matlab R2016a, Python3.6。本文采用了在动作识别领域广泛应用的 KTH 数据集^[13]、UCF Sports 数据集^[14]和 SBU Kinect Interaction 数据集^[15]。KTH 数据集有 6 类动作, 每类动作由 25 个人完成, 分别为“行走”(Walking)、“慢跑”(Jogging)、“跑步”(Running)、

“拳击”(Boxing)、“挥手”(Handwaving)和“拍手”(Handclapping), 该数据集背景较为单一。UCF Sports 数据集共有 150 段视频, 这些视频序列来自于 BBC 和 ESPN 的各种广播体育频道。该数据集包含 10 种不同类型的动作, 每一种动作又由不同的人完成。分别为“跳水”(Diving)、“高尔夫”(Golf Swing)、“踢球”(Kicking)、“举重”(Lifting)、“骑马”(Riding Horse)、“跑步”(Running)、“滑板”(Skate Boarding)、“鞍马”(Swing Bench)、“单杠”(Swing Side Angle)和“行走”(Walking)。该数据集拍摄背景较为复杂, 尺度和视角各不相同, 因此给行为识别带来一定的挑战性。SBU Kinect Interaction 数据集包含 8 种类型的动作交互, 分别为“靠近”(Approaching)、“远离”(Departing)、“交换”(Exchanging)、“拥抱”(Hugging)、“踢腿”(Kicking)、“击打”(Punching)、“推开”(Pushing)和“握手”(Shaking hands)。每个动作由 7 个不同的人在同一实验室完成, 最终组成了 230 对互动。

以 UCF Sports 数据集为例, 图 2 为部分视频帧的 SURF 提取, 图 3 为部分视频帧的 HOG 特征改进前后的对比图。改进后的 HOG 特征对图像的边缘细节信息描述得更加充分。



图 2 SURF 提取

Fig. 2 SURF extraction

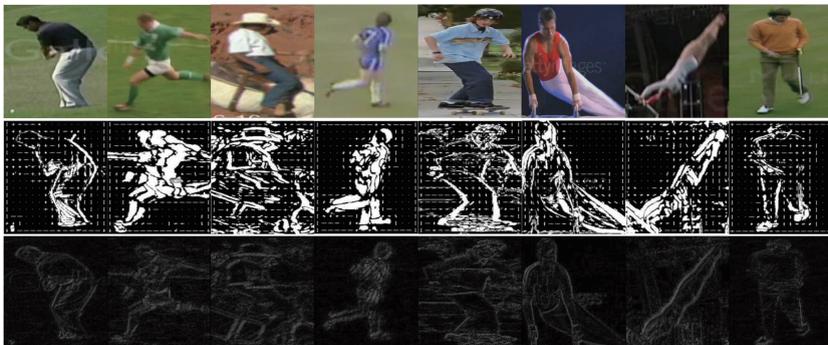


图 3 HOG 特征改进前后的对比

Fig. 3 Comparison of HOG features before and after improvement

仿真中所设置的参数如下： $D_{Cell}=9, S_{Block}=32 \times 32, S_{Cell}=16 \times 16, S'_{Block}=16, C=1$ ，当 k 分别取 1500、3000 和 1000 的时候，3 个数据集的识别效果最好。为了验证本文方法的性能，对上述提到的数据集进行测试。实验采用交叉验证进行测试，选择 70% 的数据集作为训练集，剩下的作为测试集。最后获得 KTH 数据集、UCF Sports 数据集和 SBU Kinect Interaction 数据集识别准确率的混淆图，如图 4 所示。

从混淆图中可以看出，在 KTH 数据集上，“行走”、“慢跑”和“跑步”动作之间存在一定的相似性，导致一定的误识别率；“挥手”和“拍手”之间也有 4%~7% 的误差。在 UCF Sports 数据集上，由于

