

1-3-2019

Human Action Recognition Based on Depth Image

Tang Chao

1. Department of Computer Science and Technology, Hefei University, Hefei 230601, China; ;

Miaohui Zhang

2. Energy Research Institute, Jiangxi Academy of Sciences, Nanchang 330096, China; ;

Li Wei

3. School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; ;

Cao Feng

4. School of Computer and Information Science, Shanxi University, Taiyuan 030006, China; ;

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Human Action Recognition Based on Depth Image

Abstract

Abstract: Because of the complexity and non-rigidity of human actions, traditional human action recognition based on RGB video data is a very challenging research topic. According to some deficiencies of existing recognition method based on RGB video data, *a novel human action recognition method is proposed based on depth image data. In this new method, the block mean feature in the depth difference motion historical image is fused with the Gabor feature as mixed features and then a rotation forest algorithm is used to model.* The experimental results show that the proposed method is simple, fast and efficient compared with other supervised action recognition algorithms on DHA depth datasets.

Keywords

human action recognition, depth image, depth difference motion historical image, Gabor feature, rotation forest

Authors

Tang Chao, Miaohui Zhang, Li Wei, Cao Feng, Xiaofeng Wang, and Xiaohong Tong

Recommended Citation

Tang Chao, Zhang Miaohui, Li Wei, Cao Feng, Wang Xiaofeng, Tong Xiaohong. Human Action Recognition Based on Depth Image[J]. Journal of System Simulation, 2018, 30(5): 1641-1649.

基于深度图像的人体行为识别

唐超¹, 张苗辉^{2*}, 李伟³, 曹峰⁴, 王晓峰¹, 童晓红⁵

(1. 合肥学院计算机科学与技术系, 合肥 230601; 2. 江西省科学院能源研究所, 南昌 330096; 3. 厦门理工学院计算机与信息工程学院, 厦门 361024; 4. 山西大学计算机与信息技术学院, 太原 030006; 5. 合肥职业技术学院信息中心, 合肥 238000)

摘要: 由于人体动作的复杂性和非刚性特点, 传统的基于 RGB 视频数据的人体行为识别是一个非常具有挑战性的研究课题。针对现有基于 RGB 视频数据识别方法的不足, 提出了一种基于深度图像数据的人体行为识别方法, 该方法将深度差值运动历史图像中分块均值特征与 Gabor 特征进行融合, 采用泛化能力较好的旋转森林算法进行建模。在 DHA 深度数据集上实验结果表明, 相比其它监督行为识别分类算法, 基于深度图像的方法具有简单、快速、高效的特点。

关键词: 人体行为识别; 深度图像; 深度差值运动历史图像; Gabor 特征; 旋转森林

中图分类号: TP391 文献标识码: A 文章编号: 1004-731X (2018) 05-1641-09

DOI: 10.16182/j.issn1004731x.joss.201805003

Human Action Recognition Based on Depth Image

Tang Chao¹, Zhang Miaohui^{2*}, Li Wei³, Cao Feng⁴, Wang Xiaofeng¹, Tong Xiaohong⁵

(1. Department of Computer Science and Technology, Hefei University, Hefei 230601, China; 2. Energy Research Institute, Jiangxi Academy of Sciences, Nanchang 330096, China; 3. School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; 4. School of Computer and Information Science, Shanxi University, Taiyuan 030006, China; 5. Information center, Hefei Technology College, Hefei 238000, China)

Abstract: Because of the complexity and non-rigidity of human actions, traditional human action recognition based on RGB video data is a very challenging research topic. According to some deficiencies of existing recognition method based on RGB video data, a novel human action recognition method is proposed based on depth image data. In this new method, the block mean feature in the depth difference motion historical image is fused with the Gabor feature as mixed features and then a rotation forest algorithm is used to model. The experimental results show that the proposed method is simple, fast and efficient compared with other supervised action recognition algorithms on DHA depth datasets.

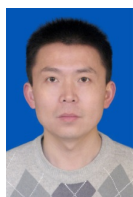
Keywords: human action recognition; depth image; depth difference motion historical image; Gabor feature; rotation forest

引言

在过去的十几年中, 人体行为识别一直是计算机视觉领域中非常活跃的研究课题, 它有着非常广

泛的应用, 包括视频监控、人机交互、辅助技术、计算行为科学和虚拟现实等, 许多学者对此进行研究与综述分析^[1-4]。

然而在过去的研究中, 由于技术的限制, 一般都是在普通的摄像机所获得的 RGB 视频图像数据上进行分析, 但是 RGB 图像对光照的强弱和环境的变化比较敏感, 同一行为在不同光照强度下, 会得到完全不同的图像信息。因此, 对图像中人体行为的信息进行提取时, 很难准确的把人



