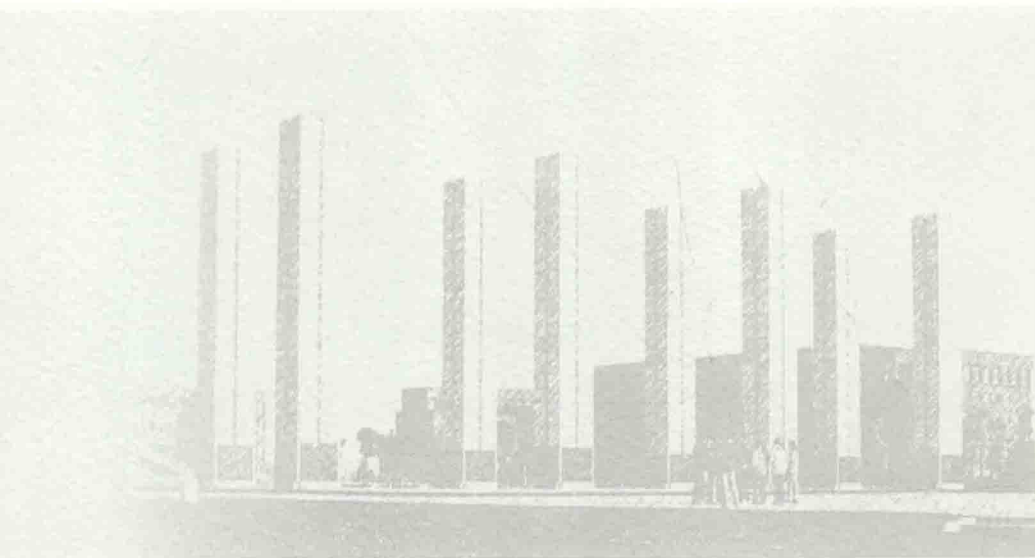


FUZA CHANGJINGXIA RENTI DONGZUO SHIBIE

复杂场景下人体动作识别

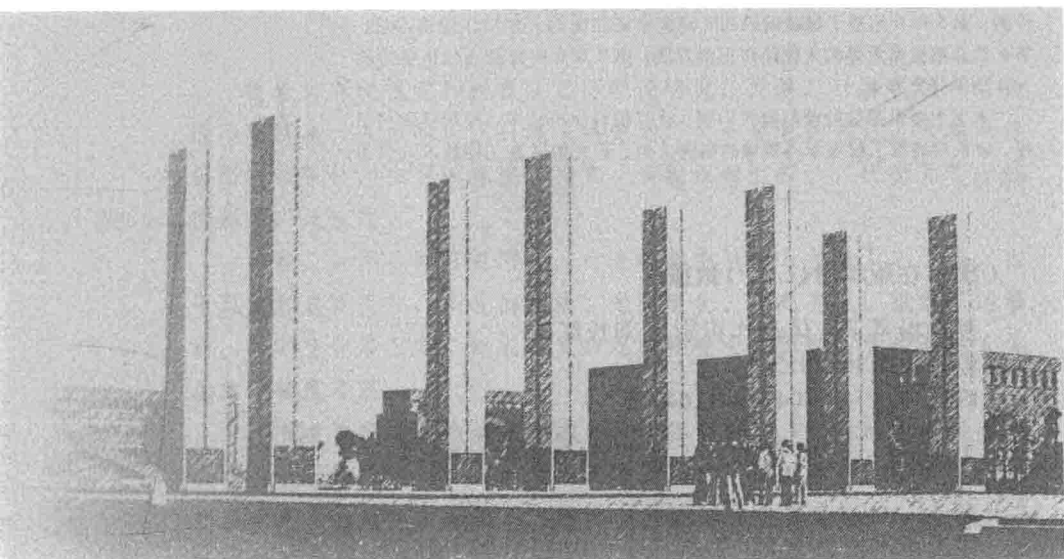


刘长红 著

HUMAN ACTION RECOGNITION
IN COMPLEX SCENE

高等教育出版社

复杂场景下人体动作识别



刘长红 著

HUMAN ACTION RECOGNITION
IN COMPLEX SCENE

高等教育出版社·北京

内容简介

本书在对人体动作识别与行为理解学习研究的基础上,从实际应用角度出发编写而成。全书共7章,重点阐述在现实复杂场景(如人体表现变化、杂乱背景和视角变化等)下的人体动作识别技术和方法。第1章介绍人体动作识别的意义、典型应用和难点问题;第2章主要对人体动作识别研究现状进行介绍;第3章论述在杂乱背景和摄像机移动场景下的时空兴趣点检测方法;第4章介绍基于视频抖动检测算法的混合时空兴趣点检测;第5章介绍基于稀疏编码的时空金字塔匹配的人体动作识别方法;第6章介绍视角无关的人体动作识别方法;第7章介绍复杂场景下鲁棒的人体动作分类方法。

