

复杂感知数据 处理与应用

于小亿 著



科学出版社

复杂感知数据处理与应用

于小亿 著

科学出版社

北 京

内 容 简 介

本书主要围绕复杂现实感知数据的分析与理解和复杂虚拟感知数据的分析与理解等问题讨论。包括将深度图像特征和 RGB 图像特征结合的视觉感知处理应用的图像分类,基于超图聚类的人体行为识别算法和将频繁序列挖掘算法应用于人体行为模式的挖掘,以及在医学上的应用——医学儿童视频行为分析数据获取方法和生物学方面的应用——果蝇运动行为分析,并探讨了包括隐私保护处理的敏感信息保护处理方法;在虚拟感知数据处理方面,本书介绍了一种用户轻度参与的智能标签的创新方式来实现 Life-log 回溯服务和基于混合推荐算法的人物推荐系统的设计与实现,最后讨论了当前信息隐藏检测领域存在的一些主要问题。本书可供本科高年级学生、研究生和相关科研工程技术人员学习和参考。

图书在版编目(CIP)数据

复杂感知数据处理与应用/于小亿著. —北京:科学出版社,2014

ISBN 978-7-03-035605-5

I. 复… II. 于… III. 数据处理 IV. TP274

中国版本图书馆 CIP 数据核字(2012)第 222231 号

责任编辑:魏英杰 / 责任校对:陈玉凤

责任印制:徐晓晨 / 封面设计:陈 敬

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京厚诚则铭印刷科技有限公司 印刷

科学出版社发行 各地新华书店经销

*

2014 年 7 月第 1 版 开本:720×1000 B5

2014 年 7 月第 1 次印刷 印张:17

字数:343 000

POD定价: 80.00元

(如有印装质量问题,我社负责调换)

前 言

随着科技的发展和社会的进步,人类社会既存在于传统意义上的物理的现实空间,也存在于虚拟的空间(如网络空间)。人类生活的空间构成一个庞大复杂的现实和虚拟交织的网络,产生复杂的社会感知数据。例如,现实空间上亿级的视频监控终端和网络空间百亿级的计算机、智能手机等共同构成了一个庞大分布式的泛在社会感知信息网络;产生的感知数据越来越庞大,数据包括文本、图形、图像、视频、音频等相互关联的不同模态,且动态变化、复杂多样。如何对这些复杂感知数据进行处理,以满足人们的需要是一个迫切而重要的问题。本书尝试对复杂感知数据进行处理,将海量庞杂、异质多源、大范围时空关联的社会感知数据化繁为简,高效地提炼出满足人们需要的、人可理解并利用的信息情报和知识资源,从而有效地服务于各种应用系统。本书主要围绕复杂现实感知数据的分析与理解、复杂虚拟感知数据的分析与理解和跨现实等问题讨论。

在视觉感知处理方面的图像分类方面,本书主要介绍了 RGB-D 特征,将深度图像特征和 RGB 图像特征结合在一起,稀疏编码技术构建字典,然后用基于空间金字塔的多级 Max Pooling 进行图像表示,使用线性核的 SVM 分类器,在 RGB-D Object 数据集、RGB-D Scene 数据集和 NYU Depth 数据集这 3 个数据集上进行图像分类实验,分类结果的表现不错。在基于视觉感知数据的人的行为识别方面,介绍了一种基于超图聚类的人体行为识别算法和将频繁序列挖掘算法应用于人体行为模式的挖掘方法。在此基础上,探索利用智能视觉监控系统的客观快速准确的儿童视频行为分析数据获取方法,同时研究系统对儿童、监护者隐私的影响,以及怎样有针对性地实施隐私保护等问题,为具有隐私保护的儿童视频行为分析的医学应用作一些关键技术方面的准备。在基于隐私保护的基础上介绍与人类视觉系统特性相吻合的敏感信息保护的理论机制,探索如何屏蔽敏感信息而不破坏视觉监控系统的可用性;探索防篡改取证分析和抗信息隐藏检测的大容量敏感信息数据的信息隐藏方法,以及可伸缩性敏感信息保护机制。最后将基于视觉的行为识别方法应用生物学的研究,利用人工智能与计算机视觉技术进行客观快速准确鲁棒的以果蝇为模型的行为分析方法,主要介绍:研究果蝇群体密度分析、高密度情况下果蝇运动多目标视觉跟踪、基于实例的果蝇运动行为识别研究、果蝇运动频繁行为模式挖掘等问题。

