

适用于单目视频的无标记三维人体运动跟踪

邹北骥^{1,2)} 陈 姝¹⁾ 彭小宁^{1,2)} 史 操¹⁾

¹⁾ (中南大学信息科学与工程学院 长沙 410083)

²⁾ (怀化学院计算机科学与技术系 怀化 418008)

(bjzou@vip.163.com)

摘 要 在无标记人体运动跟踪过程中,由于被跟踪目标缺乏明显的特征以及背景复杂而使得跟踪到的人体运动姿态与真实值偏差较大,不能进行长序列视频跟踪.针对这一现象,提出一种基于形变外观模板匹配进行单目视频的三维人体运动跟踪算法,其中所用的人体外观模型由三维人体骨骼模型及二维纸板模型组成.首先根据人体骨骼比例约束采用逆运动学计算出关节旋转欧拉角;然后利用正向运动学求得纸板模型中像素在三维空间中的坐标,将这些像素根据摄像机成像模型投影到二维图像中得到形变外观模板;最后采用直方图匹配得到人体运动跟踪结果.实验结果表明,该算法对于一些复杂的长序列人体运动能够得到较为理想的跟踪结果,可应用于人机交互和动画制作等领域.

关键词 人体运动跟踪;逆运动学;人体骨骼模型;摄像机模型
中图法分类号 TP391

Markerless 3D Human Motion Tracking for Monocular Video Sequences

Zou Beiji^{1,2)} Chen Shu¹⁾ Peng Xiaoning^{1,2)} Shi Cao¹⁾

¹⁾ (School of Information Science and Engineering, Central South University, Changsha 410083)

²⁾ (Department of Computer Science and Technology, Huaihua College, Huaihua 418008)

Abstract In markerless human motion tracking, the reconstructed human motion pose has great difference from the ground-truth value due to the absence of obvious markers and complex background. To overcome this problem, we present an approach to track 3D human motion from uncalibrated monocular video sequences based on deformable appearance template matching where the human appearance model adopted in this research contains 3D human skeleton model and 2D cardboard model. Firstly, the Euler angles of joints are estimated by inverse kinematics based on human skeleton constraint; secondly, the coordinates of pixels in the cardboard model in the scene are determined by forward kinematics, and the region of morphing appearance template in the image is obtained by projecting these pixels in the scene onto the image plane under perspective projection; finally, the human motion can be tracked by histogram matching. The experimental results show that favorable tracking results on a number of long complex human motion sequences can be generated by the method. This approach can be applied to several areas such as human-computer interaction and human animation.

Key words human motion tracking; inverse kinematics; human skeleton model; camera model

收稿日期:2007-12-25;修回日期:2008-03-31. 基金项目:国家自然科学基金(60673093);国家自然科学基金重大研究计划(90715043);湖南省自然科学基金(06JJ2065);长江学者和创新团队发展计划(IRT0661). 邹北骥,男,1961年生,博士,教授,博士生导师,主要研究方向为CAD&CG、数字图像处理等. 陈 姝,男,1976年生,博士研究生,主要研究方向为人体运动跟踪、虚拟现实(csu_cs@163.com). 彭小宁,男,1962年生,博士研究生,副教授,主要研究方向为图像处理、多媒体技术. 史 操,男,1981年生,博士研究生,主要研究方向为人体运动跟踪、模式识别与图像处理.

基于视频的人体运动跟踪是从视频序列中提取运动人体的关节三维坐标,进而恢复其三维运动姿态.由于基于视频的人体运动跟踪能够广泛应用于计算机人体动画、虚拟现实、智能监控等领域,近年来已受到越来越多的研究者重视.

按照其采用方法不同,基于视频的人体运动跟踪算法粗略地分为 2 类:自顶向下方法及自底向上方法^[1].特征跟踪法作为一种自底向上的方法,一般分 2 步顺序完成:首先从视频帧中提取特征,并在各个帧的特征间建立对应关系;然后从特征对应中恢复三维人体结构与运动信息.其中,第一步即特征跟踪部分,已有大量研究人员进行了深入的研究^[2-3].一般利用当前帧的特征位置及运动模型来预测下一帧特征的位置^[4],常用的预测方法包括卡尔曼滤波^[4-5]及粒子滤波^[6-7].陈坚等^[8]采用二阶自回归过程动力学模型对人体运动进行建模,利用粒子滤波来跟踪人体运动.Ning 等^[9]则将人体运动以高斯模型表示,在跟踪过程中通过运动参数等约束来优化跟踪结果,取得了较好的效果.第二步即姿态恢复部分,由跟踪到的特征二维坐标值恢复该特征在三维空间中的坐标,一般是根据成像模型来提取

