

文章编号: 1672-8785(2018)11-0001-06

基于计算机视觉的人体异常行为识别综述

向玉开^{1,2} 孙胜利¹ 雷林建^{1,3} 刘会凯^{1,2} 张悦^{1,2}

(1. 中国科学院上海技术物理研究所, 上海 200083;

2. 中国科学院大学, 北京 100049;

3. 上海科技大学, 上海 201210)

摘 要: 人体行为识别是计算机视觉和模式识别领域的研究热点之一。作为人体行为识别的一个重要分支, 人体异常行为检测近年来也不断得到学界及工业界的重视。人体行为识别研究从早期的依赖人体形状特征发展到基于梯度设计的特征检测, 再到当前随着神经网络的新发展, 深度学习开始广泛应用于行为识别。同时由于红外波段具有适应弱光照环境、可全天候检测等优点, 基于该波段的人体行为识别研究开始兴起, 它也必将成为人体行为识别领域中一个新的研究热点。

关键词: 人体行为识别; 异常行为检测; 深度学习; 红外

中图分类号: TP312 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2018.11.001

Review of Human Abnormal Action Recognition Based on Computer Vision

XIANG Yu-kai^{1,2}, SUN Sheng-li¹, LEI Lin-jian^{1,3}, LIU Hui-kai^{1,2}, ZHANG Yue^{1,2}

(1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. ShanghaiTech University, Shanghai 201210, China)

Abstract: Human action recognition is one of the hotspots in the field of computer vision and pattern recognition. As an important branch of human action recognition, abnormal human action detection has arrested attention of academic and business communities constantly. The research on human action recognition has developed from the research based on human shape features to the research based on gradient design. At present, with the new development of neural network, deep learning has been widely used in action recognition. Because infrared wavebands have advantages of dealing with weak light environment and 24-hour monitoring, they have been applied to the research on human action recognition. This will become a new research hotspot in the field of human action recognition.

Key words: human action recognition; abnormal action detection; deep learning; infrared

收稿日期: 2018-09-10

作者简介: 向玉开(1994-), 男, 湖南湘西人, 硕士研究生, 主要研究方向为图像处理及计算机视觉。

E-mail: yk_xiang@126.com

0 引言

人体异常行为检测属于人体行为识别的范畴,即对特定应用场景下的特定行为进行识别和检测,比如独居老人的跌倒行为等。基于计算机视觉的人体行为识别是对含有人体行为动作的视频序列生成相应的动作类型标签,目前已经广泛应用于视频监控、视频检索、人机交互、智能家居和辅助医疗等领域。

根据算法所用数据来源的不同,人体行为识别可以分为基于可见光波段的人体行为识别和基于红外波段的人体行为识别两种。随着计算机视觉和模式识别领域在可见光图像分类、识别和分割等任务中取得显著成果,基于可见光波段的人体行为识别也获得进展。当前大部分的人体行为识别系统就是基于可见光波段进行设计的。

与可见光图像相比,红外图像由于对光照变化不敏感而具有全天候监控的特点,且能够降低复杂背景带来的影响,同时还更加保护被监测者的隐私。因此,近年来基于红外图像的人体行为识别研究越来越受到重视。

人体行为识别主要分为特征提取和分类识别两部分。其中,特征提取是在视频数据中提取能够表征行为动作的特征,其好坏直接影响动作分类的结果,所以是人体行为识别的关键。分类识别则是将提取出的特征向量输入到分类器中进行学习建模,然后使用学习到的模型判定测试视频序列中的行为动作类型。

随着深度学习理论不断发展,其在识别、分类等任务中的准确率和鲁棒性逐步提高。将神经网络应用于人体行为识别已成为当前研究的热点。

1 主流数据集的介绍及对比

当前国内外有多个公开的人体行为数据集。通过提出完备的数据集,可以有效验证相关算法的可行性。同时,公开的数据集有利于研究者方便地对比不同算法的性能。

1.1 Weizman 行为数据集

Weizman 数据集由以色列 Weizman 科学研究所录制拍摄,包含走路、原地跳跃、单腿跳跃、双腿跳跃、侧身跳跃、向前跳跃、快跑、弯腰、单手挥手和双手挥手等 10 种动作。其中每个动作由不同的 10 个人演示完成。该数据集在固定背景、固定视角的条件下拍摄,且已给出场景中运动前景的轮廓^[1]。

1.2 KTH 数据集

KTH 数据集包含走、跳、跑、挥手、拍手和击拳等 6 种动作,由不同的 25 个人在 4 个不同场景下演示完成,共 599 个视频序列。该数据集中存在拍摄的远近变化以及摄像机自身的轻微抖动^[2]。