在虚拟感知数据处理方面,本书介绍了一种用户轻度参与的智能标签的创新方式来实现 Life-log 回溯服务。通过该系统对采集的 Life-log 记录进行处理,使

用户可以更加高效、便捷、准确地回溯过去的行为和事件。然后本书详细介绍了一个基于混合推荐算法的人物推荐系统的设计与实现。通过分析社交网络服务中用户之间关系的特点,基于不同的用户信息类型,使用了不同的人物推荐算法。对于系统中的新用户,通过使用用户的人口统计学信息,对其进行分类,按照用户所属的类别为其进行推荐,该算法可以解决冷启动问题。最后本书讨论了当前信息隐藏检测领域存在的一些主要问题,并详细介绍了对 LSB 信息隐藏提出了两种检测方案:一是基于各向同性的信息隐藏检测框架;二是基于优化的 LSB 检测方法。分析表明,这两种方案为彻底解决 LSB 信息隐藏检测提供了一种更一般化的方法和思路。

作 者

目 录

前言

第 1 章 绪论	1
1.1 引言	1
1.2 视觉感知数据分析与理解	4
1.2.1 基于机器视觉物体识别分类	4
1.2.2 基于视觉的人体检测及追踪	5
1.2.3 基于视觉的人体行为分析	7
1.2.4 基于视觉的人体行为频繁模式挖掘	10
1.3 虚拟感知数据分析与理解	11
1.3.1 Life-log 技术介绍	11
1.3.2 社交网络的发展现状	16
1.3.3 推荐系统简介	17
1.3.4 信息隐藏检测原理与研究现状	20
1.4 本书组织结构	21
参考文献	22
第 2 章 基于深度图像特征和稀疏编码的图像分类算法	30
2.1 RGB 深度特征提取	31
2.1.1 三种 SFIT 特征	31
2.1.2 RGB-D 特征提取	35
2.2 稀疏编码	37
2.2.1 稀疏编码基本原理	37
2.2.2 稀疏编码字典构建	40
2.3 图像表示	42
2.3.1 Average Pooling 和 Max Pooling	42
2.3.2 空间金字塔匹配	43
2.3.3 基于 PCA 降维的加权 Max Pooling	46
2.4 SVM 分类	47
2.4.1 SVM 分类器基本原理	47
2.4.2 线性 SVM 分类	49
2.5 实验分析与结果	51
2.5.1 实验环境	51

2.5.2 代码结构	52
2.5.3 实验结果	53
参考文献	57
第3章 基于视觉感知的人的行为识别方法	60
3.1 算法架构	60
3.2 局部特征点提取	61
3.2.1 局部特征点提取	61
3.2.2 局部特征点描述符	65
3.3 字典构建	68
3.3.1 字典构建介绍	68
3.3.2 基于 hMetis 的超图聚类的字典构建算法	70
3.4 视频表示	75
3.5 实验分析	76
3.5.1 编程环境	76
3.5.2 代码结构	77
3.6 实验结果	78
参考文献	81
第4章 人的行为模式挖掘	84
4.1 技术框架	84
4.2 人体运动追踪技术	87
4.2.1 背景建模技术	87
4.2.2 人体运动追踪技术	88
4.3 人的行为挖掘及识别	93
4.3.1 Cuboids	93
4.3.2 人体行为建模	94
4.3.3 频繁行为模式挖掘	94
4.3.4 模板匹配	95
4.4 实验结果与分析	95
4.4.1 实验环境	95
4.4.2 实验结论	97
参考文献	97
第5章 基于视觉的儿童医学行为数据获取	104
5.1 概述	104
5.1.1 国内外研究概况	106
5.1.2 难点与发展趋势	109