特征点的三维坐标值.采用比例正交投影模型进行特征三维坐标恢复是目前常用的方法^[10-11],比例因子 s 在该模型中对于重建效果影响极大^[11].这类算法中, s 通过一定的约束条件进行估算,由于估算的结果与真实值存在偏差,使得重建后的效果不够理想.采用透视投影模型虽然能够取得较好的重建效果^[12-13],但需要进行摄像机标定或者满足一定的约束条件,在算法的通用性及适用性方面还有待提高.本文提出一种基于形变外观模板匹配的无标记人体运动跟踪算法,该算法将特征点跟踪与三维姿态恢复这 2 个步骤融为一体,具有约束条件少、无须摄像机标定、应用简单、姿态恢复精确等优点.

1 本文算法概述

本文算法中,关节跟踪采用形变模型匹配和局部搜索方法.按照二维图像的成像过程,模板匹配经历了 3 个过程:图像中待匹配的关节 对应模型(所依附身体段中像素点)在三维空间中的坐标值 利用已求得的成像模型投影到二维图像,如图 1 所示.

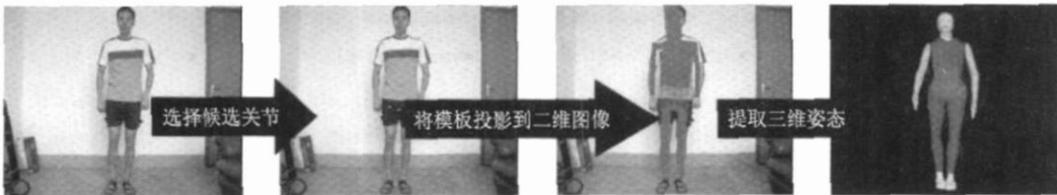


图 1 形变外观模板匹配过程示意图

本文算法的总体流程如图 2 所示,共分为 4 个部分:

1) 执行系统的初始化.用户使用系统提供的操

作界面来标定关节位置,该部分输入为首帧图像,根据用户在图像上标定的关节坐标得到各身体段(模型)的区域信息(像素值)及人体骨骼模型中各身体

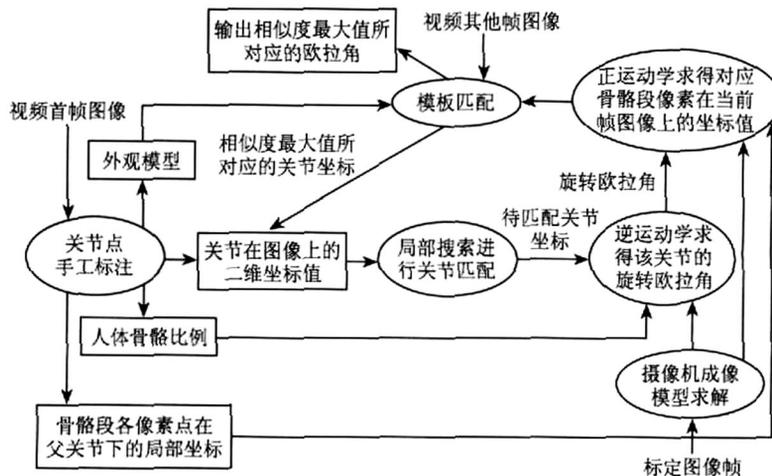


图 2 本文算法总体流程图

段的相对比例尺寸.

2) 按照层次化方式,根据待匹配关节坐标利用逆运动学求得该关节所依附身体段的旋转欧拉角.首先根据身体段比例约束在摄像机成像模型下求得该关节对应的相对三维坐标,然后利用逆运动学求得旋转欧拉角.

3) 根据已求得的关节旋转欧拉角,按照正向运动学求得该关节依附身体段(模型)中像素点旋转后的相对三维坐标值,最后根据摄像机成像模型将这些像素点投影到二维图像,从而得到待匹配关节所依附身体段在二维图像上的位置,即得到形变模型.

4) 将形变模型与初始时的外观模型作直方图匹配,最大相似度对应的关节坐标就是跟踪到的关节位置,对应所求得各关节旋转欧拉角就是恢复出的人体姿态.

2 建模

2.1 人体骨骼模型

我们将人体看成是一种树型棍状模型,如图 3 所示,该骨骼模型由 16 个关节及 15 个身体段组成.其中, J_1 为树型结构的根节点,对应人体的盆骨关节;模型中的线段(人体骨骼段)长度根据人体测量学得到,为一相对比例长度,在应用中应该根据实际测量值设置.