本书主要供图像处理与模式识别、计算机科学与技术、控制科学与工程、信息与通信工程等学科领域的研究人员、研究生以及工程技术人员参考使用。

图书在版编目(CIP)数据

复杂场景下人体动作识别 / 刘长红著. -- 北京: 高等教育出版社, 2015. 1
ISBN 978 - 7 - 04 - 041164 - 5

I. ①复… II. ①刘… III. ①图象识别 IV. ①TP391.41

中国版本图书馆 CIP 数据核字(2014)第 226128 号

策划编辑 冯 英 责任编辑 冯 英 封面设计 李卫青 版式设计 于 婕
插图绘制 郝 林 责任校对 张小楠 责任印制 赵义民

出版发行	高等教育出版社	咨询电话	400 - 810 - 0598
社 址	北京市西城区德外大街 4 号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	北京机工印刷厂	网上订购	http://www.landraco.com
开 本	787mm × 1092mm 1/16		http://www.landraco.com.cn
印 张	8.75	版 次	2015 年 1 月第 1 版
字 数	130 千字	印 次	2015 年 1 月第 1 次印刷
购书热线	010 - 58581118	定 价	27.60 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 41164 - 00

前 言

随着计算机视觉和机器学习等理论研究的发展、计算机硬件性能的不断提高,人体动作识别越来越受到广大研究者的关注,人体动作识别的相关技术在视频标注与检索、智能视频监控、人机交互等领域得到了广泛应用。

目前,网络视频日益增多,行业视频应用也日益普及。但是,由于尺度和表观变化、杂乱的背景、光照变化、视角变化、摄像机的移动、时间变化等因素,造成在复杂现实环境中,识别不受控制的人体动作准确度不高。

本书主要研究复杂场景下的人体动作识别,分别对时空兴趣点检测和表示、码书学习和人体动作描述子进行了深入分析,提出了新的相应算法,并对复杂场景下人体动作分类方法进行了研究。全书共7章,主要包括:

1. 针对杂乱背景和摄像机晃动/移动环境下难以准确进行时空兴趣点检测的问题,提出一种基于非线性各向异性扩散的时空兴趣点检测方法。在空间域上,利用 Weickert提出的边缘增强非线性各向异性扩散滤波器具有保持和增强边缘信息同时又能平滑掉噪声和图像结构内部的特点来抑制杂乱背景。在时间域上,考虑摄像机晃动/移动所导致的时间关系上的不确定性,采用单向扩散,利用与梯度成正比的抑制函数,来抑制摄像机晃动/移动所导致的影响。从而有效地解决了复杂场景下时空兴趣点检测面临的杂乱背景和摄像机晃动/移动问题。并且利用视频抖动检测算法,提出了一种改进的视频抖动自适应的时空兴趣点检测方法,结合本书基于非线性各向异性扩散的时空兴趣点检测方法和 Dollar提出的时空兴趣点检测方法进行时空兴趣点检测,可以对各种场景下的视频自适应地选择最优检测算法。

2. 针对基于 K均值聚类(K-means)的码书学习方法中存在的问题,构建更具有判别性的高层动作描述子,提出了一种基于稀疏编码的时空金字塔匹配的人体动作识别方法。利用更具有判别性的稀疏编码算法学习码书,计算每个图像块(cuboid)的最紧致且具有判别力的稀疏表示,有效解决基于 K-means的码书学习中冗余和初始化问

题,并进一步基于更符合生物视觉系统机制的最大连接池(max pooling)时空金字塔匹配学习视频中人体动作的时空关系,计算高层的动作描述子。实验表明,该方法比基于K-means的方法可获得更高的识别率,特别是在复杂现实视频的处理中,识别率有较大提高。

3. 针对复杂场景下人体动作的视角变化问题,提出了一种基于线性动态系统的人体动作识别方法。利用局部动态性作为特征,对局部块采用线性动态系统进行建模,将得到的模型参数作为局部表示,从而既考虑人体动作的动态性,又对少量人体表观、部分遮挡、光照变化保持不变。利用时序上的动态性作为局部表示,充分捕捉到人体动作变化的过程,而不是考虑人体表观的各种特征,从而有效地实现了视角无关的动作识别。实验表明,该方法对任意角度、任意运动方向的视频序列都得到较好的识别结果,而且对于交叉视角识别也可获得很好的性能。