5.2	技术框架	109
5.3	安全性、隐私性保护	111
5.4	儿童行为医学数据的获取方法	114
5.5	儿童行为医学数据的获取系统的有效性	116
5.6	实验与分析	117
5.6.1	基于 AAM 模型和柔性可控的隐私保护	117
5.6.2	基于 Kinect 的儿童哭闹行为分析	118
	参考文献	119
第 6 章	视觉监控敏感信息处理技术	123
6.1	概述	123
6.2	敏感信息屏蔽方法	126
6.3	敏感信息数据保护管理	128
6.4	基于信息隐藏的处理方法也是对敏感信息保护的重要方法	128
6.5	国外已获支持进行敏感信息保护视觉监控研究的研究机构	130
6.6	难点与发展趋势	130
6.7	技术框架	131
6.8	实验方案与关键技术	133
6.9	与人类视觉系统特性相吻合的敏感信息保护的理論	133
6.10	敏感信息屏蔽与敏感区域伪装	134
6.11	可伸缩性的敏感信息数据表示与管理系统研究	136
6.12	抗信息隐藏检测的大容量信息隐藏方法研究	137
	参考文献	138
第 7 章	基于视觉的生物行为分析方法	144
7.1	概述	144
7.2	生物行为分析方法	151
7.3	实验方案	153
7.3.1	果蝇群体密度分析与密集情况下运动视觉目标跟踪研究	154
7.3.2	果蝇运动行为识别研究	156
7.3.3	果蝇运动行为模式挖掘	160
	参考文献	161
第 8 章	Life-log 相关技术	170
8.1	概述	170
8.2	采集模块需求分析与设计	171
8.3	总体设计	178
8.4	智能标签模块设计思想	178

8.4.1 标签构成	179
8.4.2 智能标签技术	180
8.5 实验结果与分析	184
8.5.1 采集模块的实现	185
8.5.2 标签模块的实现	189
8.5.3 回溯服务模块的实现	190
8.5.4 验证实验	193
参考文献	195
第9章 基于SNS的人物推荐技术	198
9.1 概述	198
9.1.1 协同过滤	198
9.1.2 基于内容的过滤	199
9.1.3 基于人口统计学的过滤	200
9.1.4 混合推荐系统	200
9.1.5 推荐系统的评价指标	201
9.1.6 人物推荐的相关研究	202
9.2 人物推荐系统的框架	205
9.2.1 系统整体框架设计	205
9.2.2 离线系统设计	207
9.2.3 在线系统设计	209
9.3 人物推荐系统的相关算法	211
9.3.1 数据处理模块相关算法	211
9.3.2 推荐引擎模块的实现	213
9.3.3 推荐解释的生成	219
9.4 系统应用与效果评测	219
9.4.1 推荐算法评测	219
9.4.2 推荐结果展示界面	223
参考文献	224
第10章 信息隐藏检测技术	228
10.1 概述	228
10.2 基于各向同性的信息隐藏检测和估计	229
10.2.1 各向同性	229
10.2.2 LSB 信息隐藏	233
10.2.3 针对 LSB 信息隐藏的信息隐藏检测框架	233
10.2.4 提高估计信息量精度的方法	244

10.2.5 一般性 LSB 信息隐藏检测框架的实验结论	248
10.2.6 与现有信息隐藏检测方法的联系	252
10.3 基于加权的 LSB 信息隐藏检测估计	255
10.3.1 图像像素的对称性	255
10.3.2 Fridrich 优化方法	256
10.3.3 推广的优化 LSB 信息隐藏检测方法	256
10.3.4 优化方法的实验结果	259
10.3.5 分析与讨论	261
参考文献	262

