基于视觉的人体运动分析综述①

阮涛涛，姚明海，瞿心昱，楼中望

(浙江工业大学 信息工程学院，杭州 310023)

摘 要: 基于视觉的人运动分析越来越受到计算机视觉领域研究者的广泛关注，它成为图像分析、心理学、人工智能等领域的研究热点，在智能视频监控、虚拟现实、用户接口、运动分析等方面有着广泛的应用。从运动目标检测、运动目标分类、人体运动跟踪、人体行为识别与描述四个环节综述了人体运动分析的研究现状，分析了存在的一些问题和未来的研究发展方向。

关键词: 人体运动分析；运动检测；跟踪；行为识别

A Survey of Vision-Based Human Motion Analysis

RUAN Tao-Tao, YAO Ming-Hai, QU Xin-Yu, LOU Zhong-Wang

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: Vision-based human motion analysis is receiving increasing attention from computer vision researchers. It becomes the hotspot of the field of image analysis, psychology and artificial intelligence. It has a wide application in intelligent video surveillance, virtual reality, user interfaces and motion analysis. In this paper, moving target detection, classification, tracking and human action recognition are overviewed. We analyze some problems and challenges. Finally, some research directions in the future are discussed.

Keywords: human motion analysis; motion detection; tracking; action recognition

1 引言

人是社会活动的主体，在计算机上研究并且解释人的行为是一件令人兴奋且有意义的事。基于视觉的人体运动分析由于其本身的复杂性和广泛的应用前景[5]，成为当前非常活跃的研究领域[1-3]。人体运动分析是指运用某些算法跟踪人体的运动，识别并解释人的行为。人体运动分析是一个宽广的概念，从理论上来说，可以包括面部运动，手势运动和肌肉收缩引起的皮肤表面的变化[6]，其最终目的是达到对人体运动的理解并加以应用。基于视觉的人体运动分析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，是一个多学科的交叉研究方向[3]。由于人体是非刚性运动体，并且由于宽松的衣服、相互遮挡、以及人影、光照变化以及噪声的影响，对人体运动分析将会促进这些领域在理论上产生新的处理方法，并将对诸多应用领域产生潜在的影响。

基于视觉的人体运动分析具有广泛的应用前景，Gavrila[8]总结了其应用的主要领域。结合近年来的发展，我们在其基础上主要归纳为：(1)智能视频监控(Intelligent Video Surveillance)。智能视频监控主要用于对安全要求较高的场合，比如停车场、超市、百货大楼、银行、ATM 柜员机、机场、码头等。视频监控系统能够监视一定场所中人的活动，并对其行为进行分析和识别，跟踪可疑行为从而采取相应的报警措施。(2)虚拟现实(Virtual Reality)。虚拟现实是人们通过计算机对复杂数据进行可视化操作与交互的一种全新方式。虚拟现实创建一个虚拟的仿真场景，实现人与这个虚拟世界的交互。该领域的具体应用涉及视频游戏、虚拟摄影棚、视频会议、人物动画等方面。对物理空间中人的运动分析可应用于电脑游戏中逼真的人物动作、行为及军事中的单兵模拟训练。(3)用户接口(UserInterfaces)。在高级用户接口中，可以通过手势的

① 基金项目:浙江省自然科学基金(20080376)
收稿时间:2010-04-22;收到修改稿时间:2010-05-21

识别来代替传统的鼠标和键盘输入，从而实现人与计算机之间的智能交互。(4)运动分析(Motion Analysis)。人体运动分析可以应用于各种体育项目中，提取运动员的各项运动诸元，便于进行量化分析，结合人体生理学、物理学原理，研究改进的方法，可以为运动员的训练提供参考和指导，有助于提高运动员的训练水平。这使得体育训练摆脱纯粹的依靠经验的模式，进入理论化、数字化的时代。此外，运用于基于内容的视频检索，可以从大量的体育视频数据库中检索到感兴趣的内容。(5)另外在汽车行业中的安全气囊自动控制、驾驶者睡眠检测、行人探测、偏离车道检查等方面[5]，人体视觉运动分析也具有广泛的应用。

由于人体运动识别的巨大应用价值，很多研究机构、学校以及公司的研究人员投入其中。麻省理工学院、卡内基梅隆大学、英国邓迪大学和俄亥俄州立大学等都成立了专门的基于视觉的人体运动实验室。Moeslund 和 Granum[1]对 1980 年到 2000 年上半年的工作做了回顾和分析，对基于计算机视觉的人体运动捕获(Capture)，主要从初始化(Initialization)、跟踪(Tracking)、姿态估计(Pose Estimation)和识别(Recognition)四个方面作为详尽的论述，并且提出该领域的发展方向。Moeslund 等[5]接着对 2000 年到 2006 年这四个方面最新进展做了分析。Poppe[6]从基于模型(Model-Based)和无模型(Model-Free)两方面，重点对身体轮廓的构建进行分析，没有细致研究对身体运动的解释。Poppe[7]又对人体行动的识别进行了分析研究，但没有考虑环境上下文。国内方面，王亮等[2]从运动检测、目标分类、人的跟踪、行为理解与描述四个方面对人的运动作了详尽分析，该文对 2000 年以前国外有关人体运动分析领域的工作做了总结。杜友田等[3]从人运动的类别、人运动的表示方法、人运动的识别技术和方法三个方面分析了人运动分析的进展，侧重点为人的运动识别。黎松洪等[4]借鉴王亮[2]的结构，大致分析了到 2007 年以来人体运动分析研究的若干新进展。本文也借鉴王亮[2]的文章结构，介绍运动检测、目标分类、人的跟踪、行为理解与描述的近几年的发展。

