

基于紧身衣的人体动画研究*

朱 强, 庄越挺, 潘云鹤

(浙江大学 计算机科学与工程系, 浙江 杭州 310027);

(浙江大学 微软视觉感知联合实验室, 浙江 杭州 310027)

E-mail: zhuq2000@yahoo.com; yzhuang@cs.zju.edu.cn

http://www.zju.edu.cn

摘要: 传统人体动画技术计算量大, 缺乏真实感, 制作成本昂贵. 为此, 提出了一种基于紧身衣和相机定标的新的
人体动画技术. 将身穿紧身衣模特的运动拍摄成视频后, 可在图像序列中自动跟踪人体关节信息, 并建立透视投
影下的三维人体运动骨架序列, 最终通过自动生成 BVH(biovision hierarchy)文件与当今流行的 Poser, 3DMax 等
人体动画制作软件接口, 将模特的运动信息赋予新的角色. 该方法具有运动信息来源广泛、高效、制作成本低等
特点, 而且产生的人体运动非常真实. 这不仅将动画师从枯燥的工作中解放出来, 而且也可用于家庭动画短片
的制作. 最后还给出了实验结果.

关键词: 紧身衣; 色块; 聚类; 定标; BVH(biovision hierarchy)

中图法分类号: TP391 文献标识码: A

计算机的出现推动了各项事业的发展, 影视业也受到了巨大的冲击. 无论是从风靡世界的 DISNEY 动画片,
还是从像《星球大战》这样的巨片中, 人们都能体会到计算机所带来的逼真的视觉享受. 传统动画技术总体
分为两类: 一类是伴随着计算机动画技术发展起来的模型动画, 另一类是近年来兴起的基于运动获取的动画.

运动获取是当前计算机动画领域中最热门的方向之一. 它的操作过程大体是这样的: 首先让一个模特穿上
布满传感器或光学 Mark 的特制衣服表演各种动作, 然后计算机根据传感器反馈的信息获取人体表面的运动信
息, 这样就可以建构出运动中的人体三维体表模型. 这一技术虽然比关节动画产生更真实的人体运动, 但其获取
运动信息的过程往往耗资巨大, 又需要专门的工作室, 或者有电缆线附在人体上, 而这又限制了演员的自由运
动, 造成一定程度的动作失真.

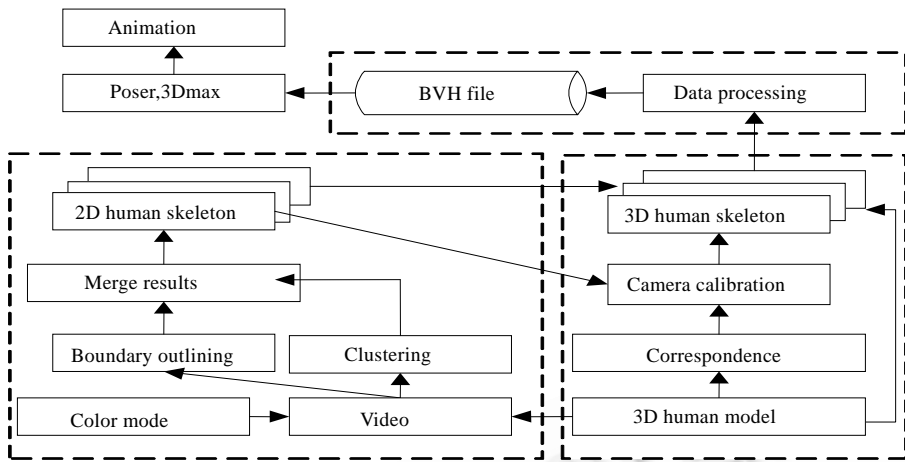
针对传统方法的各种缺陷, 我们试图开发高效的、制作成本低且运动限制少的方法来完成这个动画制作.
通过对国外一些大公司的运动捕获装置(motion capture)的比较和研究, 基于现有计算机视觉领域中对人体运动
分析的研究, 我们得到了启发, 开发出了基于紧身衣的人体动画技术. 我们特制了一件紧身衣, 其上对应于不同
人体关节的地方都贴上不同的颜色块. 当拍摄穿上该紧身衣的人体视频之后, 系统就能自动地跟踪关节的位
置, 并在透视投影下恢复人体的三维运动信息, 最终我们通过 BVH 文件将该运动信息优化后赋予不同的角色.