第 1 章 绪 论

1.1 引 言

随着科技的发展和社会的进步,人类社会既存在于传统意义上的物理的现实空间,也存在于虚拟空间中(如网络空间)。人类生活的空间构成一个庞大复杂的现实和虚拟交织的网络,产生复杂的感知数据,如现实空间上亿级的视频监控终端和网络空间百亿级的计算机、智能手机等共同构成了一个庞大分布式的泛在感知信息网络;产生的感知数据越来越庞大,如根据 IDC 的统计和预测,2009 年全球数据量达到了 0.8 ZB(1021 字节),2020 年将达到 35 ZB。复杂感知数据包括文本、图形、图像、视频、音频等相互关联的不同模态,且动态变化、复杂多样。如何对这些复杂感知数据进行处理,以满足人们的需要是一个迫切而重要的问题。本书尝试对复杂感知数据进行处理,将海量庞杂、异质多源、大范围时空关联的社会感知数据化繁为简,高效地提炼出满足人们需要的、人可理解并利用的信息情报和知识资源,从而有效服务于各种应用的系统。本书主要围绕“复杂感知数据的高效处理和语义理解”进行以下 3 个更为直接、紧密关联的问题讨论。

(1) 复杂视觉感知数据的分析与理解

视觉感知是获取物理世界信息的最主要通道。根据认知学理论,人类获得的信息中 80% 来自视觉。随着信息技术的飞速发展,与视觉相关的感知信息量的急剧膨胀。例如,全国各地为建设“平安城市”部署了数千万级的视频监控终端,构建了描述人们在物理空间社会活动的视觉传感网。海量视觉数据的及时处理和有效利用已成为重大技术瓶颈,其中的关键问题就是复杂视觉数据的分析与理解,即从大范围监控区域分布式摄像机网络的跨时空视觉感知数据中提炼出目标(包括人、车、物)、行为、事件、社会关系等语义信息,并实现整体感知与理解。这需要解决如下问题——时间、地点、人物、事件以及其因果关系等,同时需要解决大范围视觉监控区域的整体感知与理解问题。

对视觉感知数据的分析与理解,首先需要对图像信息进行识别、分类。图像分类应用非常广泛。例如,使用图像分类技术来对图片内容进行分析,判定其是否为健康合法。而在路面、大楼和机场等安全监视系统中,也需要从获取的视频图像序列中对图像进行识别分类。这些自动化的应用可以使人们从简单枯燥的操作中解放出来。遥感图像分类、生物信息图像分析以及医学图像分析都是使用图像分类技术十分前沿的科学研究领域,也在图像分类技术的帮助下得到飞速的发展。因

此,对图像分类技术的研究无论在应用上还是在科学研究上具有十分重要的意义。本书针对当前图像分类技术的研究现状和挑战,在前人研究的基础上提出了自己的图像分类算法。

(2) 复杂虚拟感知数据的分析与理解

随着信息技术特别是计算机硬件技术的飞速发展,人们日常生活和工作中倾向于使用多种计算设备进行工作、学习和娱乐,Forrester 公司最新调查数据显示,中国主要大城市中计算机的普及率达到了 40.6%,而全世界使用智能操作系统的手机用户数量 4 年内预计将达到 10 亿,尤其是随着带有多种传感器的智能手机的快速普及使得用户使用的计算设备变得丰富。用户在日常生活与工作中接触的计算系统早已经不再局限于单独的一个设备或者少数几个应用程序上,而是在多个设备中进行切换操作,相关操作行为也涉及跨越多个设备上的多种不同类型的应用程序。通过 PC 和智能手机等移动设备采集用户日常生活和工作的各种行为和事件,并以数据的形式保存下来,所有基于这些数据化的个人行为记录进行的研究属于虚拟感知数据的分析与理解。

通过 PC 和智能手机等移动设备采集用户日常生活和工作的各种行为和事件,并以数据的形式保存下来,所有基于这些数据化的个人行为记录进行的研究也称为 Life-log 研究,数据化的个人行为记录简称为 Life-log 记录。通过多种设备采集来的行为记录被整理并保存在计算机的数据库中,用户可以随时回溯自己以往的行为和经历,为生活和工作提供了便利。然而技术的发展带来了信息的爆炸式增长,个人日常生活和工作产生的信息在一段时间的累积之后,其信息量远超出了用户所能记忆的范围,需要对这些庞大的个人行为记录进行分析和处理,研究更好的 Life-log 数据组织和索引方式,为用户提供更有效的回溯服务。另一方面,记忆是人们储存、保持和回想信息的能力,对人们的日常生活非常重要,然而记忆不像那些可以被保存在物理设备上的、数据化的信息,记忆很容易被淡化和忘却,利用 Life-log 技术可以辅助用户记忆。

