第 18 卷 第 11 期
2006 年 11 月
计算机辅助设计与图形学学报
JOURNAL OF COMPUTER-AIDED DESIGN & COMPUTER GRAPHICS
Vol.18, No.11
Nov., 2006

基于视频的人体运动捕捉综述

李豪杰[1,2)] 林守勋[1)] 张勇东[1)]
[1)](中国科学院计算技术研究所虚拟现实技术实验室 北京 100080)
[2)](中国科学院研究生院 北京 100049)
(hjli@ict.ac.cn)

摘 要 系统地总结和介绍了基于视频的人体运动捕捉研究的技术方法和最新进展. 将基于视频的人体运动捕捉方法分为自顶向下的方法和自底向上的方法 2 大类,阐述了每一类中各种方法的原理、技术特点和研究现状,最后分析了该领域的难点问题和发展趋势.

关键词 计算机视觉;人体运动捕捉;人体运动跟踪;姿态估计
中图法分类号 TP391

A Survey of Video Based Human Motion Capture

Li Haojie[1,2)] Lin Shouxun[1)] Zhang Yongdong[1)]
[1)](*Virtual Reality Laboratory*, *Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100080)
[2)](*Graduate University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract A comprehensive survey of the research and recent developments of video based human motion capture (HMC) is presented. The various approaches of HMC are classified into two categories: Top-down category and Bottom-up category. The approaches in each category are described and their principles and technical characteristics are summarized. Finally, the open issues and technical trends of HMC are discussed.

Key words computer vision; human motion capture; human body tracking; pose estimation

基于视频的人的运动分析和理解是近年来计算机视觉领域广受关注的一个研究热点,其涉及的关键技术之一就是人体运动捕捉(human motion capture, HMC,本文专指基于视频的人体运动捕捉). HMC 是指从一个或多个视角已同步的图像序列中恢复出人体姿态参数(关节点位置或关节角度)的过程,这里的姿态指头、躯干和四肢的运动,不包括表情和手语等小尺度的动作[1].

HMC 和分析的研究始于 20 世纪 80 年代初[2],因其重要的学术价值和广阔的应用前景,近几年已成为计算机视觉的一个活跃的研究领域. 首先,HMC 技术涉及计算机视觉的许多基本问题,如运动检测、目标识别、刚体和非刚体跟踪、场景恢复与表示、多摄像机数据融合等,其研究成果可为计算机视觉其他领域所借鉴;HMC 还融合了图像处理、计算机图形学、人体运动学、最优化及机器学习等多学科的理论,极具挑战性. 其次,HMC 技术有着极为广泛的潜在应用,如智能视觉监控中人的跟踪和行为分析,人机交互中的身体姿势、手势识别,体育和医学康复中的人体运动分析,虚拟现实中的交互式游戏、远程控制,以及动画和影视制作等[1,3]. 因此,国内外一些重要学术机构都开展了与 HMC 技术相关的研究[4-10].

文献[1]从系统功能角度将 HMC 系统分成初始化、跟踪、姿态估计和识别 4 个阶段,并将姿态估计分为基于模型的方法、间接使用模型的方法和非模型的方法 3 类;文献[3,11]主要侧重于运动分析,分别按照人运动的分析方法和一般处理框架对相关工作进行了评述;这些文献概括的都是 2000 年以前的研究工作. 由于应用驱动和新理论的引入,HMC 技术近几年取得许多新进展,发表了大量相关学术论文,本文重点概括了对 HMC 研究近 5 年的相关

收稿日期:2005-10-17;修回日期:2006-07-28
基金项目:北京市自然科学基金重点项目(4051004);北京市科技计划项目(D0106008040291,Z0004024040231)

工作,并对今后的发展趋势进行探讨,以期对本领域研究技术人员有所裨益.

1 自顶向下的捕捉方法

HMC 是从 2D 投影图像或图像序列中恢复场景中的 3D 人体运动信息,这是一件非常困难的事情,原因如下:1)投影过程中会失掉深度信息,使恢复成为一个病态过程;2)人体是一个高自由度的关节连接结构,其运动复杂且存在自遮挡现象;3)图像噪声、摄像机运动等因素使得无法提取可靠的底层特征.

