

人运动的视觉分析综述*

王 亮 胡卫明 谭铁牛

(中国科学院自动化研究所模式识别国家重点实验室 北京 100080)

摘 要 目前, 人运动的视觉分析是计算机视觉领域中最活跃的研究主题之一, 其核心是利用计算机视觉技术从图像序列中检测、跟踪、识别人并对其行为进行理解与描述, 它在虚拟现实、视觉监控、感知接口等领域均有着广阔的应用前景。人运动的视觉分析系统一般遵从下述的处理过程: 1)运动检测; 2)运动目标分类; 3)人的跟踪; 4)行为理解与描述。本文将重点从此四个方面回顾人运动分析的目前发展水平和常用的处理方法, 并对研究难点及未来的发展趋势作了较为详细的分析。

关键词 人的运动, 视觉分析, 运动检测, 跟踪, 行为理解

中图法分类号: TP391

1 引言

人运动的视觉分析^[1]是近年来计算机视觉领域中备受关注的前沿方向, 它从包含人的图像序列中检测、识别、跟踪人并对其行为进行理解和描述, 属于图像分析和理解的范畴。从技术角度而言, 人的运动分析的研究内容相当丰富, 主要涉及到模式识别、图像处理、计算机视觉、人工智能等学科知识; 同时, 动态场景中运动的快速分割、人体的非刚性运动、人体自遮挡和目标之间互遮挡的处理等也为人的运动分析研究带来了一定的挑战。

由于人的运动分析在高级人机交互、安全监控、视频会议、医疗诊断及基于内容的图像存储与检索等方面具有广泛的应用前景和潜在的经济价值, 从而激发了世界上广大科研工作者及相关商家的浓厚兴趣, 尤其在美国、英国等国家已经开展了大量相关项目的研究。例如, 1997 年美国国防高级研究项目署 (Defense Advanced Research Projects Agency) 设立了以卡内基梅隆大学 (Carnegie Mellon University) 为首、麻省理工学院 (Massachusetts Institute of Technology) 等高校参与的视觉监控重大项目 VSAM (Visual Surveillance and Monitoring)^[2], 主要研究用于战场及普通民用场景进行监控的自动视频理解技术; 实时视觉监控系统 W⁴^[3]不仅能够定位人和分割出人的身体部分, 而且通过建立外观模型来实现多人的跟踪, 并可以检测人是否携带物体等简单行为; 英国的雷丁大学 (University of Reading)^[4]已开展了对车辆和行人的跟踪及其交互作用识别的相关研究; IBM 与 Microsoft 等公司也正逐步将基于视觉的手势识别接口应用于商业领域中^[5-6]。当前, 国际上一些权威期刊如 IJCV (International Journal of Computer Vision)、CVIU (Computer Vision and Image Understanding)、PAMI (IEEE Trans. on Pattern Analysis and Machine Intelligence)、IVC (Image and Vision Computing) 和重要的学术会议如 ICCV (International Conference on Computer Vision)、CVPR (IEEE Computer Society Conference on Computer Vision and Pattern Recognition)、ECCV (European Conference on Computer Vision)、IWVS (IEEE International Workshop on Visual Surveillance) 等将人的运动分析研究作为主题内容之一, 为该领域的研究人员提供了更多的交流机会。

基于视觉的人的运动分析在最近几年里取得了一定的进展, 如 Aggarwal 与 Cai 的文章^[7]从人体部分的运动分析、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研究现状作了回顾。该文覆盖了 1998 年以前的研究工作, 但缺乏研究难点分析及未来发展方向的介绍; Gavrilu^[1]从局限于人或者手的运动的研究工作出发, 重点回顾了目前所采用的不同分析方法 (拥有明确形状模型的二维方法、没有明确形状模型的二维方法及三

*国家自然科学基金 (69825105、60105002) 及中科院自动化研究所创新基金 (编号 1M01J02) 资助。

维方法)。同样地, 该文也缺乏具体的研究难点分析, 同时也仅仅提出三维跟踪和行为识别为未来的研究方向; Pentland^[8]围绕人脸识别、视觉监控、三维人的跟踪、智能房间/感知接口四个问题介绍了“人的观察”(Looking at People)领域目前的发展水平和常用的数学工具, 并简单总结了一些研究挑战和机遇。

上述三篇文献的共同缺点是主要从跟踪和行为识别上加以回顾, 缺少人运动分析中运动分割、目标分类、行为描述等的具体介绍, 同时文献的阐述由于没有明确的分类而显得混乱。与以上不同的是, 为了使广大相关科研人员对人的运动分析领域的研究现状有着更加清晰明确的了解, 本文将特地从人的运动分析系统的一般处理框架, 即运动检测、目标分类、人的跟踪、行为理解与描述出发, 分层次地回顾了从低级、中级处理到高级处理各个阶段的目前发展水平; 为避免对于每篇论文的详细总结, 我们重点回顾了各个处理阶段常用的处理方法, 并将每个阶段相应地划分为不同方法的子类以便于讨论; 最后, 我们对研究难点作了一定的分析, 同时对未来的发展趋势作了较为详尽的总结和阐述。

2 典型应用

人运动的视觉分析研究具有广泛的应用前景, Gavrilu^[1]总结了它的一些主要应用领域, 下面据此对其典型应用做出进一步的介绍。

1) 智能监控 (Smart Surveillance)

智能监控^[9, 10]系统的需求主要来自那些对安全要求敏感的场所, 如银行、商店、停车场等。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督作用, 因为它们通常是将摄像机的输出结果记录下来, 当异常情况(如停车场中的车辆被盗)发生后, 保安人员才通过记录的结果观察发生的事实, 但往往为时已晚。而我们需要监控系统应能够每天连续 24 小时的实时监视, 并自动分析摄像机捕捉的图像数据, 当盗窃发生或发现到具有异常行为的可疑的人时, 系统能向保卫人员准确及时地发出警报, 从而避免犯罪的发生, 同时也减少雇佣大批监视所需要的人力、物力和财力的投入; 在访问控制(Access Control)场合, 也可以利用人脸或者步态的跟踪识别技术^[11-16]以便确定来人是否有进入该安全领域的权利; 另外, 人的运动分析在自动售货机、ATM 机、交通管理、公共场所行人的拥挤状态分析^[17, 18]及商店中消费者流量统计等监控方面也有着相应的应用。

2) 感知接口 (Perceptual Interface)

在高级用户接口应用领域中, 我们希望未来的机器能像人一样与我们更加容易和便捷地交流, 如手势驱动控制、手语翻译等。人与人之间的信息交流主要是依靠语言, 并适当结合手势、身体姿势和面部表情等, 因此视觉信息可以作为语音和自然语言理解的有效补充来完成更加智能的人机交互。这就要求未来的计算机必须具备感知外部环境的能力, 即代替传统的键盘或鼠标输入模式、独立地提取周围环境的重要信息(如检测到人的存在等); 更进一步的能力是进行人的识别和行为理解, 结合面部表情、身体姿势和手势等^[19-22]的分析来与人进行相应的交流。对于机场等高噪声的场合, 基于视觉的高级用户接口能够提供比语音识别更加准确的信息输入^[23]。

3) 运动分析 (Motion Analysis)