收稿日期: 2017-08-19 修回日期: 2017-09-18;
基金项目: 国家自然科学基金(61672204, 41401521, 61602220), 合肥学院人才科研基金(15RC07), 安徽高校优秀拔尖人才培养资金(gxfx2017099), 山西省青年科技研究基金(2015J05015), 福建省自然科学基金(2016J01325, 2015J05015), 江西省自然科学基金(20161BAB21057);
作者简介: 唐超(1977-), 男, 安徽合肥, 博士, 讲师, 研究方向为机器学习和计算机视觉。

<http://www.china-simulation.com>

• 1641 •

体与背景进行完全分割,对后续的特征提取和动作分类也会造成很大的误差,降低识别率。随着科学的不断进步,微软推出了 Kinect 传感器,可以捕捉到场景的深度信息。图像的深度信息只与物体距离相机的远近有关,不会受到光照、环境和阴影的影响。

基于深度图像的人体行为识别和传统的行为识别方法比起来,它的准确率和鲁棒性都更好。因此,基于深度图像的人体行为识别具有非常高的研究价值和实用性。为了能够很好的描述人体行为的结构信息,本文采用了从深度差值运动历史图像中分别提取分块均值特征与 Gabor 特征,并进行融合作为行为描述子,然后采用旋转森林算法进行建模分类。

本文的主要贡献如下:

1) 提出了一种基于深度差值运动历史图像分块均值特征与 Gabor 特征的人体动作识别方法,通过多视角的特征融合来表征人体行为,充分利用不同特征之间的优势来提升动作表征能力;

2) 本文分块均值特征能够很好的描述人体行为的结构信息,而 Gabor 特征是鲁棒性较好的特征,因为 Gabor 方法是从图像的不同尺度、不同方向的角度去提取图像特征,即频率和旋转方向上,这样能够提取图像不同参数设定下的不同细节特征,不同动作细节信息也会不同,动作分类时就有较好的区分度;

3) 采用旋转森林进行建模分析,充分利用了不同学习器学习偏置来提升学习器的泛化能力;

4) 在公共深度数据集上的实验结果表明本文提出的基于特征融合的人体行为识别算法取得较高的识别率。

1 相关工作

目前有关行为识别的研究已经有许多的成果,但是在真实的环境下的行为识别还是困难的。近年来伴随着 RGB-D 相机的发展,特别是微软推出的

Kinect 传感器,使得最近的研究工作集中在使用深度图像来解决问题。因为,与传统的 RGB 数据相比,深度信息对照明条件的变化更为鲁棒。近年来许多研究者都开展了基于深度图像数据信息的人体行为识别工作^[5-7]。

Zhang 等^[8]从深度图像序列中提取出一种简单高效的 3D 纹理直方图特征(3D Histograms of Texture, 3DHoTs),将深度图像视频帧序列投影到三个正交直角平面上,然后从中提出纹理特征,该特征具有较好的判别性,并采用多类的 boosting 分类器来识别动作。Zhang 等^[9]利用深度图中的深度梯度信息和骨架关节信息来构造一个粗粒度的深度骨架特征(Depth-Skeleton Feature, DS),然后通过密码本学习(Codebook Learning)和稀疏编码进行对 DS 进行特征提升,得到细粒度的特征。然后采用随机决策森林(Random Decision Forests, RDF)进行建模分类,在 3 个公共的数据集上的实验结果都优于传统的监督学习方法。Yang 等^[10]构建一个多实例多任务学习系统在基于 3D 的人体骨架关节图上取得较好的识别结果。Uddin^[11]采用隐马尔可夫模型来识别深度视频数据中人体行为,人体动作特征则采用了基于关节的时空特征。由于从深度图像上提取的特征具有视角依赖性,使得现有的 3D 行为识别结果对视角变化比较敏感。为了解决这个问题,Rahmani 等^[12]提出方向主成份直方图描述子(Histogram of Oriented Principal Components (HOPC) Descriptor),该特征对噪音、视角变化、尺度变化,速度变化具有较好的鲁棒性。Akagunduz 等^[13]提出一种快速和鲁棒性好的形状序列描述子剪影方向体(Silhouette Orientation Volume, SOV)用来表征人体动作,用于人体跌倒检测。在 SDU-Fall 数据集上的实验结果取得 91.89% 正确率。

总而言之,虽然现有的基于深度图像的行为识别方法取得很大的进展,但是在实际工程应用中识别的可靠性依然不高,原因是大量动作存在类内差异,如人体动作存在速度差异等。

2 方法框架

为了提高识别系统的鲁棒性和实用性,以及更加有效地利用不同的特征的优势,本文将两种不同的行为特征进行融合作为行为描述子,然后采用旋转森林算法进行训练识别,系统流程见图 1。该方法保留了简单特征的高效计算便利,同时又保证了特征的鲁棒性和判别能力,方法框架主要包括以下几个步骤:

- 1) 输入深度视频图像序列;
- 2) 对输入的每帧图像进行灰度化处理;
- 3) 分别求出每个像素点对应像素值的最大值和最小值,计算出深度差值运动历史图像,捕获目标的运动变化过程;
- 4) 为了减小后期的计算量和运行时间,提取包含人体剪影的感兴趣区域;
- 5) 分别在人体剪影感兴趣区域中进行分块均值特征和 Gabor 特征提取,同时采用特征融合的方法来提升动作的表征能力;
- 6) 利用旋转森林算法进行建模分类。