1.3 UCF Sports 数据集

UCF Sports 数据集包含从各种体育频道(如 BBC 等)和互联网视频网站(如 YouTube 等)上收集到的视频序列。它来源于真实的生活场景,故其涉及的场景类型广泛,拍摄视角多样。该数据集包含一系列子数据集,其命名根据数据集含有动作类别数目或数据集来源的不同而存在略微差异,如 UCF YouTube、UCF50 等^[3]。

1.4 Hollywood 数据集

Hollywood 数据集是从 32 部好莱坞电影中采集到的,包含打电话、下车、握手、拥抱、接吻、坐下、坐着和站起来等 8 个类别的动作,总共有 1707 个视频。因为样本来源于不同的电影,所以数据集中含有人体姿态、着装、光照和遮挡等方面的变化^[4]。

1.5 异常行为数据集(多相机跌倒数据集)

异常行为数据集总共拍摄了 24 个室内生活场景的行为视频(共计 192 个),其中每个场景使用 8 台相机从不同视角拍摄。该数据集包含多种日常生活行为(如搬运箱子、打扫、脱外套、坐下等),同时也包含跌倒行为的多种情况,即行走跌倒(前倾跌倒、后仰跌倒、侧身跌倒)、站立失衡跌倒、静坐起身跌倒等^[5],如图 1 所示。

1.6 InfAR 红外行为识别数据集

InfAR 红外行为识别数据集在红外条件下拍摄, 包含打架、拍手、握手、慢跑、拥抱、双腿跳、单腿跳、拳击、推搡、单手挥舞、双手挥舞和走路等 12 种日常行为, 其中每个行为类型为 50 个视频序列, 总计 600 个视频序列。所有行为均由 40 个不同的人完成, 每个视频样本均由一个人或者多个人完成一个或多个行为, 其中也包含多人的交互行为。该数据集还引入了遮挡、季节与视角的变化等因素^[6], 如图 2 所示。

在公开的人体行为识别数据集中, 大部分

都是基于可见光波段构建的, 近年来才开始出现基于红外视频的行为识别数据集。同时, 针对特定应用场景中的异常行为所构建的数据集还相对较少, 无法满足对特定场景中的异常行为进行识别和检测的需要。