该数据集背景复杂、视角多变换以及动作之间存在遮挡和速度较快等问题，在“踢球”和“跑步”两个动作之间，有 8%~15% 的误识别率；对于其他具有遮挡的动作，识别率较高。在 SBU Kinect Interaction 数据集上，因为动作之间存在一定的相似性，存在较大误识别率的有“击打”和“推开”，“靠近”和“交换”的误识别率也有 14%，因为在交换物体时，人体需要相互靠近。对于其他较为明显的特征，本文识别率较高且误识别率较低。

为评价本文算法的性能，以识别准确率 (η) 为指标，与同样使用该数据集的其他算法进行比较，比较结果如表 1 所示。

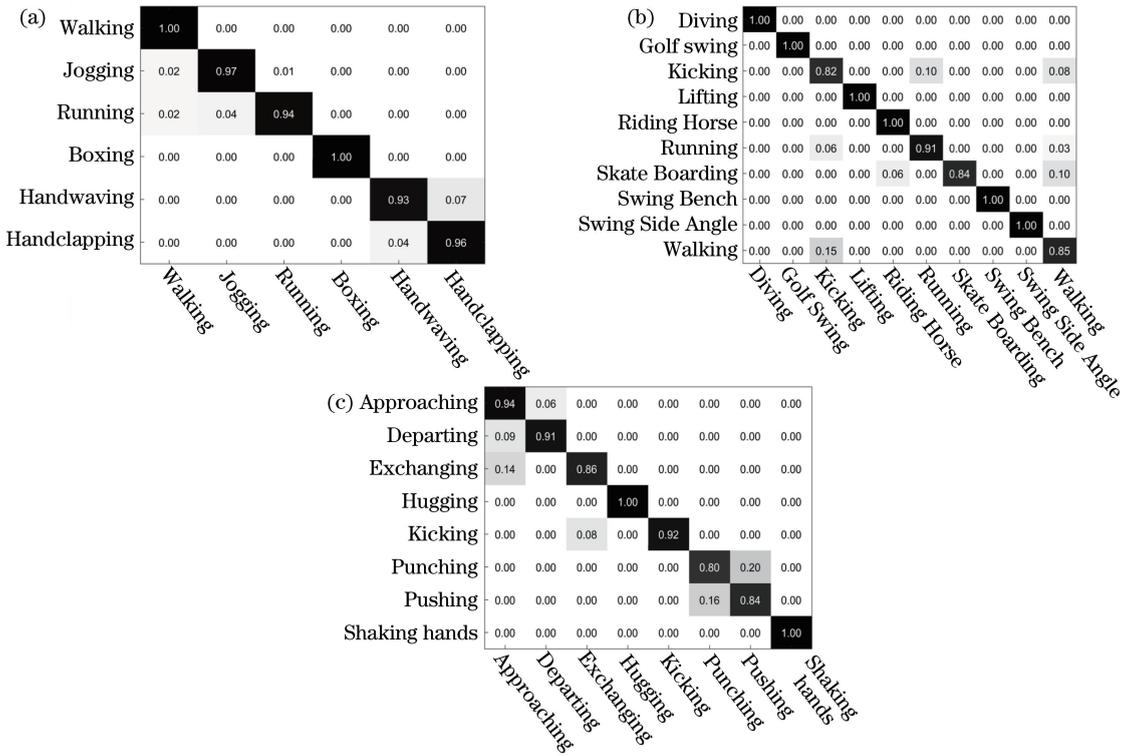


图 4 本文算法在不同数据集上的识别率。(a) KTH 数据集上的混淆矩阵图；(b) UCF Sports 数据集上的混淆矩阵图；(c) SBU Kinect Interaction 数据集上的混淆矩阵图

Fig. 4 Recognition accuracies of proposed algorithm in different datasets. (a) Confusion matrix on KTH dataset; (b) confusion matrix on UCF Sports dataset; (c) confusion matrix on SBU Kinect Interaction dataset

