

基于视觉的人的运动识别综述

杜友田, 陈 峰, 徐文立, 李永彬

(清华大学自动化系, 北京 100084)

摘 要: 目前, 基于视觉的人的运动分析是一个非常活跃的研究领域, 而对人的运动进行理解和识别更是该领域内一个富有挑战性的研究方向, 它在智能监控、感知接口和基于内容的视频检索等领域具有广泛的应用前景. 本文对人的运动识别领域近年来的发展作了比较详细的论述, 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该方向的进展情况, 并对当前该研究方向上亟待解决的问题做了比较详细的分析.

关键词: 计算机视觉; 智能监控; 运动表示; 运动识别

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2007) 01-0084-07

A Survey on the Vision-Based Human Motion Recognition

DU You-tian, CHEN Feng, XU Wen-li, LI Yong-bin

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: Vision-based human motion analysis is currently one of the most active research fields, and the human motion recognition is a challenging research topic in this field. Human motion recognition has many promising applications such as intelligent surveillance, perceptual interface and content-based video retrieval. In this paper, the improvements of this topic in the last several years are discussed in detail from three aspects: the category of human motions, motion representation methods and motion recognition methods. At the end of the survey, some vital problems of human motion recognition are presented and analyzed.

Key words: computer vision; intelligent surveillance; motion representation; motion recognition

1 引言

当前, 基于视觉的人的运动分析是一个非常活跃的研究领域, 它针对包含人的图像序列进行运动检测^[1]、目标分类、跟踪以及对人的运动进行理解和识别^[2]. 人运动的理解和识别, 属于人的运动分析的高级处理部分, 近年来越来越受到人们的关注. 人的运动识别, 可以简单地看成一个时变数据的分类问题, 包括运动的表示和运动的识别两部分. 基于视觉的人的运动识别研究既包含了图像处理以及计算机视觉等知识, 也涉及了模式识别以及人工智能的理论, 是一个多学科交叉的研究方向. 观察人的根本目的是识别和理解人的个体动作、人与人之间的以及人与外界环境之间的交互行为, 并且把它们描述出来. 然而, 客观环境的多样性以及人运动的复杂性使得人的运动识别变得非常困难.

人的运动识别有着很广阔的应用前景, 主要集中在以下三个方面: (1) 智能视觉监控. 智能视觉监控主要应用于一些对安全要求比较高的场合, 譬如银行、超市以及车库等地方, 可以实时地检测并识别外部事件尤其是人的可疑行为的发生. (2) 感知接口. 在高级用户接口领域中, 我们希望未来的计

算机能够和人更加容易地进行交流和沟通, 如能够理解人的手势以及表情等. 这就要求未来的计算机能够对外部的环境进行感知并能够理解外部的事件尤其是人的行为. (3) 基于内容的检索. 随着多媒体数据的不断增多, 视频数据尤其是大量的体育视频查找起来非常麻烦, 人的运动识别可以使得在视频数据库内检索特定的事件变得非常方便.

由于人的运动识别的挑战性以及其巨大的应用价值, 越来越多的学校、研究所以及公司的研究人员投入到该领域中来. 麻省理工学院、卡内基梅隆大学以及其他国外著名大学成立了专门的计算机视觉及多媒体方向的实验室; Nice 和 ObjectVideo 等公司已经针对飞机场、国界线等应用场合开发了一些相应的智能监控系统; 国内在该领域的研究起步相对晚一些, 但是一些研究单位也已经开始了这方面的研究并取得了许多成果. 研究人的运动识别的文献越来越多, Moeslund 和 Granum^[3]对以前的工作做了回顾和分析, 主要分成四个部分对基于视觉的人体运动捕获进行了分析: 初始化、目标跟踪、人的姿态估计和识别, 同时对当前该领域做的假设做了比较详尽的归纳. 王亮^[2]等人从检测、目标分类、跟踪、理解等四个阶段对人体的运动进行了分析, 该文分析了 2000 年以前的关

于人体运动分析领域的文章,但是对人体行为识别方面的分析较少.此外,Gavrilu^[4]也对该领域的部分文章做了阐述和分析.这些综述文章的特点是把人的运动分析的各个环节都进行了讨论,其中人运动的理解和识别部分分析得比较简单.文献[5]针对人行为的识别和理解进行了分析,但在人运动的类别分析和运动表示等方面没有对以前的文章做分析.针对一种特殊的运动——手势,文献[6]从手势建模、手势分析和手势识别三个方面对以前的研究工作进行了分析和总结.鉴于当前人的运动识别研究的不断发展和深入,对该领域的工作及进展做详细的回顾和分析显得很有必要.本文主要针对基于视觉的人的运动识别这个研究方向,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,并对当前亟需解决的问题做了详细的分析.