自顶向下的捕捉方法通常称为基于模型的方法,该方法将人体模型投影并与图像匹配,以获得人体姿态参数. 借助人体模型,可利用人体几何结构、运动模型等先验知识,有效地克服上述困难. 本文将目前文献中基于模型的人体运动跟踪归入这类方法.

自顶向下的捕捉方法采用"分析—合成"的方式,跟踪是在"预测—匹配—修正"的循环中实现的,其一般处理框架如图 1 所示. 跟踪时,根据上一帧跟踪结果和运动模型预测当前状态,将预测的人体几何模型投影到图像平面,计算模型投影特征与图像特征的匹配误差,结合运动学约束,根据误差修正预测的状态,从而得到当前帧的人体姿态.

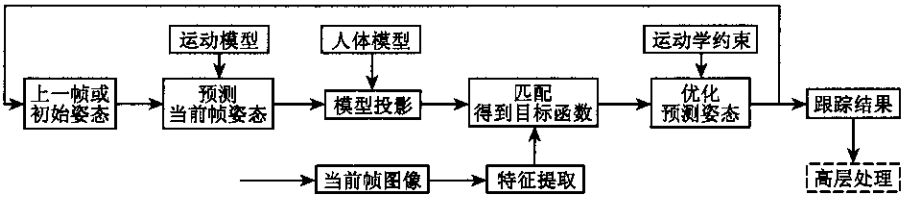

图 1 自顶向下的基于模型的人体运动跟踪处理框架

采用自顶向下方法的运动跟踪系统涉及人体几何模型、人体运动学约束和运动模型、特征提取及优化算法等研究内容,除特征提取部分将在第 3 节介绍外,其余内容均在本节详细阐述.

1.1 人体模型及其表示

研究人员提出了各种人体结构和几何形状的表示方法,主要人体模型有棍图模型、2D 模型和 3D 体模型等,如图 2 所示.

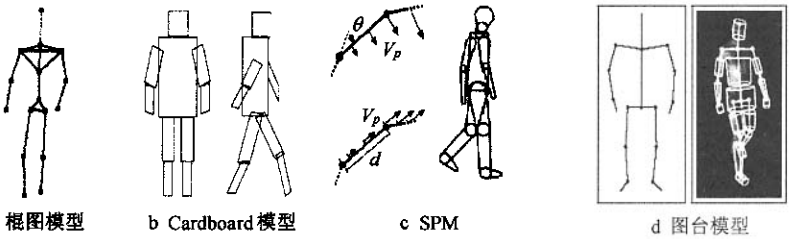

图 2 人体模型

1) 棍图模型是最简单的人体结构表示方法,它由如图 2a 所示点和线段组成,分别表示关节点和骨骼. 棍图模型可用来指导对图像特征的拟合,以获得人体姿态,也可作为运动捕捉的一种结果输出形式.

2) 2D 模型用于单目视频的跟踪,对人体朝向或拍摄角度进行了限定,如要求被跟踪者平行于摄像机平面或面向摄像机运动等. 最常用的是如图 2b 所示的 Cardboard 模型[12] 和如图 2c 所示的 SPM (scaled prismatic model)[13].

3) 3D 模型包含深度信息,能有效地解决自遮挡和自碰撞问题,一般用在多摄像机跟踪中. 该类模型包括骨架模型和形状模型 2 个部分. 骨架模型用棍图表示,定义各骨骼长度及关节间连接关系;形状模型表示附着在骨架上的人体外围组织,一般用如图 2d 所示基本体素(如圆台[5]、超二次曲面[14]等)的组合表示.

上述模型除 Cardboard 模型外,都属分层的树型结构. 一般将骨盆或躯干定义为树的根,整个人体运动可视为由根节点的全局运动和子节点绕其父节点的旋转运动组成,因此,人体姿态可用根节点的位置、方向参数和关节角参数组成的状态向量表示. 基于模型的人体运动跟踪就是在由状态向量张成的状态空间中搜索合适的点,使该点对应姿态的投影与图像特征最为匹配.

1.2 人体运动的先验知识

人体运动的先验知识是指人体运动学约束和运

动模型,在裁剪状态空间和运动预测时起重要作用.