2 基于可见光波段的异常行为识别研究的现状及趋势

由于可见光波段的人体行为数据易于获取, 基于该波段的人体行为识别研究开始较早。其研究方法从早期的基于人体几何特征、运动信息等来识别人体行为, 发展到近年来基

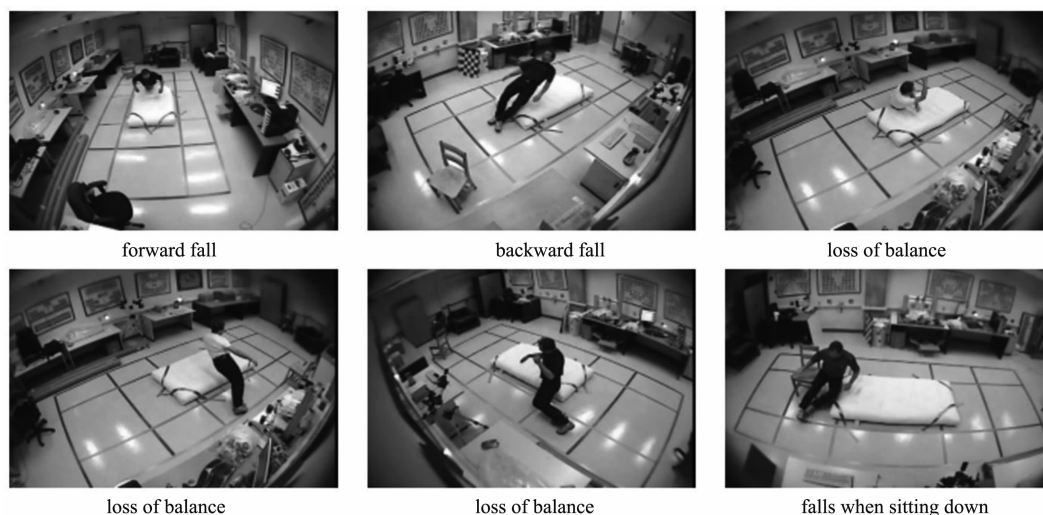


图 1 多相机跌倒数据集

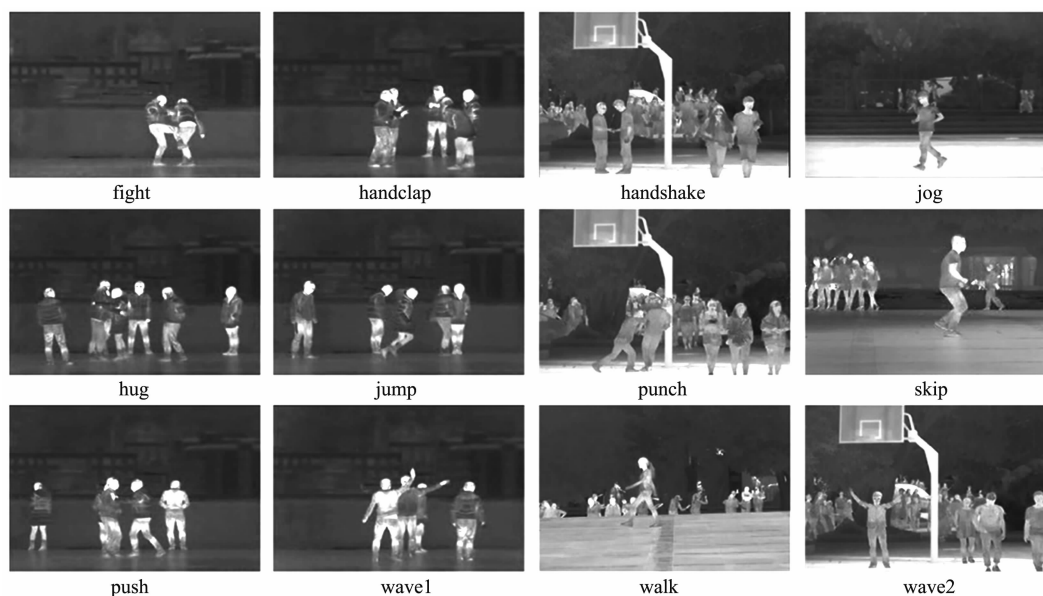


图 2 InfAR 红外行为识别数据集

于 HOG、SIFT 等图像特征(梯度等)提取方法进行行为识别,以及当前将深度学习理论应用于人体行为识别研究并取得了不错的成果。

2.1 人体行为识别研究现状及趋势

2.1.1 基于传统方法的行为识别

人体行为运动会直接引起图像中的人体几何轮廓呈现规律性变化,故可通过对人体几何特征进行分析和设计来表征人体行为运动特性的几何特征,以达到识别动作的目的。

Fujiyoshi H 等人^[7]首先对视频中的人提取人形轮廓,然后根据所得轮廓计算人体重心和四肢及头部顶点,并将其组成星状图来表征当前人体姿态,同时以由重心与顶点构成的矢量作为特征向量。人体几何特征常常与表征运动状态的运动历史图或运动能量图相结合,用于表征人体行为动作特征。Bobick A F 等人^[8]提出用运动历史图和运动能量图表征人体运动,并以模板匹配来识别人体行为。

基于人体几何轮廓变化设计的运动特征对运动的表达较为粗略。而随着 SIFT 特征和 HOG 特征的相继提出,基于图像梯度的特征提取方法开始应用于计算机视觉及模式识别,其对运动的特征描述也更为细致。

Hsiao P C 等人^[9]通过将不同帧的人体运动轮廓中心校正对齐并将运动轮廓叠加,得到了运动历史图;然后计算基于边缘方向直方图的 TSSC 特征,以反映动作在时间上的形态变化。Zhang H B 等人^[10]通过用光流表征运动信息来定位感兴趣区域,然后使用其方向梯度和光流直方图来描述该区域,最后利用支持向量机(Support Vector Machine, SVM)分类器对人体的各种动作类型进行分类。Zhang Z 等人^[11]利用基于 SIFT 特征的 ASIFT 进行关键点轨迹表达以获得关于人体行为动作的全局时空信息,再通过 LDA 描述特征,最后由 K-means 进行分类以识别人体行为。

2.1.2 基于深度学习的行为识别

与人为设计的特征表达相比,基于神经网络的深度学习模型可以学习到更加抽象的高维

特征,并具有更强的泛化性能和鲁棒性。

作为一种前馈神经网络,卷积神经网络(Convolution Neural Network, CNN)利用参数共享来减小神经网络训练中的参数膨胀问题。在卷积层对输入数据进行卷积,然后将其输出并通过一个非线性激活函数作非线性映射。在卷积层之间加入池化层,用于压缩数据和参数的量以防止过拟合。最后,在全连接层的输入分类器中进行最终识别。