2 人运动的类别

人的运动覆盖的范围很广,包括人整个身体或部分肢体的运动,表情以及唇语等等,这些运动具有不同的特点,研究方法也不尽相同.而针对人整个身体或部分肢体运动的识别是当前研究的一个热点和难点,也是本文讨论的重点.通常来说,不同的运动具有不同的复杂程度,其表示方法和识别方法可能会不同^[7].

Bobick^[8]认为人的运动可以分为三类:动作(movement)、行为(activity)和行动(action),这三类运动分别处于三个不同复杂程度的层次上.动作是运动的基元,是最基本的运动,是形成其他复杂、高级运动的基础.一般来说动作在执行过程中会持续较短的时间,其识别方法一般可以采用几何或概率统计的方法.譬如,打网球时挥动了一下球拍,这就是一个动作.行为是指若干个连续动作的序列,时间尺度较大,明显地反映了人运动的目的性.当前大部分关于人的行为识别的方法是基于概率统计的,需要考虑其中各个动作或者运动状态之间的相互关系.一个完整的手势是一个典型的行为,它可能包含了人手的几个动作,清楚地表达了人的主观性和目的性.行动是更高层次的运动,它反映了动作在大的时间尺度上的运动以及人和人、人和环境以及和其他目标之间的相互关系.理解和识别行动的方法需要用到概率统计或者人工智能的技术,如动态贝叶斯网络和文法分析等,而且在识别时常常要考虑到运动与上下文之间的关系.譬如,多人的交互就可以归到行动中.动作、行为和行动在一些场合下也可以称作事件.也就是说,事件既可以指低级的、小尺度的动作,也可以指大尺度的、中高级的运动,包括行为和行动.尽管这三类运动复杂程度不同,但有些时候难以进行严格的区分,并且各类运动识别对应的研究方法也没有太严格的限制.故在本文后面的人运动的表示方法和识别技术分析中,我们并不将这三类运动分开讨论.

人的运动识别研究也常常按照单个人的运动、两人交互以及人群的行为来进行.目前大多数工作是针对单个人的运动进行研究的,对两人交互及人群的行为研究得比较少^[9~11],但这两类运动在实际环境中都非常普遍,对其进行

研究也很有意义.

3 人运动的表示方法

人的运动表示指的是从含有人运动的视频序列中抽取出来能够合理表示运动的数据,这对于运动识别来说是至关重要的.运动的表示与运动的类别和应用场合有紧密的关系,对于不同的情况通常会选择不同的运动表示方法.譬如,在一个大的场景下进行较远距离的监控时,只需要提取运动目标的轨迹就可以满足需要了,而在识别人的手势时,对人的肢体进行2-D或3-D建模则非常必要.衡量运动表示的优劣有四个标准:紧凑性(minimalism)、完备性(completeness)、连续性(continuity)、唯一性(uniqueness)^[12,13],大部分运动的表示只能满足其中部分标准.一般地,人运动的表示方法可以分为两种:基于表观的表示方法和基于人体模型的表示方法.

3.1 基于表观(appearance)的表示方法

基于表观的方法^[14~20,22~24]不对人的物理结构进行直接建模,而是采用图像上的颜色信息或者灰度信息对人的运动进行分析.

直接利用图像的信息进行分析是一类最简单的方法,如网格特征(mesh feature).Yamato等人^[15]采用二维网格特征作为人的运动特征,首先将人从图像中提取出来,并将图像二值化,然后将整幅图像分割成若干个网格,计算每个网格中人体的像素数占整个网格的像素数的比例,以此来描述人体的运动.T. Nishimura等人^[20]也采用了这种网格特征来表达人体的运动.这种方法比较简单,但是精度较低.

第二类基于表观的表示方法是利用人的轮廓或者区域信息.基于人的轮廓来表示运动是一种很好的方法,能够比较精确地描述人的运动细节.Kale等人^[16]采用基于人体轮廓的方法来解决步态识别问题,首先将人体的轮廓提取出来,然后求出图像中每一行的轮廓宽度,以此宽度向量作为特征向量进行识别.Veeraraghavan等人^[17]采用轮廓上的标记点来分析人的运动,对于有限的标记点集合,通过Kendall形状理论^[21]来对其进行分析,最后对人的步态进行了分析.基于有限的标记点的好处是对运动的表示比较精确,而且特征空间的维数较低,计算复杂度小.