1) 人体运动学约束是指人运动时应遵从的限制,包括关节角的运动范围约束和身体各部分不能相互穿透的约束. 这些约束可作为硬约束将状态空间分成合法部分和非法部分,以减少搜索范围[15];也可作为软约束,即惩罚因子,如刘国翌等[9]在其跟踪框架中建立了一个身体自相交代价函数,以避免产生身体部分相互穿透的结果.

2) 运动模型描述人体运动规律,一般针对特定的运动模式(如行走、跑步、舞蹈等),用统计学习方式获得,它能大大地缩小解的搜索空间,且有助于解决遮挡问题,因此很多学者对此进行了研究.

根据运动复杂程度,运动模型分成单层模型和分层模型 2 种. 单层运动模型适合简单运动模式,如行走等,主要建模方法有主元分析法(principle component analysis, PCA)[16]、高斯法[7]等. Sidenbladh 等[16]用商业运动捕捉系统获得行走的 3D 运动数据,将其按运动周期手工分割对齐后用多元 PCA 降维,得到一个主元表示的、紧凑的行走运动模型来指导跟踪. 由于人的运动本质上是复杂的非线性过程,更多的学者使用分层的运动模型. 在分层模型中,底层表示为线性过程,高层表示为这些线性过程间的转移概率,常用工具有隐马尔科夫模型[17]、可交换线形动态模型[18]、有限状态机[19]等. 为对高维非线性状态空间进行建模,一般要进行降维聚类分析,主要有分层主成分分析[20]、最小描述长度法[19]等.

基于单目视频的 3D 运动跟踪对某些应用(如视觉监控及对已存档视频的运动分析)十分有用,而从单目视频恢复 3D 姿态,运动模型具有重要意义.

1.3　优化算法

在自顶向下的跟踪中,优化是通过对状态空间的搜索来完成的,即在姿态参数空间中寻找合适的点,使该点所对应模型投影的表面形状与图像中的人体形状最为相似. 然而,对由几十个自由度构成的高维状态空间搜索非常困难,除使用各种先验知识对状态空间进行剪枝外,还必须采取适当的优化算法进行搜索,这也是运动跟踪的核心内容. 主要优化算法如下:

1) 微分法[6,9]. 一种常用的优化方法,首先建立一个可微的目标函数,一般是模型投影特征与图像特征的差值函数,通过雅可比矩阵建立模型参数微分和图像坐标微分的联系;然后沿负梯度方向迭代地对模型参数修正,直到满足优化条件. 该方法收敛速度快,但无法保证收敛到全局最优点,不适合跟踪大幅度运动.

2) 动力学法[21-22]. 在预测模型投影与图像数据(如人体轮廓)间生成一种物理力,类似弹簧弹力. 力的方向由模型指向图像,当二者偏离时,该力使二者趋于重合,根据该力不断地修正模型参数,直至二者匹配为止. 该方法可以跟踪大幅度运动,但无法保证其收敛性.

3) 局部搜索法[14]. 一种传统人工智能的方法,通过生成-测试的方式,在预测位置附近进行启发式搜索,寻找与模型最匹配的位置. 为有效地搜索高维空间,一般要对状态空间进行分解,然后采用分层搜索方式. 该方法速度较慢,且易陷入局部最优点.

4) Kalman 滤波[23]. 一种常用的跟踪算法,首先用前一时刻状态进行预测;然后根据预测状态和误差协方差确定搜索范围,在该范围内查找特征进行匹配得到测量值,用测量值差和增益矩阵对预测状态修正,得到状态参数的后验估计. Kalman 滤波由于用概率来描述跟踪过程中的不确定信息(如噪声、杂乱背景等),同时保留了尽可能多的历史信息,因此跟踪更加鲁棒. 但 Kalman 滤波是线性滤波器,要求系统状态和噪声都是单峰分布,这在实际中很难满足.

5) 粒子滤波. 在人体运动跟踪中,因噪声、杂乱背景及非线性动态系统的影响,状态分布呈多峰形式. 此时 Kalman 滤波不再适用,而粒子滤波提供了一种方便、有效的计算多高斯后验概率的方法,在目标跟踪中获得广泛应用.