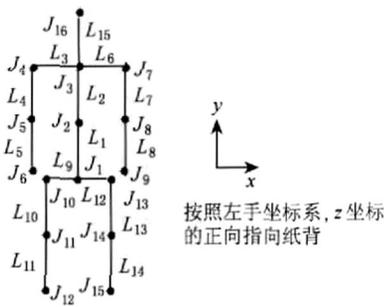


图 3 人体骨骼模型

为了叙述方便,我们在人体骨骼模型上定义了 2 种坐标系:1) 局部坐标系.固结于各个身体段,坐标系的原点为身体段的依附关节,坐标系的方向按照图 3 设置;2) 全局坐标系.坐标系的原点为 J_1 关节,坐标系的方向按照图 3 设置.

2.2 人体外观模型

图 3 所示定义的人体骨骼模型中身体段以线条表示,其长度满足骨骼比例关系.在利用模板匹配进行人体运动跟踪时,由于线条不能很好地表征人体三维运动后身体段在投影面上的变化,因而我们将

人体近似为纸板模型,作为人体运动跟踪中的匹配模板.人体外观模型由形状模型及纹理模型构成,根据人体骨骼模型,人体外观模型共由 15 个矩形块组成,每个矩形块代表人体的一个身体段,矩形块由身体段中的像素组成,包括像素点在矩形块中的相对坐标(形状)及该像素点对应的像素值(纹理).如图 4 所示,各矩形块的中心线即为人体的骨骼模型中的骨骼段,各矩形块边缘的中点即为人体的骨骼模型中的关节,在图中标注了关节 J_1 ,其他关节类似标注.

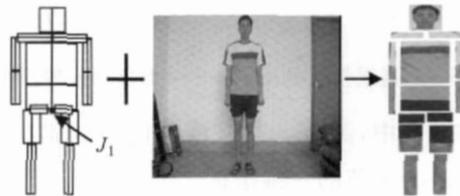


图 4 外观模型示意图

在矩形块中,各像素的坐标根据局部坐标系获得,局部坐标系的原点位于关节上,局部坐标系的方向在第 2.1 节中已说明.

2.3 摄像机成像模型

2.3.1 成像模型

由文献 [10] 可知,三维空间中某点的坐标值 (x, y, z) 与该点在二维投影平面上的坐标 (u, v) 满足方程

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{s} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (1)$$

其中,参数 s 为一比例因子,由 $s = z/f$ 求得, z 为该点在三维空间中的 z 坐标值, f 为摄像机焦距.从式 (1) 可知,当 z 发生变化时, s 值将线性地发生变化, s 的变化量 ds 相对于 z 值的变化量 dz 表示为 $T(dz) = ds$.

2.3.2 确定 dz 与 ds 之间的变换方程

首先,我们从视频序列中找到一帧包含 3 段首尾相连骨骼段的视频帧,其中第一、三骨骼段平行于投影面,一个典型的视频帧如图 5 所示.图 5 中,右肩、右前臂与投影面平行.图 6 所示为 3 段首尾相连骨骼段在透视投影模型下的成像示意图.



图 5 用于标定的图像

在图 6 中,骨骼段 ab 与 cd 平行于投影平面,其在投影平面上的成像分别为 ab 和 cd .骨骼段 bc 不平行于投影平面,其在投影平面上的成像为 $b'c'$.

假设 ab, bc 和 cd 的长度分别为 L_{ab}, L_{bc} 和 L_{cd} , 其值可以从人体骨骼模型中获取. ab, bc 和 cd 的投影长度分别为 L'_{ab}, L'_{bc} 和 L'_{cd} . 以 ab 为例, $L_{ab} = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$, 其中 (a_x, a_y) 和 (b_x, b_y) 分别是关节 a, b 在投影平面上的坐标.

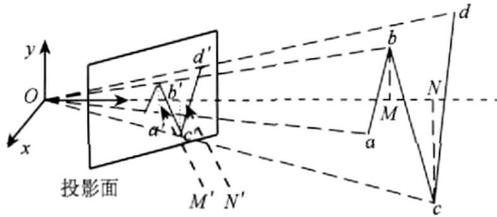


图 6 在透视投影模型下 3 个骨骼段的成像示意图

在图 6 中, 关节 a 在三维空间中的坐标为 (a_x, a_y, a_z) , b, c 关节的三维坐标类似定义.