第三类基于表观的表示方法是利用人的运动信息,如光流^[22]、目标轨迹^[23,24]以及速度等.基于光流的方法计算量大,而且不够鲁棒.Psarru等人^[24]采用时空轨迹(spatio-temporal trajectory)来表达人的行为,然后采用一阶马尔可夫过程对其进行建模.

3.2 基于人体模型的表示方法

基于人体模型的表示方法本质上是将人体及其姿态参数化,通过分析这些参数化的人体模型来识别人的运动,这种表示方法通常用来表示比较精细的运动,如手势等.和基于表观的方法相比,该表示方法能够更精确、更突出地表示出运动的细节,提高了识别的准确度.人体模型主要有三种:线图模型^[25]、2-D模型^[26]和立体模型^[27],其中线图模型和2-D模型在人的运动识别中用的较为广泛,而立体模型由于其复杂度较高、参数难以估计,故较少地用于人的运动识别,尤其是在

单摄像机的情况下.

Feng 和 Perona^[28]采用 2-D 模型来为人体建模,整个模型由十个矩形组成,每个矩形有五个自由度,然后采用当前帧和下一帧中模型的参数联合表示当前帧中的人体姿态. Arie 等人^[27]采用 3D 模型来描述人体,全身由九个圆柱体和一个球体组成,分别代表躯干、上臂、前臂(包括手)、大腿、小腿(包括脚)和头,在具体算法中作者将表示人体的 3D 模型投影到 2D 平面上,通过跟踪得出所有主要关节的 2D 笛卡尔坐标,最终人体的姿势由上肢、下肢和躯干的姿态决定,特征向量由上臂、下臂、大腿、小腿和躯干的角度及其角速度组成.

基于模型的表示方法能够准确地描述人的运动,尤其是涉及到肢体的动作,能够较为容易地解决遮挡问题,但是其特征空间的维数很高,在进行非线性优化的时候非常困难.另外,在图像分辨率低的情况下,对模型参数进行估计也很困难.

4 人运动的识别技术和方法

关于人的运动识别的研究越来越多,很多新的方法也不断地涌现出来.模板匹配技术是一种较早用在人的运动识别上的方法^[14, 27, 29].近些年来,越来越多的研究倾向于采用概率网络来解决该问题^[15, 23, 30, 31, 36].此外,最近几年,采用文法来进行人的运动识别开始崭露头角并且引起了人们的重视^[32, 33].

4.1 基于模板的方法

基于模板的方法是将运动图像序列转化成一个或者一组静态的模板,通过将待识别样本的模板和已知的模板进行匹配而获得识别的结果.

基本的模板匹配方法是将待识别模板和已知的模板标本进行直接匹配,取距离最小的已知模板所属的类别作为识别结果. Bobick 和 Davis^[14]将图像序列转化成运动能量图像(MEI)和运动历史图像(MHI),采用马氏距离(Mahalanobis distance)来度量模板之间的相似性.其中 MEI 反映了运动所覆盖的范围及其强度,而 MHI 在一定程度上反映了运动在时间上的变化.该方法计算量小,但是鲁棒性不够好,尤其对时间间隔的变化比较敏感.

由于同一模式的运动的持续时间可能不一样,所以在模板匹配的时候将模板在时间上进行规整非常有必要. Arie 等人^[27]采用在匹配前将模板序列进行线性规整,然后采用投票的方式进行匹配.通常在运动过程中,某些动作持续的时间长度是随机变化的,所以线性时间规整的方法不能很好的解决问题.动态时间规整(dynamic time warping,简称 DTW)是一种很好的非线性时间规整方法,它的目的是将待识别运动模板的时间轴非线性地映射到训练模板的时间轴上,使得二者距离最小. Bobick^[29]采用 DWT 的方法对人的手势进行了识别.动态时间规整的方法较好地解决了人的运动在时间尺度上的不确定性.

4.2 基于概率网络的方法

基于概率网络的方法是当前最主要的人的运动识别方法^[9, 18, 23, 30, 31, 34 ~ 39].采用概率网络的方法充分考虑了人运动

的动态过程,并且将时间尺度和空间尺度上的运动的微小变化采用概率的方法进行建模.所以,该方法对于运动序列在时间和空间尺度上的小的变化具有很好的鲁棒性.采用概率网络来进行运动识别主要是把运动的每一种静态姿势定义为一个状态或者一个状态的集合,通过网络的方式将这些状态连接起来,状态和状态之间的切换采用概率来描述,一个运动序列可看成一次这些状态或者状态集合的遍历过程.目前,在人的运动识别领域内采用的概率网络主要有两种:隐马尔可夫模型(Hidden Markov Models,简称 HMMs)和动态贝叶斯网络(Dynamic Bayesian Networks,简称 DBNs)^[9, 18, 34, 35],如图 1 和图 2,前者是后者的一种特殊形式.近两年,作为一种新的概率网络模型,条件随机场(Conditional random field)^[40]被用于人的行为识别中^[41].条件随机场对观察值不进行建模,所以避免通常概率模型中的独立性假设,比较符合实际情况.目前,HMMs 及其改进模型是人的运动识别中应用最广泛的数学模型.