表 1 不同方法的识别准确率

Table 1 Recognition accuracies of different methods

Method	η on KTH / %	Method	η on UCF sports / %	Method	η on SBU Kinect Interaction / %
Method in Ref. [16]	95.6	Ref. [16]	88.5	Ref. [15]	80.3
Method in Ref. [17]	95.6	Ref. [18]	92.7	Ref. [22]	89.4
Method in Ref. [18]	96.8	Ref. [20]	95.8	Ref. [23]	94.8
Method in Ref. [19]	97.8	Ref. [21]	97.5	Ref. [24]	98.5
Proposed method	96.7	Proposed	94.4	Proposed	90.8

由表 1 可以看出，在 KTH 数据集上，本文算法的识别率为 96.7%，比文献[16-17]分别提高了

1.1%、1.1%，接近于文献[18]的识别准确率，但是文献[18]中稠密轨迹提纯过程较为复杂，而且对硬

件的要求较高,相比文献[19]深度学习的算法,本文准确率较低,深度学习算法虽然具有很多优势,但是在参数选择、模型训练等方面较为复杂。在 UCF Sports 数据集上,本文算法的识别率为 94.2%,比文献[16,18]分别提高了 5.9%、1.7%,在一定程度上提高了动作的识别率,但是远低于文献[20-21]的识别准确率,再一次证明了深度学习方法的优势。因为 SURF 对视角变换和旋转保持不变性的优势,在 UCF Sports 数据集中,大部分动作都是在不同视角和不同背景下完成的,因此相较于单一背景的 KTH 数据集,识别准确率提高较为明显。在 SBU Kinect Interaction 数据集上,本文算法的识别率为 90.8%,比文献[15,22]分别提高了 10.5%、1.4%,远比文献[23-24]的深度学习方法低,但是文献[23]的循环神经网络算法训练需要分为多个阶段、步骤繁琐、对硬件要求较高。文献[24]的算法需要获取人体骨骼节点信息之间的关联、对时空域进行建模而且 LSTM(Long Short-Term Memory)的训练复杂度、解码时延高。本文算法在保证识别准确率的同时,易于训练且对硬件要求低。

4 结 论

本文改进了一种基于全局特征和局部特征的人体行为识别算法。利用 SURF 对遮挡、视角变换和旋转保持不变性的优势,融合改进的 HOG 特征进行人体行为识别。优点是不仅保留了 HOG 特征的全局属性和 SURF 的局部信息,还减少了遮挡、视角变换等因素对动作识别的影响。在 KTH 数据集、UCF Sports 数据集和 SBU Kinect Interaction 数据集的评估证明,本文提出的方法能够识别视频中的各种动作。实验结果表明,在 3 个数据集上的识别率分别为 96.7%、94.2%和 90.8%。未来工作将专注于显著对象检测方法检测视频帧中的显著对象,并且仅提取这些对象的特征来识别人体动作。

参 考 文 献

- [1] Zhao H Y, Jia B X. Human action recognition using image contour[J]. Computer Science, 2013, 42(2): 312-315.
赵海勇, 贾保先. 基于轮廓特征的人体行为识别[J]. 计算机科学, 2013, 42(2): 312-315.
- [2] Cai J X, Feng G C, Tang X, et al. Human action recognition based on local image contour and random forest[J]. Acta Optica Sinica, 2014, 34(10): 1015006.
蔡加欣, 冯国灿, 汤鑫, 等. 基于局部轮廓和随机森林的人体行为识别[J]. 光学学报, 2014, 34(10): 1015006.
- [3] Li Y D, Xu X P. Human action recognition by decision-making level fusion based on spatial-temporal features[J]. Acta Optica Sinica, 2018, 38(8): 0810001.
李艳获, 徐熙平. 基于空-时域特征决策级融合的人体行为识别算法[J]. 光学学报, 2018, 38(8): 0810001.
- [4] Zhang Z M, Hu Y Q, Chan S, et al. Motion context: a new representation for human action recognition[M]// Forsyth D, Torr P, Zisserman A. Computer vision-ECCV 2008. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2008, 5305: 817-829.
- [5] Laptev I. On space-time interest points [J]. International Journal of Computer Vision, 2005, 64(2/3): 107-123.
- [6] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]// 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, October 15-16, 2005, Beijing, China. New York: IEEE, 2005: 65-72.
- [7] Yang S, Li S Y, Shao Y Y, et al. Building recognition method based on improved HOG feature [J]. Computer Engineering and Applications, 2018, 54(7): 196-200.
杨松, 李盛阳, 邵雨阳, 等. 基于改进 HOG 特征的建筑物识别方法[J]. 计算机工程与应用, 2018, 54(7): 196-200.
- [8] Shao Y H, Guo Y C, Gao C. Infrared human action recognition using dense trajectories-based feature[J]. Journal of Optoelectronics • Laser, 2015, 26(4): 758-763.
邵延华, 郭永彩, 高潮. 基于稠密轨迹特征的红外人体行为识别[J]. 光电子·激光, 2015, 26(4): 758-763.
- [9] Huang Y W, Wan C L, Feng H. Multi-feature fusion human behavior recognition algorithm based on convolutional neural network and long short term memory neural network[J]. Laser & Optoelectronics Progress, 2019, 56(7): 071505.
黄友文, 万超伦, 冯恒. 基于卷积神经网络与长短期记忆神经网络的多特征融合人体行为识别算法[J]. 激光与光电子学进展, 2019, 56(7): 071505.
- [10] Bay H, Tuytelaars T, van Gool L. SURF: speeded