本文第 1 节将介绍我们所用的身体模型和颜色模型. 第 2 节介绍人体运动跟踪的具体算法. 第 3 节运用计
算机视觉的技术建立透视投影下的三维人体运动骨架序列. 第 4 节介绍了 BVH 文件的格式和如何将我们的三
维空间数据转换为旋转数据. 最后给出几段运动获取的结果, 并对文章进行总结并探讨未来的工作方向.

* 收稿日期: 2000-06-14; 修改日期: 2000-10-23

基金项目: 国家自然科学基金资助项目(69803009, 69733030); 国家教育部高等学校博士学科点专项科研训练资助项目
(20010335049); 国家教育部高等学校骨干教师资助计划项目; 国家教育部优秀青年教师基金资助项目

作者简介: 朱强(1976 -), 男, 浙江湖州人, 博士生, 主要研究领域为人体动画; 庄越挺(1965 -), 男, 浙江慈溪人, 教授, 博士生导师, 主要研究领域为人体动画, 多媒体数据库, 智能 CAD; 潘云鹤(1946 -), 男, 浙江宁波人, 教授, 博士生导师, 中国工程院院士, 主要研究
领域为计算机美术, 形象思维, 智能 CAD 系统, GIS, 计算机动画.



动画, Poser, 3Dmax, BVH 文件, 数据处理, 二维人体骨骼, 三维人体骨骼, 结果合并, 相机定标, 边界勾勒, 聚类, 对应, 颜色模型, 视频, 三维人体模型.

Fig.1 The architecture of our approach

图 1 系统的整体框架

1 人体模型和颜色模型

1.1 人体模型

我们将三维人体看成是由关节连接的刚体的集合^[1].如上肢是由肘关节连接的上下臂两个刚体,上臂与躯干是由肩关节连接的.该人体模型由 16 个关节点构成^[2].如图 2 所示,我们在白色为底紧身衣的各个关节点处附上色块(头部可视为自然的黑色块),共选择了 8 种对比较明显的颜色.因为颜色数目小于关节数目,所以两个不同关节点的色块可能具有相同颜色.为此,在首帧我们利用了一些合理约束对这样两个色块加以区别,罗列如下(假设在不平卧并且不完全下蹲的情况下):

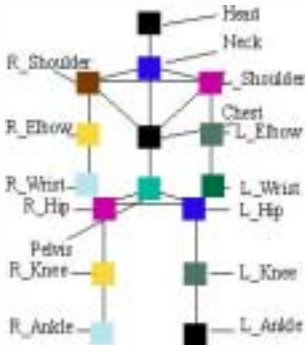


Fig.2 2D human model
图 2 二维人体色块模型

- 头部<高于>颈部<高于>胸部<高于>腹部
- 上肢关节点高于下肢关节点(手腕大腿间可有例外)

以上两点足以区别开模型中的相同色块.在后续帧我们可以用运动连续性将色块区分开来,无须加上上述运动约束.实际上我们将特征提取与特征对应这两步合二为一来完成,即在寻找图像序列特征点的过程中完成了

对应的建立.