另一方面,网络空间中也产生大量的感知数据,随着博客、微博、社交网站的流行,以及多媒体终端设备包括网络音视频摄像头、数码相机、数码摄像机、智能手机的日益普及,人们可以很方便地在网络发表文字、图像、图形、视频、语音并茂的社会媒体。感知数据总体上具有开放性、时效性、互动性、匿名化、群体化、跨媒体、非结构化等特点。赋予机器类似人类认知系统的跨媒体统一感知能力是智能信息处理的一个核心和热点科学问题,其中的关键问题是要将纷繁庞杂的跨媒体网络感知数据映射到高层语义空间实现统一的内容理解和信息聚合以满足人们的需要。

网络感知数据分析与理解,本书主要关注社交网络服务(social networking service, SNS)的研究。SNS 强调以用户为中心的模式,通过网络应用促进网络上人与人之间的信息交换和协同合作,随着 Facebook, Twitter, LinkedIn 等网站的

高速发展,社交网络服务在人们的生活中扮演着越来越重要的角色。用户是社交网络服务最基本的元素,如何帮助用户在社交网络中找到自己感兴趣的人,对于用户和社交网络服务本身都有着重要的意义。

最后,虚拟感知数据的处理也涉及信息安全方面的处理。计算机网络通讯技术的发展,特别是互联网技术的发展,使数据的交换和传输变成了一个相对简单的过程。人们借助于计算机、数字扫描仪、打印机等电子设备可以方便、迅捷地将数字信息传达到世界各地。信息媒体的数字化和网络技术的发展,也使人们日益关注数字媒体的安全性。信息隐藏(information hiding)技术正是在这种情况下从20世纪90年代迅速发展起来。信息隐藏技术是指在一定的多媒体中隐藏其他的信息,这种思想具有一系列的应用,如数字水印,隐秘通信等领域。早期人们对信息隐藏的研究与应用主要在于数字水印(watermarking),数字水印应用于数字媒体的版权保护和完整性认证。随着技术的发展,信息隐藏应用于隐秘通信(covert communication)以及寻找隐秘通信的存在的研究与应用也逐渐引起了国际学术界、企业界以及政府有关部门的广泛关注。

一般把信息隐藏在隐秘通信中的应用称为隐写术(steganography)。隐写术是发信方往多媒体数据(如图像、声音、视频信号等)中添加要传送的信息而不影响原数据的视听效果,然后利用公共信道传送出去,接收方接到数据后解码出隐藏的信息,以达到隐蔽通信的目的。在不引起混淆的情况下,本书把隐写术直接称为信息隐藏。信息隐藏(隐写术)不仅提供了非密码的安全途径,更重要的是突破了“加密”的思维方式,它将带来信息战乃至情报战的革命。

与信息隐藏相对应,信息隐藏检测(steganalysis)是指分析互联网上的公开网站和多媒体数据流,检测可疑的隐蔽通信渠道,从而可以通过主动加扰等攻击手段切断可疑的隐蔽信道。信息隐藏检测技术在军事上称为隐蔽通信的对抗。在某种程度上讲,隐蔽通信的对抗技术甚至比隐蔽通信更重要,隐蔽通信及其对抗技术能侦探敌方的通信渠道,堵塞其通信方式,防止国家机密的泄漏,对国家安全和国防安全意义重大。

