人体三维运动实时跟踪与建模系统[1)]

徐一华 李京峰 贾云得
(北京理工大学计算机系 北京 100081)
(E-mail: {yihuaxu, lijingfeng, jiayunde}@bit.edu.cn)

摘 要 提出了一种新的人体三维运动实时跟踪与建模系统设计方法，并基于此实现了一套鲁棒的参考应用系统. 针对人机交互等对跟踪精度要求不是很高的应用场合，系统在跟踪精确性和简易性与可推广性之间做了很好的折中. 系统使用多个摄像头采集图像，实时计算场景深度信息，然后结合使用深度和颜色信息进行人体跟踪. 应用一个简易的人体上半身三维模型，并使用基于颜色直方图的粒子滤波算法对头部和手部进行跟踪，从而恢复出模型的各个参数. 系统以人脸检测和人手肤色聚类算法为初始化方法. 大量实验证明，该系统能在复杂背景下进行人体上半身的跟踪和三维模型恢复，能进行完全自动的初始化，有较强的抗干扰能力和自动错误恢复能力. 系统在 2.4GHz PC 机上能以 25 帧 / 秒的速度运行.

关键词 人体跟踪，人体建模，人体三维模型，人机交互
中图分类号 TP391

A Real-time 3D Human Tracking and Modeling System

XU Yi-Hua LI Jing-Feng JIA Yun-De
(*Department of Computer Science, Beijing Institute of Technology, Beijing* 100081)
(E-mail: {lijingfeng,yihuaxu,jiayunde}@bit.edu.cn)

Abstract We propose a new method for real-time 3D human tracking and modeling, and implement a robust reference application system. Focusing on applications, e.g. desktop human computer interface, which require low tracking accuracy, the method makes a good compromise between the accuracy and system simplicity. It uses multiple cameras and recovers the depth maps in real-time, and then uses both color and depth information for tracking. It adopts a simple 3D human upper body model, and uses color-histogram-based particle filtering to track human head and hands, and hereby reconstructs the 3D model. By using face detection and hand color clustering algorithms, the system initialization is fully automatic. Extensive experiments demonstrate that the system can robustly track and model human motion from complex background, and can automatically recover from losing of tracking object. The system runs at 25Hz on a 2.4GHz PC.

Key words Human tracking, human modeling, human 3D model, human computer interaction

1 引言

基于视觉的人体运动分析在人机交互、数字娱乐、视频会议、智能监控、虚拟现实等领域有着很大的应用前景. 其主要任务是从一系列图象中检测、跟踪和识别人体，将人体

1) 国家自然科学基金项目 (60473049) 和国防基础研究项目 (60107211) 资助
Supported by National Natural Science Foundation of P. R. China (60473049), National Defence Science Foundation of P. R. China (60107211)
收稿日期 2005-5-18 收修改稿日期 2006-1-4
Received May 18, 2005; in revised form January 4, 2006

在图象中的运动转化为计算机可以理解的数字化符号，以此对人的运动行为进行分析理解．但是，人体是一个复杂的联合体，具有很高的自由度；在某些情况下运动速度很快，自遮挡严重；对于实际应用来讲，人所处的背景环境可能很复杂，因此进行鲁棒跟踪是一项很困难的任务．目前实际应用的人体运动捕捉系统大多对光照等环境有严格控制，并要求被跟踪者穿戴带特殊标记的服装，难以应用于人机交互等一般场合[1]．本文工作核心内容是采用一种相对简便的方法，能够在一般环境下实时跟踪和恢复人体上半身的三维运动参数，为后续的各种上层应用提供较好的数据．

在前人的工作中，如 Zivkovic 等[2] 的基于颜色直方图的类期望最大化的跟踪算法，能够实现人体等各种非刚体目标的跟踪．但由于其不特定区分人体各子部分，所以无法完成人体模型恢复工作．Krahnstoever 等[3] 、Zhang 等[4] 、Toyama 等[5] 通过基于样本的概率方法或跟踪过程中得到的人体运动等信息自动获取二维人体模型或模板．这类算法不需要先验模型知识，但模型获取的鲁棒性相对较差，同时这种获取的模型参数存在难以理解其语义的问题．Chen 等[6] 和 Ju 等[6] 的工作分别使用了线状模型和纸板人模型等拟合人体，跟踪得到的模型运动参数可以用于识别理解等．Wren 等[7] 采用了一个简单的人体上半身三维模型，通过非线性的递归滤波器进行模型恢复，准确性较高，并且能部分解决自遮挡问题．本文也使用一种简易有效的三维模型对人体上半身进行建模与跟踪．