由于骨骼 ab 平行于投影面, 因而其在 z 轴上的投影相交于点 M , 点 M 对应的可变因子 s 按照 $s_{ab} = OM/f = L_{ab}/L'_{ab}$ 计算得到. 类似地, 我们可以得到骨

$$\begin{aligned} \sqrt{(J_i^x - J_{i+1}^x)^2 + (J_i^y - J_{i+1}^y)^2 + (J_i^z - J_{i+1}^z)^2} &= L_i \Rightarrow \\ \sqrt{(J_i^x - J_{i+1}^x)^2 + (J_i^y - J_{i+1}^y)^2 + (z)^2} &= L_i \Rightarrow \\ \sqrt{[s_{J_i} \cdot J_i^x - (s_{J_i} + s) \cdot J_{i+1}^x]^2 + [s_{J_i} \cdot J_i^y - (s_{J_i} + s) \cdot J_{i+1}^y]^2 + (z)^2} &= L_i \Rightarrow \\ \sqrt{[s_{J_i} \cdot J_i^x - (s_{J_i} + T(z)) \cdot J_{i+1}^x]^2 + [s_{J_i} \cdot J_i^y - (s_{J_i} + T(z)) \cdot J_{i+1}^y]^2 + (z)^2} &= L_i \end{aligned} \quad (3)$$

其中, L_i 是骨骼段 L_i 的相对长度; (J_i^x, J_i^y, J_i^z) 和 $(J_{i+1}^x, J_{i+1}^y, J_{i+1}^z)$ 分别是关节 J_i 和 J_{i+1} 的三维坐标值; (J_i^x, J_i^y) 和 (J_{i+1}^x, J_{i+1}^y) 分别是关节 J_i 和 J_{i+1} 在投影平面上的坐标值; z 是关节 J_{i+1} 和 J_i 之间的相对深度值, $z = J_{i+1}^z - J_i^z$. 式(3)是一个一元二次方程, 其解共有 3 种情况: 1 个实数解, 2 个实数解和虚数解. 在 2 个实数解的情况下, 我们需要从这 2 个解中选择 1 个解作为真实解. 建立如图 7 所示的投影坐标系, 坐标原点为 O , z 轴正向通过图像中心指向运动目标. 根据 z 和 s 的定义, 当 J_{i+1} 离投影平面的距离小于 J_i 时, 我们选择负实数作为 z ; 否则, 将正实数作为 z . 图 7 所示为投影射线与球面相交的情形, 在这种情况下, z 是负实数值.

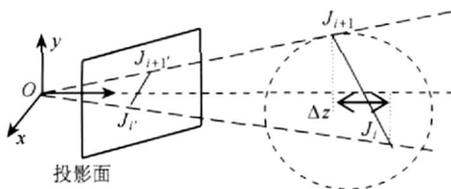


图 7 身体段成像示意图

骼 cd 在 z 轴上的投影点 N 对应的可变因子 s , $s_{cd} = L_{cd}/L'_{cd}$. 由于骨骼 ab 与骨骼 cd 平行于投影面, 因而点 M 与点 N 的距离满足 $d_z = c_z - b_z$.

根据空间几何知识, L_{bc} 满足

$$\begin{aligned} \sqrt{(c_x - b_x)^2 + (c_y - b_y)^2 + (c_z - b_z)^2} &= L_{bc} \Rightarrow \\ \sqrt{(c_x - b_x)^2 + (c_y - b_y)^2 + d_z^2} &= L_{bc} \Rightarrow \\ \sqrt{(s_{cd}c_x - s_{ab}b_x)^2 + (s_{cd}c_y - s_{ab}b_y)^2 + d_z^2} &= L_{bc} \end{aligned} \quad (2)$$

由于 $s_{cd}, s_{ab}, c_x, c_y, b_x, b_y$ 和 L_{bc} 已知, 因而 d_z 通过式(2)可以计算得到. 相对于 z 的绝对变化量 $(|d_z|)$, 其对应的 s 绝对变化量 $|ds| = abs(s_{cd} - s_{ab})$.

2.3.3 人体关节相对三维坐标值估算

设已知某关节 J_i 对应的 s 值为 s_{J_i} , 下面介绍如何计算其子关节 J_{i+1} 在摄像机成像空间下的三维坐标值.

根据骨骼长度不变性, 可以得到

由于可以通过式(3)计算出 z 的值, 因而关节 J_{i+1} 对应的可变因子 $s_{J_{i+1}} = s_{J_i} + T^{-1}(z)$. J_i 和 J_{i+1} 在投影空间中的三维坐标 计算公式为

$$\begin{aligned} J_i^x &= s_{J_i} \cdot J_i^x, \\ J_i^y &= s_{J_i} \cdot J_i^y, \\ J_{i+1}^x &= s_{J_{i+1}} \cdot J_{i+1}^x, \\ J_{i+1}^y &= s_{J_{i+1}} \cdot J_{i+1}^y. \end{aligned}$$

3 系统初始化

系统初始化包括 3 个部分: 人体骨骼模型中身体段之间相对比例, 外观模型初始化以及根关节对应的可变因子 s 值. 系统提供了图形化的用户界面,



图 8 标注后的视频首帧

可以对首帧中人体关节进行手工标注. 一个标注后的图像如图 8 所示, 其中, 圆点为标注的关节, 矩形框为标注的身体段. 用作标注的图像中人体姿态做如下约束: 身体面向摄像机垂直站立, 双手垂直与身体平行.