一般地,信息隐藏技术应具有如下的特性:安全性(隐藏的信息对人的感觉器官和统计意义上应是不可觉察的,或者说是透明的。任何第三方都不能感知隐秘信息的存在)、负载(不同于数字水印,信息隐藏要求有一定的负载容量)以及鲁棒性(隐秘信息对有意或无意的图像操作与失真具有一定的抵抗力)。对信息隐藏的检测与密码的分析相类似,一旦隐秘信息或者隐秘信道的存在被发现,信息隐藏就被破获了。信息隐藏与信息隐藏的检测技术犹如矛和盾的发展,一方的进步促使另一方的发展。

下面将详细介绍相关感知数据处理技术。

1.2 视觉感知数据分析与理解

1.2.1 基于机器视觉物体识别分类

物体识别、图像分类是一项极具挑战的任务,因为获取的图像可能受到视角变化、照度、遮挡和背景的影响。为了解决这些问题,当前主要有两类算法,分别是 Part-based 的方法^[1-3],还有基于 Bag-of-features 的方法^[4-8]。来自北卡罗来纳大学教堂山分校的 S. Lazebnik^[8]提出了基于 Bag-of-features 的 Spatial Pyramid Matching 框架,该框架克服了 Bag-of-features 方法中丢失的空间信息,极大地提高了图像分类的准确率。

最近一种称为稀疏编码(sparse coding)的方法成为图像分类领域最为热门的技术,2009、2010 以及 2011 年每年的 CVPR 上均出现 20 余篇与稀疏编码相关的论文,并且在 Spatial Pyramid Matching 方法的基础之上,稀疏编码的引入能够进一步提高物体识别的准确率。文献[9]和文献[10]提出的稀疏编码的方法是当前引用最多的稀疏编码方法。

然而,这些方法都忽略了物体或者场景的深度信息,而深度信息被认为对于图像分类是非常重要的线索,因为深度信息可以在一定程度上反映物体或者场景的三维信息。为了有效地利用深度信息,基于深度信息的图像分类算法开始变得流行。这些算法大多是采用 Kinect^[15]来获取深度图像。特别地,在文献[16]中作者第一次提出了两大包含深度图像的数据集,分别是 RGB-D 物体数据集^[17]和 RGB-D 场景数据集^[17]。并且该作者提出了一种基于文献[18]的图像分类算法,该算法从核方法的角度来提取图像特征和进行图像分类。然而该算法的缺陷在于它需要首先对物体进行三维建模,这将非常耗费时间。并且,该算法仅使用了物体的深度信息,却忽略了图片的 RGB 色彩信息。此外,该算法依赖于通过预先准备好的掩码从图像背景中提取出物体,在大多数情况下这往往是无法做到的,这种掩码是无法预先获取到的。

在文献[19]中,作者提出了 Berkeley 三维物体数据库用于物体检测,虽然并不是图像分类算法,但作者给出了一个关于深度图像特征的重要结论,即基于深度图像的 Histogram of Gradients 特征的分类算法比起基于彩色图像的 Histogram of Gradients 特征的分类算法效果要差。这是因为基于梯度的特征描述符对于深度图像是不合适的。该结论对于基于深度特征和稀疏编码的图像分类算法具有指导意义,尤其是深度图像的特征提取。

在文献[20]中,作者采用了 RGBD-SIFT 特征(将深度图像和 RGB 彩色图像的 SIFT 特征结合在一起)来对 NYU 深度数据集来进行图像分类,该数据集可以在文献[21]中找到。结果显示基于 RGBD-SIFT 特征的图像分类结果比起单独的

RGB-SIFT 和 Depth-SIFT 都要好。这表明将 RGB 图像的特征和对应的深度图像的特征相结合可以获得更好的效果。然而,该算法存在一个缺陷,即是 SIFT 特征对于深度图像而言并不适合,因为 SIFT 特征的描述符是基于梯度的,这违背了在文献[19]中得到的结论。因为,深度图像中的深度值对于视角十分敏感,并且和 RGB 图像的像素值不同,深度图像的深度值要么在边缘区域变化强烈,要么在背景区域几乎保持不变。这样的话,SIFT 中基于梯度的描述符并不能很好地描述深度特征。