2 运动目标检测

运动目标检测是整个人体运动视觉分析的最底层，是后续处理如目标分类、运动跟踪、行为理解与描述等的基础。它是在一段图像序列的每帧图像中找到运动目标所在位置，其难点在于如何快速精确地找出匹配目标。赵文哲等[9]对运动目标检测方法进行了对比分析。

2.1 背景减除法(Background Subtraction)

背景减除法是目前运动检测常用的方法，Elhabian 等[10]对背景减除法做了大概综述。这种方法一般以摄像头固定为前提，利用当前图像与背景图像的差分来检测运动目标。背景减除法定位精确、速度快，但对于动态场景的变化，缺乏合理的背景更新方法，并且对光照条件或外部环境的变化比较敏感。背景减除法一般包括背景的提取、背景更新和背景差分。常用背景提取方法有背景统计法，卡尔曼滤波法(Kalman Filtering)和背景模型法。其中，背景模型法应用广泛，研究人员通常研究如何获取一个背景的模型，使其能满足场景的动态变化，一般是用一定的算法来动态生成一个合适的背景模型，并按一定的时间来更新此模型。当前研究较多的模型主要有高斯模型、混合高斯模型、非参数化模型等。

Hironobu 等[11]提出了图像差分阈值判定法。Shoushtarian 等[12]提出了三种动态背景减除算法并进行了比较。Herrero-Jaraba 等[13]提出了在动态场景下基于背景减除法的物体检测法，他们使用了长时背景和短时背景的双背景方法来分别处理长时间不变和短时改变的背景。Stauffer 等[14-15]提出了基于混合高斯模型，利用在线估计更新模型。文献[16-17]提出了对高斯混合模型的一些改进方法。Kim 等[18]将背景值量化成编码本，用编码本描述长视频中背景模型的压缩形式。Elagammal 等[19]通过核心密度估计建立了一种非参数化的背景模型，这种方法能适应小扰动场合下的背景。Wu 等[20]提出使用编码本的时空环境(Spatio-temporal Context)下的动态背景减除法。

2.2 时间差分法(Temporal Difference)

时间差分法又可以称为帧间差分法、帧差法，也是最常用的目标检测和分割方法之一。帧差法用相邻两帧或三帧的像素差分值来提取图像中的运动区域，如果差的绝对值小于某一阀值，则没有运动；反之，则有运动，所以阈值的选择是比较重要的。帧差法具有更新速度快、算法复杂度低、计算量小、可连续处理等优点，但存在对环境噪声敏感、一般难以获得目标完整轮廓、提取运动目标位置不精确(可能在目标内部产生空洞)等缺点。

VSAM 项目组[21]提出了用三帧差分法来进行运动检测，并使用自适应背景减除法，以消除空洞现象。Spagnolo 等[22]提出用领域相关系数结合帧差和背景减除法，有效地抑制了光照变化对检测结果的影响，并解决阴影、重影等问题。

2.3 光流法(Optical Flow)

光流法是运动目标检测的重要方法之一。它研究的是利用图像序列中的像素强度数据的时域变化和相关性来确定各自像素位置的“运动”。光流法大致可分为基于匹配的、频域的和梯度的三类方法。部分早期研究工作可参考 Barron 等[23]的工作。光流法的缺点是运动边界和多运动问题、对噪声敏感、计算方法复杂、计算量大。有时由于遮挡、人体影子、光照条件变化等影响，难以有效检测运动人体。

Ahmad 等[24]在图像序列中使用了一种组合局部整体(Combined Local-global)光流法来提取运动特征，用偏离光流不变矩来提取整体形状光流特征。Gehrig 等[25]用 Lucas 和 kanade[26]的方法的锥形描述来计算光流，并用光流梯度直方图来描述特征。熊静旖等[27]引入了一种初始运动估计器 (扩展相位相关法) 来改善光流法的性能，并能计算某些复杂的运动模式。这种改进的光流法可显著提高配准精度，特别是对存在大尺度位移的图像，并且对随机噪声不敏感。Ahad 等[28]用四个方向的光流来计算运动模板，并用该模板的 Hu 矩来创造特征向量。Efros 等[29]对远距离、低分辨率的行为主体提出用校正和模糊的方法提高光流法对噪声的鲁棒性。Li 等[30]用光流方向直方图的方法增强了光流特征的稳定性。

3 运动目标分类

目标分类是基于视觉的人体运动分析研究课题中一个重要的内容。场景中的行为分析一般要依靠目标分类的结果。目标分类主要涉及两方面的问题，一是特征提取表示问题；二是分类准则的定义问题。对给定的可能包含运动目标的视频图像来说，不同的前景区域可能对应不同的运动目标。比如说室外的监控摄像头所捕捉的图像中可能有晃动的树枝、跑过的小动物等。所以目标分类的目的就是正确地从检测到的运动区域中将人的运动区域提取出来。这类问题也可以看作特定目标与其他目标的两类分类问题。当然，这个步骤在某一些约束情况下可能是不必要的，比如说室内监控摄像头所捕捉到的图像，基本上知道场景中运动的物体仅仅是人在运动。目标分类最常见的分类方法是按特征的分类，可分为基于形状信息的目标分类和基于运动特征的目标分类和以上两种方法的混合。

3.1 基于形状信息的分类(Shape-Based Classification)