粒子滤波是一种非参数的方法,用离散的带权重的样本(即粒子,一个粒子对应状态的一个假设)集合表示后验概率分布,利用概率传播采样点来模拟状态分布随时间的演化. 将粒子滤波算法用于人体运动跟踪是近年来一个重要进展[5-8,16,24-26]. 但在跟踪高自由度的人体运动时,该算法所需样本数随自由度增加呈指数增长,致使其效率极低. 因此,目前基于粒子滤波的人体运动跟踪研究都集中在对算法效率的改进上,主要方法有:a. 改进重采样方式[5],提高重采样样本的有效性;或用局部优化获得的协方差指导采样[6]以提高采样效率. b. 与局部优化相结合. 在前一时刻只保留少量局部最优假设,当前时刻仅从这些假设出发用梯度下降法优化[25],从而降低所需样本数目. c. 分解状态空间,降低搜索维数,如将整个人体分成头/躯干和四肢,先对头/躯干采样,再对四肢部分采样[8]. d. 使用辅助信息,如用底层特征检测获取人体姿态的一部分参数,以降低搜索维数[26];或将底层信息融入重要性函数[51],提高采样效率.

粒子滤波具有全局搜索的性质,能避免局部

极值点,且允许底层特征提取有噪声;缺点是计算量太大,且跟踪结果不如前 4 种方法精确.

1.4 存在的问题

自顶向下的捕捉方法,通过人体模型可方便地使用先验知识,解决遮挡问题,且跟踪结果较精确,适合运动分析方面的应用. 但缺点也很明显:1)目前文献中大多采取手工标定方式进行人体模型的初始化,很难自动初始化;2)跟踪过程中会因误差累积、噪声等影响致使跟踪失败,难以自动恢复;3)优化速度慢,难以达到实时要求.

2 自底向上的捕捉方法

自底向上的捕捉方法是一类非模型匹配的姿态估计方法,通过对底层特征的跟踪、分析或推理获取人体姿态信息. 主要有特征跟踪法、形状分析/标记法、识别法和推理法.

2.1 特征跟踪法

特征跟踪法的捕捉原理是对输入图像序列,首帧标注或检测出人体关键点位置,后续帧中通过跟踪建立起帧间特征的对应关系,即可获得各时刻人体的姿态参数. 该方法一般要求有明显的特征,如罗忠祥等[10]要求被跟踪者穿上特制的紧身衣,其各关节点处为不同的色块,通过对色块的跟踪得到 2D 关节点位置序列;Pfinder 系统[27]通过特定的姿势分析出头、手、脚的位置,对每个区域用空间位置和颜色建模,然后用颜色对这些区域进行跟踪,从而获得各帧的姿态.

与基于模型的跟踪不同,在基于特征的跟踪中,各特征点间独立进行匹配,不考虑全局结构约束,因此处理简单、速度快,但人体不规则运动常使特征点出现重叠、自遮挡等现象,致使跟踪很不可靠.

2.2 形状分析/标记法

该方法在文献[1]中被称为间接使用模型的方法,意指人体模型不直接用来匹配图像数据,而是作为启发式知识指导对图像特征(人体剪影或轮廓)进行分析和标记,以获得姿态信息. 主要步骤如下:

Step1. 获取人体剪影或轮廓(获取方法见第 3 节).

Step2. 对轮廓进行分析,找出轮廓上的特征点[28],如凸凹点(对应肢体末端,如头、手、脚、肘、膝等);或将轮廓分解成不同的部分[29](对应肢体及躯干).

Step3. 对分析出的特征点或部分进行标记或识别. 通常有 2 种方法:a. 基于规则的方法,针对特定姿势用人体模型的结构约束进行标记,如 Haritaoglu 等[28]用轮廓的水平和垂直投影人体分成 4 种基本姿势,然后根据特定姿势下人体轮廓上身体特征点之间的相对距离和拓扑顺序,将找出的凸凹点映射为身体不同部位;b. 基于概率的方法,如 Zhao[29]使用局部形状和全局几何关系作为相似度量,用贝叶斯方法对分割的各部分进行识别.

形状分析/标记法无需初始化,对智能监控和人机交互等应用基本可达到实时要求. 但其仅限对简单动作进行分析,所得结果亦不精确,且需提取准确的人体区域,对环境、光照要求较严格.