- up robust features [M] // Leonardis A, Bischof H, Pinz A. Computer vision-ECCV 2006. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2006, 3951: 404-417.
- [11] Freeman W T, Adelson E H. The design and use of steerable filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13 (9): 891-906.
- [12] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE, 2005: 8588935.
- [13] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach [C] // Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., August 26-26, 2004, Cambridge, UK. New York: IEEE, 2004: 8380901.
- [14] Soomro K, Zamir A R. Action recognition in realistic sports videos [M] // Moeslund T, Thomas G, Hilton A. Computer vision in sports. Advances in computer vision and pattern recognition. Cham: Springer, 2014: 181-208.
- [15] Yun K, Honorio J, Chattopadhyay D, et al. Two-person interaction detection using body-pose features and multiple instance learning [C] // 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 28-35.
- [16] Ding S T, Qu S R. An improved interest point detector for human action recognition [J]. Journal of Northwestern Polytechnical University, 2016, 34 (5): 886-892.
丁松涛, 曲仕茹. 基于改进时空兴趣点检测的人体行为识别算法 [J]. 西北工业大学学报, 2016, 34(5): 886-892.
- [17] Lu T R, Yu F Q, Chen Y. A human action recognition method based on LSDA dimension reduction [J]. Computer Engineering, 2019, 45(3): 237-241, 249.
- 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.
- [18] Cheng H S, Li Q W, Qiu C C, et al. Human action recognition algorithm based on improved dense trajectories [J]. Computer Engineering, 2016, 42 (8): 199-205.
程海粟, 李庆武, 仇春春, 等. 基于改进密集轨迹的人体行为识别算法 [J]. 计算机工程, 2016, 42(8): 199-205.
- [19] Lin B, Fang B, Yang W B, et al. Human action recognition based on spatio-temporal three-dimensional scattering transform descriptor and an improved VLAD feature encoding algorithm [J]. Neurocomputing, 2019, 348: 145-157.
- [20] Rahmani H, Mian A, Shah M. Learning a deep model for human action recognition from novel viewpoints [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (3): 667-681.
- [21] Tu Z G, Xie W, Qin Q Q, et al. Multi-stream CNN: learning representations based on human-related regions for action recognition [J]. Pattern Recognition, 2018, 79: 32-43.
- [22] Song S J, Lan C L, Xing J L, et al. An end-to-end spatio-temporal attention model for human action recognition from skeleton data [C] // Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. USA: AAAI, 2017: 4263-4270.
- [23] Wang H S, Wang L. Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3633-3642.
- [24] Cui R, Hua G, Zhu A C, et al. Hard sample mining and learning for skeleton-based human action recognition and identification [J]. IEEE Access, 2019, 7: 8245-8257.