基于形状信息的分类是对所检测出来的运动目标，根据它们的形状轮廓信息来进行分类。该方法采用区域的宽高比、投影特性、轮廓变化、直方图、面积信息等特征做为物体分类的依据。Toth 和 Aach[31]采用了傅里叶描述子(Fourier Descriptors)作为特征向量来描述场景中不同的物体，利用向前回馈式神经网络(Feed-forward Neural Net)来分类人、车和背景的扰动。

3.2 基于运动特征的分类(Motion-based Classification)

人体的运动是非刚体运动，呈现一定的周期性，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。这类方法可以进行时频分析，利用周期性出现的自相似性来实现分类，并且可以与光流法相结合，通过计算运动区域的残余光流来分析运动实体的刚性和周期性[32]。Ran 与 Weiss[33]使用周期性分类人和车辆，并与人体运动的典型序列进行比较，适用于图像分辨率低、目标较小的红外和航空图片。还有一种分类方法是将以上两种结合起来。Bogomolov 等[34]就使用基于目标形状特征和运动特性相结合进行目标分类，他把同类目标的静态轮廓的的相似性和身体的倾斜角、脚之间的距离等运动的特征用支持向量机进行分类，提高了鲁棒性的精确度。

4 人的运动跟踪

人的运动的跟踪是人体运动分析过程的关键，是进一步识别和理解人体运动行为的基础。Yilmaz 等[35]对人的跟踪算法做了比较详尽的综述与归纳。在文献[2][36]中，将人的跟踪方法分为四类，分别是基于模型的跟踪、基于区域的跟踪、基于活动轮廓的跟踪和基于特征的跟踪。

4.1 基于模型的跟踪(Model-based Tracking)

基于模型的跟踪是最近使用较多的跟踪方法。基于模型的方法能够比较准确地描述人的运动，能够较为容易地解决遮挡问题。缺点是运动分析的精度取决于模型的精度，模型太过精细维数较高，运算也比较复杂，另外，在图像分辨率低的情况下，模型参数的估计比较困难。对人体进行跟踪时，通常有三种形式的模型，线图(Stick Figure)模型、2D 模型和 3D 模型。线图模型将人体骨骼化，用直线来代表人的各位部分。2D 模型将人体投影到二维的平面区域。3D 模型利用球，椭球，圆柱等三维模型描述人体结构。线图模型与 2D 模型早期在人的运动跟踪中运动较为广泛，3D 模型由于复杂度较高，算法复杂，一般不适合单摄像头的情况。Wu 等[37]使用二维的多关节模型对人体进行跟踪。Kehl 等[38]使用超椭球的方法对人体建模跟踪，他建立了有 10 个连接部分 24 个自由度的人体模型。Roberts 等[39]使用超二次曲面对人体进行建模跟踪。

4.2 基于区域的跟踪(Region-based Tracking)

基于区域的跟踪是对运动对象相应区域进行跟踪，它将人体划分为不同的小块区域，通过跟踪小区域来完成人的跟踪。基于区域的跟踪首先要得到包含目标的模板，模板可以略大于目标的矩形，也可以为不规则形状，然后在序列图像中，运用相关算法跟踪目标。基于区域的跟踪在当目标未被遮挡时，跟踪精度比较高；其缺点是比较费时，区域的合并和分割存在着不准确性，并且目标形变不能太大。早期相关研究可参考文献[40-41]。

Wren 等[42]利用了小区域的特征跟踪单人的室内行为，他把人的身体分为头、躯干、四肢等小区块，并且用高斯分布把人体各部分和背景建模，最后通过像素属于人体不同部位来进行跟踪，也就是通过跟踪人的不同部位来对整个人进行跟踪。McKenna 等[43]提出使用颜色和梯度信息的自适应背景减除法来处理阴影的影响，分成区域、人、群体三个层次来执行跟踪。每个区域都可以混合和分离，人体由一个或多个区域在一定的几何约束条件下构成，因此，通过跟踪区域和独立的颜色表面模型，即使在有遮挡的情况下，单人或人群也可以得到有效跟踪。

4.3 基于活动轮廓的跟踪(Active Contour based Tracking)

活动轮廓是图像范围内的曲线或表面，基于活动轮廓的跟踪是利用曲线或表面来表达运动目标，并且此轮廓可以自动更新，以便实现对目标的连续跟踪。此算法的优点在于计算量低，缺点是存在初始化困难，并且在阴影下效果欠佳。董春利等[44]对活动轮廓模型目标的跟踪算法有着较详尽的综述，并把活动轮廓模型按轮廓曲线的表达形式不同分为参数活动轮廓模型(snake 模型)和几何活动轮廓模型。Snake 模型[45]是一个基于参数的变形轮廓线，轮廓曲线的能量由内部能量和外部能量两部分组成，内部能量描述曲线平滑性，外部能量使目标的边界达到最小值。在参数活动轮廓模型中，Xu 等[46]提出的梯度矢量流模型(Gradient Vector Flow，GVF)受到研究人员的重视，它克服了基本 snake 模型捕获范围小的问题。几何轮廓模型可以认为是 snake 模型的扩展，但这种模型的轮廓曲线运动是基于轮廓曲线的几何度量参数。近年来的研究集中于水平集方法和基于粒子滤波的方法得到活动轮廓的运动目标，相关文献可参考[47-48]。张晓燕等[49]提出一种基于改进活动轮廓的视频对象自动分割及跟踪算法，在通过有关算法分割出轮廓后，通过使用新三步搜索(New Three-Step Search，NTSS)算法估计出运动向量进行运动补偿来得到视频对象在下一帧的初始曲线，再使用梯度向量流场作为外力的改进的活动轮廓算法进行分割，以便实现目标跟踪，该方法可以得到目标的精确轮廓，并且可以进行多目标的跟踪。