分割图像中的人体部分并在图像序列中跟踪分析感兴趣的关节运动, 对于建立人体的几何模型、解释人体的运动行为机制从而提高它的运动性能有着积极的推动作用, 这可以应用于体育运动、舞蹈等训练中; 传统的步态分析是肌电图学、关节动力学的研究范畴, 而目前的医学步态分析^[24-26]是一个旨在提供诊断和治疗支持的研究领域, 它可以提供人体正常步态建模的线索, 有助于开发生物反馈系统来分析病人的步态, 判断其腿步受伤情况或者畸形程度, 从而做出积极的整形补偿或有效的治疗; 当然, 步态分析也可以作为独特的生物特征用于人的远距离的身份鉴别^[16]; 另外, 人的运动分析和行为识别也可以应用于从大量的体育

活动数据库中进行基于内容的快速搜索等。

4) 虚拟现实 (Virtual Reality)

目前许多电脑游戏中人的形体、运动和行为交互的设计逼真性实际得益于物理空间中人的运动分析,包括人体模型和关节运动机制的获取及姿势的恢复等;基于互联网络的交互式空间的开发刚刚起步,如虚拟聊天室,它在通过文本交流的同时可以通过二维图标来导航用户,如果增加手势、头的姿势、面部表情等线索,将会给参与者们提供更加丰富的交互形式;另外,人的运动分析在视频会议、人物动画、虚拟工作室等其它虚拟现实场合也有着相当广泛的应用。

3 研究现状

人运动的视觉分析主要是针对包含人的运动图像序列进行分析处理,它通常涉及到运动检测、目标分类、人的跟踪及行为理解与描述几个过程,其一般性处理框架如图 1 所示。其中,运动检测、目标分类、人的跟踪属于视觉中的低级和中级处理部分 (Low-level and Intermediate-level Vision),而行为理解和描述则属于高级处理 (High-level Vision)。当然,它们之间也可能存在交叉(比如有时跟踪过程中运动检测的使用)。下面将重点从处理方法入手回顾这几个方面国内外的发展现状。

3.1 运动检测

运动检测的目的是从序列图像中将变化区域从背景图像中提取出来。运动区域的有效分割对于目标分类、跟踪和行为理解等后期处理是非常重要的,因为以后的处理过程仅仅考虑图像中对应于运动区域的像素。然而,由于背景图像的动态变化,如天气、光照、影子及混乱干扰等的影响,使得运动检测成为一项相当困难的工作。下面归纳出目前几种常用的方法。

1) 背景减除 (Background Subtraction)

背景减除^[3, 27-33]方法是目前运动分割中最常用的一种方法,它是利用当前图像与背景图像的差分来检测出运动区域的一种技术。它一般能够提供最完整的特征数据,但对于动态场景的变化,如光照和外来无关事件的干扰等特别敏感。最简单的背景模型是时间平均图像,大部分的研究人员目前都致力于开发不同的背景模型,以期减少动态场景变化对于运动分割的影响。例如 Haritaoglu 等^[3]利用最小、最大强度值和最大时间差分值为场景中每个像素进行统计建模,并且进行周期性地背景更新;McKenna 等^[27]利用像素色彩和梯度信息相结合的自适应背景模型来解决影子和不可靠色彩线索对于分割的影响;Karmann 与 Brandt^[28]、Kilger^[29]采用基于卡尔曼滤波 (Kalman Filtering) 的自适应背景模型以适应天气和光照的时间变化;Stauffer 与 Grimson^[30]利用自适应的混合高斯背景模型(即对每个像素利用混合高斯分布建模),并且利用在线估计来更新模型,从而可靠地处理了光照变化、背景混乱运动的干扰等影响。

2) 时间差分 (Temporal Difference)

时间差分^[34, 35]方法是在连续的图像序列中两个或三个相邻帧间采用基于像素的时间差分并且阈值化来提取出图像中的运动区域。例如 Lipton 等^[34]利用两帧差分方法从实际视频

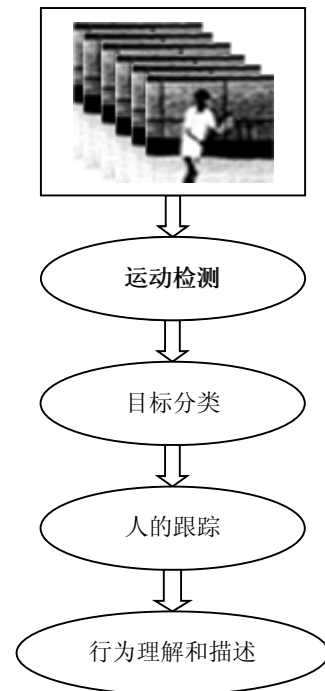


图 1 人运动分析的一般处理框架

图像中检测出运动目标,进而用于目标的分类与跟踪;一个改进的方法是利用三帧差分代替两帧差分,如 VSAM^[2]开发了一种自适应背景减除与三帧差分相结合的混合算法,它能够快速有效地从背景中检测出运动目标。时间差分运动检测方法对于动态环境具有较强的自适应性,但一般不能完全提取出所有相关的特征像素点,在运动实体内部容易产生空洞现象。

3) 光流 (Optical Flow)

基于光流方法^[26, 36, 37]的运动检测采用了运动目标随时间变化的光流特性,如 Meyer 等^[26]通过计算位移向量光流场来初始化基于轮廓的跟踪算法,从而有效地提取和跟踪运动目标。该方法的优点是在摄像机运动存在的前提下也能检测出独立的运动目标。然而,大多数的光流计算方法相当复杂,且抗噪性能差,如果没有特别的硬件装置则不能被应用于全帧视频流的实时处理。关于光流更加详细的讨论可参见 Barron 等的文章^[36]。

当然,在运动变化检测中还有一些其它的方法,如 Friedman 与 Russell^[38]利用扩展的 EM^[39] (Expectation Maximization) 算法,为每个像素建立了混合高斯分类模型,该模型可以自动更新,并能自适应地将每个像素分类为背景、影子或者运动前景,在目标运动速度缓慢的情况下亦能较好地完成运动区域的分割,并可以有效地消除影子的影响;另外,Stringa^[40]也提出了一种新颖的基于数学形态学的场景变化检测算法,在变化的环境条件下获得了相对稳定的分割效果。

3.2 目标分类

目标分类的目的是从检测到的运动区域中将对应于人的运动区域提取出来。不同的运动区域可能对应于不同的运动目标,比如交通道路上监控摄像机所捕捉的序列图像中可能包含行人、车辆及其它诸如飞鸟、流云、摇动的树枝等运动物体,为了便于进一步对行人进行跟踪和行为分析,运动目标的正确分类是完全必要的。注意,这个步骤在一些情况下可能是不必要的(比如已经知道场景中仅仅存在人的运动时)。下面仅给出两种常用的目标分类方法。

1) 基于形状信息的分类 (Shape-based Classification)

基于形状信息的分类是利用检测出的运动区域的形状特征进行目标分类的方法。例如 VSAM^[2]采用区域的分散度、面积、宽高比等作为特征,利用三层神经网络方法将运动目标划分为人、人群、车和背景干扰;Lipton 等^[34]利用分散度和面积信息对二维运动区域进行分类,主要是区分人、车及混乱扰动,时间一致性约束使其分类更加准确;Kuno 与 Watanabe^[41]使用简单的人体轮廓模式的形状参数从图像中检测运动的人。

2) 基于运动特性的分类 (Motion-based Classification)