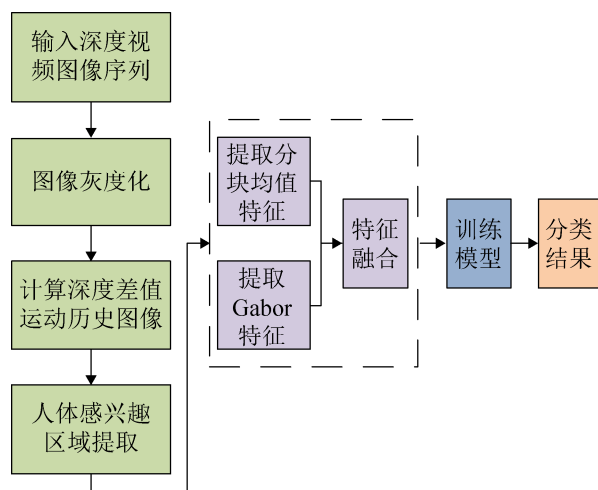


图 1 人体行为识别系统图

Fig. 1 System diagram of human action recognition

3 特征提取与融合

本节主要介绍人体感兴趣区域的提取过程,并详细给出分块均值特征与 Gabor 特征提取方法。其中,3.1 节给出了深度图像预处理过程;3.2 节描述

了分块均值特征描述子;3.3 节介绍 Gabor 特征提取方法;在 3.4 节将给出多特征混合方法。

3.1 预处理

输入的深度视频图像序列要进行一些前期的处理,然后运用深度像素值差值方法描述运动过程,生成深度差值运动历史图像,为了后续特征提取减少计算量,提高运算效率,同时需要对深度差值运动历史图像提取有效地感兴趣区域。

对于一个动作的运动过程,可以把一个动作视频序列分为 N 帧(N 的大小不影响深度差值计算过程),对于每帧图像中每个像素值不为零的像素点,计算该点对应位置在整个运动过程中间距最大值以及最小值,为了描述整个动作变化过程,将计算所得各点位置的间距最大值减去对应的间距最小值,以其差值来表征该像素点的动作变化过程,整个图像就变成了差值图像,这样便生成了深度差值运动历史图像。具体运算过程如下公式:

$$I_{\max}(i, j) = \max\{D(i, j, t) : D(i, j, t) \neq 0, t \in [1 \cdots N]\} \quad (1)$$

$$I_{\min}(i, j) = \min\{D(i, j, t) : D(i, j, t) \neq 0, t \in [1 \cdots N]\} \quad (2)$$

$$I_{diff} = I_{\max} - I_{\min} \quad (3)$$

式中: $D(i, j, t)$ 表示深度图像; I_{\max} 表示运算得到的整个动作变化过程各点间距最大值; I_{\min} 表示运算得到的整个动作变化过程各点间距最小值; I_{diff} 代表能表征整个运动过程的深度差值图像。计算生成的差值历史图像能比较好的表示人体行为移动变化过程。

在得到表征行为移动变化的差值历史图像后,发现区域中存在着很多值是 0 的像素点,而这些在后面的处理中是没有用,甚至会增加特征提取的计算量,因此可以利用深度差值运动历史图像转化的二值图像(阈值设为默认值 0),然后找到像素值非零的数量最大的连通区域,获取该区域的矩形位置,按照这样的矩形位置数据对深度差值运动历史图像进行截取。即为不包含过多的多余像素,又不丢失有效区域的差值历史图像即人体感兴趣区域。

3.2 分块均值特征

在获得包含人体轮廓剪影的矩阵数据后, 通过分块的方法将该矩阵划分为不同大小不同层级的块, 划分从大到小依次是 $8 \times 8=64$ 块、 $4 \times 4=16$ 块、 $2 \times 2=4$ 块、 $1 \times 2=2$ 块和 $2 \times 1=2$ 块, 这样, 它们就一共被分成了 88 块, 再对每一块进行操作就能够很

好的得到这一小块数据中的空间信息, 这种分块的划分方法的具体划分如图 3 所示。最后, 将得到的 88 块中每一小块内的像素的平均值作为它的特征值, 所有小块的均值特征联合在一起就构成了一个行为的特征向量, 这样, 这个特征向量的维数是 88, 记为分块均值特征(Block Mean Feature, BMF)。