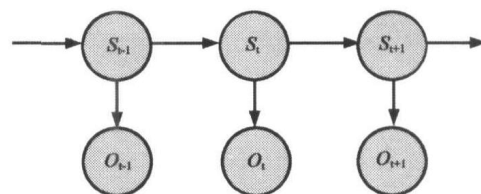


图 1 隐马尔可夫模型(HMM)

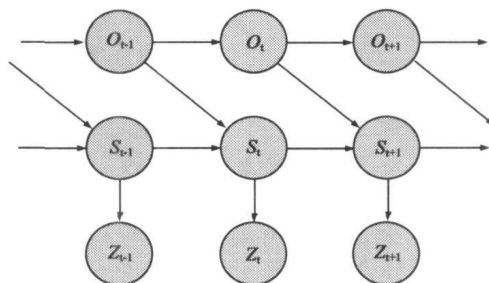


图 2 一个动态贝叶斯网络(DBN)的例子

自从 Yamato 等人^[15]将 HMMs 引入到人的运动识别中来,HMMs 及其改进模型逐渐成为主流的方法^[23, 30, 31, 36 ~ 38],其隐含的假设是运动的当前状态的取值只依赖于前一时刻的状态.基于 HMMs 的识别方法通常是先提取出表示运动的特征向量序列,通过学习算法将 HMMs 的参数训练出来,根据训练好的模型对未知的运动序列进行分类.一般来说人的运动并不严格符合马尔可夫特性,并且 HMMs 的结构导致了它只能为单一的动态过程进行较好地建模,这些原因使得传统的 HMMs 在识别较为复杂的运动时不能得到很好的结果.针对较为复杂的交互行为譬如手势,Brand 等人^[36]提出了耦合隐马尔可夫模型(Coupled Hidden Markov model,简称 CHMM),并对手语进行了识别.CHMM 能够合理地地为两个或多个有相互关系的动态过程进行建模,并且将多个动态过程的特征空间分解开来,大大减少了状态的个数,降低了算法的计算复杂度.Oliver^[30]等人提出了一个交互行为识别系统,系统采用 HMMs 和 CHMMs 对广场等地方的人的交互行为进行识别,实验结果表明 CHMMs 要比 HMMs 的识别性能优越得多.Ren 和 Xu^[31]把基元特征和 CHMM 结合起来,对智能教室里的多种行

为了进行了识别。Luh^[42]和 Nguyen^[23]采用分层隐马尔可夫模型(Hierarchical Hidden Markov models,简称 HHMMs)对长时间的人的运动做了分析。HHMMs 具有多层的隐马尔可夫模型的结构,所以能够更为清楚地表达出人的运动中不同层次的行为细节。Duong 等人^[43]除了考虑行为的层次结构外,还将运动状态的典型持续时间融合到模型中去,提出了 S-HSMM(Switching Hidden semi Markov Model)。实验结果表明在对复杂的运动进行识别时 S-HSMM 性能要比 HHMM 好,但是该算法复杂度较高。除此以外, HMM 的其它改进模型如 AHMM^[44]和 LHMM^[45]等也被用于该领域中。

动态贝叶斯网络对于动态过程的建模和分析、多信息的融合、复杂动态系统的分析识别更为有效和灵活。近些年来,动态贝叶斯网络在人的运动识别研究中逐渐成为了一个有力的数学工具。动态贝叶斯网络的每一个时间切片都是一个有向无环的概率图,不同的切片之间包含相等或不等数目的状态节点。同隐马尔可夫模型一样,动态贝叶斯网络可以对任何分布建模,在很多应用中,这种灵活性是本质的。和隐马尔可夫模型相比,动态贝叶斯网络的优点是它能够将一个复杂的动态系统分解成一些变量,充分利用了时序概率空间模型中的稀疏性,降低了计算的复杂度^[46]。同时,它更能够反映出复杂动态系统的本质。Pavlovic 等人^[18]将隐马尔可夫模型和线性动态系统(LDS)耦合起来,构成一个具有混合状态的动态贝叶斯网络,以此对手势进行识别。该文认为目标的运动由两部分组成,一部分是外界环境对目标的操作,用 HMM 对不同类型的操作进行建模,另一部分是符合牛顿定律的物理运动,用 LDS 表示。Luo 等人^[35]采用动态贝叶斯网络对体育运动进