基于运动特性的分类是利用人体运动的周期性进行目标分类的方法^[42-44]。例如 Cutler 与 Davis^[42]通过跟踪感兴趣的运动目标,计算出目标随着时间变化的自相关特性,而人的周期性运动使得其自相关也是周期性的,因此通过时频化方法分析目标是否存在周期性的运动特性而将人识别出来;Lipton^[43]通过计算运动区域的残余光流 (Residual Flow) 来分析运动实体的刚性和周期性,非刚性的人的运动相比于刚性的车辆运动而言具有较高的平均残余光流,同时它也呈现了周期性的运动特征,据此可以将人区分出来。

上述两种常用的目标分类方法有时可以结合起来使用^[3],甚至可以考虑运动物体色彩或速度等特征,以期得到更加准确的分类结果。另外,Stauffer^[45]提出了利用时间共生矩阵进行分层分类的方法,该方法不仅可以用来区分物体,还可以用来区分行为。

3.3 人的跟踪

跟踪等价于在连续的图像帧间创建基于位置、速度、形状、纹理、色彩等有关特征的对应匹配问题,常用的数学工具有卡尔曼滤波^[46] (Kalman Filtering)、Condensation 算法^[47]及

动态贝叶斯网络^[48] (Dynamic Bayesian Network) 等。其中 Kalman 滤波是基于高斯分布的状态预测方法,不能有效地处理多峰模式 (Multi-mode) 的分布情况; Condensation 算法是以因子抽样为基础的条件密度传播方法,结合可学习的动态模型,可完成鲁棒的运动跟踪。目前,就跟踪对象而言,有跟踪如手、脸、头、腿等身体部分^[12, 21, 22, 50-54]与跟踪整个人体的^[55-83];就跟踪视角而言,有对应于单摄像机的单一视角^[49-66]、对应于多摄像机的多视角^[67-71]和全方位视角^[72];当然还可以通过跟踪空间 (二维或三维)、跟踪环境 (室内或户外)、跟踪人数 (单人、多人、人群)、摄像机状态 (运动或固定) 等方面进行分类。下面仅依据不同的跟踪方法来加以分类介绍。

1) 基于模型的跟踪 (Model-based Tracking)

传统的人体表达方法有如下三种^[81]: a) 线图法 (Stick Figure): 人运动的实质是骨骼的运动,因此该表达方法将身体的各个部分以直线来近似,例如 Karaulova^[66]建立了人体运动学的分层模型,用于单目视频序列中人体的跟踪。b) 二维轮廓 (2-D Contour): 该人体表达方法的使用直接与人体在图像中的投影有关,如 Ju 等^[81]提出的纸板人模型,它将人的肢体用一组连接的平面区域块所表达,该区域块的参数化运动受关节运动 (Articulated Movement) 的约束,该模型被用于关节运动图像的分析; Niyogi 与 Adelson^[65]利用时空切片方法进行人的跟踪: 首先观察由人的下肢轨迹所产生的时空交织模式,然后在时空域中定位头的运动投影,接下来识别其它关节的轨迹,最后利用这些关节轨迹勾画出一个行人的轮廓。c) 立体模型 (Volumetric Model): 它是利用广义锥台、椭圆柱、球等三维模型来描述人体的结构细节,因此要求更多的计算参数和匹配过程中更大的计算量。例如 Rohr^[82]使用 14 个椭圆柱体模型来表达人体结构,坐标系统的原点被定位在躯干的中心,目的是想利用该模型来产生人的行走的三维描述; Wachter 与 Nagel^[63]利用椭圆锥台建立三维人体模型,通过在连续的图像帧间匹配三维人体模型的投影来获得人运动的定量描述,其中,它利用了迭代的扩展卡尔曼滤波方法,结合边缘、区域信息及身体解析约束确定的身体关节运动的自由度,实现单目图像序列中人的跟踪。

2) 基于区域的跟踪 (Region-based Tracking)

基于区域的跟踪方法目前已有较多的应用,例如 Wren 等^[31]利用小区域特征进行室内单人的跟踪,文中将人体看作由头、躯干、四肢等身体部分所对应的小区域块所组成,利用高斯分布建立人体和场景的模型,属于人体的像素被规划于不同的身体部分,通过跟踪各个小区域块来完成整个人的跟踪。基于区域跟踪的难点是处理运动目标的影子和遮挡,这或许可利用彩色信息以及阴影区域缺乏纹理的性质来加以解决,如 McKenna 等^[27]首先利用色彩和梯度信息建立自适应的背景模型,并且利用背景减除方法提取运动区域,有效地消除了影子的影响;然后,跟踪过程在区域、人、人群三个抽象级别上执行,区域可以合并和分离,而人是由许多身体部分区域在满足几何约束的条件下组成的,同时人群又是由单个的人组成的,因此利用区域跟踪器并结合人的表面颜色模型,在遮挡情况下也能够较好地完成多人的跟踪。

3) 基于活动轮廓的跟踪 (Active Contour Based Tracking)

基于活动轮廓的跟踪^[75-80]思想是利用封闭的曲线轮廓来表达运动目标,并且该轮廓能够自动连续地更新。例如 Paragios 与 Deriche^[76]利用短程线的活动轮廓、结合 Level Set 理论在图像序列中检测和跟踪多个运动目标; Peterfreund^[78]采用基于卡尔曼滤波的活动轮廓来跟踪非刚性的运动物体; Isard 与 Blake^[79]利用随机微分方程去描述复杂的运动模型,并与可变形模板相结合应用于人的跟踪。相对于基于区域的跟踪方法,轮廓表达有减少计算复杂度的优点,如果开始能够合理地分开每个运动目标并实现轮廓初始化的话,即使在有部分遮挡存在的情况下也能连续地进行跟踪,然而初始化通常是很困难的。

4) 基于特征的跟踪 (Feature-based Tracking)

基于特征的跟踪包括特征的提取和特征的匹配两个过程。Polana 与 Nelson 的文章^[83]就是一个很好的点特征跟踪的例子，文中将每个行人用一个矩形框封闭起来，封闭框的质心被选择作为跟踪的特征；在跟踪过程中若两人出现相互遮挡时，只要质心的速度能被区分开来，跟踪仍能被成功地执行；该方法的优点是实现简单，并能利用人体运动来解决遮挡问题，但是它仅仅考虑了平移运动，如果结合纹理、彩色及形状等特征可能会进一步提高跟踪的鲁棒性。另外，Segen 与 Pingali 的跟踪系统^[57]使用了运动轮廓的角点作为对应特征，这些特征点采用基于位置和点的曲率值的距离度量在连续帧间进行匹配。以卡尔曼滤波形式的点、线特征的跟踪技术在计算机视觉领域中已被很好地开发，如 Jang 与 Choi^[54]利用区域的形状、纹理、色彩和边缘特征信息建立了活动模板，结合卡尔曼滤波的预测方法，使特征匹配能量函数最小化来完成运动目标的跟踪过程，该活动模型对于非刚性物体的跟踪具有很好的自适应性。

需要指出，基于多摄像机的人的运动跟踪也得到了一定的研究，如 Utsumi^[69]利用多摄像机对人进行跟踪，通过选择最好的视点来解决人与人之间的互遮挡和自身遮挡问题；Cai 等^[70]提出了在不同摄像机之间进行目标匹配的概率方法，并且对相邻摄像机之间的自动切换问题作了探讨。不过，目前有关多摄像机研究的文章并不是很多。