4.4 基于特征的跟踪(Feature-based Tracking)

基于特征的跟踪算法通过抽取特征和匹配特征来实现，该方法利用了特征位置的变化信息来跟踪目标，通常分为三步: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。基于特征的跟踪算法可以根据选择的特征细分为三类[50]：整体特征算法(Global Feature-based Algorithms)，局部特征算法(Local Feature-based Algorithms)和依靠图形的算法(Dependence-graph-based Algorithms)。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被看到，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。文献[51,52]对特征点的选择问题进行了讨论。Nickels 等[53]使用了基于 Sum-of-Squared-Differences (SSD)的特征跟踪。Tissainayagam 等[54]提出了基于贝叶斯多重假设跟踪(Multiple Hypothesis Tracking，HMT)的方法，先用 MHT 算法进行基于边缘图形的轮廓分割，再用 MHT 算法对选择的目标进行时间上的跟踪，描述对象主要是提取它的关键点(角点)，然后这些通过跟踪这些关键点进行物体的跟踪。

5 行为理解与描述

与运动检测、目标分类和人的跟踪研究相比，越来越多的研究人员投入到对人的行为理解与描述的研究当中。人的行为理解与描述是指对人的行为进行分析和识别，并用自然语言加以描述，这是一个模式识别问题。这种技术从视频序列中抽取相关的视觉信息，用合适的方法进行表达，然后将抽取的序列与事先的模板序列的参考行为进行匹配，然后进行行为分类，并解释这些视觉信息，实现人的行为的识别理解。人的行为理解与描述是基于视觉人运动分析的高级处理环节。在许多文献当中，表示行为的单词很多，如 movement，activity，action，behavior 等。在不同的应用背景下，行为的含义也不尽相同，就在同一篇文章下，“行为”这个词的含义也不同。Bobick[55]用动作识别(Movement recognition)、行为识别(activity recognition)和行动识别(Action recognition)来分类行为

识别。Moeslund 等[5]用基元行为(Action primitive)，行动(action)和行为(activity)三方面从低级到高级三方面来分类行为。如果根据 Moeslund[5]的分类方法，目前大部分的行为分析都处于 action 阶段。

5.1 模板匹配方法(Template Matching)

模板匹配方法首先从给定的序列图像中抽取相关特征，接着将图像序列转换为一组静态形式模板，再接着通过测试序列的模板与事先存储着的代表“正确”行为的模板匹配来获得识别结果。基于模板匹配的算法计算量少，但对行为时间间隔和噪声比较敏感。

Bobick 等人[55,56]最早提出时空模板方法，并用运用能量图像(Motion Energy Images，MEI)和运动历史图像(Motion History Images，MHI)来表示图像序列。MEI 是运动图像随着时间累积形成的二值化图像。MHI 是 MEI 的增强。MHI 像素强度是是 MEI 像素运动历史的一个函数，并基于 Hu 矩进行匹配模板，采用马氏距离(Mahalanobis Distance)来度量模板之间的相似性。Bradski 等[57]提出时间运动历史图像(tMHI)来进行运动分割，tMHI 能够确定正常的光流，并基于运动对象的轮廓和运动方向来分割运动，用 Hu 矩来对轮廓边界二分值并识别姿势。Weinland 等[58]提出用 Motion History Volumes(MHV)模板来描述基于视角自由的人的行为，他把 2D 的模板扩展到 3D，其实相当于 3D 的 MHI 模板。他在圆柱坐标系中进行 Fourier 变换，然后以 Fourier 特征描述行为。Weinland 等人的工作有助于解决视角问题。Wang 等[59]人不使用明确的特征跟踪和复杂的人体运动概率模型，而是直接用平均运动形状(Average Motion Energy)和平均运动能量(Mean Motion Shape)两个模板来描述行为，并利用监督模式分类的不同距离测量来进行行为分类。Yilmaz 等[60]提出使用 Spatio-Temporal Volume(STV)方法来识别行为，把人的 3D 轮廓投影成 2D 轮廓，这个投影轮廓是 STV 确定的，行为主体的轮廓随着时间变化的轮廓。识别描述子是通过分析 STV 的差分几何特性得到的，然后通过识别描述子来识别行为。Dimitrijevic 等人[61]用时空模板(Spatio-Temporal Templates)匹配来进行人体姿势识别。动态时间规整[62,63](Dynamic Time Warping，DTW)也常用来匹配运动序列。DTW 是使得输入序列的时间轴映射到训练模板上的时间轴上，使总累积失真最小。DTW 具有算法鲁棒，识别率高的优点，但有运算量较大和缺乏有效聚类训练方法的缺点。

5.2 状态空间方法

状态空间法又称为基于概率网络的方法，这种方法可以避免行为时间间隔建模，但模型训练复杂。因为人的运动具有马尔可夫性，当前的状态只受前一个状态的影响，这种方法将人的运动看成不可直接观测的马尔可夫过程，充分考虑到了人行为的动态过程，将人的运动序列看成状态间的一次遍历，概率地识别人的运动时空序列。此方法是目前使用较多的人体运动识别方法。它的优点是对时间和空间尺度上的运动微小变化的鲁棒性较好，可以避免行为时间间隔建模，运动持续时间得到很好的解决。缺点是计算比较复杂，需建立非线性模型，模型训练复杂，没有固定解决方法，需选择合适的状态数和特征矢量的维数[64]。目前在人的运动识别中使用的状态空间法主要有隐马尔可夫模型(Hidden Markov Models，HMMs)和动态贝叶斯网络(Dynamic Bayesian Networks，DBNs)。