行建模,每个时间片包含五个隐含的状态节点和四个观察节点,状态节点分别表达人的四肢和头部的位置信息,而观察节点表示目标端点的位置坐标。Gong 和 Xiang^[9]采用动态贝叶斯网络对人群的行为进行建模和分析,并对特定的人群行为进行识别,其网络的结构通过 Schwarz 贝叶斯信息准则(BIC)学习得到。

4.3 基于文法技术(syntactic technique)的方法

文法技术在文本分析和语音识别领域用的非常多。计算机视觉领域中的基于文法技术的分析大部分是用于静态图片的纹理识别,近些年来逐渐被用于人的运动识别并引起了人们的注意。文法分析主要的优势在于它对复杂结构的理解和对先验信息的利用。

Brand^[32]采用简单的确定性语法来对离散的事件进行识别。Ivanov 和 Bobick^[33]采用随机文法技术对多智能体的长时间行为和交互进行了检测和识别,其思想是将识别问题分成两层,底层是用独立概率事件检测器来检测候选特征,其输出为上层随机上下文无关的句法分析机制服务。Cho 等人^[47]应用统计语法推理来解决人的行为的自动识别,其中人的行为定义为多个关节体运动的组合。Kitani^[48]将文法技术和图模型结合起来,在统计上下文无关语法(SCFG)的基础上实现了一个分层的贝叶斯网络,这种方法对于解决单智能体的行为重叠很有效。

其他如神经网络^[49]和有限状态机^[50]等方法也经常被用于人的运动识别领域中。从上面的阐述可以看出,各种方法都有自己的优点和不足,表 1 是在文献[2,6,7,46,51]基础上分析得到的几种具体识别方法的比较结果。

表 1 几种识别方法的比较

	适合的运动类别	时间尺度鲁棒性	空间尺度鲁棒性	先验知识需求程度	计算复杂度
模板匹配	movement	low	low	low	low
DTW	movement, activity	high	low	low	medium
HMM	movement, activity	high	high	medium	high
DBN	activity, action	high	high	medium	medium
文法技术	activity, action	high	—	high	high

5 存在的问题

尽管基于视觉的人的运动识别在近些年已经取得了很大的研究进展,但是这些研究还处在一个比较基础的阶段,还很不成熟,有许多问题和难点亟需解决。

(1) 更本质的运动表示方法

计算机视觉的研究目的是从 2D 的图像去理解 3D 的客观世界,在由 3D 到 2D 的成像过程中,丢失了很多信息。摄像机与目标的相对视角、距离的变化,以及外界环境的改变等都可能使得同一种运动有不同的投影。所以,在表示人的运动时如何能够找到更本质的运动表示方法是很有意义的。尽管从 2D 图像中能够对人的姿态进行三维重建,但是需要做较多的假设,并且复杂度太高,难以用于人的运动识别。目前基

于视角不变的运动识别的研究很少^[12,52,53], Rao 和 Shah^[52]将轨迹表示为一系列的动态瞬间和间隔,动态瞬间是时空曲率的极值点,是与视角变化无关的特征,它表示了行为的速度、方向和加速度等物理量的重要变化。另外,对于表示运动的特征来说,不同复杂程度的运动会采用不同的运动表达方法,其特征向量的长度也不一样,特征向量维数太低则不能准确地表示运动,维数太高则导致计算量大,信息冗余,而且学习起来很困难,一个有效的方法是采用多分辨率的方法来提取特征或者用识别的结果来指导特征提取。

(2) 连续运动的实时识别

目前的大多数运动识别方法是基于分割后的行为序列进行分析的,而对连续的、长时间的运动序列进行分析的研究工作比较少。其中有两个需要解决的子问题:一是如何判断一个行为的起始点和终止点;二是如何学习长运动序列中的新的

行为.在较为简单的运动中,第一个已经取得了较好的结果^[54].对于第二个子问题,半监督学习是一种行之有效的方法,但是如何更好地采用半监督学习的方法来学习新的行为,目前研究得还很少^[55,56].尤其是在可疑行为的检测和识别方面,由于可疑行为的稀少性,难以获取大量样本来进行监督学习,所以基于半监督学习的方法就显得更为重要.