3.4 行为理解与描述

运动检测、目标分类与人的跟踪是人运动分析中研究较多的三个问题，而人的行为理解与描述是近年来被广泛关注的研究热点，它是指对人的运动模式进行分析和识别，并用自然语言等加以描述。行为理解可以简单地被认为是时变数据的分类问题，即将测试序列与预先标定的代表典型行为的参考序列进行匹配。由此可见，行为理解的关键问题是如何从学习样本中获取参考行为序列，并且学习和匹配的行为序列必须能够处理在相似的运动模式类别中空间和时间尺度上轻微的特征变化。

匹配时变数据的技术通常有：a) 动态时间规整 DTW^[85] (Dynamic Time Warping)：DTW 具有概念简单、算法鲁棒的优点，早期被广泛地应用于语音识别中，并且最近才被用于匹配人的运动模式^[86, 87]；对 DTW 而言，即使测试序列模式与参考序列模式的时间尺度不能完全一致，只要时间次序约束存在，它仍能较好地完成测试序列和参考序列之间的模式匹配。b) 隐马尔可夫模型 HMMs^[88] (Hidden Markov Models)：HMMs 是更加成熟的匹配时变数据的技术，它是随机状态机器。HMMs 的使用涉及到训练和分类两个阶段，训练阶段包括指定一个隐马尔可夫模型的隐藏状态数，并且优化相应的状态转换和输出概率以便于产生的输出符号与在特定的运动类别之内所观察到的图像特征相匹配。对于每一个运动类别，一个 HMM 是必须的。匹配阶段涉及到一个特定的 HMM 可能产生相应于所观察图像特征的测试符号序列的概率计算。HMMs 在学习能力和处理未分割的连续数据流方面比 DTW 有更好的优越性，当前被广泛地应用于人的运动模式匹配中^[89-92]。c) 神经网络 NN (Neural Network)：它同样也是目前比较感兴趣的匹配时变数据的方法，如 Guo 等^[93]用其分析人的运动模式；Rosenblum 等^[94]使用径向基函数网络从运动中识别人的情感。

对于人的行为识别而言，目前主要有下述两种方法。

1) 模板匹配方法 (Template Matching)

采用模板匹配技术^[22, 83, 95, 96]的行为识别方法首先将图像序列转换为一组静态形状模式，然后在识别过程中和预先存储的行为标本相比较。例如 Bobick 与 Davis^[95]采用运动能量图像 MEI (Motion Energy Images) 和运动历史图像 MHI (Motion History Images) 来解释图像序列中人的运动。序列中的运动图像首先经差分运算并二值化；而后这些包含运动区域的二值化运动图像随着时间累加形成 MEI；最后 MEI 增强为 MHI，MHI 中每个像素的值与该位置的持续运动时间成比例。每个行为由不同视角下图像序列的 MEI 和 MHI 所组成，从

中可以提取出基于矩的行为特征用于识别阶段的模板匹配。Polana 与 Nelson^[83]利用二维网格的特征进行人的运动识别，首先计算连续帧间的光流场，并将每个光流帧在空间栅格上沿 X 和 Y 方向分解，每个单元格的幅度被累加，从而形成一高维特征向量用于识别；为了归一化运动的持续时间，他们假设人的运动是周期性的，并将整个序列分解为该行为的许多循环过程；最后，采用最近邻算法进行行为识别。模板匹配技术的优点是计算复杂度低、实现简单，然而它对于噪声和运动时间间隔的变化是敏感的。

2) 状态空间方法 (State Space Approaches)

基于状态空间模型^[89-92, 97-98]的方法定义每个静态姿势作为一个状态，这些状态之间通过某种概率联系起来。任何运动序列可以看作作为这些静态姿势的不同状态之间的一次遍历过程，在这些遍历期间计算联合概率，其最大值被选择作为分类行为的标准。目前，状态空间模型已经被广泛地应用于时间序列的预测、估计和检测，最有代表性的是 HMMs。每个状态中可用于识别的特征包括点、线或二维小区域。例如 Yamato 等^[91]利用二维小区域块的运动、彩色、纹理等特征进行人的行为识别：人的运动区域块的网格特征被用作学习和识别的低级特征；学习是利用 HMMs 来为每个类别产生符号模式；模型参数的优化是利用 Baum-Welch 算法实现的；识别是以给定序列图像下前向计算 (Forward Calculation) 的结果而确定。Bregler^[97]基于人体动力学在不同抽象等级的统计分解提出了一个综合性的网络用来识别人的运动：识别过程起始于低级处理，基于运动、彩色相似度和空间接近程度等的小区域块以混合高斯模型来估计，不同的身体部分区域在图像序列中被跟踪；在中级处理阶段，具有一致运动特性的区域被匹配为动力学系统中的简单运动 (如行走被认为是两个简单运动的组成，一个是腿的支撑，一个是腿在空中的摆动)；高级阶段，HMMs 被用来作为这些中级动力系统的混合模型来表达复杂的运动，识别过程通过最大化 HMMs 的后验概率来完成。状态空间方法虽然能克服模板匹配的缺点，但通常涉及到复杂的迭代运算。

人的行为的语义描述^[99-102]近来得到了一定的研究，它是应用自然语言的概念，选择一组运动词语或短句来报告场景中运动目标的行为。比如 Kojima 等^[100]提出了一种新的方法，用于产生实时视频图像中人的行为的自然语言描述：首先，从每帧图像中提取出代表整个人体的头部区域，并利用基于模型的方法，估计其三维姿势和位置；然后，这些参数轨迹被分解为单一运动的基元，评估每个基元的诸如姿势和位置的变化度、与环境中其它一些目标的相对距离等概念特征，并且选择最合适的词语和其它语义元素；最后，利用机器翻译技术，产生人行为描述的自然语言文本。Remagnino 等^[101]提出了一个包含行人与车辆的基于事件的视觉监控系统，该系统提供了三维场景中目标动态活动的文字性描述。

相比于车辆运动而言，图像序列中人的运动描述是更加复杂的；而且，关于行为、事件、状态有着不同的概念，因此如何选择有效充分的表达方式来表达场景的内容是很困难的。目前人的行为描述还只局限于简单的语义解释，对复杂场景中人的行为的语义描述工作还相当艰巨。

4 难点与发展趋势

尽管人的运动分析研究已经取得了一定的成果，但下述几个方面仍是今后研究的难点问题，迫切需要引起广大科研工作者的关注。

1) 运动分割 (Motion Segmentation)

快速准确的运动分割是个相当重要但又是比较困难的问题。这是由于动态环境中捕捉的图像受到多方面的影响，比如天气的变化、光照条件的变化、背景的混乱干扰、运动目标的影子^[103]、物体与环境之间或者物体与物体之间的遮挡、甚至摄像机的运动等，这些都给准确有效的运动分割带来了困难。就以运动目标的影子为例，它可能与被检测的目标相连，也可能与目标分离。在前者情况下，影子扭曲了目标的形状，从而使得以后基于形状

的识别方法不再可靠；在后者情况下，影子有可能被误认为场景中一个完全错误的目标。尽管目前图像运动分割主要利用背景减除方法，但如何建立对于任何复杂环境的动态变化均具有自适应性的背景模型仍是相当困难的问题。一个可喜的发展是，一些研究者们正利用时空统计的方法构建自适应的背景模型，这也许对于不受限环境中的运动分割而言是个更好的选择。

2) 遮挡处理 (Occlusion Handling)