2.3 识别法

该类方法需要训练数据,对事先标注的运动库建立索引或从中学习特征与姿态的对应关系,通过最近邻检索或函数映射获得人体姿态参数,从而将姿态估计问题转换为模式识别问题. 识别法可分为模板匹配法和映射法 2 类.

1) 模板匹配法

先建立模板库,模板包括特征及相应的姿态. 对测试图像提取特征后,用某种度量与模板库中的样例比较,用最相似样例的姿态作为估计结果. Mori 等[30]以边缘为特征,用形状关联匹配从模板库中找到最相似样例,从该样例获得输入图像的 2D 关节点位置,再用 Taylor 法恢复出 3D 人体姿态.

对复杂人体运动,不同姿态的图像投影特征可能非常接近,即特征与姿态的对应是 1∶N 关系. 为此,Howe[31]从模板库中检索出相近的多个样例,对候选样例序列用时域相似性约束选出每帧的最佳匹配. 除静态图像外,还可对运动序列进行匹配,如 Efros 等[32]用光流作为描述子,从运动库中检索出最相近的动作姿态序列作为输出.

2) 映射法

通过监督学习获得映射函数,建立底层特征与姿态间的映射关系. 该方法直接将底层特征映射为相应的姿态参数,无需保留样本库. 常用映射工具有:神经网络[33]、相关向量机[34]、贝叶斯估计[35]等. Agarwal 等[34]用人体轮廓的形状关联作特征,用 Relevance Vector Machine 作回归器,对输入特征直接输出相应的 3D 姿态参数;Elgammal 等[33]用 Local Linear Embedding 技术对高维轮廓特征空间降维后,用广义径向基函数来学习映射关系.

识别法无需人体模型和对高维空间的搜索,速度快、鲁棒性好,可获得 3D 姿态. 但也存在如下缺陷:a. 只能识别已知动作,且所得结果可能不精确;b. 对于复杂运动,其训练样本集的覆盖性和代表性很难保证.

2.4 推理法

在自顶向下的捕捉方法中使用的是全耦合的树型结构,树结构的缺点是难以融入底层信息,如人脸或身体其他部分的检测结果等,因此不易做到自动

初始化和失败恢复;且树结构的状态空间很难分解,导致计算量大. 为克服这一困难,近年来很多学者用松耦合的图模型表示人体结构,采取自底向上方式,先独立检测或预测身体各部分位置,再通过概率推理获得满足人体几何结构约束的姿态配置. 该类方法也可看作是对第 1.3 节中粒子滤波算法的改进.

在图结构中,每个身体部分用变量 X_i 表示,相应的图像观测特征为 Z_i,相邻部分间的连接是软约束,用势函数 $\psi_{ij}(X_i, X_j)$ 表示,前后帧同一部分间的时域约束也用势函数表示. 这样,身体部分 X_i 的后验概率 $p(X_i \mid Z)$ 的计算便包括:X_i 本身的先验分布 $\psi_i(X_i)$、相似度 $p(Z_i \mid X_i)$ 和其近邻的势函数对该部分产生的约束 3 部分,因此可对 X_i 进行局部滤波. 尽管如此,直接计算 $p(X_i \mid Z)$ 仍很复杂,一般采用近似推理算法,如中值场蒙特卡罗算法[36]、信念传播算法[37]、动态规划[38]等.

这种基于身体部分检测结果推理的方法包括 2 个步骤:

Step1. 身体部分的检测或预测,如人脸检测、肤色区域检测、基于模板的肢体检测[37]及帧间预测[36]等. 这里不要求各部分的检测结果十分可靠,往往是含噪声的多个候选位置.

Step2. 人体结构配置的推理,即通过贝叶斯推理求解满足约束的最大后验分布.

由于对各部分进行局部滤波,推理法计算复杂度远小于对整个树模型的滤波方法,其局限性在于很难使用高层知识来处理自遮挡现象.

3　特征提取

本节介绍 HMC 中常用底层特征及其提取方法. 特征提取是从图像中提取出边界、区域、轮廓、光流等底层特征,将前景与背景分离的过程. 特征提取对 HMC 极其重要,在自顶向下的跟踪中,特征的选择决定了可采取的搜索策略及匹配的准确性和鲁棒性;在自底向上的方法中,姿态恢复的准确性则直接取决于特征提取结果. 目前常用特征及提取方法如下:

1) 边缘. 一般用来与模型投影边界进行匹配,用边缘作匹配能得到较精确的位置. 提取边缘不要求背景固定,可用简单边缘提取算子,但易受背景和人体本身纹理的干扰,因此一般需要在相似度函数中考虑噪声的统计特性[6].