Ji S 等人^[12]提出了一种用于行为识别的三维卷积神经网络模型。该模型运用三维卷积分别在空间和时间上提取特征,然后将两者特征结合,获得最后的特征表示,从而获得多个相邻帧间的运动信息。

长短期记忆(Long Short-Term Memory, LSTM)网络是一种特殊的循环神经网络(Recurrent Neural Network, RNN)。RNN 是一类用于处理时间序列数据的神经网络,而 LSTM 网络尤其适合处理和预测时间序列中间隔和延迟相对较长的任务。Gammulle H 等人^[13]提出用卷积神经网络学习显著的空间特征,然后将卷积神经网络卷积层和全连接层学到的特征信息输入 LSTM 网络中,接着学习这些空间特征的时序相关性,最终输出人体行为识别数据。

2.2 人体异常行为识别研究现状及趋势

随着人体行为识别的不断进步,针对特定应用场景的异常行为识别也越来越受到研究者的重视。

Chua J L 等人^[14]根据人体形状特征,设计和计算三个分别表征人体头部、上身及腿的点,然后由三点间的距离、方向等变化表征人体动作,并通过应用不同动作在这些特征中的阈值差异来识别动作类别。Rougier C 等人^[15]根据人处于跌倒等动作时,其动作在空间和时间上都会有很大幅度的变化这一特点,利用运动历史图检测出大的运动动作,然后根据人体形状特征在跌倒与行走之间的差别,区分行走和跌倒等行为。Wang S 等人^[16]提出利用

PCANet 识别人体异常行为,即以浅层的 PCANet 深度学习网络提取人体行为特征,然后通过用提取的特征训练线性 SVM 来分类识别跌倒等行为动作。

Núñez-Marcos A 等人^[17]利用卷积神经网络判断跌倒行为。首先根据 RGB 计算光流图以表达运动信息,然后将得到的光流图输入到基于 VGG-16 的卷积神经网络中以得到用来训练分类器的特征向量,最后将表征动作行为的特征向量输入到由全连接层和 softmax 层组成的网络中得到类别概率,并由此判定相应动作的类别。

近年来,针对特定应用场景的异常行为识别也逐渐引入了深度学习理论,以使模型学习到具有更强泛化性能的特征并增强鲁棒性。

3 基于红外波段的人体异常行为识别研究的现状及趋势

由于基于可见光波段的人体行为识别系统容易受到光照变换的影响,而且在弱光照环境下不能很好地进行人体行为识别,同时还存在遮挡以及可能的隐私泄露等其他影响因素,因此近年来基于红外波段的人体行为识别研究越来越受到学界的重视。

3.1 人体行为识别研究现状及趋势

Han J 等人^[18]提出在红外运动人体轮廓提取的基础上引入步态能量图,将人体行为序列表示在一张图中以保存人体行为的时空信息,然后再将所提取的时空信息输入贝叶斯分类器进行分类识别。

基于步态能量图的方法虽然能够表征人体运动的时空信息,但是依据人体动作轮廓叠加形成的运动图样并不能详尽地表达运动信息,而基于光流的方法则可表征人体不同部位的不同速度变化,因而具有更为详尽的运动表征。Gao C 等人^[19]提出了一个新的红外人体行为识别数据集——InfAR Dataset,并利用光流-运动历史图-卷积神经网络(Optical-Flow Motion History CNN, OF-MHI-CNN)模型得到了表征运动信息的红外运动图像的光流图像。在此基

础上获得了运动历史图,并将此输入神经网络,最终输出了人体行为类别。

为了处理时序数据同时又能利用卷积神经网络良好的学习特征的性能,三维卷积神经网络的出现很好地解决了以上问题。它可以同时在时间和空间维度上进行卷积操作;在获取单帧特征的同时,也能表达相邻帧之间运动状态随时间变化相匹配的空间变化。

Jiang Z 等人^[20]首次将三维卷积神经网络应用于红外人体行为识别,即采用双流结构分别将红外图像序列以及基于红外图像序列获得的光流图像序列分别输入到三维卷积神经网络中,然后对双流得到的输出进行融合,经 Softmax 层将其输出后得到行为类型。Liu Y 等人^[21]提出了基于光流累计图的全局时空信息特征表达,即分别基于光流图、光流运动历史图和光流累计图以卷积神经网络获取局部、空间、全局时序的信息,最后将三层流的信息融合输入 SVM 中得到最终的行为类别,实现了 InfAR 数据集上 79.25% 的准确率。

3.2 人体异常行为识别研究现状及趋势

Mastorakis G 等人^[22]利用 Kinect 的红外传感器获得深度信息,从而得到人体的三维检测框。然后根据三维信息计算高度方向上的运动速度,同时将宽度和深度方向上的向量综合为一个向量。最后根据高度方向上的速度变化并结合宽度-深度向量的值的变化,判断人体是否处于跌倒状态中。