3.1 计算人体骨骼模型中骨骼段相对长度

虽然根据人体测量得到的骨骼段相对长度比较精确, 但在对模型初始化时标注的关节坐标与真实的关节坐标不一定完全一致, 则标注出来的骨骼比例与人体测量得到的骨骼比例有偏差, 因而需要重新计算人体骨骼比例. 我们以已标注图像上的 L_5 骨骼段作为参考来求取其他骨骼段的相对长度.

设 L_5 骨骼段在人体骨骼模型中的相对长度为 L_5 , 其值由测量真实人体长度得到, L_5 骨骼段在投影图像上的长度为 L_5 , $L_5 = \sqrt{(J_5^x - J_6^x)^2 + (J_5^y - J_6^y)^2}$, 其中 (J_5^x, J_5^y) 和 (J_6^x, J_6^y) 分别为关节 J_5 和 J_6 在投影图像上的坐标. 由于整个身体平行于投影面, 故该帧中人体所有关节对应的 s 值都相同, $s = L_5/L_5$.

设其他身体段在图像上的长度为 L_i , 则这些身体段在人体骨骼模型中的相对长度按 $L_i = s \cdot L_i$ 计算.

3.2 外观模型初始化

外观模型由包含空间信息的身体段像素组成, 身体段像素值在图像上标注出身体段(矩形框)后则可以获得, 像素点的空间信息则以身体段局部坐标系下的坐标表征. 设根据某像素在图像上的坐标求得的其在局部坐标系下的二维坐标为 (x, y) , 则在摄像机成像空间下该像素点在其局部坐标系中的三维坐标为 $(s \cdot x, s \cdot y, 0)$.

4 利用形变外观模板匹配进行人体运动跟踪

我们采用局部搜索及模板匹配的方法进行人体运动跟踪, 首先利用粒子滤波对下一帧关节的坐标进行预测, 将预测结果作为关节搜索区域的中心点. 整个跟踪分为 3 个步骤:

Step1. 由候选关节坐标计算出在此坐标下关节的旋转欧拉角.

Step2. 由计算出的旋转欧拉角, 根据正向运动学计算出外观模型旋转后的三维坐标值.

Step3. 将外观模型旋转后的三维坐标值投影到二维成像面得到形变后的外观模型区域, 将形变后的外观模型与对应图像上的投影面进行匹配.

我们参照图 3 根据树型深度遍历, 依次恢复人体的三维运动姿态及关节在图像上的坐标位置, 即先确定父关节的坐标, 再采用局部搜索确定子关节的坐标, 故需先确定根关节 J_1 在图像中的坐标.

由于以根关节 J_1 为中心的某一邻近区域在人体运动过程中基本上不发生形变, 因而以根关节 J_1 为中心的一个矩形作为模板, 在视频各帧中采用局部搜索并根据直方图匹配来确定根关节的坐标.

本文算法假定待恢复视频中的人体关节 J_1 只作水平位移, 不作前后位移, 故在第 3.1 节中计算出的 s 值可以作为整个视频序列帧中根关节对应的 s 值.

4.1 计算候选关节坐标下的关节旋转欧拉角

设已知搜索区间中某一候选关节的坐标, 该候选关节的父关节对应的 s 值及其在图像上的坐标已求得, 根据第 2.3.3 节可以计算出该候选关节与其父关节之间的深度差 z 及它们在三维空间中的 x, y 坐标.

由图 3 可得各关节在其依附局部坐标系中的向量值: J_2 在 J_1 局部坐标系中的向量为 $(0, L_1, 0)$, J_3 在 J_2 局部坐标系中的向量为 $(0, L_2, 0)$, J_4 在 J_3 局部坐标系中的向量为 $(-L_3, 0, 0)$, 类似可以得到其他关节在其依附局部坐标系中的向量值.

关节的旋转方向为先绕 z 轴旋转 α 角, 再绕 y 轴旋转 β 角, 最后绕 x 轴旋转 γ 角. 基坐标系(参考坐标系)原点固结于关节 J_1 , 其方向始终与人体初始状态下关节 J_1 局部坐标系的方向一致.