目前已有很多采用单个摄像机实现人体跟踪的系统[5~8]．使用单摄像机的缺点是很难获得可靠的深度信息．深度信息可以显著提高人体跟踪算法的鲁棒性，同时也为三维人体模型的建立提供基础．本文采用多个摄像机同步采集现场视频，利用实时立体视觉恢复得到深度图像，从而提高跟踪的鲁棒性并获取三维运动参数．

由于人体运动内在的复杂性，使用卡尔曼滤波方法有一定的局限性．再如 Mean-shift[9]，CAMShift[10] 等基于局部直方图的迭代方法，无法处理目标快速运动的情况，一旦跟踪目标在前后两帧中没有重叠，就会丢失跟踪目标．本文采用基于颜色直方图的粒子滤波方法[11]，能够处理复杂人体运动，能够做到快速运动目标的跟踪．

自动初始化以及能够自动从错误中恢复的机制是人体跟踪应用系统的重要组成部分．目前大部分系统的初始化过程需要人工介入，并且不存在发生跟踪错误时的恢复机制．本文使用人脸检测和肤色聚类算法完成这一工作．此外系统会在检测到跟踪错误时启动上述过程以重定位人体位置，从而实现错误恢复的功能．

2　人体模型

在基于模型的人体跟踪方法中，模型设计对跟踪性能有决定性影响，必须同时考虑到对人体描述的准确性和计算复杂性．本文采用了一种适当简化的三维人体上半身模型，如图 1 所示．该模型共使用 8 个控制点，分别是：头、左肩、左肘、左手、右肩、右肘、右手、躯干．描述了人体上半身的 6 个主要部分：头，躯干，左右大臂、左右小臂与手．这个 24 维的向量描述了人体上半身的模型．通过后续的约束，我们对这个向量进一步降维．

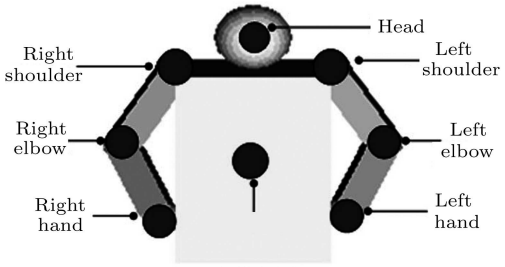

图 1　人体上半身三维模型

Fig. 1　3D human upper body model

本模型对人体各部分施加一定约束，如骨骼的连接方式、长度等．参照 [12] 中对大量人体的统计分析，估算各部分比例关系如下：$W_{torso}=1.4W_{head}$，$H_{torso}=1.4H_{head}$，$W_{arm}=0.5W_{head}$，$H_{arm}=2.2H_{head}$，$W_{forearm}=0.46W_{head}$，$H_{forearm}=1.9H_{head}$，其中 Wtorso, Htorso, Whead, Hhead,

Warm, Harm, Wforearm, Hforearm 分别为躯干、头、大臂和小臂的宽度和高度. 该模型假设人体上半身基本保持直立, 身体、肩膀和头部处在同一个深度, 并且相对位置恒定. 对于人体的一般运动, 这些假设都是可以得到满足的. 根据上述身体各部分的比例关系, 以及头及双手的位置, 可以大致估算出身体、双肩及手肘的位置. 这样就可以恢复出整个三维模型. 计算手肘的迭代公式如下, 我们通过牛顿迭代法求解手肘位置.

$$H^2_{arm}=(Hand_x\text{-}Elbow_x)^2+(Hand_y\text{-}Elbow_y)^2+(Hand_x\text{-}Elbow_y)^2$$

$$H^2_{forearm}=(Shoulder_x\text{-}Elbow_x)^2+(Shoulder_y\text{-}Elbow_y)^2+(Shoulder_x\text{-}Elbow_y)^2$$

$$Elbow_x=\max(Set(Elbow_x))$$

3 系统实现

系统首先进行人的整体分析, 用背景减除的方法得到人体整体图像. 而后进行局部分析, 得到人体的头、躯干、手、以及手肘等各个部分的局部参数. 最后根据获得的运动参数进行人体模型恢复, 得出人的运动分析结果. 系统的结构框图如下.