HMMs 及其改进方法是目前人体行为识别中用得比较多的方法。Yamato 等[65]最早使用基于隐马尔可夫模型的方法对人体行为进行识别，并用 Baum-Welch 算法获得 HMM 训练参数。Ahmad 等[24]考虑多维的 HMM 来处理不同从不同视角得到的联合特征。Brand 等[66]提出耦合隐马尔可夫模型(Coupled Hidden Markov Model，CHMM)用于手语行为的识别。Ren 等[67]用 PCHMM(Primitive-based CHMM)用于双手的行为识别。Bui 等[68]提出了抽象隐马尔可夫(Abstract Hidden Markov Model，AHMM)的识别策略。Nguyen 等[69]提出用抽象隐马尔可夫记忆模型(Abstract Hidden Markov mEmory Model，AHMEM)识别复杂室内人的行为。基于 AHMM 的人体运动行为识别关键难点在于学习以及推理两方面[70]。钱堃等人[70]提出了一种采用级联形式的基于抽象隐马尔可夫模型的人运动行为识别方法，采用具有较高计算效率的 Rao-blackwellised 粒子滤波(Rao-Blackwellised Particle Filter，RBPF)近似推理方法识别人运动的时空序列。李宁等人[71]提出的基于“从左到右三状态半连接 HMM”(Begin-Middle-End Semi-Connected HMM，BME-SCHMM)的人体行为识别方法，为每个状态的输出概率引入了权重的概念，降低了运算复杂度。Nguyen 等人[72]使用层级隐马尔可夫模型(Hierarchical HMM，HHMM)对人体复杂行为进行识别。Peursum 等人[73]使用改进的 HHMM——因子状态层级隐马尔可夫模型(Factored-State Hierarchical HMM，FS-HHMM)对人体行为进行识别。Duong 等人[74]提出 Switching Hidden Semi-Markov Model (S-HSMM)，这是对隐半马尔可夫模型的(Hidden Semi-Markov Model，HSMM)的双层扩展，底层用 HSMM 表示简单动作和它们的持续时间，顶层表示由简单动作序列组成的高级行为。Caillette 等

人[75]用可变长马尔可夫模型(Variable Length Markov model，VLMM)对 3D 姿势的每个动作建模，进而识别高层次的行为。Natarajan 等人[76]使用 Hierarchical Variable Transition Hidden Markov Model (HVT-HMM)分三层对人体建模，最顶层的 HVT-HMM 对人的综合行动(Composite Actions)建模，并包含一个单独的马尔可夫链；中间层和底层分别对基元行为(Primitive Actions)、身体姿势(Body Pose)建模。另外，分层隐马尔可夫模型(Layer HMM)[77]，最大熵马尔可夫模型(Maximum Entropy Markov Models，MEMM)[78]也被用于人的行为识别。

动态贝叶斯网络(DBNs)早期应用于语音识别研究，近年来成为人体行为识别的重要工具。DBNs 是以时间展开的贝叶斯网络，将复杂的运动分解成一些简单的变量。Luo 等人[79]首先将 DBNs 引入行为识别当中，该文提出，相比于 HMM，DBNs 能提供更多的对象描述细节:每个时间片上，HMM 有一个隐含节点和一个观测点 ，DNBs 却有五个隐含节点和四个观测点。李妍婷等[80]用贝叶斯网络解决多视觉行为识别方法。Ren 等[81]提出基元 DBNs(Primitive-based DBNs)来解决特定主题的行为识别，所谓的基元是由描述上下文信息的特征组成，DBNs 能整合不同的弱信息特征并把它加强，提高了识别的效率和鲁棒性。Park 等人[82]使用分级贝叶斯网络(Hierarchical Bayesian Network)来识别双人行为，并用 DNNs 进行多部位姿势识别。应当指出，DNBs 的训练相比 HMM 要简单，但设计却比 HMM 复杂 。

另外，条件随机场(Conditional Random Fields，CRF) [83-85]、神经网络(Neural Networks)[86,87]、有限状态机(Finite State Machines)[88]、置信网络(Belief Networks)[89]等方法也用于人体行为识别。

5.3 行为语义描述

行为语义描述是人体运动分析的高级层次，近年来也取得一定的研究。Kojima 等[90]提出一种从视频序列中自动产生自然语言注释的方法。Kojima 等[91]按照他先前提出的方法[90]做了进一步的研究，把人的行为概念通过简单的语义进行分类，通过使行为概念和从视频序列中抽取的语义特征的相似性，恰当的句法成分被转化成自然语言。Ryoo 等人[92]提出用基于上下文文法(Context-Free Grammar，CFG)来描述复杂行为和交互。他把识别框架分为四层，部分身体抽取层(the Body-Part Extraction Layer)，姿势层(The Pose Layer)，姿态层(The Gesture Layer)和行为与交互层(The Action And Interaction Layer)。首先使用多像素水平技术来抽取部分身体特征，然后用 DBNs 进行人体姿势估计，接着用 HMMs 进行体姿态估计，最后用 CFG 进行行为识别。Ogale 等人[93]使用概率上下文无关文法(Probabilistic Context-Free Grammar，PCFG)自动构建行为语法。Guerra-Filho 等人[94]提出由 syntax，morphology 和 kinetology 三层结构的人体行为语言(Human Activity Language，HAL)，使用行为语言可以描述并理解行为。