4. 提出了一种基于稀疏表示的更具有判别性和鲁棒性的动作分类方法。利用稀疏表示的判别性能力和自适应选择 K 近邻数,解决 K 最近邻(K -NN)分类方法中对近邻数 K 值的敏感性问题,而且提高了分类性能。通过扩展基本的稀疏表示求解方程,包含一个错误项,使得在动作分类的同时还能有效地处理噪声、遮挡、干扰等的影响。

人体动作识别的研究具有重要的科学意义,同时在现实生活中具有广泛的应用前景,目前的研究仍在不断深入中。

本书由江西师范大学博士文库出版资助项目资助出版。在编写过程中得到江西师范大学计算机信息工程学院的大力支持,感谢学院领导、同事在各方面给予的关心和帮助,感谢我的学生。同时,我还要感谢我的家人,是他们的理解与默默支持,使得我能全心写作,顺利完成本书。

由于作者水平有限,书中不妥之处在所难免,恳请读者批评指正。

刘长红

2014年11月

目 录

第1章 引言	1
1.1 应用背景	1
1.1.1 视频标注与检索	1
1.1.2 智能视频监控	2
1.1.3 人机交互	2
1.2 研究的难点	3
1.3 研究内容和结构安排	5
1.3.1 研究内容及工作进展	5
1.3.2 结构安排	6
第2章 人体动作识别的相关综述	8
2.1 视频图像表示	9
2.1.1 全局表示	10
2.1.2 局部表示	12
2.2 人体动作建模与分类	20
2.2.1 维数约简	20
2.2.2 直接分类	20
2.2.3 状态空间模型方法	22
2.3 方法的特点	25
2.4 本章小结	26
第3章 杂乱背景和摄像机移动下的时空兴趣点检测方法	28
3.1 引言	28
3.2 非线性各向异性扩散滤波器及分析	31
3.2.1 Perona-Malik扩散模型	33
3.2.2 非线性各向异性扩散滤波	34
3.3 杂乱背景和摄像机移动下的时空兴趣点检测	35
3.3.1 基于非线性各向异性扩散滤波的时空兴趣点检测算法	36
3.3.2 计算方法	39
3.4 实验与分析	40

3.4.1 实验数据集和实验设置	41
3.4.2 摄像机轻微晃动下的时空兴趣点检测方法比较	41
3.4.3 杂乱背景和摄像机快速运动下的时空兴趣点检测方法比较	43
3.4.4 计算复杂度分析	45
3.5 本章小结	46
第4章 改进的视频抖动自适应的时空兴趣点检测方法	48
4.1 引言	48
4.2 视频抖动检测算法	48
4.2.1 视频画面抖动检测算法	50
4.2.2 金字塔 Lucas-Kanade 稀疏光流特征	52
4.2.3 稀疏光流的前-后向误差估计	53
4.2.4 基于光流运动熵的视频画面抖动检测算法	54
4.3 基于视频抖动检测算法的混合时空兴趣点检测	55
4.4 实验与分析	56
4.4.1 实验数据集和实验设置	56
4.4.2 视频抖动检测验证	56
4.4.3 静态背景下的时空兴趣点检测方法比较	57
4.4.4 摄像机轻微晃动下的时空兴趣点检测方法比较	58
4.4.5 杂乱背景和摄像机快速运动下的时空兴趣点检测方法比较	59
4.5 本章小结	61
第5章 基于稀疏编码的时空金字塔匹配的人体动作识别	62
5.1 引言	62
5.2 稀疏编码	63
5.2.1 数学描述	64
5.2.2 稀疏编码算法	64
5.3 基于 BoF 的人体动作识别框架	66
5.4 基于稀疏编码的人体动作识别方法	67
5.4.1 特征检测和表示	68
5.4.2 基于稀疏编码的码书学习	69
5.4.3 基于 max pooling 的人体动作描述子	70
5.5 基于稀疏编码的时空金字塔匹配	71
5.5.1 时空金字塔匹配	71
5.5.2 基于 max pooling 的时空金字塔匹配	72
5.6 实验与分析	73
5.6.1 实验设置	74