1) 求根关节 J_1 的旋转矩阵

由于三维数据没有考虑扭矩的作用, 可以认为在 y 轴上的旋转分量为 0, 对应的旋转矩阵为单位矩阵, 由正向运动学可得 J_1 坐标系相对于基坐标系的旋转矩阵为

$${}^0_1R = \begin{bmatrix} \cos & -\sin & 0 \\ \sin & \cos & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos & -\sin \\ 0 & \sin & \cos \end{bmatrix} = \begin{bmatrix} \cos & -\sin & \cos & \sin & \sin \\ \sin & \cos & \cos & -\cos & \sin \\ 0 & \sin & \cos & \end{bmatrix} \quad (4)$$

建立 J_2 关节局部坐标与世界坐标之间的变换方程为

$$\begin{bmatrix} J_2^x - J_1^x \\ J_2^y - J_1^y \\ J_2^z - J_1^z \end{bmatrix} = {}^0_1R \begin{bmatrix} 0 \\ L_1 \\ 0 \end{bmatrix} \quad (5)$$

其中,向量 $[0, L_1, 0]$ 为 J_2 在 J_1 局部坐标系中的坐标, $[J_2^x - J_1^x, J_2^y - J_1^y, J_2^z - J_1^z]$ 为 J_2 在基坐标系中的坐标. 根据式(5)可得方程组为

$$\begin{cases} \sin \cos = (J_2^x - J_1^x)/L_1 \\ \cos \cos = (J_2^y - J_1^y)/L_1, \\ \sin = (J_2^z - J_1^z)/L_1 \end{cases}$$

解之得 $\alpha = \arcsin\left(\frac{z}{L_1}\right), \beta = -\arctan\left(\frac{J_2^x - J_1^x}{J_2^y - J_1^y}\right)$.

2) 求第 i 个关节 J_i 的旋转矩阵

设已求得 J_{i-1} 坐标系相对基坐标系的旋转矩阵 ${}_{i-1}^0R = {}_1^0R_1^1R_2^2R \dots {}_{i-1}^{i-2}R$; 其中 0R 等矩阵为 J_1 到 J_{i-1} 链通路中各关节局部坐标系之间的旋转矩阵(只是坐标系的姿态变换,不涉及位置变换).

关节 J_{i+1} 其坐标变换方程为 $P = {}_{i-1}^0R {}^{i-1}R P_{i+1}$; 其中, P_{i+1} 为初始状态下关节 J_{i+1} 在 J_i 局部坐标系中的向量值, P 为当前状态下关节 J_{i+1} 在 J_i 局部坐标系中的向量值, ${}^{i-1}R$ 为待求的旋转矩阵, 其形式与求 0R 过程中 0R 定义类似. 如果 J_i 关节为 J_4, J_7 , 由于 x 轴上的旋转分量为 0, 则 ${}^{i-1}R$ 定义为

$${}^{i-1}R = \begin{bmatrix} \cos & -\sin & 0 \\ \sin & \cos & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos & 0 & \sin \\ 0 & 1 & 0 \\ -\sin & 0 & \cos \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$P = \begin{bmatrix} J_{i+1}^x - J_i^x \\ J_{i+1}^y - J_i^y \\ J_{i+1}^z - J_i^z \end{bmatrix}.$$

4.2 根据正向运动学求外观模型旋转后在图像上的投影区域

设由第 4.1 节求得的旋转角度为 (α, β) , 父关节在摄像机成像三维空间下的 x, y 坐标为 (J_{i-1}^x, J_{i-1}^y) , 对应的 s 值为 $s_{J_{i-1}}$, 则该身体段中像素的三维坐标值经过旋转后的值计算过程如下:

设模型中的像素初始时(身体段未旋转时)在局部坐标系下的坐标为 $(x_l, y_l, 0)$, 其值求解详见第 3.2 节, 则旋转后其在局部坐标系下的坐标为

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} \cos & -\sin & 0 \\ \sin & \cos & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos & 0 & \sin \\ 0 & 1 & 0 \\ -\sin & 0 & \cos \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos & -\sin \\ 0 & \sin & \cos \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ 0 \end{bmatrix};$$

模型中的像素在摄像机成像空间下的 x, y 坐标为

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} J_{i-1}^x \\ J_{i-1}^y \end{bmatrix} + \begin{bmatrix} x_r \\ y_r \end{bmatrix}.$$

模型中的像素旋转后对应的 s 值为 $s_p = s_{J_{i-1}} + T(z_r)$. 因而模型中的像素在二维图像上的坐标为 $(x_p/s_p, y_p/s_p)$.

4.3 模板匹配

由第 4.2 节可以得到外观模型在此候选关节下其在二维图像上的投影区域, 采用直方图匹配可以得到该候选关节对应的相似度, 直方图采用 Bhattacharyya 距离作为匹配准则. 相似度最大的候选关节则为跟踪到的关节位置, 对应的旋转欧拉角就是恢复出来的人体运动姿态.