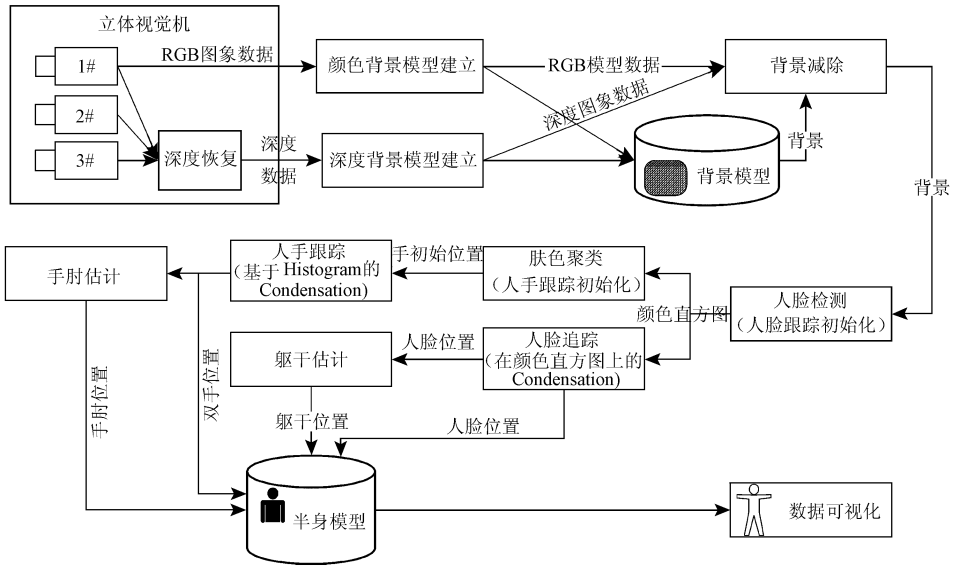

图 2 系统框图

Fig. 2 System diagram

3.1 深度恢复

本文目前采用 [13] 提出的立体视觉机作为数据来源. 它采用三个摄像头同步采集图像, 通过立体视觉处理获得高精度稠密深度图. 它采用嵌入式硬件板卡设计, 能够以 25 帧 / 秒的速度提供 640×480 分辨率、8 位精度的深度图像及同步的三路彩色图像. 图 3(2,4) 为该立体视觉机的输出图像.

3.2 背景减除

实验表明, 在场景变化不大的情况下, 采用合适的背景模型有利于跟踪的进行[14]. 但在单摄像机系统中, 光照变化等因素将会影响背景减除效果, 可能会导致跟踪错误. 由于仅用颜色信息进行背景减除不稳定, 因此我们结合深度和颜色信息共同进行背景减除, 从而获得更可靠的人体区域.

假设初始的若干帧不出现前景, 作为背景帧进行背景模型的建立. 系统首先对深度图

进行预处理以去除部分错误点，然后计算这些深度图像的平均深度，以得到场景的深度背景模型．在背景减除过程时，当图像中出现深度小于背景深度的点时，即判断该点为前景点，这样得到基于深度的前景模板．在进行基于颜色的背景减除时，为了减少亮度对颜色的扭变，建模过程在 HSI 颜色空间中进行．在背景图像序列中，统计其中每个点的颜色分布的数学期望和方差，即可建立颜色背景高斯模型．通过颜色背景减除即得到基于颜色的前景模板．采用深度和颜色信息分别生成前景模板之后，使用求与的方法求解两个模板的交集，然后进行形态学滤波以得到前景模板．这样的融合操作消除了很多错误像素．如图三所示：单纯基于颜色或者深度信息的前景模板有很多噪声及边缘不清晰，但是两者交集中的错误量显著减少．最后使用前景模板生成前景图像，如图 3(7).