虽然上述方法都不能很好的利用深度信息,但它们的共同之处在于都是用了统一的 Spatial Pyramid Matching 框架和基于直方图的图像表示方式。Spatial Pyramid Matching 框架被证明了在图像分类领域非常的有效,原因在与其充分地利用了传统 Bag-of-features 方法忽略的空间信息,因而被传统 Bag-of-features 方法和 sparse coding 等方法采用。而对于基于直方图的图像表示方式,则存在缺陷。在文献[9]中,作者指出在线性 SVM 分类器的条件下,基于 Max Pooling 的图像表示方式要优于基于直方图的图像表示方式。此外,在文献[26]中,作者更是通过了多组实验进一步证明了对于线性分类器,基于 Max Pooling 的图像表示方式能够极大地提高图像分类算法的准确性。甚至,当传统的 Bag-of-features 算法和 Max Pooling 结合使用时效果会比 sparse coding 和直方图结合使用的效果还要好。

因此,基于上述国内外的最新研究现状,可以总结出 3 条结论,基于这 3 条结论可以更好地利用图像的深度信息,进一步提高图像分类算法的准确性。这 3 条结论如下。

① 首先,SIFT 特征描述符是不适于直接用来描述深度图像特征的。

② 其次,将深度图像的特征和对应的 RGB 图像特征相结合可以获得更好的结果。

③ 最后,采用稀疏编码和空间金字塔匹配框架和基于 Max Pooling 的图像表示方式会有利于图像分类算法准确率的提高。

1.2.2 基于视觉的人体检测及追踪

运动目标的检测与运动目标的跟踪在人体行为分析过程中耦合性非常强,往往运动目标的检测算法的选择决定运动目标跟踪算法的选取。

目标跟踪是计算机视觉研究领域重要的任务。跟踪目标可以表示为它们的形状或者外观,可以用如下几个形状或者形状的组合来表示。

(1) 点。对象可以表示为一个点或者点的集合。例如,在文献[27]中 Veenman 使用中心点,在文献[28]中 Serby 用一个点的集合表示。用点表示对象时比较适合跟踪目标只占整张图片一小部分的情况。

(2) 主几何形状。目标的形状经常表示为矩形、椭圆^[29]等,目标的动作可以用这种形式的表示来建模变幻、仿射变换、射影变换等。主几何形状方法不仅适合简单刚性目标的表示,也适合非刚性目标的表示。

(3) 目标轮廓和剪影。可以用目标的边缘定义轮廓,在目标轮廓里面即为剪影,剪影和轮廓表示适用于跟踪一些非刚性形状^[30]。

(4) 骨骼模型。目标的骨骼可以从目标剪影中用中轴变换来抽取^[31],这种模型经常用做形状表示来识别目标^[32]。建模刚性和柔性目标都适用于骨骼模型。

目前有很多方法用来表示对象的外观特征。下列是目标跟踪时使用的外观表示,应该注意到跟踪时可以将形状表示可以跟外观表示结合^[33]。

(1) 概率密度表示。目标外观特征的概率密度可以通过上述的形状模型从特定图像区域计算出来^[34]。

(2) 模板表示。模板可以通过简单的几何形状和剪影的组合构成,因为模板表示只能从一个视角产生,因此模板方法也只能适用于目标姿势改变不大的情况。

(3) 活动外观模型。活动外观模型可以从目标的形状和外观的同时建模产生^[35]。

(4) 多视角外观模型。这种模型对一个目标不同的视角进行编码,通过给定的视角采用产生子空间的方法来表示目标不同的视角。目前比较成熟的子空间方法有主成分分析法(PCA)和独立成分分析法(ICA)^[36,37]

(5) 外观特征模型。基本思想是利用物体表面的一些特征来描述运动物体,然后利用该特征实现对运动物体的跟踪,它包括特征提取和特征匹配两个过程。

Hu^[38]利用人的主轴特征进行跟踪。在目标跟踪过程中,使用什么特征能够比较好地将前景和背景区分开来,是特征提取的重要任务。现在一般使用特征有角点、颜色、纹理形状以及运动信息等^[39-41]。