6 存在问题和发展趋势

由于人的着装、运动方向和非刚性、摄像头视角和环境(阴影、光照等)的多变性，人体运动分析是一个具有挑战性的问题。在过去几十年的研究里，人体运动分析在各个层面上都取得了很大的进展。研究人员对人体运动的物理特征的了解已经很深入，人体的建模也从二维的大部分转成三维的。确定性的线性跟踪技术已经被采样跟踪技术取代。机器学习在人体运动分析中扮演越来越重要的角色，并且会一直持续下去。但就整体而言，人体运动分析还存在很多难点与问题，需要做进一步的研究。

6.1 存在的问题

(1) 遮挡和视角: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。人的行为极其复杂多变，遮挡时，只有部分人可见，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。解决遮挡问题一般采用三维的行为描述方法或者采用多视角的技术，也可以用统计的方法从所获得的图像信息中进行人的位置与姿势的预测等。对于多视角系统，此方法对人体的跟踪识别的优势是很明显的，它可以从不同的角度与方向解决遮挡问题，但是我们需要解决多摄像头之间的标定与选择问题、信息融合问题，另外，存储量与所需要的运算量与运算时间也会加大，会影响实时检测。

(2) 行为识别：相对于检测与跟踪技术的发展，行为理解的研究进展比较缓慢，虽然有一些研究成果，但目前对行为分析的研究往往只进行到 3 层分类法[5]的第 2 个阶段——Action Recognition。也就是简单的日常标准动作，如走、站、坐跑、跳、蹲等，还有就是在具体场景中处理特定的行为，对这些行为的分析跟实际应用中的行为分析有比较大的差距，行为分析在真实场景中的应用，仍然存在很多的问题。行为分析不仅要识别人的行为，还要结合所处的环境理解人的行为。行为识别的技术难点还在于特征选择与向量维数问题，如果选择的特征过多，那么特征向量的维数就会过大，会增加计算的复杂度；反之，特征过少，

又可能不能充分表达人的行为，不能进行有效的行为识别。因此，选择合适的特征与向量维数也是一个需要在具体情况下解决的问题。

另外，以鲁棒性、准确度和速度三个基本要求的人运动分析的性能评估[3]，也是人运动分析需要解决的。

6.2 未来可能的方向

基于视觉的人运动分析本质上是人工智能的问题，涉及机器视觉、图像处理、模式识别、人工智能等多个学科。未来的人体运动视觉分析研究必须解决一些开放的问题，以满足潜在的应用，未来的发展应推广到对复杂场景下人与事的理解以及复杂行为的高层次理解，在此基础上，应将现有的简单行为识别推广到复杂场景下的自然语言描述，并且能够根据外部的环境，进行自主发育学习与理解，机器学习将会持续扮演重要的角色。另外，目前相当多的识别研究侧重于单人的行为，以后的研究方向可向交互识别如双人行为、群体交互行为、人机交互甚至人与物体的交互发展。

7 结束语

人的运动分析已经成为计算机视觉领域一个重要的研究方向。它在智能监控、虚拟现实，人机交互等方面的广泛应用前景引起了科研人员的兴趣。本文从运动目标检测、运动目标分类、人体运动跟踪、人体行为识别与描述四个方面总结了近年来人行为理解研究现状和进展，并对存在的问题和发展趋势做了简要阐述，希望对有关研究人员有所帮助。

参考文献

1 Moeslund TB, Granum E. A survey of computer vision-based human motion capture. CVIU, 2001,81(3):231－268.

2 王亮,胡卫明,谭铁牛.人运动的视觉分析综述.计算机学报, 2002,25(3):225－237.

3 杜友田,陈峰,徐文立,李永彬.基于视觉的人的运动识别综述.电子学报,2007,35(1):84－90.

4 黎洪松,李达.人体运动分析研究的若干新进展.模式识别与人工智能,2009,22(1):70－78.

5 Moeslund TB, Hilton A, Krüger V. A survey of advances in vision-based human motion capture and analysis. CVIU, 2006,104:90－126.

6 Poppe R. Vision-based human motion analysis: an overview. CVIU, 2007,108(1-2):4－18.

7 Poppe R.A survey on Vision-based human action recognition. IVC, 2010.

8 Gavrila DM. The visual analysis of human movement: a survey. CVIU, 1999,73(1):82－98.

9 赵文哲,秦世引.视频运动目标检测方法的对比分析.科技导报,2009,27(10):64－70.

10 Elhabian SY, El-Sayed KM, Ahmed SH. Moving object detection in spatial domain using background removal techniques-state-of-art. Recent Patents on Computer Science, 2008(1):32－34.

11 Hironobu AL, Lipton AJ, Fujiyoshi H, Patil R. Moving target classification and tracking from real-time video. IEEE, 1998:8－14.

12 Shoushtarian B, Bez HE. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. Pattern Recognition Lett, 2005,(26):5－26.

13 Herrero-Jaraba E, Orrite-Urenuela C, Senar J. Detected motion classification with a double-background and a neighborhood-based difference. Pattern Recognition Letters, 2003,(24):2079－2092.

14 Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. CVPR, 1999,2:23－25.

15 Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. PAMI, 2000,22(8):747－757.

16 KaewTraKulPong P, Bowden R. An Improved adaptive background mixture model for real-time tracking with shadow detection. Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems.

17 Li L, Huang W, Gu IY, Tian Q. Foreground object detection in changing Background based on color co-occurrence statistics. IEEE Workshop on Applications of Computer Vision. 2002. 269－274.