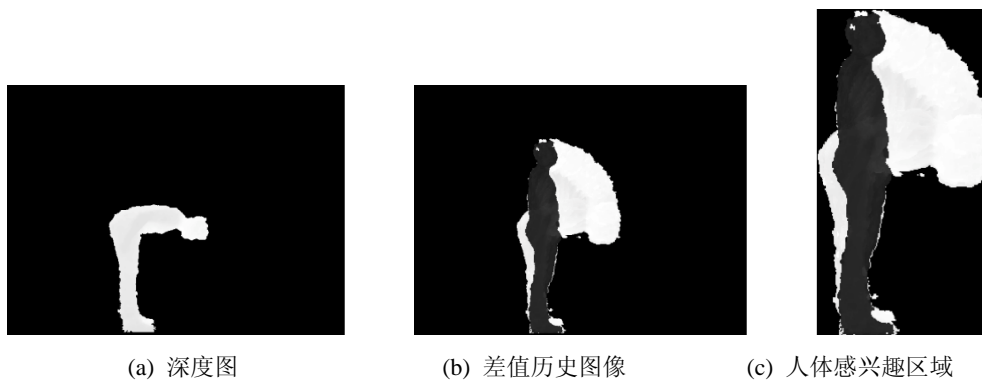


图 2 深度图像预处理
Fig. 2 Preprocessing of depth image

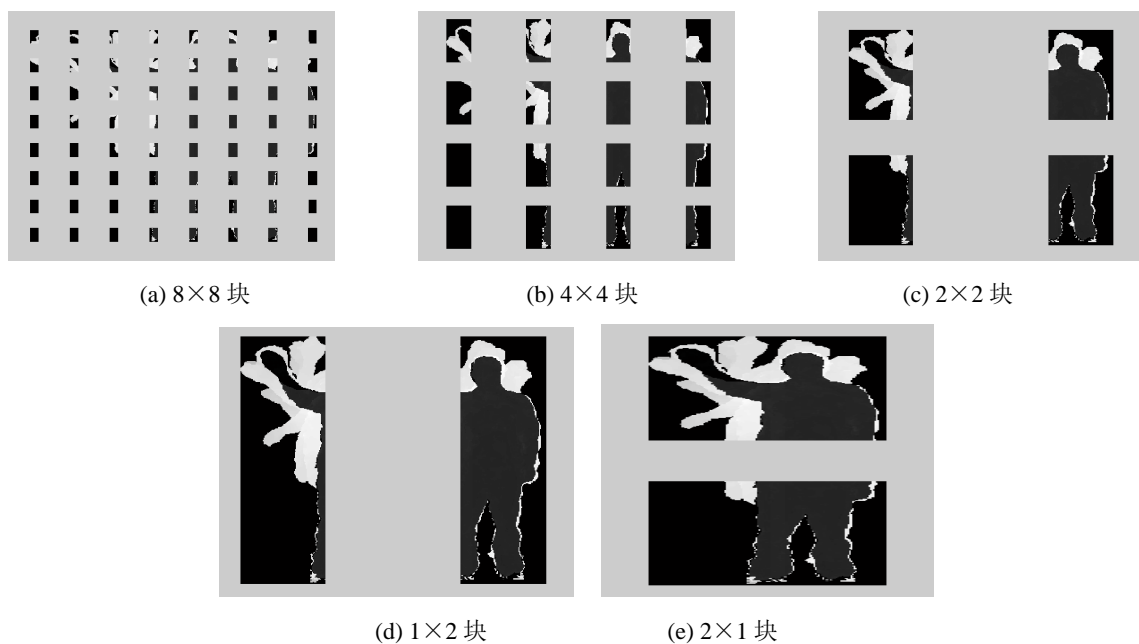


图 3 图像分块示意图

Fig. 3 Diagram of image block

3.3 Gabor 特征

本文采用 10 个不同尺度($v=0, 1, \dots, 9$)和 8 个不同方向($\mu=0, 1, \dots, 7$)即 80 个滤波器来生成 Gabor 特征, 具体运算过程是: 将剪裁后的有效深度差值

运动历史图像 I 分别与各个尺度, 各个方向的滤波器进行卷积, 输出取卷积后的模值:

$$G_{\mu,v}(x,y) = I(x,y) \times \psi_{\mu,v}(x,y) \quad (4)$$

式中: $\psi_{\mu,v}(x,y)$ 是 Gabor 函数, 由于卷积的计算量

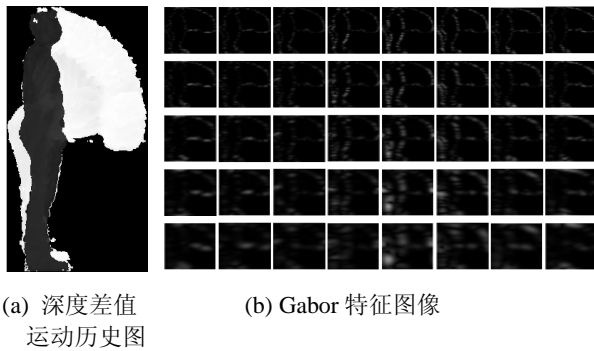
较为复杂, 使得时间复杂度很高, 因此用卷积定理来变换运算过程, 先用快速傅立叶变换计算, 可得:

$$F\{G_{\mu,\nu}(x,y)\} = F\{I(x,y)\}F\{\psi_{\mu,\nu}(x,y)\} \quad (5)$$

对运算结果再进行傅立叶反运算, 可得:

$$G_{\mu,\nu}(x,y) = F^{-1}\{F\{I(x,y)\}F\{\psi_{\mu,\nu}(x,y)\}\} \quad (6)$$

经过上面的变换后, 一幅深度差值图像就生成了 80 幅 Gabor 特征图像, 如果把 80 幅特征图像作为一个行为运动过程的分类标准, 那时间复杂将会很高, 所以, 本文对每一幅特征图像计算其均值和方差, 以均值和方差进行向量连接作为向量维度, 这样 10 个尺度 8 个方向的 80 幅特征图像形成的特征向量共 $10 \times 8 \times 2 = 160$ 维, 记为 Gabor 特征(Gabor Feature, GF), 如图 4 所示。



(a) 深度差值运动历史图 (b) Gabor 特征图像

图 4 Gabor 特征提取

Fig. 4 Gabor feature extraction

3.4 特征混合

为了使人体动作特征具有较高的区分辨别能力, 特征融合是一个有效的方法。目前, 主要的特征融合方法有: 像素级融合、特征级融合、以及决策级融合^[14-15]。本文采用特征级的特征融合方法。特征级融合是介于像素级融合和决策级融合的中间层融合。采用该方法可以最大程度保留待融合特征的有效辨别信息, 也可以有效消除各特征之间的冗余特征信息和辨别能力较差的特征。在具体操作时, 本文采用计算简单的串行特征融合方法, 具体地本文将分块特征与 Gabor 特征进行融合, 将两种不同的特征组合在一起作为行为描述特征, 即:

$$F_{fusion} = \{BMF, GF\} \quad (7)$$

4 识别方法

旋转森林(Rotation Forest, RF)^[16]是一种基于多分类器的集成学习系统。该系统中的基学习器是通过将同一学习算法应用于不同训练集来产生的。具体算法如下所示。

旋转森林算法

训练部分:

输入:

X : 训练样本数据集($N \times n$ 矩阵)

Y : 训练样本数据对应的标签集($N \times 1$ 矩阵)

L : 旋转森林要训练的基分类器数目

K : 特征子集的数目

F : 特征集

$\{\omega_1, \omega_2, \dots, \omega_c\}$: 标签类别集合

过程:

FOR $i=1 \dots L$

生成旋转矩阵 R_i^a :

把特征集 F 划分为 K 个特征子集:

$F_{i,j}(j=1, 2 \dots K)$

FOR $j=1 \dots K$

从训练集 X 中挑出特征子集 $F_{i,j}$ 对应的样本集 $X_{i,j}$

从 $X_{i,j}$ 中随机删除某类标签的样本

对 $X_{i,j}$ 做 bootstrap 采样, 样本容量为 $X_{i,j}$ 的 75%, 新样本集为 $X'_{i,j}$

对 $X'_{i,j}$ 做 PCA 主成分分析, 得到分量矩阵 $C_{i,j}$

END FOR

把得到的分量矩阵 $C_{i,j}(j=1 \dots K)$ 组合成一个对角矩阵(系数旋转矩阵) R_i

按原来的特征排列顺序重新调整矩阵 R_i 的列, 得到旋转矩阵 R_i^a

利用新的训练集 $[XR_i^a, Y]$ 生成基分类器 $BASE_i$

END FOR

分类部分:

对于一个待分类样本 x , 分类结果为:

$$BASE^*(x) = \operatorname{argmax}_y \frac{1}{M} \times BASE_i(x)$$

$$BASE_i(\mathbf{x}) = d_{ij}(\mathbf{x}R_i^\alpha)$$

要注意的是, D_i 对待识别样本 \mathbf{x} 进行分类时, \mathbf{x} 也要先通过对应的旋转矩阵进行变换。旋转森林的输入参数中, 要事先确定 K 特征子集数目和 L 基分类器的数目。

5 实验与结果

本节将介绍在 DHA 数据集上实验结果和分析。

5.1 度量标准

实验中, 采用了交叉验证的方法 (Cross-Validation) 来训练识别模型和测试性能。同时采用精确度 (Precision)、召回率 (Recall) 和 F 值 (F-measure) 作评价手段来衡量算法的效果, 如公式 (8)~(10) 所示:

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$F = \frac{2RP}{R + P} \quad (10)$$

对于二分类问题来说, TP 对应于被分类模型正确预测的正样本数; FP 对应于被分类模型错误预测为正类的负样本数; FN 为对应于被分类模型错误预测为负类的正样本数。这些公式可以推广到多分类问题中。