5.6.2 KTH数据集上的实验	75
5.6.3 YouTube数据集上的实验	78
5.7 本章小结	79
第6章 视角无关的人体动作识别的研究	81
6.1 引言	81
6.2 视角无关的人体动作识别的研究现状	82
6.3 线性动态系统	84
6.3.1 模型参数估计	85
6.3.2 距离度量	86
6.4 基于线性动态系统的视角无关的人体动作识别	88
6.4.1 基于 LDSs模型参数的局部描述子	89
6.4.2 模型参数空间中的码书学习	90
6.4.3 动作描述子计算	91
6.5 实验与分析	91
6.5.1 数据集与实验设置	91
6.5.2 单/多目识别	92
6.5.3 交叉视角的识别	95
6.5.4 基于 BoF的视角无关的动作识别方法比较	96
6.6 本章小结	98
第7章 复杂场景下鲁棒的人体动作分类方法	99
7.1 引言	99
7.2 超完备基的图像稀疏表示	100
7.3 基于稀疏表示的人体动作分类算法	100
7.3.1 人体动作的稀疏表示	101
7.3.2 基于 l_1 最小化的稀疏表示求解方法	101
7.3.3 l_1 最小化优化算法	102
7.3.4 分类方法	102
7.4 在噪声、干扰、遮挡情况下的人体动作分类算法	103
7.5 实验与分析	104
7.5.1 基于 l_1 与 l_2 最小化算法的比较	105
7.5.2 分类方法 SR与 SVM、K-NN的比较	106
7.5.3 鲁棒性测试	107
7.6 本章小结	110
参考文献	111
名词索引	126

第1章 引言

随着计算机视觉和机器学习理论研究的发展,计算机硬件性能的不断提高,如何让计算机能够看见人、认识人、解释及识别人的手势和行为,近年来受到了广大研究者的关注。根据传感器数据来识别和理解人的动作和行为成为“以人为中心的计算”的关键^[1]。人体动作识别,即看见人、识别他们在做什么、然后模仿他们的动作或做出进一步的响应^[2],是我们日常生活中的基本组成部分。对于人类来说,动作识别是件很容易的事,是最基本的技能,但要提出相应的方法使得计算机具有和人相似的能力却是一项艰难而具有挑战性的问题。人体动作识别的研究内容相当丰富,涉及计算机视觉、图像处理、模式识别、神经网络以及机器学习等多学科领域,所以人体动作识别的研究具有重要的科学意义,能够促进多学科领域的发展,而且在现实生活中具有广泛的应用前景。

1.1 应用背景

1.1.1 视频标注与检索

随着多媒体、计算机和网络技术的快速发展以及大容量存储技术的不断提高,网络上出现了大量的图像和视频数据,传统的手工标注方法无法满足人们对图像、视频信息的检索需求。传统的视频标注主要由手工完成,费时费力,而且标注的内容不准确,可能存在二义性和标注不完整现象。因此,为了能够有效地对网络上图像、视频信息进行检索,人们提出了基于内容的视频信息检索。基于内容的视频标注根据视频的内容按照不同的语义概念标注关键字,建立视频索引,进而实现高效视频检索。新闻视频和人体运动视频的标注也是其中的必要部分,然而在新闻视频和人体运动视频的标注中,主要感兴趣的是视频中人的运动特征,通过对视频中的人体运动进行分析获取其运动数据,采用人体动作或行为类别进行描述。在这种情况下,人体动作识别对视频的标注和检索具有极大的帮助。近年来,苏黎世理工学院的计算机

视觉实验室开发了基于姿态分析的视频检索系统^[3],国内也在这方面开展了很多相关的研究,清华大学开发的TV-FI(Tsinghua Video Find It)系统,中国科学院计算所研究的“面向体育训练的三维人体运动仿真与视频分析系统”^[4]等。

1.1.2 智能视频监控

传统视频监控系统集中应用于消防、交通、公共设施等行业,随着视频处理技术的发展和发展的需要,视频监控系统呈现快速发展趋势,智能视频监控系统逐渐扩展到各行各业乃至家庭用户中,成为我们日常生活的一部分。

目前监控摄像机在商业应用中已经普遍存在,但并没有充分发挥其实时主动的监控作用,多数场合尽管装有摄像机,但还是需要专门工作人员监视显示器以便及时发现异常事件,或者往往只是将摄像机拍摄的视频记录下来,当异常情况(如银行ATM机上的违法行为)发生之后,人们才通过记录的视频信息来查看当时发生的事实。然而存储设备的容量是有限的,能保存的视频信息也是有限的,不能长时间地保留记录,所以人们迫切需要一个真正系统化、智能化的视频监控系统,能够连续不间断地进行实时监视,并自动实时分析摄像机捕捉到的视频数据,检测、跟踪其中的人体运动,分析、识别其动作和行为,当异常情况或犯罪行为发生时,系统能准确及时地发出警报,从而有效防止此类事件的发生。因此,人体动作和行为识别在智能视频监控中具有重要的作用,有助于发现异常事件(如盗窃、伤人等)、行人检测等,提高视频监控水平。