2) 剪影. 即图像中的人体区域,用来与模型投影区域匹配. 与边缘相比,剪影不易受噪声影响,但易丢失细节信息,一般可通过减背景、去除阴影、形态学滤波等操作得到[39].

3) 外轮廓. 即剪影的边界,也是一种边缘特征. 它可通过主动轮廓[21]变形得到,在能量函数中可融入多种测度,因此比边缘方法鲁棒,但计算量大.

4) 图像模板. 像素的灰度或颜色是一种最直接的特征,可将人体表示为一个或多个图像模板,跟踪时用模板匹配的方法将模板对齐到当前帧[12],但该方法存在模板更新问题.

5) 光流. 一种时序数据,反映帧间点或特征的一致运动,光流提取不要求背景固定,但计算量大、抗噪性差,该方法可用来估计模型的运动参数[26].

6) 小区域特征. 如颜色块[10]、Blob[27]等,一般用在基于特征的跟踪中,用颜色模板或者颜色统计特征来表示.

7) 3D 数据. 近年来很多研究者用 3D 特征数据进行跟踪和分析,如利用多摄像机立体匹配得到密集的 3D 点云[40],或使用轮廓重建技术得到人体的体数据表示[22]等. 使用 3D 特征可避免优化中对模型多次投影,提高效率,但目前 3D 重建算法还不很成熟.

4　难点问题和技术发展趋势

经过 20 多年的研究,HMC 技术虽已取得很大进展,但仍处于实验室研究阶段,现有系统一般都限于简单场景,且仅能捕捉单人或几个人的简单动作,其鲁棒性、精确性和速度离实用还有很大距离. 很多难题还有待解决:1)不加约束的人与背景的自动分离问题. 在动态环境中,因受摄像机运动、光照变化、杂乱背景及人与背景颜色相近等影响,该问题一直未得到很好解决;2)遮挡问题. 在遮挡情况下,特别是当上肢与躯干服饰的颜色纹理相近时,很难从图像中将它们区分开来;3)人体建模问题. 不同个体间身体结构、外形、衣着等差异使得人体形状的精确建模非常困难,而身体模型直接影响跟踪的精确性;4)人体模型的自动获取和失败恢复问题. 目前算法多用手工初始化,因误差累积等原因无法对长序列跟踪,因此算法的"自举"性是实际应用时必须解决的难题;5)处理速度问题. 人体姿态包含几十个自由度,如何对这一高维空间快速搜索以达到实时要求,也是亟待解决的.

尽管面临众多难题,重要理论价值和广阔的应用前景仍是 HMC 研究的强大动力. 我们认为,一个

鲁棒、精确的运动捕捉系统取决于测量技术和估计技术的进步，下面从这两方面分析 HMC 的发展趋势.

在测量技术方面，使用多摄像机仍将是未来的技术趋势，借助多个摄像机，不仅可解决遮挡问题，还能恢复深度信息得到 3D 姿态；另外，结合多种类型的感知设备(如红外摄像机、3D 扫描仪器、高清晰度摄像机等)，将大大地提高观测特征的准确性和可靠性. 在这方面，多视角及多类型数据的融合是重要研究方向.

在估计技术方面，基于模型的 3D 跟踪能获取精确的 3D 姿态信息，具有更广泛的用途，仍将是研究重点，针对基于模型跟踪的难点，以下几个方面值得关注：1)结合计算机图形学建立更精细的人体模型，如 Plankers 等[40]用人体动画模型从不同层次对人体进行表示，以提高逼真性和跟踪精度；2)研究将自底向上和自顶向下的方法结合起来以弥补二者缺陷的方法，如 Curio 等[41]将底层特征映射得到的人体姿态，用于跟踪中模型的自动初始化和失败自动恢复；3)使用多种特征提高跟踪的鲁棒性. 综合使用多种图像特征(如颜色、纹理、边缘、剪影及运动信息)能取长补短，将有效地提高跟踪的鲁棒性；此外，采用 3D 数据也是一个发展趋势；4)运动建模技术. 使用运动模型能大大地提高跟踪效率和鲁棒性，但对任意人体运动或对人的大多数动作建模非常困难，借鉴语音识别方法，建立一些类似字母表的基本动作单元，从而构建动作库是一种可行的途径[24]；5)为提高搜索速度以达到实时性要求，除期待新的理论算法外，采用并行处理是一种可行的方式[39]；6)鉴于日常生活中人的交互行为的重要性，对多人体运动进行捕捉、理解也是重要趋势.