5.2 数据集

DHA (Depth-included Human Action) 行为数据集^[17] 包含 17 类人体行为, 每类的人体行为分别由 21 个人完成, 这 21 个人中包含有 12 个男性人物和 9 个女性人物。17 类人体行为有以下动作: 行为 1 (bend)、行为 2 (front-clap)、行为 3 (jack)、行为 4 (jump)、行为 5 (run)、行为 6 (leg-kick)、行为 7 (side)、行为 8 (pjump)、行为 9 (walk)、行为 10 (skip)、行为 11 (rod-swing)、行为 12 (two-hand-wave)、行为 13 (one-hand-wave)、行为 14 (arm-swing)、行为 15 (side-boxing)、行为 16 (tai-chi)、行为 17 (pitch)。

本次研究对象 DHA 数据集包括深度视频分帧图像序列、二值图像序列和 RGB 图像, 而且这 357 组图像序列的分辨率都是 480×640 , 如图 5 所示。



图 5 DHA 数据集样例
Fig. 5 DHA dataset sample

5.3 实验和结果

在本节中, 我们进行 2 个实验来验证方法的可行性和高效性。第 1 个实验在 DHA 数据集上测试单一特征和采用混合特征后的识别率以及 Precision, Recall 和 F-Measure 情况。第 2 个实验是本文方法与其它算法对比情况。

实验 1 我们以混淆矩阵的方式给出本文方法的识别结果, 图 6~8 以混淆矩阵的方式展示了本文提出的方法分别在基于分块均值特征、基于 Gabor 特征和基于混合特征上的行为识别正确率结果。混淆矩阵的第 (i, j) 个元素表示第 i 类行为被分类为第 j 类行为的比例, 因此对角线上的值越大, 分类效果就越好。从上述图中可以看到当采用混合特征时, 17 个动作类别的识别准确率都要高于前面的单一特征。

从图 6 中可以看出, 在 DHA 数据集上, 当采用分块均值特征时, 17 种动作的识别率普通不高, 由于 front-clap、run、side、pjump、two-hand-wave 和 tai-chi 这 6 种动作的相似度较高, 所以采用单一特征时这 6 种动作更容易分错混淆。从图 7 中可以看出, 在采用 Gabor 特征时, jump, side 和 two-hand-wave 这 3 种动作误识率也较低, 识别的时候易混淆, 而从图 8 中可以看出, 将二者结合的混合特征则很好地避免了这点, 其中 jack、run、pjump、side、skip two-hand-wave、tai-chi 和 pitch 八个动作的识别率达到了 100%。