目前，大部分人的运动分析系统都不能很好地解决目标之间互遮挡和人体自遮挡问题，尤其是在拥挤状态下，多人的检测和跟踪问题更是难于处理。遮挡时，人体只有部分是可见的，而且这个过程一般是不可训练的，简单依赖于背景减除进行运动分割的技术此时将不再可靠，为了减少遮挡或深度所带来的歧义性问题，必须开发更好的模型来处理遮挡时特征与身体各部分之间的准确对应问题。另外，一般系统也不能完成何时停止和重新开始身体部分的跟踪，即遮挡前后的跟踪初始化缺少自举方法。当然，可喜的进步是利用统计方法从可获得的图像信息中进行人体姿势、位置等的预测；不过，对于解决遮挡问题最有实际意义的潜在方法应该是基于多摄像机的跟踪系统。

3) 三维建模与跟踪 (3-D Modeling and Tracking)

二维方法在早期人的运动分析中证明是很成功的，尤其对于那些不需要精确的姿势恢复或低图像分辨率的应用场合（如交通监控中的行人跟踪）。二维跟踪有着简单快速的优点，主要的缺点是受摄像机角度的限制。而三维方法在不受限的复杂的人的运动判断（如人的徘徊、握手与跳舞等）、更加准确的物理空间的表达、遮挡的准确预测和处理等方面的优点是二维方法所不能比拟的^[51, 68, 71]；它能提供更加有意义的与身体姿势直接相关的可视化特征用于行为识别；同时，三维恢复对于虚拟现实中的应用也是必需的。目前基于视觉的三维跟踪研究仍相当有限，三维姿势恢复的实例亦很少，且大部分系统由于要求鲁棒性而引入了简化的约束条件。三维跟踪也导致了从图像中人体模型的获取^[104]、遮挡处理、人体参数化建模、摄像机的标定等一系列难题。以建模为例，人体模型通常使用许多形状参数所表达。然而，目前的模型很少利用了关节的角度约束和人体部分的动态特性；而且过去的一些工作几乎都假设 3-D 模型依据先验条件而提前被指定，实际上这些形状参数应当从图像中估计出来。总之，3-D 建模与跟踪在未来工作中应值得更多的关注。

4) 多摄像机的使用 (Use of Multiple Cameras)

使用单一摄像机的三维人的跟踪研究还很缺乏，身体姿势和运动在单一视角下由于遮挡或深度影响而容易产生歧义现象，因此使用多摄像机进行三维姿势跟踪和恢复的优点是很明显的。同时，多摄像机的使用不仅可以扩大监视的有效范围，而且可以提供多个不同的方向视角以用于解决遮挡问题。很明显，未来的人运动分析系统将极大受益于多摄像机的使用。对于多摄像机跟踪系统而言，我们需要确定在每个时刻使用哪一个摄像机或哪一幅图像。也就是说，多摄像机之间的选择和信息融合是一个比较关键的问题。

5) 性能评估 (Performance Evaluation)

一般而言，鲁棒性、准确度、速度是人运动分析系统的三个基本要求。例如，系统的鲁棒性对于监控应用特别重要，这是因为它们通常被要求是自动、连续地工作，因此这些系统对于如噪声、光照、天气等因素的影响不能太敏感；系统的准确度对于控制应用特别重要，例如基于行为或姿势识别的接口控制场合；而系统的处理速度对于那些需要实时高速的监控系统而言更是非常关键。因此，如何选择有效的工作方案来提高系统性能、降低计算代价是个特别值得考虑的问题。同时，如何利用来自不同用户、不同环境、不同实验条件的大量数据测试系统的实时性、鲁棒性亦相当重要。

随着人运动分析研究和其它相关技术的发展，下述几个方面已经成为未来的发展趋势：

1) 音频与视觉相结合的多模态接口

人的相互交流主要是依据语言，过去的许多工作是语音理解，但语音识别受距离和环境噪声的限制，尤其在机场等高噪声环境，将会严重影响语音识别的性能。人的可视化描述与语音解释一样重要，研究者们正逐渐将语音与视觉信息集成起来^[105]以产生更加自然的高级接口。当前一些接口系统在视觉方面仅仅做了如脸的表情、身体姿势等的大尺度分析，但还不能分析大多数人的正常姿势，这意味着人机之间的通讯仅局限于几个特定的姿势，这个局限是人的姿势结构的不易理解造成的，而且跟踪多人的系统由于来自摄像机的分辨率、计算机处理能力和视角的影响而不能准确的估计身体姿势。为了完成优化尺度和广域的分析，可以寻求准确实时的多摄像机的信息融合方法，以便机器更好地理解人的通信行为。目前音频和视频的信号处理相对独立，如何更好地集成音频和视频信息用于多模态用户接口是一个严峻的挑战。

2) 人的运动分析与生物特征识别相结合

在智能房间的门禁系统、军事安全基地的视觉监控系统、高级人机交互等应用中，人的运动分析与生物特征识别相结合的研究日益显得重要。在人机交互中不仅需要机器能知道人是否存在、人的位置和行为，而且还需要利用特征识别技术来识别与其交流的人是谁。远距离的身份识别已经越来越重要，比如 2000 年 DARPA 赞助的重大项目——HID (Human Identification at a Distance) 计划。目前的研究主要集中在人的脸像识别、步态识别或特定行为的识别。近距离时一般可以通过跟踪人脸来加以身份识别；如果是远距离的监控，脸的特征可能被隐藏，或者分辨率太低不易识别，然而进入监控领域的人的运动步态是可见的，这激活了步态作为一个独特的生物行为特征应用于人的身份鉴别。由于人的步态具有易于感知、非侵犯性、难于伪装等优点，近来已引起了计算机视觉研究者们浓厚的研究兴趣^[14-16]。

3) 人的运动分析向行为理解与描述高层处理的转变

人的行为理解是需要引起高度注意并且是最具挑战的研究方向，因为观察人的最终目标就是分析和理解人的个人行为、人与人之间及人与其它目标的交互行为等，如 W⁴ 系统^[3]可以分析人是否携带物体、放置物体、交换物体等简单的行为。目前人的运动理解还是集中于人的跟踪、标准姿势识别、简单行为识别等问题，如人的一组最通常的行为（跑、蹲、站、跳、爬、指等）的定义和分类。近年来利用机器学习工具构建人行为的统计模型的研究有了一定的进展，但行为识别仍旧处于初级阶段，连续特征的典型匹配过程中常引入人运动模型的简化约束条件来减少歧义性，而这些限制与一般的图像条件却是不吻合的，因此行为理解的难点仍是在于特征选择和机器学习。目前，用于行为识别的状态空间方法和模板匹配方法通常在计算代价和运动识别的准确度之间进行折中，故仍需要寻找和开发新的技术以利于提高行为识别性能的同时，又能有效地降低计算的复杂度。另外，如何借助于先进的视觉算法和人工智能等领域的成果，将现有的简单的行为识别与语义理解推广到更为复杂场景下的自然语言描述，是将计算机视觉低、中层次的处理推向高层抽象思维的关键问题。

5 结束语

人的运动分析，尤其是人的行为理解与描述，已经成为计算机视觉中一个重要的研究领域，它在智能监控、虚拟现实、用户接口等方面的应用前景引起了广大科研人员的浓厚兴趣。近来的技术发展更是证明了视觉系统具有成功处理人的复杂运动的能力，而且许多研究者们正逐步将其成果推向更加智能的应用场合。本文从人运动分析的一般处理过程出发，回顾了各个处理阶段的目前发展情况和常用的方法，同时对研究难点和未来的发展趋势作了较为详细的阐述，希望能对相关领域的研究人员和工程技术人员有所裨益。