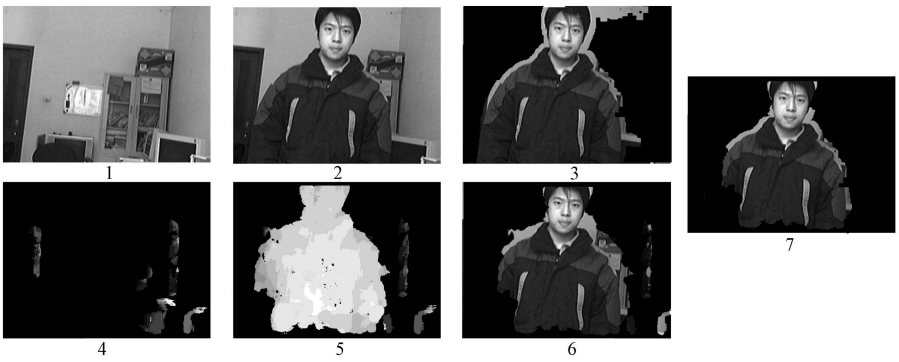

图 3　结合颜色和深度的背景减除 (1 、2 为颜色和深度背景，3 、4 为输入的彩色和深度图像，5 、6 为单独进行颜色和深度背景减除结果，7 为结合颜色和深度进行背景减除结果)

Fig. 3　The depth and color cues are combined to subtract background

3.3　头部定位与跟踪

经过背景减除得到良好的前景图像后，进行头部定位和跟踪工作．前人的跟踪系统经常要求使用者配合系统进行初始化过程，如把目标放在图像的正中央等．为了提供更友善的使用方式，本文使用自动人脸检测算法来初始化跟踪系统．假定在初始化过程中使用者正面朝向摄像机，这时在人体上半身的各个部分中，人脸的特征最为明显．本文采用 [15] 所提出的基于 Haar 特征的级联式 Boost 分类器进行人脸的定位．该算法具有很强的尺度适应性以及一定的旋转适应性，运行速度很快，鲁棒性很高．

获取初始的人脸位置之后，开始进行人脸的跟踪过程．本文采用基于肤色直方图的粒子滤波算法．粒子滤波方法是构建在贝叶斯概率理论框架基础上的对非线性非高斯问题的最优参数估计法．它能够处理复杂背景下目标的非线性运动，适用于人体各个部分的跟踪．本文跟踪计算人脸的 4 个参数：中心点位置和两个方向上的缩放尺度．由于人脸肤色组成特殊，在颜色直方图上呈现一定的分布特征．该特征具有缩放和旋转不变性，有较好的抗干扰能力．系统首先计算人脸检测得到的区域在 HSI 颜色空间中直方图．为去除图像大小的影响，需要对计算得到的直方图做归一化处理．然后计算初始人脸直方图和通过因子采样得到的直方图之间的距离，即可以获得每一个因子的观测权重．跟踪算法整体流程如下：

基于肤色直方图的粒子滤波跟踪算法：

1) 计算初始位置直方图 Hist_init

2) 从已有时间 t 的样本集 $\{S_{t-1}^n, \pi_{t-1}^n, n = 1 \cdots N\}$ 中，进行重采样，产生时间 t 的新样本，其中 N 为样本数：

a) 产生均匀分布随机数 $r \in [0, 1]$

b) 通过二分法查找最小的 j, 使得 $C_{t-1}^{j} \geqslant r$

c) 使 $S_t'^n = S_{t-1}^j$

3) 预测：使用匀速直线运动模型进行预测，使得 $S_t^n = AS_t'^n + W_t^n$, 其中 A 是运动模型矩阵, W 表示标准的正态分布噪声.

4) 观测

a) 计算每一个样本向量对应的图像区域归一化直方图 $\{H_{\text{sample}}(n), n = 1 \cdots N\}$.

b) 通过公式 7 计算初始人脸区域和样本区域直方图的距离 $B(H_{\text{init}}, H_{\text{sample}})$

c) 计算样本的置信度：$\pi_{t-1}^n = 1/B(H_{\text{init}}, H_{\text{sample}})$

d) 归一化样本置信度使得 $\sum_{n=1}^{N} \pi_t^{(n)} \equiv 1$

5) 参数估计：计算样本的加权平均值作为参数估计的值, $\vec{X}_t = \sum_{i=1}^{N} \pi_t^{(n)} \vec{S}_t^{(n)}$, $\vec{X}$ 是对目标新的位置的预测.