1.2 颜色模型

RGB 是一种广泛采用的模型,但它主要是基于色彩易于硬件实现为目的,往往同一种颜色在强弱不同的光照下 RGB 的值相差较大.在我们的应用中,由于帧间明暗存在差别,同一色块因光照不均也存在高光或暗区域,所以我们需要的一种能去掉亮度信息的颜色模型,即采用色度空间.为此我们选择了[r,g]模型^[3],下面是[r,g]到[R,G,B]的变换矩阵:

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} 1 / (R + G + B), 0, 0 \\ 0, 1 / (R + G + B), 0 \\ 0, 0, 1 / (R + G + B) \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} . \tag{1}$$

2 骨架的跟踪

为了使代表各关节色块的跟踪更具鲁棒性,我们利用了两个层次的信息.其中基于自适应策略的聚类利用了色块的内部区域信息,由于在色块中存在着高光区域,导致仅由本算法得到的边界不完整.同时,基于 Robert 算子的边界勾勒利用了色块的边界信息,但由于边界模糊带,即色块与衣服底色交接处的存在,该算法也存在着同样缺点.我们综合上述两种算法,提出了层次化的跟踪:利用基于自适应策略的聚类算法避免了色块边界的噪声;而基于 Robert 算子的边界勾勒算法又避免了色块内部区域噪声对跟踪的影响.我们在每个算法跟踪完成之后,对两块结果区域求并,使得源于不同层次的噪声互相抵消.最后,对合并后的新区域,我们作出区域边界上点 X, Y 坐标的直方图,选取相距最远的峰值作为最后跟踪所得的矩形色块边界.对算法进一步的描述可以参考文献[4].

3 三维人体运动骨架重建的实现

首先我们需要根据跟踪出的二维关节位置与三维人体模型之间的对应,通过基于 5 点的 Tsai 单相机线形定标^[5]获得摄像机参数.然后,利用透视投影关系和人体骨骼的比例关系,求解人体模型上各个三维特征点的坐标.在随后的序列中,我们利用运动连续性和人体关节的合理性有效地消除了三维运动信息的二义性.

3.1 基于 5 点的 Tsai 单相机线形定标

由计算机图形学可知,世界坐标系中的任何一点 $P_w(X_w, Y_w, Z_w)$ 都可以经过以下两个变换(见式(2)和式(3))转化到投影平面上的点 $P(u, v)$ (R 为旋转矩阵, T 为平移矩阵).

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \times \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} - T, \quad (2)$$

$$(u, v) = \left(f \cdot X_c / Z_c, f \cdot Y_c / Z_c \right). \quad (3)$$

摄像机定标问题可以描述如下:给定一些三维人体模型上的特征对象以及其对应的二维图像上的特征对象,求得投影参数 R, T 以及 f . 本文在以后的表述中分别将该平面的各个特征点称为关键节点.跟踪结果出来后,关键节点在二维平面上的投影是已知的.关键节点在世界坐标系中的位置是由系统来指定的.因为只要关键节点间的距离的比例符合人体解剖学,那么我们总能在世界坐标系中确定摄像机的位置和方向,使得系统指定点的透视投影与二维图像帧上躯干上的对应点趋向重合.下面我们详细描述该算法.

Tsai 算法有两种形式:一是基于共面的 5 点定标法;二是基于不共面的 7 点定标法.下面我们就第 2 种形式进行讨论.最后我们将解释如何从 7 点定标法简化为我们所需要的 5 点定标法.因为 $R(r_1, r_2, \dots, r_9), T(T_x, T_y, T_z)$ 和 f 一共有 12 个不独立的变量,而我们根据 5 点只能获得 5 个对应方程.所以我们必须引入中间变量: $Ty^{-1}r_i, Ty^{-1}T_x$, $i \in [1, 6]$. 具体步骤如下:

(1) 这些中间变量可以通过式(4)的线形方程组加以计算:

$$\begin{pmatrix} Ty^{-1}r_1 \\ Ty^{-1}r_2 \\ Ty^{-1}r_3 \\ Ty^{-1}T_x \\ Ty^{-1}r_4 \\ Ty^{-1}r_5 \\ Ty^{-1}r_6 \end{pmatrix} \times \begin{pmatrix} Y_{di}X_{wi}, Y_{di}Y_{wi}, Y_{di}Z_{wi}, Y_{di}, -X_{di}X_{wi}, -X_{di}Y_{wi}, -X_{di}Z_{wi} \end{pmatrix} = X_{di}. \quad (4)$$

(X_{wi}, Y_{wi}, Z_{wi}) 为第 i 点的世界坐标,而 (X_{di}, Y_{di}) 为 i 点在图像平面上对应的点.如果我们有超过 7 对的对应点,那么我们就可以通过解一个超定方程组来解得这 7 个未知数.