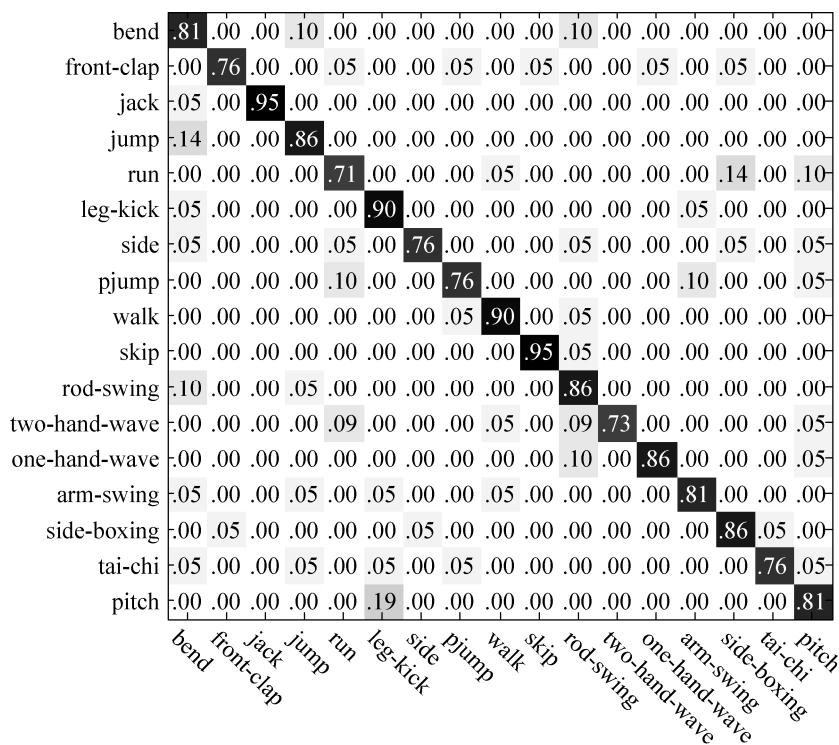


图 6 使用 BMF 特征识别结果混淆矩阵
Fig. 6 Confusion matrix results using BMF features

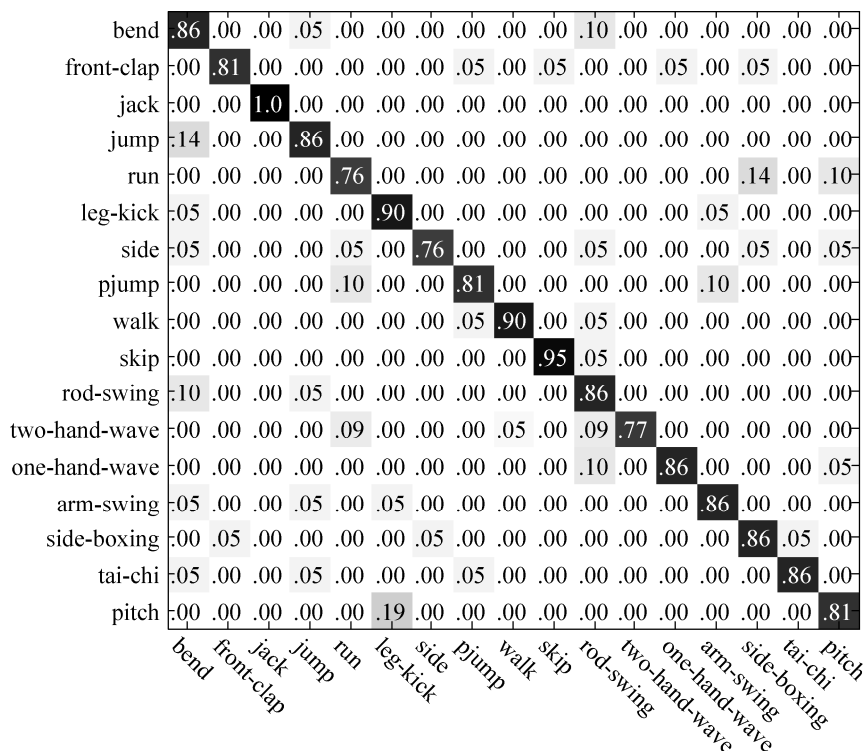


图 7 使用 GF 特征识别结果混淆矩阵
Fig. 7 Confusion matrix results using GF features

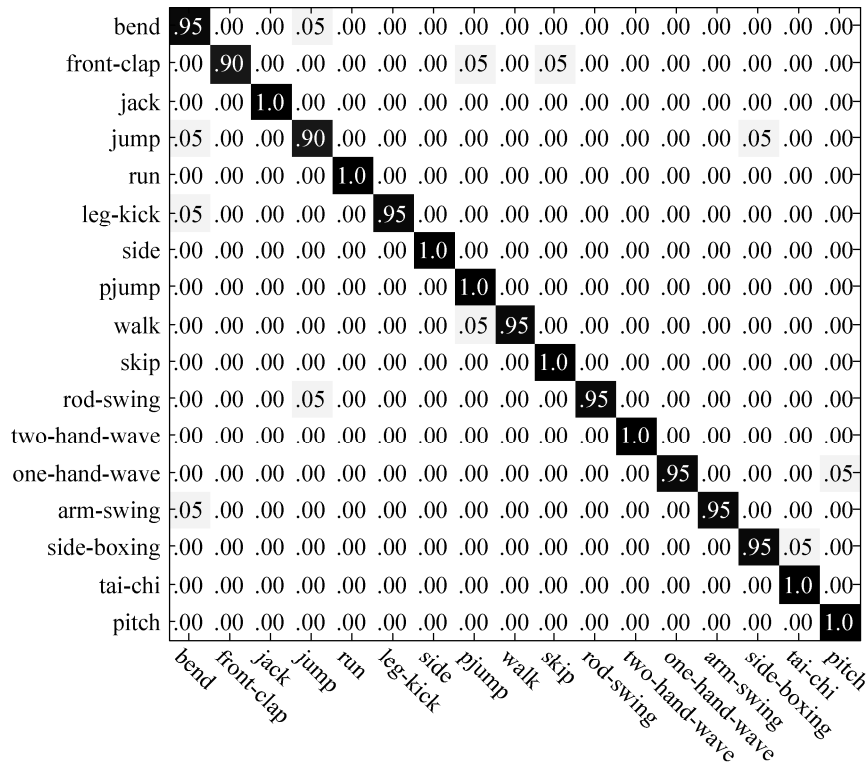


图8 使用混合特征识别结果混淆矩阵

Fig. 8 Confusion matrix results using mixed features

同时我们给出了采用精确率、召回率和 F 值来评估不同特征及其组合情况下的行为识别率见表 1。从表 1 中可以看到采用组合特征 F_{fusion} 的行为识别率高于采用单一特征 BMF 和 GF 特征表示下的行为识别率,可见多特征组合的行为表示有着较高的识别率。

在第 2 个实验中,我们将本文方法与其它方法进行了对比(如表 2 所示)。表 2 列举了本文旋

转森林算法与最近邻(KNN)、Boosting、Bagging、支持向量机(SVM)和人工神经网络(ANNs)识别方法的比较。从表中的比较结果中,我们可以看出,基于 F_{fusion} 特征组合的旋转森林识别算法取得最高的识别率达到 97%。同时,旋转森林识别算法整体性能也优于 KNN、Boosting、Bagging 和 ANNs 方法,取得了与达到了 88% 的平均正确率。