(3) 人的运动识别算法的评价

对于一种方法,鲁棒性、准确率以及算法速度等都是很重要的指标.目前,准确率是当前研究工作中最为关注的,而另外两项指标讨论的比较少,尤其是算法的鲁棒性,该领域内几乎没有工作对其进行分析.在人的运动识别中,目前还没形成一些标准的数据集来对算法进行检验和评价,大部分研究工作是基于各自的测试序列上进行的.这样导致的结果是很多算法没有一个统一的评价和衡量标准.另外,算法的鲁棒性和速度是将理论应用于实际时必须考虑的问题.一个算法适合于什么样的环境,在多么恶劣的环境下仍能够保持一定的正确率,这在目前很多研究工作中很少被讨论到.

(4) 智能技术的应用

目前绝大多数人的运动识别研究集中在简单的动作和行为识别上,譬如手势,走,跑等简单而规范的运动.如何能采用人工智能领域的技术和成果将简单的运动识别推广到对复杂场景下的事件的理解,是解决该领域理论问题和应用问题的关键.另外,计算机视觉主要解决的问题是对外部世界的理解,对于一个实际的视觉系统来说,如何根据外部的环境进行长时间的自主学习,如何对学习知识进行组织、储存和记忆,这也是一个重要的问题.

6 结束语

基于视觉的人的运动识别实质上是一个人工智能问题,它涉及了计算机视觉、图像处理、模式识别、人工智能以及生理学等领域的知识.该方向的研究成果在智能监控等领域有着重要的应用意义和广泛的应用前景.目前,该方向的研究还处在一个相对基础的阶段,客观环境的多样性以及人运动的复杂性使得人的运动识别变得比较困难,还有很多亟待解决的问题摆在研究者面前.

参考文献:

- [1] 魏志强,纪筱鹏,冯业伟.基于自适应背景图像更新的运动目标检测方法[J].电子学报,2005,33(12):2261-2264.
WEI Zhi-qiang,JI Xiao-peng,FENG Ye-wei. A moving object detection method based on self-adaptive updating of background[J]. Acta Electronica Sinica,2005,33(12):2261-2264. (in Chinese)
- [2] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述[J].计算机学报,2002,25(3):225-237.
WANG Liang,Hu Wei-ming,TAN Tie-niu. A survey of visual analysis of human motion[J]. Journal of Computers,2002,25(3):225-237. (in Chinese)
- [3] Moeslund T,Granum E. A survey of computer vision-based human motion capture[J]. CVIU,2001,81(3):231-268.
- [4] Gavrilu D M. The visual analysis of human movement: A survey[J]. CVIU,1999,73(1):82-98.
- [5] 刘相滨,向坚持,王胜春.人为识别与理解研究探讨[J].计算机与现代化,2004,(12):1-5.
Liu Xiang-bin,XIANG Jian-chi,WANG Sheng-chun. Exploration of recognition and understanding of human behavior[J]. Computer and Modernization,2004,(12):1-5. (in Chinese)
- [6] 任海兵,祝远新,徐光耀.基于视觉手势识别的研究——综述[J].电子学报,2000,28(2):118-121.
REN Hai-bing,ZHU Yuan-xin,XU Guang-you. Vision-based recognition of hand gestures: a survey[J]. Acta Electronica Sinica,2000,28(2):118-121. (in Chinese)
- [7] Aggarwal J K,Park S. Human motion: Modeling and recognition of actions and interactions[A]. In Proc 3DPVT[C]. Greece:IEEE Computer Society,2004. 640-647.
- [8] Bobick A F. Movement,activity,and action:The role of knowledge in the perception of motion[J]. Phil Trans Royal Society London B,1997,352(1358):1257-1265.
- [9] Gong S,Xiang T. Recognition of group activities using dynamic probabilistic networks[A]. In Proc ICCV[C]. Washington:IEEE Computer Society Press,2003. 742-749.
- [10] Vaswani N,Chowdhury A R,Chellappa R. Statistical shape theory for activity modeling[A]. In Proc ICME[C]. Baltimore:IEEE Computer Society Press,2003. 181-184.
- [11] Vaswani N,Chowdhury A R,Chellappa R. "Shape activity": A continuous-state HMM for moving/deforming shapes with application to abnormal activity detection[J]. IEEE Trans Image Processing,2005,14(10):1603-1616.
- [12] Parameswaran V,Chellappa R. Quasi-invariants for human action representation and recognition[A]. In Proc ICPR[C]. Quebec:IEEE Computer Society Press,2002,307-310.
- [13] Faugeras O. Three-Dimensional Computer Vision: A Geometric Viewpoint[M]. Massachusetts,USA: The MIT press,1993.
- [14] Bobick A F,Davis J W. The recognition of human movement using temporal templates[J]. IEEE Trans. PAMI,2001,15(3):257-267.
- [15] Yamato J,Ohya J,Ishii K. Recognizing human action in time-sequential images using Hidden Markov model[A]. In Proc CVPR[C],IEEE,1992. 379-385.
- [16] Kale A,Rajagopalan A N,Cuntoor N,Kruger V. Gait-based recognition of humans using continuous HMMs[A]. In Proc. IEEE International Conference on Automatic Face and Gesture Recognition[C]. Washington:IEEE Computer Society Press,2002. 321-326.
- [17] Veeraraghavan A,Chowdhury A R,Chellappa R. Matching shape sequences in video with applications in human movement analysis[J]. IEEE Trans. PAMI,2005,27(12):1896-