由于基于红外视频的异常行为识别缺乏公开的行为数据集,当前基于红外波段的异常行为检测研究还相对较少。

4 面临的挑战

(1) 缺乏针对特定场景的异常行为识别的公开数据集,制约了相应研究的发展;

(2) 与可见光波段相比,红外波段缺少颜色、纹理等信息,使得数据输入端的信息量减少;

(3) 在利用三维卷积神经网络处理时序信息时,模型不易收敛且模型参数过于庞大。

5 结束语

基于计算机视觉的人体行为识别具有广阔的应用前景。当前深度学习理论已在基于可见光波段的人体行为识别中得到了应用,并取得了良好的成果。近年来,随着基于红外波段的人体行为数据集的提出,红外人体行为识别研究越来越受到学界的重视。同时,针对特定应用场景的异常行为识别研究方兴未艾。以后基于红外波段的人体行为识别必将被引入到特定应用场景下的异常行为检测中。由于可实现高泛化性和鲁棒性,该方法必将成为人体行为识别领域中一个新的研究热点。

参考文献

- [1] Gorelick L, Blank M, Shechtman E, et al. Actions as Space-time Shapes [C]. Beijing: 10th IEEE International Conference on Computer Vision, 2005.
- [2] Schuldt C, Laptev I, Caputo B. Recognizing Human Actions: A Local SVM Approach [C]. Cambridge: 17th International Conference on Pattern Recognition, 2004.
- [3] Soomro K, Zamir A R. Action Recognition in Realistic Sports Videos [M]. Lausanne: Springer, 2014.
- [4] Marszalek M, Laptev I, Schmid C. Actions in Context [C]. Miami: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [5] Auvinet E, Rougier C, Meunier J, et al. Multiple Cameras Fall Dataset [R]. DIRO: Technical Report 1350, 2010.
- [6] Gao C, Du Y, Liu J, et al. InfAR Dataset: Infrared Action Recognition at Different Times [J]. *Neurocomputing*, 2016, **212**: 36–47.
- [7] Fujiyoshi H, Lipton A J. Real-time Human Motion Analysis by Image Skeletonization [C]. Breckenridge: Fourth IEEE Workshop on Applications of Computer Vision, 1998.
- [8] Bobick A F, Davis J W. The Recognition of Human Movement Using Temporal Templates [M]. Washington, DC: IEEE Computer Society, 2001.
- [9] Hsiao P C, Chen C S, Chang L W. Human Action Recognition Using Temporal-state Shape Contexts [C]. Monaco: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [10] Zhang H B, Li S Z, Guo F, et al. Real-time Human Action Recognition Based on Shape Combined with Motion Feature [C]. Atlanta: IEEE International Conference on Intelligent Computing and Intelligent Systems, 2010.
- [11] Zhang Z, Liu J. Recognizing Human Action and Identity Based on Affine-SIFT [J]. *Electrical & Electronics Engineering*, 2012, **237**(6): 216–219.
- [12] Ji S, Xu W, Yang M, et al. 3D Convolutional Neural Networks for Human Action Recognition [C]. Washington, DC: Ninth International Conference on Machine Learning and Applications, 2010.
- [13] Gammulle H, Denman S, Sridharan S, et al. Two Stream LSTM: A Deep Fusion Framework for Human Action Recognition [C]. Hawaii: IEEE Winter Conference on Applications of Computer Vision, 2017.
- [14] Chua J L, Chang Y C, Lim W K. A Simple Vision-based Fall Detection Technique for Indoor Video Surveillance [J]. *Signal Image & Video Processing*, 2015, **9**(3): 623–633.
- [15] Rougier C, Meunier J, St-Arnaud A, et al. Fall Detection from Human Shape and Motion History Using Video Surveillance [C]. Niagara Falls: International Conference on Advanced Information Networking and Applications, 2007.
- [16] Wang S, Chen L, Zhou Z, et al. Human Fall Detection in Surveillance Video Based on PCANet [J]. *Multimedia Tools & Applications*, 2016, **75**(19): 11603–11613.
- [17] Núñez-Marcos A, Azkune G, Arganda-Carreras I. Vision-Based Fall Detection with Convolutional Neural Networks [J]. *Wireless Communications & Mobile Computing*, 2017, **2017**(1): 1–16.
- [18] Han J, Bhanu B. Human Activity Recognition in Thermal Infrared Imagery [C]. San Diego: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

(下转第 33 页)