6) 回到 2)

3.4 人手定位与跟踪

人手关节众多，自由度很高. 自由运动的人手姿态差异很大，在图像中的特征较难提取，人手运动的跟踪和初始化都是比较困难的问题.

本系统利用人脸跟踪得到的位置和颜色信息来指导人手的初始定位. 由于同一个人的肤色大致相同，在得知人脸肤色信息之后，即可以得到图像中所有的类肤色区域，生成肤色区域二值图. 然后通过肤色聚类和位置约束，求得人手的初始位置. K-mean 算法聚类效率较高，但需要事先确定类别数量. 由于假定初始场景中只存在一个人，因此肤色区域一般只有人脸和左右手区域. 但是由于可能会有遮挡或区域合并等原因，可能的肤色区域会是三个、两个或者一个. 因此我们对 K-mean 算法做了一定改进，使其能够在一定的条件下自动减少类别数量，从而自动决定实际的肤色区域个数.

人手初始定位完成后，可以获得人手的肤色直方图信息，然后同样使用的基于肤色直方图的粒子滤波器进行人手跟踪. 同人脸跟踪一样，跟踪左右手位置及缩放各 4 个参数. 由于人脸运动速度相对较慢，因此在设置运动模型时，速度矩阵 A 的时候需要设置较小的运动速度，并且在预测的过程中添加的随机分布高斯噪声 W 的方差也应该较小. 而人手运动速度可能很快，因此在人手运动模型的设置中，速度矩阵 A 应允许跟踪目标有较大的位移，并且添加的随机噪声方差也较大，以保证跟踪的正常进行.

3.5 三维模型恢复

得到了脸和手的二维参数之后，可以通过这些二维参数与已知的深度图像获取身体各个部分的三维参数. 头和手的三维参数估计相对简单. 由于目标较小，可以假定同一个目标的深度相同. 在深度图中求出目标区域的平均深度 Z_{obj}. 通过摄像机标定的参数，在给定深度的情况下，可以得到物理坐标系和图像坐标系之间的转换关系. 这样就计算得到目标的三维坐标.

本文假定躯干与头部处于同一个平面，深度相同. 施加前面讲的人体模型约束，可以估算出躯干的三维坐标. 最后进行手肘的参数估算. 由于手肘在图像中的特征不明显，对手肘进行跟踪定位比较困难. 但是由于已知手的位置，并可以由头的位置估算出肩的位置，再施加骨骼长度等约束，即可大致估计出手肘的位置.

3.6 错误恢复

由于各种干扰等因素，跟踪器有时会发生跟踪错误的情况，因此在一个实际的应用系统中，某种程度的错误恢复机制是必不可少的. 本系统采用人脸检测和人手肤色聚类方法

解决这一问题. 当系统检测到跟踪错误时, 就会重新启动人脸检测和人手肤色聚类过程, 并以此指导整个跟踪器重新启动. 同时跟踪器也会对检测过程进行指导, 防止由于检测出多处人脸目标而发生错误跟踪情况. 此外, 当发生遮挡情况或多个跟踪目标重合时, 系统会进行重新初始化过程从而进行恢复错误. 实验证明本系统的错误恢复机制是鲁棒有效的.

4　实验结果

按照上述方法, 我们构建了一套完整的人体跟踪系统, 并在多种应用环境包括复杂背景和光照条件下进行了大量实验. 实验结果表明, 本系统能够实时地实现人体运动的鲁棒跟踪和三维模型恢复. 系统能够在跟踪目标进入时做到自动初始化; 能够处理快速的人体运动; 能够有效抵制复杂背景中其它运动目标和类肤色区域的干扰; 能够处理局部被遮挡的情况和跟踪目标丢失的情况, 有较强的抗干扰能力. 图 5 是一组跟踪实验结果以及对应的三维模型恢复结果. 其输入图像序列分辨率为 640×480, 其中前 30 帧为单纯的背景图像序列, 跟踪过程中有他人进入场景作为干扰. 人脸跟踪和手的跟踪均在 HSI 颜色空间中进行, 颜色直方图的分桶数为 20. 系统工作时同时运行三个跟踪器跟踪人脸和左右手, 并同时进行人体模型建立过程. 在主频为 2.4GHz 的普通 PC 机上, 系统以 25 帧 / 秒的速度实时运行. 下面的分别为定性和定量的实验结果.