参 考 文 献

[1] Moeslund T B, Granum E. A survey of computer vision-based human motion capture [J]. Computer Vision and Image Understanding, 2001, 81(3): 231-268

[2] O'Rourke J, Badler N I. Model-based image analysis of human motion using constraint propagation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, 2(6): 522-536

[3] Wang Liang, Hu Weiming, Tan Tieniu. Recent developments in human motion analysis [J]. Pattern Recognition, 2003, 36(3): 585-601

[4] Bregler C, Malik J, Pullen K. Twist based acquisition and tracking of animal and human kinematics [J]. International Journal of Computer Vision, 2004, 56(3): 179-194

[5] Deutscher J, Blake A, *et al*. Articulated body motion capture by annealed particle filtering [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Hilton Head, 2000, 2: 126-133

[6] Sminchisescu C, Triggs Bill. Covariance scaled sampling for monocular 3D body tracking [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Hawaii, 2001: 447-454

[7] Ning Huazhong, Wang Liang, *et al*. Articulated model based people tracking using motion models [C] //Proceedings of IEEE International Conference of Multimodal Interfaces, Pittsburgh, 2002: 383-388

[8] Chen Rui, Liu Guoyi, *et al*. 3D human motion tracking based on sequential Monte Carlo method [J]. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(1): 85-92 (in Chinese)
(陈 睿, 刘国翌, 等. 基于序列蒙特卡罗方法的 3D 人体运动跟踪[J]. 计算机辅助设计与图形学学报, 2005, 17(1): 85-92)

[9] Liu Guoyi, Chen Rui, *et al*. Video-based 3D human body motion capture [J]. Journal of Computer-Aided Design & Computer Graphics, 2006, 18(1): 82-88 (in Chinese)
(刘国翌, 陈 睿, 等. 基于视频的三维人体运动跟踪[J]. 计算机辅助设计与图形学学报, 2006, 18(1): 82-88)

[10] Luo Zhongxiang, Zhuang Yueting, Pan Yunhe, *et al*. Video based motion capture [J]. Journal of Image and Graphics: A, 2002, 7(8): 752-758 (in Chinese)
(罗忠祥, 庄越挺, 潘云鹤, 等. 基于视频的运动捕获[J]. 中国图像图形学报: A 版, 2002, 7(8): 752-758)

[11] Gavrila D M. The visual analysis of human movement: a survey [J]. Computer Vision and Image Understanding, 1999, 73(1): 82-98

[12] Ju S, Black M, Yacoob Y. Cardboard people: a parameterized model of articulated motion [C] //Proceedings of International Conference on Automatic Face and Gesture Recognition, Killington, 1996: 38-44

[13] Morris D D, Rehg J. Singularity analysis for articulated object tracking [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Santa Barbara, 1998: 289-296

[14] Gavrila D M, Davis L S. 3-D model-based tracking of humans in action: a multi-view approach [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, 1996: 73-80

[15] Moeslund T B, Granum E. Pose estimation of a human arm using kinematic constraints [OL]. [2005-10-17]. http://www.cvmt.dk/projects/puppet/html/publications/publications.html

[16] Sidenbladh H, Black M, Fleet D. Stochastic tracking of 3D human figures using 2D image motion [C] //Proceedings of European Conference on Computer Vision, Dublin, 2000: 702-718

[17] Karaulova I A, Hall P M, Marshall A D. A hierarchical model of dynamics for tracking people with a single video camera [C] //Proceedings of British Machine Vision Conference, Bristol, 2000: 262-352