1.1.3 人机交互

普适计算促进了人机交互系统的普及和发展,从个人计算机到视频游戏机、从智能手机到数字电视无处不在。传统的人机交互是通过计算机键盘和鼠标进行的,而智能人机交互让目前的计算机尽量摆脱键盘、鼠标等传统交互设备的局限,直接通过语言、手势、动作、表情等人们习惯的自然交流方式进行交互,使计算机能像人一样与我们更加自然便捷地交流。在2003年,日本索尼公司推出的新产品EyeToy^[5]成功地打破了传统键盘、鼠标的交互方式,展示出基于视觉的人机交互方式的巨大市场潜力。日本另外一家著名游戏厂商任天堂继续延续肢体控制游戏的理念,于2006年末推出全新设计的游戏机Wii^[6](图1-1(a))。Wii采用运动感应手柄替代传统的机械手柄,避开计算机视觉

作为传感器在技术上的不足,使得 Wii 的肢体控制能力明显优于 Eye-Toy,从而提供更丰富的游戏内容。随着人体运动视频分析技术的发展,人机交互方式将发生革命性的变化,使得控制机器的方式更容易为人类接受,交互过程自然友好。微软 2010 年末将推出一款动作识别游戏装置 Project Natal^[7](图 1-1(b)),作为 Xbox 游戏机的新控制设备。该设备使用 3D 摄像头与动作识别软件,让用户通过自然的身体运动控制游戏,该款游戏设备的诞生将使得人机交互系统发生突破性的飞跃,完全摆脱了以往机械式控制,达到真正的智能人机交互。因此,要成功进行人机自然而自动地交互,人体动作识别是非常关键的。

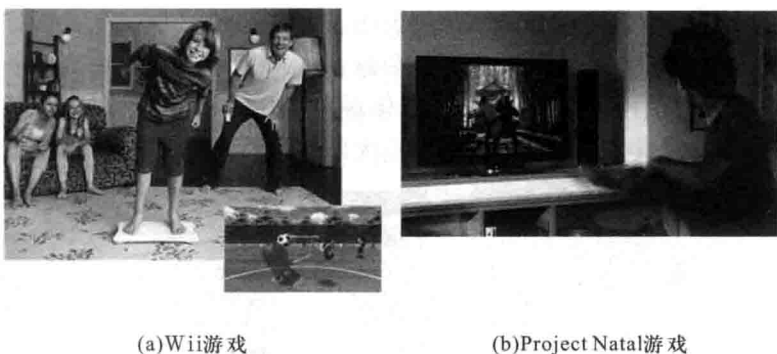


图 1-1 新型的人机交互游戏

除了游戏与娱乐方面的应用,在未来社会发展所需的智能家居,老年人看护等应用中,人体动作识别和行为理解都将起着至关重要的作用。

1.2 研究的难点

最近几年,虽然研究者们提出了大量的人体动作识别方法,但随着互联网上不同视频源(如手持设备、个人收集)的视频日益增多以及视频在各行各业应用的普及,人体动作识别面对处理各种复杂环境下的视频课题。然而由于重大的类内变化、杂乱背景和视角变化等因素,使得复杂场景下现实视频中的人体动作识别成为非常具有挑战性的研究课题。

1. 尺度和表现变化

由于摄像机与目标间距离的变化以及人体自身的个体差异,人体图像呈现不同的大小,而且不同着装和风格导致人体外观的巨大变化,使得在利用全局表示进行人体动作识别时,提取的底层特征(如侧影、轮廓等)具有很大的差异,而在局部表示中,衣服的飘动将产生很多干

扰的兴趣点,所以大小和外观的变化对人体动作识别产生很大的影响,使得同一动作类中存在很大的变化(如图 1-2 所示),而不同的动作可能导致相同的观测。因此人体动作模型的设计非常具有挑战性,既要识别出人体动作的本质特性又要考虑对各种变化的不变性。而且随着人体动作种类数的增加,难度越来越大。

2. 杂乱背景

人们的生活环境各种各样,可能是草地、大海、沙漠等景像比较单一的地方,也可能是运动场、马路、森林等比较复杂的环境,甚至是动态的背景环境(如图 1-2 所示)。背景中的物体可能外观上与人体某些部位相接近,也可能颜色与人体颜色或衣服颜色类似,从而无法很好地将人体与背景区分开来,难以定位图像中的人和观测运动;特别对于全局表示,杂乱的背景将大大影响人体跟踪或背景减除,而在局部表示中,动态的背景将使得检测的兴趣点大部分为噪声或位于背景物体上。