- 1909.
- [18] Pavlovic V, Grey B J, Huang T S. Time-series classification using mixed-stated dynamic Bayesian networks [A]. In Proc. CVPR[C]. USA:IEEE Computer Society Press, 1999. 609 - 615.
- [19] Rosales R, Sclaroff S. A framework for heading-guided recognition of human activity[J]. CVIU, 2003, 91 (3): 335 - 367.
- [20] Nishimura T, Sogo T, Oka R, Ishiguro H. Recognition of human motion behaviors using Omni-directional vision sensors [A]. In Proc. Annual Conference of the IEEE[C]. France: IEEE Industrial Electronics Society, 2000, 4: 2553 - 2558.
- [21] Kendall D G, Barden D, Carne T K, Le H. Shape and Shape Theory[M]. New York: Wiley, 1999.
- [22] Babu R V, Ramakrishnan K R. Recognition of human actions using motion history information extracted from the compressed video[J]. Image and Vision Computing, 2004, 22(8): 597 - 607.
- [23] Nguyen N T, Phung D Q, Venkatesh S, Bui H. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model [A]. In Proc CVPR[C]. San Diego: IEEE Computer Society Press, 2005. 955 - 960.
- [24] Psarrou A, Gong S G, Walter M. Recognition of human gestures and behaviour based on motion trajectories [J]. Image and Vision Computing, 2002, 20(5 - 6): 349 - 358.
- [25] Gonzalez J, Varona J, Roca F X, et al. ASpaces: Action spaces for recognition and synthesis of human actions [J]. Lecture Notes in Computer Science, 2002, 2492: 189 - 200.
- [26] Ju S X, Black M J, Yacoob Y. Cardboard people: a parameterized model of articulated image motion [A]. In Proc. Automatic Face and Gesture Recognition[C]. Vermont: IEEE Computer Society Press, 1996. 38 - 44.
- [27] Arie J B, Wang Z, Pandit P, Rajaram S. Human activity recognition using multidimensional indexing [J]. IEEE Trans. PAMI, 2002, 24(8): 1091 - 1104.
- [28] Feng X, Perona P. Human action recognition sequence of movelet codewords [A]. In Proc. 3DPVT[C]. Italy: IEEE Computer Society Press, 2002. 717 - 723.
- [29] Bobick A F, Wilson A D. A state-based approach to the representation and recognition of gesture [J]. IEEE Trans PAMI, 1997, 19(12): 1325 - 1337.
- [30] Oliver N M, Rosario B, Pentland A P. A Bayesian computer vision system for modeling human interactions [J]. IEEE Trans PAMI, 2000, 22(8): 831 - 843.
- [31] Ren H, Xu G. Human action recognition with primitive-based coupled-HMM [A]. In Proc. ICPR[C]. Quebec: IEEE Computer Society Press, 2002. 2: 494 - 498.
- [32] Brand M. Understanding manipulation in video [A]. In Proc Second Int'l Conf Automatic Face and Gesture Recognition [C]. Vermont: IEEE Computer Society Press, 1996. 94 - 99.
- [33] Ivanov Y A, Bobick A F. Recognition of visual activities and interactions by stochastic parsing [J]. IEEE Trans PAMI, 2000, 22(8): 852 - 872.
- [34] Locoço N M, Bremond F, Thonnat M. Recurrent Bayesian network for the recognition of human behaviors from video [A]. In Proc ICVS[C]. Graz: IEEE Computer Society Press, 2003. 68 - 77.
- [35] Luo Y, Wu T D, Hwang J N. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks [J]. CVIU, 2003, 92(2): 196 - 216.
- [36] Brand M, Oliver N, Pentland A. Coupled hidden markov models for complex action recognition [A]. In Proc CVPR[C]. USA: IEEE Computer Society Press, 1997. 994 - 999.
- [37] Oliver N, Horvitz E. A comparison of HMMs and dynamic Bayesian networks for recognizing office activities [J]. Lecture Notes in Artificial Intelligence, 2005, 3538: 199 - 209.
- [38] Kolonias I, Christmas W, Kittler J. Use of context in automatic annotation of sports videos [J]. Lecture Notes in Computer Science, 2004, 3287: 1 - 12.
- [39] Park S, Aggarwal J K. A hierarchical Bayesian network for event recognition of human actions and interactions [J]. Multimedia Systems, 2004, 10(2): 164 - 179.
- [40] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data [A]. In Proc ICML[C]. Massachusetts: IEEE press, 2001. 282 - 289.
- [41] Sminchisescu C, Kanaujia A, Li Z, Metaxas D. Conditional models for contextual human motion recognition [A]. In Proc ICCV[C]. Beijing: IEEE Computer Society Press, 2005. 2: 1808 - 1815.
- [42] Luhr S, Bui H H, Venkatesh S, West G A W. Recognition of Human Activity through Hierarchical Stochastic Learning [A]. In Proc. PerCom[C]. Texas: IEEE Computer Society Press, 2003. 416 - 422.
- [43] Duong T V, Bui H H, Phung D Q, Venkatesh S. Activity recognition and abnormality detection with the switching hidden semi-Markov model [A]. In Proc CVPR[C]. San Diego: IEEE Computer Society Press, 2005. 838 - 845.
- [44] Nguyen N T, Venkatesh S, West G A W. Learning people movement model from multiple cameras for behaviour recognition [J]. Lecture Notes in Computer Science, 2004, 3138: 315 - 324.
- [45] Oliver N, Garg A, Horvitz E. Layered representations for learning and inferring office activity from multiple sensory channels [J]. CVIU, 2004, 96(2): 163 - 180.
- [46] Russell S, Norvig P. Artificial Intelligence: A Modern Approach (Second Edition) [M]. Prentice Hall. 2004.
- [47] Cho K, Cho H, Um K. Human action recognition by inference of stochastic regular grammars [J]. Lecture Notes in Computer