[18] Pavloska V, James M R, Cham T J. A dynamic Bayesian network approach to tracking using learned switching dynamic models [C] //Proceedings of International Workshop on Hybrid Systems: Computation and Control, Pittsburgh, 2000: 366-380

[19] Zhao Tao, Wang Tianshu, Shum Harry. Learning a highly structured motion model for 3D human tracking [C] //Proceedings of Asian Conference on Computer Vision, Melbourne, 2002: 144-149

[20] Ong Eng-Jon, Gong Shaogang. Tracking hybrid 2D-3D human models from multiple views [C] //Proceedings of International Conference on Computer Vision, Kerkyra, 1999: 11-18

[21] Delamarre Q, Faugeras O. 3D articulated models and multi-view tracking with physical forces [J]. International Journal of Computer Vision and Image Understanding, 2001, 81(3): 328-357

[22] Luck Jason P, Debrunner Christian, *et al*. Development and analysis of a real-time human motion tracking system [C] //Proceedings of Workshop on Applications of Computer Vision, Orlando, 2002: 196-202

[23] Wachter S, Nagel H-H. Tracking persons in monocular image sequences [J]. International Journal of Computer Vision and Image Understanding, 1999, 74(3): 174-192

[24] Green R D, Guan L. Quantifying and recognizing human movement patterns from monocular video images—Part I: a new framework for modeling human motion [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(2): 179-190

[25] Choo K, Fleet D. People tracking using hybrid Monte Carlo filtering [C] //Proceedings of International Conference on Computer Vision, Vancouver, 2001: 321-328

[26] Lee M W, Cohen I. Human body tracking with auxiliary measurements [C] //Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Nice, 2003: 112-119

[27] Wren C, Azarbayejani A, darrell T, *et al*. Pfinder: real-time tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780-785

[28] Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809-830

[29] Zhao Liang. Dressed human modeling, detection, and parts localization [D]. Pittsburgh: Carnegie Mellon University, 2001

[30] Mori G, Malik J. Estimating human body configurations using shape context matching [C] //Proceedings of European Conference on Computer Vision, Copenhagen, 2002: 666-680

[31] Howe N R. Silhouette lookup for automatic pose tracking [C] //Proceedings of IEEE Workshop on Articulated and Nonrigid Motion, Los Alamitos, 2004: 15-22

[32] Efros Alexei A, Berg Alexander C, *et al*. Recognizing action at a distance [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Madison, 2003: 726-733

[33] Elgammal A, Lee Ch S. Inferring 3D body pose from silhouettes using activity manifold learning [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Washington, D C, 2004: 681-688

[34] Agarwal A, Triggs B. 3D human pose from silhouettes by relevance vector regression [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Washington, D C, 2004: 882-888

[35] Grauman K, Shakhnarovich G, *et al*. Inferring 3D structure with a statistical image-based shape model [C] //Proceedings of International Conference on Computer Vision, Nice, 2003: 641-648

[36] Wu Ying, Hua Gang, Yu Ting. Tracking articulated body by dynamic Markov network [C] //Proceedings of International Conference on Computer Vision, Nice, 2003: 1094-1101

[37] Sigal L, Bhatia S, *et al*. Tracking loose-limbed people [C] //Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Washington, D C, 2004: 421-428

[38] Ronfard R, Schmid C, Triggs B. Learning to parse pictures of people [C] //Proceedings of European Conference on Computer Vision, Copenhagen, 2002: 700-714

[39] Theobalt Christian, Carranza Joel, *et al*. A parallel framework for silhouette-based human motion capture [C] //Proceedings of Vision, Modeling and Visualization, Munich, 2003: 207-214

[40] Plankers R, Fua P. Articulated soft objects for multiview shape and motion capture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1182-1187

[41] Curio C, Giese M A. Combining view-based and model-based tracking of articulated human movements [C] //Proceedings of IEEE Workshop on Motion and Vision Computing, Breckenridge, 2005: 261-268

李豪杰　男，1973 年生，博士研究生，主要研究方向为人体运动跟踪、视频内容分析等.

林守勋　男，1948 年生，博士，研究员，博士生导师，主要研究方向为视频编码、视频分析和检索、统计机器翻译、语言技术评测.

张勇东　男，1973 年生，博士，副研究员，主要研究方向为视频内容分析、视频数据压缩等.