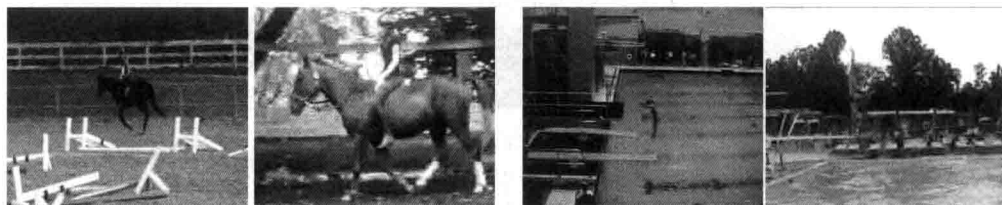


图 1-2 表观变化和杂乱背景

3. 光照变化

在各种不同的环境(室内和室外)和天气的光照条件下,视频图像中的颜色和对比度发生很大变化(如图 1-3 所示),进一步影响了人体的表现,严重情况下,甚至无法辨别人体动作。

4. 视角变化

对于现实环境下的视频,我们无法约束人体关于摄像机的视角和方向,从而导致一个姿态或运动存在无数可能的观测(如图 1-3 所示),所以现实复杂场景下的人体动作识别,必须解决视角变化的问题。



图 1-3 光照变化与视角变化

5. 摄像机移动

摄像机的移动进一步加剧了人体动作识别的难度,无论是在使用局部表示的方法中还是使用全局表示的方法中,同一人体动作都将产生差异很大的特征。摄像机的移动将改变人与摄像机之间的视角和方向,从而得到很大差异的侧影和轮廓,而且摄像机的移动使得在局部兴趣点检测中,背景物体和噪声同样产生很大的局部响应,导致所提取的兴趣点大部分为背景物体和噪声。

6. 时间上的变化

通常人体动作识别的对象是假定已经分割好的视频,但在现实场景下进行人体动作识别时,动作的分割和识别是同时进行的。所以,对连续的人体动作建模也是有待解决的问题。

在复杂现实场景下的人体动作识别中,特别是对于网络视频的标注,所有上述难点都是并存的,而且是不受控制的。所以,如何设计图像表示和建立人体动作模型,保持对以上变化的不变性,使人体动作识别真正满足现实的需求,还有待进一步研究。

1.3 研究内容和结构安排

1.3.1 研究内容及工作进展

本书主要研究复杂场景下的人体动作识别,针对以上存在的难点,采用基于特征袋(BoF)的人体动作识别框架,对基于 BoF 方法中的时空兴趣点检测和表示、码书学习和动作描述子几个关键技术深入展开研究,以提高复杂场景下人体动作识别的性能,期望得到一个更加通用的人体动作识别系统;并针对复杂场景下人体动作分类进一步进行了研究。本书的研究内容主要涉及以下几个方面。

针对杂乱背景和摄像机晃动/移动环境下难以准确进行时空兴趣点检测的问题,提出一种基于非线性各向异性扩散的时空兴趣点检测方法。在空间域上,利用 Weickert 提出的边缘增强的非线性各向异性扩散滤波器具有保持和增强边缘信息同时又能平滑掉噪声和图像结构内部的特点,来抑制杂乱背景。在时间域上考虑摄像机晃动/移动所导致的时间关系上的不确定性,采用单向扩散,利用与梯度成正比的抑制函数,来抑制摄像机晃动/移动所导致的影响。从而更有效地解决复杂场景下时空兴趣点检测面临的杂乱背景和摄像机移动问题。并且利用视频抖动检测算法,提出了一种改进的视频抖动自适应的时空兴趣点

检测方法,结合基于非线性各向异性扩散的时空兴趣点检测方法和 Dollar 提出的时空兴趣点检测方法进行时空兴趣点检测,从而可以对各种场景下的视频自适应地选择最优的检测算法。

针对基于 K-means 的码书学习方法中存在的问题和构建更具有判别性的高层动作描述子,提出一种基于稀疏编码的时空金字塔匹配的人体动作识别方法。利用更具有判别性的稀疏编码算法学习码书和计算每个 cuboid 的最紧致且具有判别力的稀疏表示,从而有效地解决基于 K-means 的码书学习中冗余和初始化问题;而且进一步基于 max pooling 的时空金字塔匹配学习视频中人体动作的时空关系,计算高层的动作描述子,避免传统的基于 BoF 的人体动作识别框架未考虑视频中时空关系的问题,并且更符合生物视觉系统机制。