- Science, 2004, 3138: 388 - 396.
- [48] Kitani K M, Sato Y, Sugimoto A. Deleted interpolation using a hierarchical Bayesian grammar network for recognizing human activity[A]. In Proc IEEE workshop on VSPETS [C]. Beijing: IEEE Computer Society Press, 2005. 239 - 246.
- [49] Hu W, Xie D, Tan T. Learning activity patterns using fuzzy self-organizing neural network[J]. IEEE Trans. Systems Man and Cybernetics Part B-Cybernetics, 2004, 34 (3): 1618 - 1626.
- [50] Hongeng S, Nevatia R, Bremond F. Video-based event recognition: activity representation and probabilistic recognition methods[J]. CVIU, 2004, 96(2): 129 - 162.
- [51] Wang L, Hu W, Tan T. Recent developments in human motion analysis[J]. Pattern Recognition, 2003, 36(3): 585 - 601.
- [52] Rao C, Shah M. View-invariance in action recognition[A]. In Proc. CVPR[C]. Hawaii: IEEE Computer Society Press, 2001. 316 - 322.
- [53] Parameswaran V, Chellappa R. View invariance for human action recognition[J]. IJCV, 2006, 66(1): 83 - 101.
- [54] Deng J W, Tsui H T. An HMM-based approach for gesture segmentation and recognition [A]. In Proc ICPR [C]. Barcelona: IEEE Computer Society Press, 2000. 679 - 682.
- [55] Zhang D, Perez D G, Bengio S, Mccowan I. Semi-supervised adapted HMMs for unusual event detection[A]. In Proc CVPR [C]. San Diego: IEEE Computer Society Press, 2005. 611 - 618.
- [56] Cohen I, Sebe N, Cozman D F, Huang T S. Semi-supervised learning for facial expression recognition [A]. In Proc 5th ACM SIGMM International Workshop on Multimedia Information Retrieval[C]. Toronto: ACM press, 2003. 17 - 22.

作者简介:



杜友田 男, 1980 年 6 月出生于山东省日照市, 2002 年 7 月毕业于西安交通大学电气工程学院, 现为清华大学自动化系博士研究生, 主要从事计算机视觉、智能监控、模式识别方面的研究。

E-mail: dyt02@mails.tsinghua.edu.cn



陈 峰 男, 1971 年 12 月出生于辽宁省鞍山市, 副教授, 1994 和 1996 年在俄罗斯圣彼得堡国立技术大学获得学士和硕士学位, 2000 年在清华大学获得博士学位, 主要研究方向为计算机视觉和视频处理。

E-mail: chenfeng@mail.tsinghua.edu.cn