按照第 4.2 节的方法可以求得此旋转角度下该关节在三维空间中对应的 x, y 和 s 值, 作为求其子关节位置时的输入值.

5 实验及结果分析

为了验证本文算法的有效性, 我们采用 VC++ 6.0 和 OpenGL 实现了一个原型系统. 系统测试视频为 2 段用三星 S600 数码相机在室内及室外环境下拍摄的人体运动视频共 300 帧, 相机拍摄速率为 30 帧/s, 图像分辨率为 640 × 480. 在拍摄过程中, 人体根关节平行于摄像面作水平移动, 在 z 坐标方向上不发生位移.

用作实验的视频未做任何预处理, 人体运动跟踪在原始图像上进行. 首先系统通过手工标注获得人体外观模型, 然后进行自动跟踪恢复人体三维运动姿态. 人体骨骼比例长度由手工测量得到, 如图 5 所示, 在实验中我们选择右肩、右上臂、右前臂作为 3 个首尾相连的骨骼段来估计一个单位的 z 变化对应的 s 变化量, 右肩对应骨骼段 ab , 右前臂对应骨骼段 cd .

图 9, 10 所示分别为室内和室外环境下跟踪的人体运动姿态, 图中显示的是从第 1 帧开始每隔 30 帧的结果图像. 其中, 图 9 b, 10 b 所示为跟踪到的二维关节坐标结果, 图 9 c, 10 c 所示为重建后的三维人体运动姿态正视图, 图 9 d, 10 d 所示为重建后的三维人体运动姿态旋转 45 后的侧视图.

在图 9 中, 人体着装简单, 其中上衣部分颜色与背景非常相似, 人体做体操动作. 这种情况下如果采用二维平面上的形变块模板匹配, 会造成跟踪结果与真实值偏差较大, 这可以从图 11 中得到相应的结果. 由于本文算法不仅考虑了二维平面上的形变, 而且兼顾了 z 方向上的模型形变, 因而取得了较好的跟踪结果.

在图 10 中,人体做剧烈的全身运动,手臂和脚部姿态都出现了大范围的变化,其中有些姿态非常具有挑战性,并且出现了手臂与上半身自遮挡现象.由于在本文算法中,相重叠的两躯干根据其在 z 坐

标上的位置,在将模型投影到二维平面上做模板匹配时,可以将被遮挡部分排除在外,因而在自遮挡发生时,依然可以取得较好的跟踪效果,能够较好地解决自遮挡造成的跟踪困难.