(2) 然后通过下面的方法计算 $(r_1, r_2, \dots, r_9, Tx, Ty)$. 令

$$A1=Ty^{-1}r_1, A2=Ty^{-1}r_2, A3=Ty^{-1}r_3, A4=Ty^{-1}Tx, A5=Ty^{-1}r_4, A6=Ty^{-1}r_5, A7=Ty^{-1}r_6.$$

$$\text{计算 } |Ty| = \left(A_5^2 + A_6^2 + A_7^2 \right)^{-1/2}. \quad (5)$$

决定 Ty 的符号可以用试探法:先取一点 $I(Xwi, Ywi, Zwi)$, 它离图像中心最远;然后假设 Ty 为正, 计算下列值: $r_1=A1*Ty, r_2=A2*Ty, r_4=A5*Ty, r_5=A6*Ty, Tx=A4*Ty, x=Xw*r_1+Yw*r_2+Tx, y=Xw*r_4+Yw*r_5+Ty$. 如果 x 和 Xd, Y 和 Yd 有相同的符号, 则 Ty 取正, 反之取负.

按下列式子计算: $r_1=A1*Ty, r_2=A2*Ty, r_3=A3*Ty, r_4=A5*Ty, r_5=A6*Ty, r_6=A7*Ty, Tx=A4*Ty$. 因为 R 矩阵是正交的, 所以第 3 行可以用前两行的交乘来计算.

(3) 最后, 在忽略镜头失真的情况下, 我们可以按式(6)来计算 Tz 和 f :

$$(yi, -dyYi) \times \begin{pmatrix} f \\ Tz \end{pmatrix} = widyYi, \quad (6)$$

其中

$$\begin{aligned} yi &= r_4 * xwi + r_5 * ywi + r_6 * zwi + Ty, \\ wi &= r_7 * xwi + r_8 * ywi + r_9 * zwi. \end{aligned} \quad (7)$$

我们可以从 5 点对应中得到 5 个方程来求解这两个变量.

对于共面的点, 如果我们令每个点的 Z 坐标为 0, 那么式(4)就成为 5 个未知量的线形方程组了. 在我们的方法中, 我们选取了三维人体模型上 5 个点: 颈部、胸部、腹部、左肩、右肩. 通过对人体初始位置的限制, 我们可以保证这 5 个点处于人体上身同一个平面内.

3.2 首帧图像对应的三维人体运动模型的建立

在上面的工作中, 已经求得了 7 个投影参数. 但此时人体模型的三维特征点中除了 5 个关键节点以外都不确定. 我们可以选择这 5 个关节点中胸部、左肩、右肩组成一个稳固的三角形, 然后下一步就是从该已知的三角形的 3 个二维特征点坐标 (U, V) 出发, 求出与该点对应的人体模型上三维特征点的坐标 $Pc(Xc, Yc, Zc)$. 但从透视投影的针孔模型可知, 连接摄像机光心和成像点得到一条直线, 该直线上的所有点在投影平面上都成像于同一点. 为了在这条直线上找到三维特征点, 本文利用人体骨骼长度知识, 从一个与所求点相邻的已知三维点出发, 寻找该直线上的一个点, 使得该点到已知三维点的距离等于两点间相应骨骼的长度.