根据对连续 600 帧有效跟踪的统计, 系统的准确性分别为头部: 97.51%, 平均误差 1.81 像素; 左手: 87.86%, 平均误差 1.89 像素; 右手, 84.73%, 平均误差 1.78 像素.

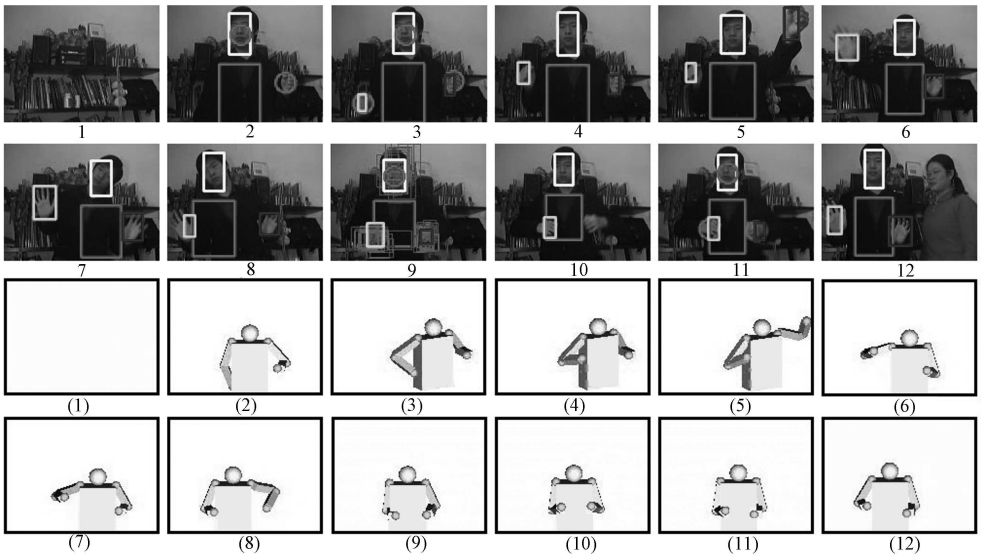

图 4　跟踪及三维模型恢复的可视化结果. 1 为初始复杂背景. 2 、3 为人脸和手初始化过程, 4∼12 为单人上半身跟踪过程; 可以允许快速运动 (4∼8), 头部偏转 (7,8), 图像模糊 (6), 错误恢复 (10,11), 抗多人干扰 (12). 9 显示了部分跟踪器的粒子. 图中, 圆形表示聚类结果中心, 方框分别表示脸、左右手和身体位置

Fig. 4　The result of tracking and 3D model rebuilding

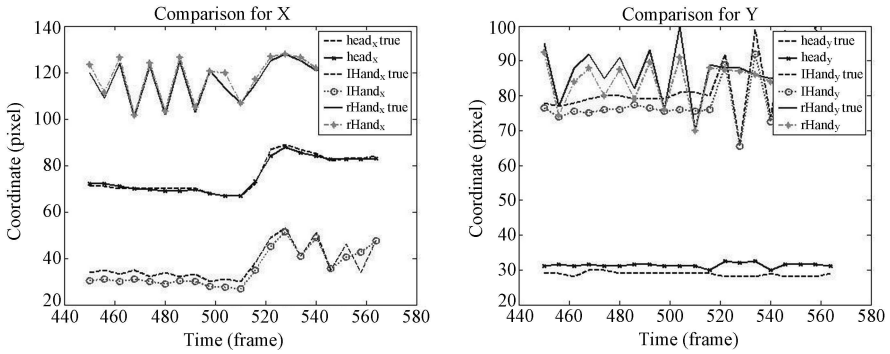

图 5 左图为脸及双手 X 方向跟踪值与实际值对比，右图为对应点 Y 方向比较

Fig. 5 The comparison of tracking result and ground truth

5 总结和讨论

本文提出了一种新的实时人体跟踪及建模方法，并实现了一套参考系统. 实验证明，该系统能在复杂背景下对人体上半身的一般运动进行实时鲁棒的跟踪和建模，并很好地解决了自动初始化和错误恢复等问题. 该系统对光照环境和被跟踪者的着装等没有严格要求，可以广泛运用在包括人机交互、数字娱乐等多种场合.