表 1 单特征与混合特征识别准确率

Tab. 1 Recognition accuracy using single feature and mixed features

数据集	特征	Precision	Recall	F-Measure
DHA dataset	F_{fusion}	97	97	97
	BMF	83	84	83
	GF	85	85	85

表 2 不同行为识别方法比较

Tab. 2 Comparison between our proposed method and other methods

数据集	描述子	RF	KNN	Boosting	Bagging	SVM	ANNs
DHA dataset	F_{fusion}	97	91	94	92	96	91
	BMF	83	80	80	79	80	80
	GF	85	81	81	80	80	74
	Avg.	88	84	85	83	85	82

6 结论

本文提出了一种基于深度视频的人体动作识别方法, 通过利用深度差值运动历史图像分别提取了分块均值特征和 Gabor 特征进行融合作为行为描述特征, 并使用旋转森林学习框架作为训练识别工具, 取得较好的学习效果。与传统单一行为特征相比, 本文方法具有较高的识别率, 实时性, 鲁棒性好的优点。行为识别是一个充满挑战性的开放性课题, 虽然本文方法取得了较好的实验结果, 但仍有许多问题需要深入研究。为了获得泛化能力强的分类模型, 通常需要大量的标记训练视频样本, 这需要大量的人工标记劳力, 这给建模带来实际的困难。如何利用大量触手可及的未标记视频样本来提升学习系统的性能成为一个值得研究的方向。

参考文献:

- [1] LO PRESTI L, LA CASCIA M. 3D skeleton-based human action classification: A survey[J]. *Pattern Recognition* (S0031-3203), 2016, 53: 130-147.
- [2] DAS DAWN D, SHAIKH S H. A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector[J]. *Visual Computer* (S0178-2789), 2016, 32(3): 289-306.
- [3] ZIAEEFARD M, BERGEVIN R. Semantic human activity recognition: A literature review[J]. *Pattern Recognition* (S0031-3203), 2015, 48(8): 2329-2345.
- [4] LIM C H, VATS E, CHAN C S. Fuzzy human motion analysis: A review[J]. *Pattern Recognition* (S0031-3203), 2015, 48(5): 1773-1796.
- [5] CHEN C, JAFARI R, KEHTARNAVAZ N. A survey of depth and inertial sensor fusion for human action recognition[J]. *Multimedia Tools and Applications* (S1380-7501), 2017, 76(3): 4405-4425.
- [6] CAI Z, HAN J, LIU L, et al. RGB-D datasets using microsoft kinect or similar sensors: a survey[J]. *Multimedia Tools and Applications* (S1380-7501), 2017, 76(3): 4313-4355.
- [7] ZHANG J, LI W, OGUNBONA P O, et al. RGB-D-based action recognition datasets: A survey[J]. *Pattern Recognition* (S0031-3203), 2016, 60: 86-105.
- [8] ZHANG B, YANG Y, CHEN C, et al. Action Recognition Using 3D Histograms of Texture and A Multi-Class Boosting Classifier[J]. *IEEE Transactions on Image Processing* (S1057-7149), 2017, 26(10): 4648-4660.
- [9] ZHANG H, ZHONG P, HE J, et al. Combining depth-skeleton feature with sparse coding for action recognition[J]. *Neurocomputing* (S0925-2312), 2017, 230: 417-426.
- [10] YANG Y, DENG C, GAO S, et al. Discriminative Multi-instance Multitask Learning for 3D Action Recognition[J]. *IEEE Transactions on Multimedia* (S1520-9210), 2017, 19(3): 519-529.
- [11] UDDIN M Z. Human activity recognition using segmented body part and body joint features with hidden Markov models[J]. *Multimedia Tools and Applications* (S1380-7501), 2017, 76(11): 13585-13614.
- [12] RAHMANI H, MAHMOOD A, DU H, et al. Histogram of Oriented Principal Components for Cross-View Action Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2016, 38(12): 2430-2443.
- [13] AKAGUNDUZ E, ASLAN M, SENGUR A, et al. Silhouette Orientation Volumes for Efficient Fall Detection in Depth Videos[J]. *IEEE Journal of Biomedical and Health Informatics* (S2168-2194), 2017, 21(3): 756-763.
- [14] YANG J, YANG J-Y, ZHANG D, et al. Feature fusion: Parallel strategy vs. serial strategy[J]. *Pattern Recognition* (S0031-3203), 2003, 36(6): 1369-1381.
- [15] LLINAS J, HALL D L. Introduction to multi-sensor data fusion[C]// *Proc of the 1998 IEEE International Symposium on Circuits and Systems*. NEW YORK, NY: IEEE, 2002: 537-540.
- [16] RODRIGUEZ J J, KUNCHEVA L I. Rotation forest: A new classifier ensemble method[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2006, 28(10): 1619-1630.
- [17] LIN Y C, HU M C, CHENG W H, et al. Human action recognition and retrieval using sole depth information[C]// *Proc of the 20th ACM International Conference on Multimedia*. Nara, Japan, 2012: 1053-1056.