参 考 文 献

- 1 Gavrilu D. The visual analysis of human movement: a survey. *Computer Vision and Image Understanding*, 1999, 73 (1): 82-98.
- 2 Collins R et al. A system for video surveillance and monitoring: VSAM final report. Carnegie Mellon University, Technical Report: CMU-RI-TR-00-12, 2000.
- 3 Haritaoglu I, Harwood D and Davis L. W⁴: real-time surveillance of people and their activities. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (8): 809-830.
- 4 Remagnino P, Tan T and Baker K. Multi-agent visual surveillance of dynamic scenes. *Image and Vision Computing*, 1998, 16 (8): 529-532.
- 5 Maggioni C and Kammerer B. *Gesture Computer: history, design, and applications*. Computer Vision for Human-Machine Interaction, Cambridge Univ. Press, 1998.
- 6 Freeman W and Weissman C. Television control by hand gestures. In: *Proc International Conference on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995, 179-183.
- 7 Aggarwal J and Cai Q. Human motion analysis: a review. *Computer Vision and Image Understanding*, 1999, 73 (3): 428-440.
- 8 Pentland A. Looking at people: sensing for ubiquitous and wearable computing. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (1): 107-119.
- 9 Collins R, Lipton A and Kanade T. Introduction to the special section on video surveillance. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (8): 745-746.
- 10 Maybank S and Tan T. Introduction—special section on visual surveillance. *International Journal of Computer Vision*, 2000, 37 (2): 173-173.
- 11 Steffens J, Elagin E and Neven H. Person Spotter-fast and robust system for human detection, tracking and recognition. In: *Proc IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, 516-521.
- 12 Yang J and Waibel A. A real-time face tracker. In: *Proc IEEE Workshop on Applications of Computer Vision*, Sarasota, USA, 1996, 142-147.
- 13 Moghaddam B, Wahid W and Pentland A. Beyond eigenfaces: probabilistic matching for face recognition. In: *Proc IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, 30-35.
- 14 Little J and Boyd J. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, The MIT Press, 1998, 1 (2): 2-32.
- 15 Shutler J D, Nixon M and Harris C. Statistical gait description via velocity moments. In: *Proc IEEE Southwest Symposium on Image Analysis and Interpretation*, Austin, Texas, 2000, 291-295.
- 16 Huang P, Harris C and Nixon M. Human gait recognition in canonical space using temporal templates. In: *Proc IEE Vis. Image Signal Process*, 1999, 146 (2): 93-100.
- 17 Boghossian B and Velastin S. Image processing system for pedestrian monitoring using neural classification of normal motion patterns. *Measurement and Control*, 1999, 32 (9): 261-264.
- 18 Boghossian B and Velastin S. Motion-based machine vision techniques for the management of large crowds. In *IEEE 6th International Conference on Electronics, Circuits and systems*, Pafos, Cyprus, 1999.
- 19 Li Y, Ma S and Lu H. Human posture recognition using multi-scale morphological method and Kalman motion estimation. In: *Proc IEEE International Conference on Pattern Recognition*, Brisbane, Australia, 1998, 175-177.
- 20 Segen J and Kumar S. Shadow gestures: 3D hand pose estimation using a single camera. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, 1999, 479-485.
- 21 Yang M-H and Ahuja N. Recognizing hand gesture using motion trajectories. In: *Proc IEEE Conference on*

- Computer Vision and Pattern Recognition, Fort Collins, Colorado, 1999, 468-472.
- 22 Cui Y and Weng J. Hand segmentation using learning-based prediction and verification for hand sign recognition. In: Proc IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997, 88-93.
 - 23 Turk M. Visual interaction with lifelike characters. In: Proc IEEE International Conference on Automatic Face and Gesture Recognition, Killington, Vermont USA, 1996, 368-373.
 - 24 Lakany H., Hayes G, Hazlewood M and Hillman S. Human walking: tracking and analysis. In: Proc IEE Colloquium on Motion Analysis and Tracking, Savoy Place, London, 1999, 5/1-5/14.
 - 25 Köhle M, Merkl D and Kastner J. Clinical gait analysis by neural networks: issues and experiences. In: Proc IEEE Symposium on Computer-Based Medical Systems, Maribor, Slovenia, 1997, 138-143.
 - 26 Meyer D, Denzler J and Niemann H. Model based extraction of articulated objects in image sequences for gait analysis. In: Proc IEEE International Conference on Image Processing, Santa Barbara, California 1997, 78-81.
 - 27 McKenna S et al, Tracking groups of people. *Computer Vision and Image Understanding*, 2000, 80 (1): 42-56.
 - 28 Karmann K and Brandt A. Moving object recognition using an adaptive background memory. In: V Cappellini, *Time-varying Image Processing and Moving Object Recognition*. 2. Elsevier, Amsterdam, The Netherlands, 1990.
 - 29 Kilger M. A shadow handler in a video-based real-time traffic monitoring system. In: Proc IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, 1992, 1060-1066.
 - 30 Stauffer C and Grimson W. Adaptive background mixture models for real-time tracking. In: Proc IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, 1999, 2: 246-252.
 - 31 Wren C, Azarbayejani A, Darrell T and Pentland A. Pfinder: real-time tracking of the human body. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, 19 (7): 780-785.
 - 32 Arseneau S and Cooperstock J. Real-time image segmentation for action recognition. In: Proc IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, Canada, 1999, 86-89.
 - 33 Sun H, Feng T and Tan T. Robust extraction of moving objects from image sequences. In: Proc the Fourth Asian Conference on Computer Vision, Taiwan, 2000, 961-964.
 - 34 Lipton A, Fujiyoshi H and Patil R. Moving target classification and tracking from real-time video. In: Proc IEEE Workshop on Applications of Computer Vision, Princeton, NJ, 1998, 8-14.
 - 35 Anderson C, Bert P and Vander Wal G. Change detection and tracking using pyramids transformation techniques. In: Proc SPIE Conference on Intelligent Robots and Computer Vision, Cambridge, MA, 1985, 579: 72-78.
 - 36 Barron J, Fleet D and Beauchemin S. Performance of optical flow techniques. *International Journal of Computer Vision*, 1994, 12 (1): 42-77.
 - 37 Verri A, Uras S and DeMicheli E. Motion Segmentation from optical flow. In: Proc the 5th Alvey Vision Conference, Brighton, UK, 1989, 209-214.
 - 38 Friedman N and Russell S. Image segmentation in video sequences: a probabilistic approach. In: Proc the Thirteenth Conference on Uncertainty in Artificial Intelligence, Rhode Island, USA, 1997.
 - 39 McLachlan G and Krishnan T. *The EM Algorithm and Extensions*. Wiley Interscience, 1997.
 - 40 Stringa E. Morphological change detection algorithms for surveillance applications. In: *British Machine Vision Conference*, Bristol, UK, 2000, 402-411.
 - 41 Kuno Y, Watanabe T, Shimosakoda Y and Nakagawa S. Automated detection of human for visual surveillance system. In: Proc IEEE International Conference on Pattern Recognition, Vienna, 1996, 865-869.