针对复杂场景下人体动作识别中的视角变化问题,提出一种基于线性动态系统的人体动作识别方法。整个识别框架基于局部表示的方法,能够对少量人体表观、部分遮挡、光照变化保持不变,而且对局部 cuboid 采用线性动态系统建模,利用时序上的动态性作为局部表示来解决表观特征对视角变化的敏感性。动态性特征充分捕捉人体动作变化的过程,而不是考虑人体表观的各种特征,从而有效地实现人体动作识别中对视角变化的不变性。

针对目前常用分类方法对噪声、遮挡、干扰很敏感和 K-NN 分类方法受近邻数 K 值影响的问题,得到更具有判别性而且鲁棒的人体动作分类,提出一种基于稀疏表示的人体动作分类方法。稀疏表示方法能自适应地在每类或多类中选择最好表示测试样本的最少训练样本,得到最紧致且具有判别性的表示,避免 K-NN 方法中对参数 K 值的敏感性而且提高了分类性能。通过扩展基本的稀疏表示求解方程,包含一个错误项,能够有效处理噪声、遮挡、干扰等导致的影响。

1.3.2 结构安排

第1章首先介绍本书研究的背景及意义,阐述视频人体动作识别的典型应用以及复杂场景下人体动作识别所存在的难点,然后简要描述本书的主要内容和结构安排。

第2章对人体动作的图像表示和动作建模与分类进行讨论,介绍近几年来国内外在人体动作识别方面研究现状,并给出了目前人体动作识别方法中存在的问题,指明研究的方向。

第3章介绍本书提出的杂乱背景和摄像机移动场景下的时空兴趣点检测方法。首先介绍和分析目前人体动作识别中时空兴趣点检测存

在的问题,然后阐述非线性各向异性扩散模型并对其进行分析。在此基础上详细阐述本书提出的基于非线性各向异性扩散的时空兴趣点检测方法,并给出有杂乱背景和摄像机移动场景下的现实视频中时空兴趣点检测的实验结果和分析。

第4章介绍基于视频抖动检测算法的混合时空兴趣点检测。首先介绍视频抖动检测算法;然后阐述基于视频抖动检测算法的混合时空兴趣点检测;最后在公开数据集上进行实验和分析,在静态视频和抖动视频上验证提出的混合算法,并与 Dollar 提出的兴趣点检测方法和基于非线性各向异性扩散的时空兴趣点检测方法进行对比分析。

第5章介绍基于稀疏编码的时空金字塔匹配的人体动作识别方法。首先介绍传统基于 K-means 的人体动作识别框架中存在的问题,然后对稀疏编码算法做简单的介绍;接着阐述基于 BoF 的人体动作识别框架和基于稀疏编码的时空金字塔匹配的人体动作识别算法;最后在公开数据库上给出了实验结果,并对这两种方法进行对比分析。

第6章介绍本书提出的视角无关的人体动作识别方法。首先分析人体动作识别中目前局部图像特征表示在视角变化场景下存在的问题,提出基于线性动态系统的局部描述子;然后介绍线性动态系统的模型参数估计方法和距离度量方法,在此基础上详细介绍基于线性动态系统的人体动作识别方法;最后在多视角数据集上对视角无关的人体动作识别方法进行验证,给出实验结果和对比分析。

第7章介绍复杂场景下鲁棒的人体动作分类方法。首先分析目前常用分类方法在复杂场景下人体动作分类中存在的问题;然后详细阐述基于稀疏表示的人体动作分类方法,介绍人体动作的稀疏表示、基于最小化 l_1 范数的稀疏表示求解方法和分类方法,并且进一步论述在噪声、遮挡、干扰等情况下的人体动作分类;最后在鲁棒性测试数据集上进行验证和分析。

第2章 人体动作识别的相关综述

人体动作识别是人体运动分析中的重要组成部分,与人体运动分析的其他步骤紧密相关。近年来已有多位研究者对基于视频的人体运动分析和识别进行了回顾和综述。Gavrila 从二维方法、三维方法和识别回顾了人体运动分析的各种方法^[8]。Aggarwal 和 Cai 从人体结构分析、单一视角或多摄像机下人的运动跟踪及图像序列中人的行为识别三个方面对较早前的研究工作做了回顾^[9]。王亮等人从运动检测、运动目标分类、人的跟踪、行为理解与描述几个方面详细介绍了 2000 年以前的研究状况和各种处理方法,并对研究难点及未来的发展趋势作了较为详细的分析^[10]。Moeslund 等从系统功能角度将基于视觉的人体运动分析分为几个不同的阶段:初始化(initialization)、跟踪(tracking)、姿态估计(pose estimation)、识别(recognition),然后从这几方面来分别详述了近期的研究进展、各种具体技术和方法,并对进一步有待解决的问题进行了讨论^[11,12]。杜友田等从人运动的类别、运动表示方法和运动识别方法三个方面分析了近些年来国内外的研究工作及最新进展,并对当前亟须解决的问题做了详细的分析^[13]。最近,Poppe 对近几年的视频人体动作识别的进展和方法作了详细的概述^[14]。