可以利用角点或者纹理特征位置的变化信息来跟踪目标,通常包括3步:特征提取、特征匹配和运动信息计算。基于特征的跟踪算法关键在于特征的检测、表达和相似性度量。基于特征匹配的跟踪方法通常不考虑运动目标的整体特征,只通过对目标的显著特征来进行跟踪。这种算法的优点在于即使目标的某一部分被遮挡,但如果有一部分特征可以看到,就可以完成跟踪任务,另外,它对于运动目标的亮度等变化不敏感。缺点是对噪声和图像模糊比较敏感。

上述都是运动目标检测的方法,这些检测的算法对于后续的行为表示及识别都有局限性,如遮挡问题就是软肋。遮挡包括人与自身身体部位、人与人之间、人与物体之间的遮挡。由于人体是非刚性运动体,其行为极其复杂多变,遮挡时,只有部分可见,这个过程一般是不可训练的,并会带来歧义性问题,这会对后期的行为识别带来影响,需要开发更好的模型来处理遮挡时的特征与模型的匹配问题。另外视角的转换也让这类方法的实用性大大降低。

当运动对象检测结束之后就是进行数据关联,也就是将检测得到的目标与预测得到的目标进行匹配的过程,一般都采取最优化算法处理,目前最优秀的方法就是粒子滤波算法。

近年来,粒子滤波算法在视觉跟踪中取得了巨大的成功,Toyama 和 Blake^[42]采用混合度量空间的方法,将基于粒子的方法并入到概率框架下,并应用于人体和嘴唇的跟踪。基于粒子群跟踪算法的优势在于可以很好地解决移动摄像机下运动混乱的问题。中科院自动化研究所发表的文献[43]将跟踪问题用鸟群捕食过程来模拟,提出了一种基于序列粒子群优化的跟踪算法框架。

利用迁移学习技术可以提升跟踪的训练效率。文献[44]提出基于迁移学习框架的 Boosting 方法利用先前帧的先验知识迁移学习用以识别当前帧的前景目标,实验证明相当有效。文献[45]提出一种基于最小化连续能量函数的多目标跟踪算法,实验表明该方法可以提高基于卡尔曼滤波的跟踪方法。另外,文献[46]提出使用简单的贪心与动态规划思想,他们提出了一种近似算法使时间复杂度随目标的数目是线性增长,这种近似算法可以提升跟踪的性能。

1.2.3 基于视觉的人体行为分析

最近几年,人体行为分析吸引了许多计算机视觉研究者的眼光,众多公司、学校以及研究机构进入该领域的探索与研究。Microsoft 公司研发的基于行为分析的体感游戏设备 Kinect+XBOX,充分体现了人机交互的理念,逐步将行为手势识别接口应用于商业市场中。Vidient 公司的 Smart Catch 智能视频监控平台在全美最大的国际机场之一的旧金山机场部署后,进一步提高了发现对公众安全有潜在威胁的事件的效率。麻省理工学院(Massachusetts Institute of Technology, MIT)、卡耐基梅隆大学(Carnegie Mellon University, CMU)、南加利福尼亚州大学(University of Southern California, USC)、法国国家信息与自动化研究所(French National Institute for Research in Computer Science and Control)以及中国科学院自动化研究所模式识别国家重点实验室(National Laboratory of Pattern Recognition, NLPR)等大学和研究机构也在人体行为识别领域进行了深入的研究,在人体运动的视觉分析,民用场景监控,交通行为事件分析,人与人之间交互场景的理解等方面取得了大量的成果,为以后的人体行为识别的进一步研究提供了一定的经验与可借鉴的地方。

经过对以上研究进入深入分析发现,尽管传统的行为方法取得了不错的成果,由于人体运动的行为纷繁众多、遮挡问题、背景杂乱、拍摄装置的拍摄角度不同以及人与人之间的差异等因素,人体行为识别的研究也有许多需要解决的问题。

当前的人体行为识别算法大致可以分为 3 种,分别是基于骨骼的人体行为识