下面以从已知的右肩部点 Pc' 推出未知的右肘部点 Pc 为例, 说明本文的算法. 设 $Pc(Xc, Yc, Zc)$ 为右肘部点在摄像机坐标系中的坐标, (u, v) 为 Pc 在投影平面中的坐标. 由于 Pc 和 Pc' 的距离 Len 等于人体模型中上臂的长度, 结合式(2)和式(3)可以得到仅含有未知量 Zc 的方程. 该方程可以形象地表示为求投影直线与以 Pc' 为球心, Len 为半径的球面的交点. 方程的解有 3 种可能情况. 当有两个解时说明肘部有两种可能位置. 这种从二维向三维建模时产生的二义性是由于该问题本身的病态性造成的. 本文结合人体解剖学中的各种约束和亮度信息来消除二义性; 当只有一个解时, 该点就是所求的三维右肘部节点的位置; 当无解时, 可能有两个原因: 作为已知条件的二维特征点坐标存在误差, 或是作为先验知识的人体骨骼长度过短, 此时, 系统会适当调整骨骼长度之后再行计算. 至此, 通过以上所述的从已知二维点计算未知三维特征点的方法, 按从中间到四周、从上到下的扩展顺序可以求出所有的三维特征点.

3.3 后续帧对应的三维人体运动模型的建立

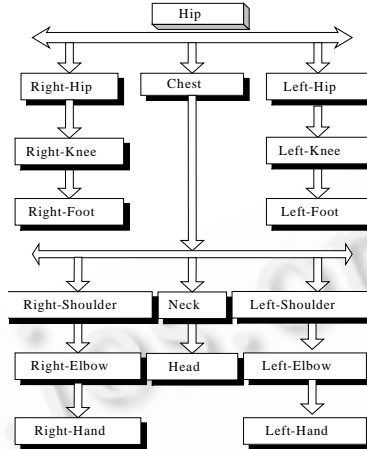
在第 3.2 节中, 我们已经对第 1 个图像帧建立了相应的三维人体模型. 事实上, 对于任一帧只要知道 3 个人体关键节点的坐标, 就可以用上节的方法求出该帧所对应的三维人体模型. 下面讨论在后续帧中如何确定第 3.2 节所述的三角形 3 个关键节点的坐标.

设第 n 帧中关键节点在摄像机坐标系中的坐标为 $Pi'(Xi', Yi', Zi')$ ($i=1\sim3$) 已知, 要计算第 $n+1$ 帧中相应的关键节点的摄像机坐标 $Pi(Xi, Yi, Zi)$ ($i=1, 2, 3$). Pi 在投影平面中对应的二维特征点坐标为 (Ui, Vi) ($i=1, 2, 3$), 则由式(8)可知 Pi 与 (Ui, Vi) 的换算关系为

$$P_i = \left(\frac{U_i \cdot Z_i}{f}, \frac{V_i \cdot Z_i}{f}, Z_i \right), (i=1,2,3). \tag{8}$$

同时,由刚体的属性可知,人体模型中各骨骼的长度保持不变,则有以下式:

$$d(P^i, P^j) = d(P_i, P_j), (i, j=1,2,3) \text{ 且 } i \neq j. \tag{9}$$



臀部, 右臀, 胸部, 左臀, 右膝, 左膝, 右脚, 左脚, 右肩, 颈部, 左肩, 右肘, 头部, 左肘, 右手, 左手.

Fig.3 The tree structure of BVH
图3 BVH文件的树结构

将 P^i 和式(8)代入上式后得到一个非线性方程组.它有3个未知数,可以通过梯度法来求出一组解.这样就得到了本帧中3个关键节点的摄像机坐标 P_c .此时再利用第3.2节中所述方法可以求出该帧对应的人体模型中全部三维特征点的坐标.本文假设人体运动具有连续性,在出现二义性的时候计算两个可能坐标到上一帧中该三维特征点的距离,并取距离较短的点作为该匹配点的最终位置.实验表明,这种消除二义性的方法有很出色的效果,在一个长序列的人体运动中充分体现了运动的连贯性与真实性.