18 Kim K, Chalidabhongse T, Harwood D, Davis L. Real-time foreground-background segmentation using codebook model. Real Time Imaging, 2005,11(3):172－185.

19 Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction. European Conference on Computer Vision. 2000. 751－767.

20 Wu MJ, Peng XR. Spatio-temporal context for codebook-based dynamic background subtraction. 2009.

21 Collins, Lipton, Kanade. A system for video surveillance and monitoring: VSAM Final Report. Technical Report, 2000.

22 Spagnolo P, Orazio TD, Leo M, Distante A. Moving object segmentation by background subtraction and temporal analysis. IVC, 2006,24(5):411－423.

23 Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. IJCV, 1994,12(1):43－77.

24 Ahmad M, Lee SW. Human action recognition using shape and CLG-motion flow from multi-view image sequences. Pattern Recognition, 2008,(41):2237－2252.

25 Gehrig D, Kuehne H, Woerner A, Schultz T. HMM-based Human Motion Recognition with Optical Flow Data. 9th

IEEE-RAS International Conference on Humanoid Robots. 425－430.

26 Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. 1981.

27 熊静旖,罗予频,唐光荣.一种应用于图像配准中大尺度位移估计的改进光流法.自动化学报, 2008,34(7):760－764.

28 Ahad MAR, Ogata T, Tan JK, et al. Motion recognition approach to solve overwriting in complex actions. Proc. of the International Conference on Automatic Face and Gesture Recognition. 2008. 1－6.

29 Efros AA, Berg AC, Mori G, Malik J. Recognizing action at a distance. ICCV, 2003:726－733.

30 Li X. HMM based action recognition using oriented histograms of optical flow field. Electronics Letters, 2007, 43(10):560－561.

31 Toth D, Aach T. Detection and recognition of moving objects using statistical motion detection and Fourier descriptors. ICIAP, 2003. 430－435.

32 Lipton AJ. Local Application of optic flow to analyse Rigid versus non-rigid motion. ICCV, 1999.

33 Ran Y, Weiss T. An efficient and robust human classification algorithm using finite frequencies probing. DVPRW, 2004. 132－138.

34 Bogomolov Y, Dror G. Classification of moving targets based on motion and appearance Machine Vision Conference. 2003. 429－438.

35 Yilmaz A, Javed O, Shah M. Object tracking: a survey. ACM Computing Surveys, 2006,38(4):13－57.

36 Wang L, Hu WM, Tan TN. Recent developments in human motion analysis. Pattern Recognition, 2003,36(3):585－601.

37 Wu Y, Hua G, Yu T. Tracking articulated body by dynamic Markov network. ICCV. 2003,2:1094－1101.

38 Kehl R, Gool LV, Markerless tracking of complex human motions from multiple views. CVIU. 2006. 190－209.

39 Roberts TJ, McKenna SJ, Ricketts IW, Online appearance learning for 3D articulated human tracking. Proc. of ICPR. 2002. 425－428.

40 Meyer F. Bouthemy P. Region-based tracking using affine motion models in long image sequences. CVGIP: Image Understanding, 1994,60(2):119－140.

41 Bascle B, Deriche R. Region tracking through image sequences. Proc. of IEEE ICCV. 1995. 302－307.

42 Wren CR, Azarbayejani A, Darrell T, Pentland A.P, Pfinder:real-time tracking of the human body. IEEE Trans. Pattern Analysis Machine Intelligence, 1997,(19):780－785.

43 McKenna S, Jabri S, Duric Z, Rosenfeld A, Wechsler H. Tracking groups of people. CVIU. 2000. 42－56.

44 董春利,董育宁,王莉.活动轮廓模型目标跟踪算法综述.计算机工程与应用,2008,44(34):208－212.

45 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. 1988.

46 Xu C, Prince JL. Snakes, shapes and gradient vector flow. IEEE Trans. on Image Processing, 1998,(7):359－369.

47 Rathi Y, Vaswani N, Tannenbaum A. Tracking deforming objects using particle filtering for geometric active contours. PAMI, 2007,29(8):1470－1475.

48 Rathi Y. Viswani N, Tannenbaum A. A generic framework for tracking using particle filter with dynamic shape prior. IEEE Trans. on Image Processing, 2007.

49 张晓燕,赵荣椿,马志强. 基于改进活动轮廓的视频对象自动分割及跟踪算法.中国图像图形学报,2007,12(3):438－443.

50 Hu WM, Tan TN, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. IEEE Trans. on Systems, Man and Cybernetics, 2004,(34):334－352.

51 Kaneko T, Hori O. Feature selection for reliable tracking using template matching.CVPR,2003(1):796－802.

52 MItra P, Murthy C, Pal S. Unsupervised feature selection using feature similarity. PAMI, 2002,24(3):301－312.

53 Nickels K, Hutchinson S. Estimating uncertainty in SSD -based feature tracking. IVC, 2002,20(1):47－58.

54 Tissainayagam P, Suter D. Object tracking in image sequences using point feature. Pattern Recognition, 2005, 38(1):105－113.

55 Bobick A. Movement, activity, and action: the role of knowledge in the perception of motion. Philosophical Trans. of the Royal Society of London, 1997,(352):1257－1265.

56 Bobick A, Davis J. The recognition of human movement using temporal templates. PAMI. 2001,23(3):257－267.

57 Bradski G.R, Davis J.W, Motion segmentation and pose recognition with motion history gradients. Machine Vision and Applications. 2002,13(3):174－184.