- 42 Cutler R and Davis L. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (8): 781-796.
- 43 Lipton A. Local application of optic flow to analyse rigid versus non-rigid motion. In: <http://www.eecs.lehigh.edu/FRAME/Lipton/iccvframe.html>.
- 44 Selinger A and Wixson L. Classifying moving objects as rigid or non-rigid without correspondences. In: *Proc DAPRA Image Understanding Workshop*, Monterey, CA, 1998, 1: 341-358.
- 45 Stauffer C. Automatic hierarchical classification using time-base co-occurrences. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition*, 1999, pp. 333-339.
- 46 Welch G and Bishop G. An introduction to the Kalman filter. In: <http://www.cs.unc.edu>, UNC-Chapel Hill, TR95-041, 2000.
- 47 Isard M and Blake A. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1998, 29 (1): 5-28.
- 48 Pavlović V, Rehg J, Cham T-J and Murphy K. A dynamic Bayesian network approach to figure tracking using learned dynamic models. In: *Proc IEEE International Conference on Computer Vision*, Corfu, Greece, 1999, 94-101.
- 49 Haritaoglu I, Harwood D and Davis L. W⁴: who? when? where? what? A real time system for detecting and tracking people. In: *Proc IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, 222-227.
- 50 Goncalves L, Bernardo E, Ursella E and Perona P. Monocular tracking of the human arm in 3D. In: *Proc the 5th International Conference on Computer Vision*, Cambridge, 1995, 764-770.
- 51 Rehg J and Kanade T. Visual tracking of high DOF articulated structures: an application to human hand tracking. In: *Proc European Conference on Computer Vision*, Stockholm, Sweden, 1994, 35-46.
- 52 Meyer D et al. Gait classification with HMMs for Trajectories of body parts extracted by mixture densities. In: *British Machine Vision Conference*, Southampton, England, 1998, 459-468.
- 53 Fieguth P and Terzopoulos D. Color-based tracking of heads and other mobile objects at video frame rate. In: *Proc IEEE Conference on Computer vision and Pattern Recognition*, Puerto Rico, 1997, 21-27.
- 54 Jang D-S and Choi H-I. Active models for tracking moving objects. *Pattern Recognition*, 2000, 33 (7): 1135-1146.
- 55 Azarbayejani A and Pentland A. Real-time self-calibrating stereo person tracking using 3D shape estimation from blob features. In: *Proc International Conference on Pattern Recognition*, Vienna, 1996, 627-632.
- 56 Cai Q, Mitiche A and Aggarwal J. Tracking human motions in an indoor environment. In: *Proc International Conference on Image Processing*, Washington, DC, USA, 1995, 1: 215-218.
- 57 Segen J and Pingali S. A camera-based system for tracking people in real time. In: *Proc International Conference on Pattern recognition*, Vienna, 1996, 63-67.
- 58 Rehg J and Kanade T. Model-based tracking of self-occluding articulated objects. In: *Proc the 5th International Conference on Computer Vision*, Cambridge, 1995, 612-617.
- 59 Chang I-C and Huang C-L. Ribbon-based motion analysis of human body movements. In: *Proc International Conference on Pattern Recognition*, Vienna, 1996, 436-440.
- 60 Cham T-J and Rehg J. A multiple hypothesis approach to figure tracking. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, 1999, 239-245.
- 61 Ricquebourg Y and Boutheymy P. Real-time tracking of moving persons by exploiting spatio-temporal image slices. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (8): 797-808.
- 62 Darrell T, Gordon G, Harville M and Woodfill J. Integrated person tracking using stereo, color, and pattern detection. *International Journal of Computer Vision*, 2000, 37 (2): 175-185.

- 63 Wachter S and Nagel H-H. Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 1999, 74 (3): 174-192.
- 64 Rossi M and Bozzoli A. Tracking and counting people. In: *Proc International Conference on Image Processing*, Austin, 1994, 212-216.
- 65 Niyogi S and Adelson E. Analyzing and recognizing walking figures in XYT. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, USA, 1994, 469-474.
- 66 Karaulova I, Hall P and Marshall A. A hierarchical model of dynamics for tracking people with a single video camera. In: *British Machine Vision Conference*, Bristol, UK, 2000, 352-361.
- 67 Chang T, Gong S and Ong E. Tracking multiple people under occlusion using multiple cameras. In: *British Machine Vision Conference*, Bristol, UK, 2000, 566-575.
- 68 Kakadiaris I and Metaxas D. Model-based estimation of 3-D human motion with occlusion based on active multi-viewpoint selection. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1996, 81-87.
- 69 Utsumi A, Mori H, Ohya J and Yachida M. Multiple-view-based tracking of multiple humans. In: *Proc IEEE International Conference on Pattern Recognition*, Brisbane, Australia, 1998, 597-601.
- 70 Cai Q and Aggarwal J. Tracking human motion using multiple cameras. In: *Proc International Conference on Pattern Recognition*, Vienna, 1996, 68-72.
- 71 Gavrilu D and Davis L. 3-D model-based tracking of humans in action: a multi-view approach. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1996, 73-80.
- 72 Boulton T. Frame-rate multi-body tracking for surveillance. In: *Proc DARPA Image Understanding Workshop*, Monterey, Calif. San Francisco: Morgan Kaufmann, 1998.
- 73 Goddard N. Incremental model-based discrimination of articulated movement from motion features. In: *Proc IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994, 89-94.
- 74 Fujiyoshi H and Lipton A. Real-time human motion analysis by image skeletonization. In: *Proc IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, 1998, 15-21.
- 75 Zhong Y, Jain A and Dubuisson-Jolly M. Object tracking using deformable templates. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (5): 544-549.
- 76 Paragios N and Deriche R. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (3): 266-280.
- 77 Bertalmio M, Sapiro G and Randall G. Morphing active contours. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (7): 733-737.
- 78 Peterfreund N. Robust tracking of position and velocity with Kalman snakes. *IEEE Trans Pattern Analysis and Machine Intelligence*, 2000, 22 (6): 564-569.
- 79 Isard M and Blake A. Contour tracking by stochastic propagation of conditional density. In: *Proc European Conference on Computer Vision*, Cambridge, 1996, 343-356.
- 80 Baumberg A and Hogg D. An efficient method for contour tracking using active shape models. In: *Proc IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994, 194-199.
- 81 Ju S, Black M and Yacobi Y. Cardboard people: a parameterized model of articulated image motion. In: *Proc IEEE International Conference on Automatic Face and gesture Recognition*, Killington, Vermont USA, 1996, 38-44.
- 82 Rohr K. Towards model-based recognition of human movements in image sequences. *CVGIP: Image Understanding*, 1994, 59 (1): 94-115.
- 83 Polana R and Nelson R. Low level recognition of human motion. In: *Proc IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, TX, 1994, 77-82.