4 BVH文件的格式与制作

4.1 BVH(biovision hierarchy)文件格式简述

BVH 是一种通用的人体特征动画文件格式,广泛地被当今流行的各种动画制作软件(Poser,3DMax 等)支持.它对人体的表达采用的就是图3的骨架模型,通过树结构来实现:用节点表示关节,用连接表示肢体.同时,身体的各个部分形成了子树的形式,包括:arm-tree,leg-tree,torso-tree.BVH文件的头部给出了上述的关节树,并指明关节与关节之间的相对位置,即定义了人体各部分的骨骼长度.由于人体存在的差异,可以通过改变相对位置的数值来区别男人、女人与小孩.对于每一个关节,都有3个旋转参数(相对于X,Y,Z轴的旋转角度)来描述运动信息.而 Hip 关节点还含有三维空间位置参数,也就是说,Hip 节点有6个自由度.从而完成了对人体的完整描述.BVH文件的尾部按照头部定义的关节顺序提供每帧数据,形成了对一个复杂动作的运动信息的表述与存储.

4.2 BVH文件的制作

我们从基于视频的运动获取系统中得到的原始数据是各关节的绝对三维坐标,需要转变为关节的旋转度.当把人体视为多关节体后,我们注意到:人体的各个部位通过关节体互相连接,当某一关节转动时将带动其他关节的转动.在 BVH 文件中,关节点的排列是树的形式,所以改变任意关节点的旋转参数都将影响该关节点的子树的所有节点,即考虑的是分析和求解各部位相对变化的角度.关节树的根节点是 Hip,在已确定的世界坐标系

下,它与 Chest 之间形成的旋转角可以认为是方向数(0 1 0),即由与 Y 轴同向的向量绕 X 轴与 Z 轴旋转得到;对于多关节,部位体的联合转动,实质上是相对坐标系下的旋转,是子节点相对于父节点的旋转.利用图形学中一系列的旋转变换,三维点信息转化为各个关节的旋转度.注意:旋转轴的选取顺序与结果有关.

5 实验结果

按上文所述算法,我们用 VC++6.0 与 OpenGL 实现了一个原型程序 Tight-Clothing Based Human Animation(TBHA).TBHA 能自动建立三维人体骨架序列和生成 BVH 文件.我们以实验室中录下的视频片段作为 TBHA 的素材.图 4 中显示的视频是我们从正面拍摄的模特跳跃动作序列.在首行中分别是这个视频片段的第 1,8,16,24,32 帧.图 5 中显示的视频是我们从正面拍摄的模特拳击动作序列.在首行中分别是这个视频片段的第 1,20,40,60,80 帧.模特身上的 16 个亮点代表了我们的跟踪结果.可以看到,跟踪的精度完全可以满足人体动画的需要.第 2 行是 TBHA 建立的透视投影下的三维人体的骨架序列.最后,我们通过制作 BVH 文件可以看到当前角色被赋予模特运动信息后的效果.

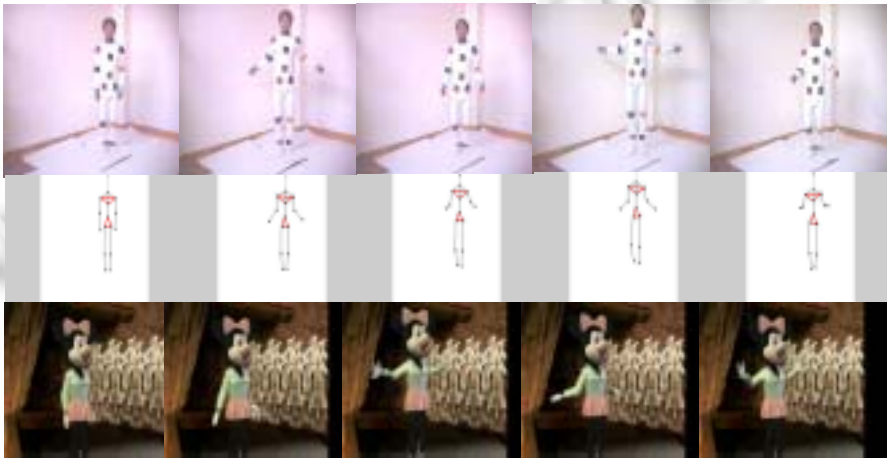


Fig.4 Jump clip (the first row is original video, the middle row is stick model, the last row is Mi_Mouse animation)

图 4 跳跃视频(第 1 行为原始视频,第 2 行为火柴棍模型,第 3 行为被赋予运动信息的米老鼠)