58 Weinland D, Ronfard R, Boyer E. Free viewpoint action recognition using motion history volumes. CVIU, 2006, 104(2-3):249－257.

59 Wang L, Suter D, Informative shape representations for human action recognition. ICPR. 2006. 1266－1269.

60 Yilmaz A, Shah M, Actions sketch: a novel action representation. CVPR. 2005.

61 Dimitrijevic M, Lepetit V, Fua P. Human body pose recognition using spatio-temporal templates. Workshop on Modeling People and Human Interaction. 2005.

62 Myers C, Rabiner L, Rosenberg A. Performance tradeoffs in dynamic time warping algorithms for isolated word

recognition. IEEE Trans. Acoustic, Speech, and Signal Procesing, 1980,28(6):623－635.

63 Veeraraghavan A, Roy-Chowdhury AK, Chellappa R. Matching shape sequences in video with applications in human movement analysis. PAMI, 2005,27(12):1896－1909.

64 刘相滨,向坚持,王胜春.人行为识别与理解研究探讨.计算机与现代化,2004(12):1－5.

65 Yamato J, Ohya J, Ishii K, Recognizing human action in time sequential images using hidden Markov model. CVPR. 1992. 379－385.

66 Brand M, Oliver N, Pentland A. Coupled hidden Markov models for complex action recognition. CVPR. 1997. 994－999.

67 Ren HB. Xu GY. Human action recognition with primitive -based coupled-HMM. ICPR, 2002,(5):494－498.

68 Hung H. Bui, Venkatesh S, West G. Policy recognition in the abstract hidden Markov model. Journal of Artificial Intelligence Research, 2002,17:451－499.

69 Nguyen T, Hung H Bui, Venkatesh S, West G. Recognising and monitoring high-level behaviours in complex spatial environments.CVPR, 2003.

70 钱堃,马旭东,戴先中. 基于抽象隐马尔可夫模型的运动行为识别方法.模式识别与人工智能,2009,(3):433－439.

71 李宁,须德,傅晓英,袁玲.结合人体运动特征的行为识别.北京交通大学学报,2009,33(2):6－16.

72 Nguyen NT, Phung DQ, Venkatesh S. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model. CVPR. 2005. 955－960.

73 Peursum P, Venkatesh S, West G, Tracking-as-recognition for articulated full-body human motion analysis. CVPR. 2007. 1－8.

74 Duong TV, Hung H. Bui, Dinh Q, et al. Activity recognition and abnormality detection with the switching hidden semi-Markov model. CVPR. 2005. 838－845.

75 Caillette F, Galata A, Howard T. Real-time 3-D human body tracking using learnt models of behavior. CVIU, 2008,109 (2):112－125.

76 Natarajan P, Nevatia R. Online, real-time tracking and recognition of human actions. WMVC. 2008. 1－8.

77 Zhang D, Gatica-Perez D, Bengio S, et al. Modeling Individual and Group Actions in Meetings: a Two-Layer HMM Framework. IEEE Trans. on Multimedia, 2006, 8(3):509－520.

78 Sminchisescu C. Kanaujia A, Li ZG, et al. Conditional models for contextual human motion recognition. ICCV. 2005.

79 Luo Y, Wu TW, Hwang JN, Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks. CVIU, 2003,(23):196－216.

80 李妍婷,罗予频,唐光荣.单目视频中的多视角行为识别方法.计算机应用,2006,26(7):1592－1594.

81 Ren HB, Xu GU, Kee S. Subject-independent natural action recognition. International Conference on Automatic Face and Gesture Recognition. 2004.

82 Park S, Aggarwal JK. A hierarchical Bayesian network for event recognition of human actions and interactions. ACM, Journal of Multimedia Systems, Special Issue on Video Surveillance, 2004,10(2):164－179.

83 Sminchisescu C, Kanaujia A, Metaxas D.N, Conditional models for contextual human motion recognition. CVIU, 2006,104(2-3):210－220.

84 Wang Y, Mori G. Max-margin hidden conditional random fields for human action recognition. CVPR. 2009. 1－8.

85 Zhang JG, Gong SG. Action categorization with modified hidden conditional random field. Pattern Recognition, 2010,43(1):197－203.

86 Yu H, Sun GM, Song WX, Li X. Human motion recognition based on neural networks. International Conference on Communications, Circuits and Systems. 2005.

87 Buccolieri F, Distante C, Leone A. Human posture recognition using active contours and radial basis function neural network. Proc. of Conference on Advanced video and Signal Based Surveillance, 2005.

88 Hong PY, Huang TS, Turk M. Gesture modeling and recognition using finite state machines. Proc. of IEEE Conference on Face and Gesture Recognition. 2000.

89 Intille SS, Bobick AF. Representation and Visual Recognition of Complex, Multi-agent Actions using Belief Networks. MIT Media Lab, Technical Report: 454, 1998.

90 Kojima A, Tamura T. Natural language description of human activates from video images based on concept hierarchy of actions. IJCV, 2002,(50):171－184.

91 Kojima A, Aoki S, Miyamoto T, et al. Generating Natural Language Annotation from Video Sequences Taken by Handy Camera. ICICIC, 2007.

92 Ryoo MS, Aggarwal JK. Recognition of composite human activities through context-free grammar based representation. CVPR, 2006. 1709－1718.

93 Ogale AS, Karapurkar A, Aloimonos Y. View-invariant modeling and recognition of human actions using grammars. ICCV, 2005.

94 Guerra-Filho G, Aloimonos Y. A language for human action. 2007.