- 84 Aggarwal J, Cai Q, Liao W and Sabata B. Non-Rigid motion analysis: articulated & elastic motion. *Computer Vision and Image Understanding*, 1998, 70 (2): 142-156.
- 85 Myers C, Rabinier L and Rosenberg A. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Trans Acoustics, Speech, and Signal Processing*, 1980, 28 (6): 623-635.
- 86 Bobick A and Wilson A. A state-based technique for the summarization and recognition of gesture. In: *Proc International Conference on Computer Vision*, Cambridge, 1995, 382-388.
- 87 Takahashi K, and Seki S et al. Recognition of dexterous manipulations from time varying images. In: *Proc IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994, 23-28.
- 88 Poritz A. Hidden Markov Models: a guided tour. In: *Proc IEEE International Conference on Acoustics, Speech and Signal Processing*, New York City, NY, 1988, 7-13.
- 89 Rabinier L. A tutorial on hidden Markov models and selected applications in speech recognition. In: *Proc IEEE* 1989, 77 (2): 257-285.
- 90 Starner T and Pentland A. Real-time American Sign Language recognition from video using hidden Markov models. In: *Proc International Symposium on Computer Vision*, Coral Gables, Florida, 1995, 265-270.
- 91 Yamato J, Ohya J and Ishii K. Recognizing human action in time-sequential images using hidden Markov model. In : *Proc IEEE Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, 1992, 379-385.
- 92 Brand M, Oliver N and Pentland A. Coupled hidden Markov models for complex action recognition. In: *Proc IEEE Conference Computer Vision and Pattern Recognition*, Puerto Rico, 1997, 994-999.
- 93 Guo Y, Xu G and Tsuji S. Understanding human motion patterns. In: *Proc International Conference on Pattern Recognition*, Jerusalem, Israel, 1994, 325-329.
- 94 Rosenblum M, Yacoob Y and Davis L. Human emotion recognition from motion using a radial basis function network architecture. In: *Proc IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994, 43-49.
- 95 Bobick A and Davis J. Real-time recognition of activity using temporal templates. In: *Proc IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, 1996, 39-42.
- 96 Davis J and Bobick A. The representation and recognition of action using temporal templates. MIT Media Lab, Perceptual Computing Group, Technical report: 402, 1997.
- 97 Bregler C. Learning and recognizing human dynamics in video sequences. In: *Proc IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997, 568-574.
- 98 Campbell L and Bobick A. Recognition of human body motion using phase space constraints. In: *Proc International Conference on Computer Vision*, Cambridge, 1995, 624-630.
- 99 Intille S and Bobick A. Representation and visual recognition of complex, multi-agent actions using belief networks. Perceptual Computing Section, MIT Media Lab, Technical Report: 454, 1998.
- 100 Kojima et al. Generating natural language description of human behaviors from video images. In: *Proc IEEE International Conference on Pattern Recognition*, Barcelona, 2000, 728-731.
- 101 Remagnino P, Tan T and Baker K. Agent orientated annotation in model based visual surveillance. In: *Proc IEEE International Conference on Computer Vision*, Bombay, India, 1998, 857-862.
- 102 Herzog G and Rohr K. Integrating vision and language: Towards automatic description of human movements. In: *Proc the 19th Annual German Conference on Artificial Intelligence*, Bielefeld, 1995, 257-268.
- 103 Onoguchi K. Shadow elimination method for moving object detection. In: *Proc IEEE International Conference on Pattern Recognition*, Brisbane, Australia, 1998, 583-587.
- 104 Kakadiaris I and Metaxas D. Three-dimensional human body model acquisition from multiple views. *International Journal of Computer Vision*, 1998, 30 (3): 191-218.

A Survey of Visual Analysis of Human Motion

WANG Liang HU Wei-Ming TAN Tie-Niu

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100080)

Abstract Visual analysis of human motion is currently one of the most active research topics in the domain of computer vision. This strong interest is driven by a wide spectrum of promising applications in many areas such as virtual reality, smart surveillance, perceptual user interface, content-based image storage and retrieval, athletic performance analysis, etc. Human motion analysis aims at attempting to detect, track and identify people, and more generally, to understand human behaviors, from image sequences involving humans. This paper provides a comprehensive survey of recent developments of vision-based human motion analysis, and it keeps up with the latest research ranging mainly from 1995 to 2000. Different from previous reviews, our emphasis is on four major issues involved in a general framework of human motion analysis, namely motion detection, moving object classification, human tracking, and activity recognition and description. This paper focuses more on overall methods and general characteristics involved in the above four issues, so each issue is accordingly divided into sub-processes and categories of approaches so as to provide more detailed discussions. First, we introduce some potential applications of human motion analysis. Then, various existing methods for each key issue are clearly discussed to examine the state-of-the-art in human motion analysis. Motion detection provides a focus of attention for later processes because only those changing pixels need be considered. Three types of techniques are addressed, namely background subtraction, temporal differencing and optical flow. As far as moving object classification is concerned, shape-based or motion-based methods are presented. Tracking is equivalent to establishing correspondence of image features between consecutive frames, and four approaches studied intensively in past work are described: model-based, active-contour-based, region-based and feature-based. The task of recognizing human activity in image sequences assumes that feature tracking for recognition has been accomplished. Two types of techniques, template matching and state-space approaches, are reviewed. Although a large amount of work has been done in the field of human motion analysis, many issues still remain open such as segmentation, modeling and occlusion handling. At the end of this survey, some detailed discussions on research challenges and future directions in human motion analysis are also provided. Past achievements show to some extent that vision systems have considerable ability to cope with complex human movements, so we are looking forward to more new techniques to devote this field.

Keywords human motion, visual analysis, motion detection, tracking, behavior understanding.



作者简介:

王亮, 男, 1977年9月出生, 博士研究生, 主要的研究方向是模式识别、计算机视觉、数字图像分析和处理、视觉监控等。

About the author—Wang Liang received his B. Sc. (1997) and M. Sc. (2000) in the Department of Electronics Engineering and Information Science from Anhui University, China. He is currently a Ph. D. candidate in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than 6 papers on major national journals and international conferences. His main research interests include pattern recognition, computer vision, digital image processing and analysis, visual surveillance, etc.

胡卫明, 男, 1968年11月出生, 博士, 副研究员, 硕士生导师, 主要的研究方向是动态场景的视觉监控、神经网络、图像处理、计算机图形学等。

About the author—Hu Weiming received his Ph. D. Degree from the Department of Computer Science and Engineering, Zhejiang University, China. From April 1998 to March 2000, he worked as a Postdoctoral Research Fellow at the Institute of Computer Science and Technology, Founder Research and Design Center, Peking University. From April 2000, he worked at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, as an Associate Professor. He has published more than 20 papers on major national journals, such as Science in China, Chinese Journal of Computers, Chinese Journal of Software and Chinese Journal of Semiconductors. His research interests are in visual surveillance and monitoring of dynamic scenes, neural network, image processing, 3D computer graphics, etc.

谭铁牛, 男, 1964年1月出生, 博士, 研究员, 博士生导师, 主要的研究方向是语音与图像处理、计算机视觉、模式识别、多媒体与机器人等。

About the author—Tan Tieniu received his B. Sc. (1984) in electronic engineering from Xi'an Jiaotong University, China, and M. Sc. (1986), DIC (1986) and Ph. D. (1989) in electronic engineering from Imperial College of Science, Technology and Medicine, London, U. K. In October 1989, he joined the Computational Vision Group at the Department of Computer Science, The University of Reading, England, where he worked as Research Fellow, Senior Research Fellow and Lecturer. In January 1998, he returned to China to join the National Laboratory of Pattern Recognition, the Institute of Automation of the Chinese Academy of Sciences, Beijing, China. He is currently Professor and Director of the National Laboratory of Pattern Recognition and Director of the Institute of Automation. He is a Senior Member of the IEEE and was an elected member of the Executive Committee of the British Machine Vision Association and Society for Pattern Recognition (1996-1997). He serves as referee for many major national and international journals and conferences. He is an Associate Editor of the International Journal of Pattern Recognition, the Asia Editor of the International Journal of Image and Vision Computing and is a founding co-chair of the IEEE International Workshop on Visual Surveillance. His current research interests include speech and image processing, machine and computer vision, pattern recognition, multimedia, and robotics.

联系方式:

电 话: 010-62647441, 传 真: 010-62551993

电子邮件: {lwang, wmhu, tnt}@nlpr.ia.ac.cn