Fig.5 Boxing clip (the first row is original video, the middle row is stick model, the last row is QinYong animation)

图 5 拳击视频(第 1 行为原始视频,第 2 行为火柴棍模型,第 3 行为被赋予运动信息的秦俑)

6 结论与展望

本文介绍了一种新的基于紧身衣的人体动画技术.拍摄下模特的运动视频之后,我们能对视频中人体的各个关节点进行自动跟踪,然后在透视投影下恢复其三维运动信息,并可通过制作 BVH 文件将我们的运动信息赋予不同的新角色.相比于国外需要特别的实验室、耗资上百万的运动捕获系统,我们的方法具有素材来源广、制作成本低廉、简单、高效等特点.

因为我们的系统是基于单相机的运动获取,在自遮挡严重时我们的跟踪算法将无法进行.由于跟踪、定标时误差的积累导致最终捕获的信息存在着误差,而我们却没有有效的机制对原始数据进行优化和编辑^[6].以上几点也是我们今后继续工作的方向.

致谢 本实验室的诸晓琴、郑科、刘丰、毛翌等同学为本系统的实现做了部分工作,在此表示感谢.

References:

- [1] Aggarwal, J.K., Cai, Q. Human motion analysis: a review. In: Proceedings of the IEEE Nonrigid and Articulated Motion Workshop 1997. Piscataway, NJ: IEEE, 1997. 90~102.
- [2] Zhuang, Yue-ting, Liu, Xiao-ming, Pan, Yun-he. Video motion capture using feature tracking and skeleton reconstruction. In: Proceedings of the IEEE International Conference on Image Processing. 1999. 232~236.
- [3] Liu, Ming-bao, Yao, Hong-xun, Gao, Wen. Real-Time human face tracking in color images. Chinese Journal of Computers, 1998,21(6):527~532 (in Chinese).
- [4] Zhuang, Yue-ting, Zhu, Qiang, Pan, Yun-he. Hierarchical model based human motion tracking: In: Proceedings of the IEEE International Conference on Image Processing 2000. 2000. 347~352.
- [5] Tsai, R.Y. An efficient and accurate camera calibration technique for 3D machine vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1986. 364~374.
- [6] Gleicher, M. Animation from observation: motion capture and motion editing. Computer Graphics, 1998,33(4):51~54.

附中文参考文献:

- [3] 刘明宝,姚鸿勋,高文.彩色图象的实时人脸跟踪方法.计算机学报,1998,21(6):527~532.

Tight-Clothing Based Human Animation Research*

ZHU Qiang, ZHUANG Yue-ting, PAN Yun-he

(Department of Computer Science and Engineering, Zhejiang University, Hangzhou 310027, China);

(Laboratory of Microsoft Visual Perception, Zhejiang University, Hangzhou 310027, China)

E-mail: zhuq2000@yahoo.com; yzhuang@cs.zju.edu.cn

http://www.zju.edu.cn

Abstract: The traditional approaches are suffering from the problems of expensive computation, low efficiency, and lack of vitality. A new approach called tight-clothing and camera calibration based human animation is proposed in this paper. After videotaping the model dressed in tight-clothing, people can track human joint in image sequence automatically, then construct corresponding three-dimension human motion skeleton sequence under the perspective projection, finally the mapping of human motion onto the motion of other roles is completed by making BVH (biovision hierarchy) file. This approach has the characteristics of the rich source material, efficient production, low cost and realistic animation result. It is possible to release the animators from the tedious work. Also such a technique can step into families for its low cost.

Key words: tight-clothing; color-block; clustering; calibration; BVH (biovision hierarchy)

* Received June 14, 2000; accepted October 23, 2000

Supported by the National Natural Science Foundation of China under Grant Nos.69803009, 69733030; the National Research Foundation for the Doctoral Program of Higher Education Ministry of China under Grant No.20010335049; the College Key Teachers Supporting Plan of the Ministry Education of China; the Foundation of the Ministry of Education of China for Excellent Young Teachers