由于运动识别可执行在各个层次,根据 Moeslund 的分类方法^[12],将运动分为:动作基元(action primitive)、动作(action)和行为(activity)。动作基元指肢体层的原子运动,动作包含动作基元和描述了一个可能周期的整体运动,行为则包含许多连续的动作,给出了一个所做运动的完整解释,比如“伸腿”是一个运动基元,而“跑”是一个动作,“跨栏”则是一个行为,包含了开始、跳和跑的动作。根据这个划分原则,本书主要研究人体动作识别,不考虑环境上下文、人之间或对象之间的交互。

人体动作识别可认为是由图像表示、人体动作建模和分类两个部分组成^[2,14]。图像表示就是从图像或视频中提取有意义的特征,而人体动作建模和分类则是对图像特征的统计特性和分布进行建模和分类。根据图像表示中是否使用了人体模型,分为基于模型的表示和无模型的表示,而无模型的表示进一步分为全局表示和局部表示。根据

建模和分类过程是否考虑时间上的动态性,将人体动作分类方法分为直接分类方法和状态空间模型方法,而状态空间模型方法主要包括隐马尔可夫模型(hidden Markov models, HMMs)、条件随机场(conditional random fields, CRF)、线性动态系统(linear dynamic systems, LDSs)等,如图 2-1 所示。下面分别在第 2.1 节和第 2.2 节对人体动作识别中的图像表示方法、动作建模和分类方法做概述。

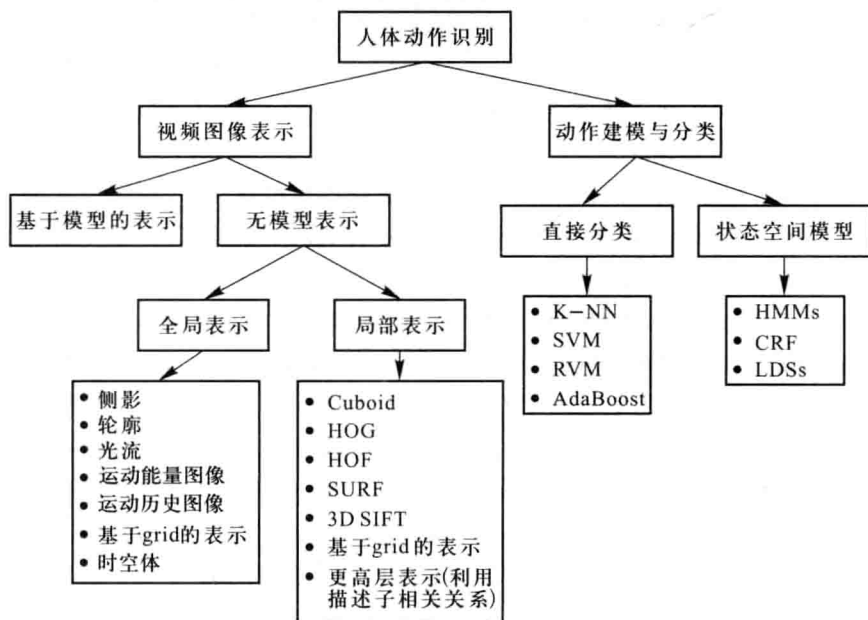


图 2-1 人体动作识别方法

2.1 视频图像表示

由于基于模型的表示主要用于人体姿态估计,所以不作阐述,仅对无模型表示的全局表示方法和局部表示方法进行描述。

全局表示采用自顶向下的方法,将图像观测序列作为一个整体进行编码。首先使用背景减除或跟踪的方法对图像中的人进行定位,提取人所在的位置;然后将感兴趣的人体区域作为一个整体进行编码,得到人体图像特征表示。这种表示具有充分的编码信息,但是需要依赖人体定位、背景减除或跟踪,对人体表观变化、杂乱背景、视角变化、噪声和遮挡非常敏感,所以只有在对这些因素得到很好控制的环境下(如静态背景),全局表示可获得很好的性能。

局部表示采用自底向上的处理方式,将图像观测序列表示为一系