除整体跟踪系统设计方案外，本文其它主要的创新和贡献如下：1) 应用了一种简化的人体上半身模型. 使用 8 个控制点对人体上半身建模，大大降低了人体跟踪的复杂度，同时能够很好的模拟人上半身的一般动作. 2) 有效引入深度信息进行人体跟踪，结合使用深度和彩色信息进行人体跟踪. 这是本系统能够获得高度鲁棒性的重要基础. 实验证明，这两种信息的结合有利于处理复杂背景和复杂光照情况下的人体分割. 3) 利用人体脸部和手部特征，提出了全自动的初始化及错误恢复方法.

基于本文的跟踪系统，可以在以下方向进行深入研究：1) 利用 PC 机实现立体视觉恢复，使本系统具有更好的可推广性. 2) 跟踪方法的进一步优化，以提高鲁棒性和运行速度. 包括更有效的利用深度信息等信息. 3) 减小系统对于人体必须正对摄像机等姿势的依赖. 4) 基于该跟踪系统开发上层应用，如人机交互、视频游戏、虚拟现实等.

References

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

2 Zoran Zivkovic, Ben Krose. An EM-like algorithm for color-histogram-based object tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, D.C., USA: IEEE Press, 2004. **1**: 798∼803

3 Krahnstoever N, Sharma R. Articulated models from video. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, D.C.: IEEE Press, 2004. **1**: 894∼901

4 Zhang J Y, Collins R, Liu Y X. Representation and matching of articulated shapes. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, D.C., USA: IEEE Press, 2004. **2**: 342∼349

5 Toyama K, Blake A. Probabilistic tracking in a metric space. In: IEEE International Conference on Computer Vision, Vancouver, Canada: IEEE Press, 2001. 50

6 Aggarwal J K, Cai Q. Human motion analysis: A review. *Computer Vision and Image Understanding*, 1999, **73**(3): 428∼440

7 Wren C R, Pentland A P. Understanding purposeful human motion. In: Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, France: IEEE Press, 2000. 378
8 Wang L, Hu W M, Tan T N. A survey of visual analysis of human motion. *Chinese Journal of Computer*, 2004, **25**(3): 225∼237
9 Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina: IEEE Press, 2000. 142∼151
10 Paul A Viola, Michael J. Jones: Robust real-time face detection. *International Journal of Computer Vision*, 2004, **57**(2): 137∼154
11 Prez P, Vermaak J, Blake A. Data fusion for visual tracking with particles. *Proceedings of the IEEE*, 2004, **92**(3): 495∼513
12 Bruderlin A. Calvert T W. Goal-directed, dynamic animation of human walking. *Computer Graphics*, 1989, **23**(3): 233∼242
13 Jia Y D, Xu Y H, Liu W C, Yang C, Zhu Y W, Zhang X X, An L P. A miniature stereovision machine for real-time dense depth mapping. In: Proceedings of International Conference on Computer Vision Systems, Graz, Austria: Springer, 2003. 268∼277
14 Stauffer C, Grimson W E L. Adaptive background mixture models for real time tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Ft. Collins, CO, USA: IEEE Press, 1999. **2**: 246∼252
15 Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection. In: Proceedings of the IEEE Conference on Image Processing, New York, USA: IEEE Press, 2002. 155∼162

徐一华 北京理工大学计算机系，博士研究生，于 1998 年和 2001 年分别获学士和硕士学位，后加入微软亚洲研究院任助理研究员．研究方向为计算机视觉、媒体计算和嵌入式计算．

(**XU Yi-Hua** Ph. D. candidate at Beijing Institute of Technology. Received his bachelor and master degrees in 1998 and 2001, respectively, then joined Microsoft Research Asia as an assistant researcher. His research interests include computer vision, media computing, and embedded computing.)

李京峰 北京理工大学计算机系硕士研究生．研究方向为计算机视觉和人机交互．

(**LI Jing-Feng** Master student at Beijing Institute of Technology. His research interests include computer vision and human computer interaction.)

贾云得 北京理工大学计算机系教授，博士生导师．研究方向是计算机视觉、媒体计算和智能系统．

(**JIA Yun-De** Professor of Computer Science at Beijing Institute of Technology. His research interests include computer vision, media computing, and intelligent systems.)