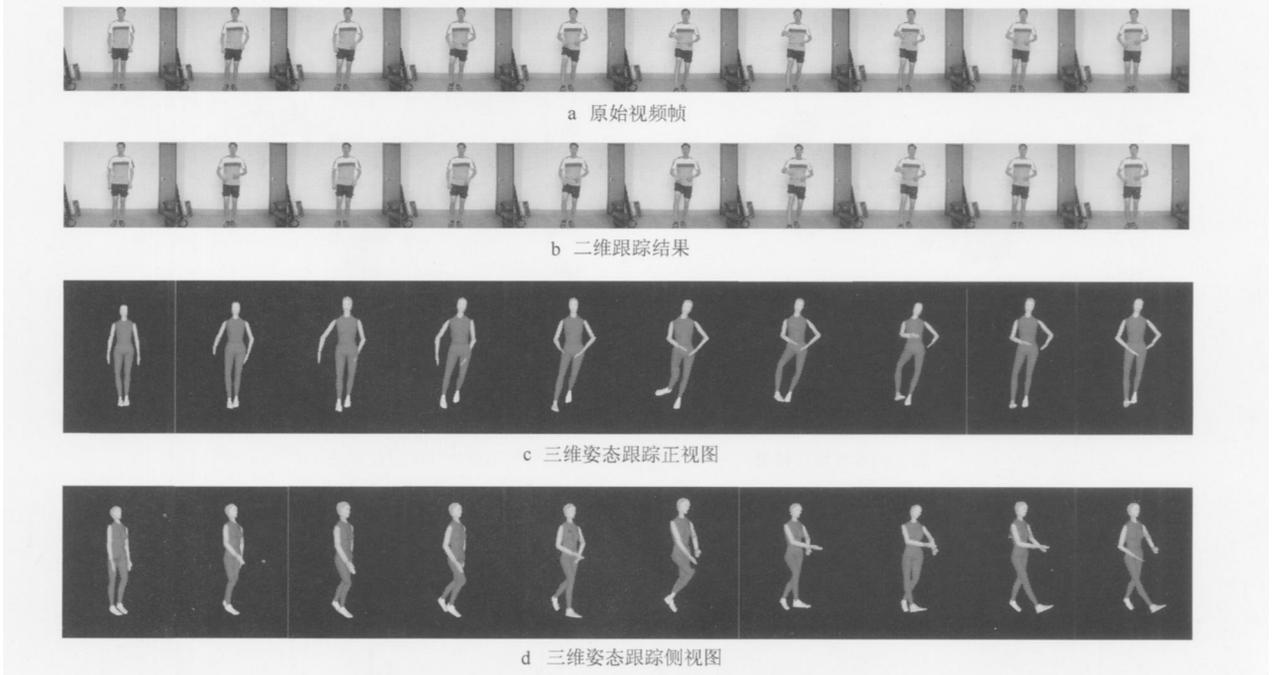


图 9 室内视频人体运动三维姿态跟踪结果

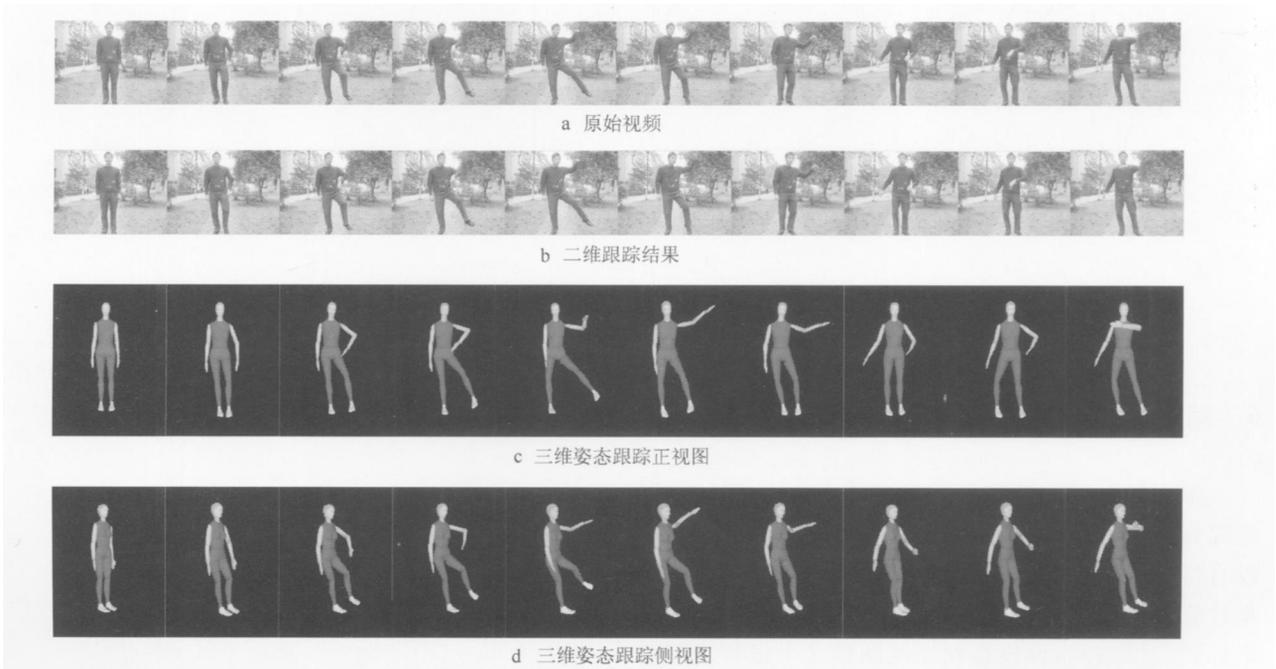


图 10 室外视频人体运动三维姿态跟踪结果

图 11 所示为整个室内视频跟踪过程中本文算法在二维图像上的跟踪结果,以及基于二维平面形变块匹配算法的跟踪结果与真实值之间的比较.在二维平

面形变块匹配跟踪算法中,采用 Condensation 算法对特征进行预测^[14],粒子数为 200.由图 11 可知,本文算法对关节的跟踪误差大致限制在 ± 20 之内,

而二维平面形变块匹配跟踪算法对关节的跟踪误差则变化较大. 图 12 所示为右大腿及右小腿的旋转欧拉角运动轨迹, 可以看出, 整个运动轨迹比较平滑, 没有明显的跳跃点. 虽然由于光照变化及 z 正负

号选取错误使得某些帧中计算的欧拉角发生一定程度的偏差, 但对于后续帧中的人体姿态恢复不会造成影响, 也就是说不会产生积累误差. 因而, 本文算法具有自恢复功能, 鲁棒性较好.

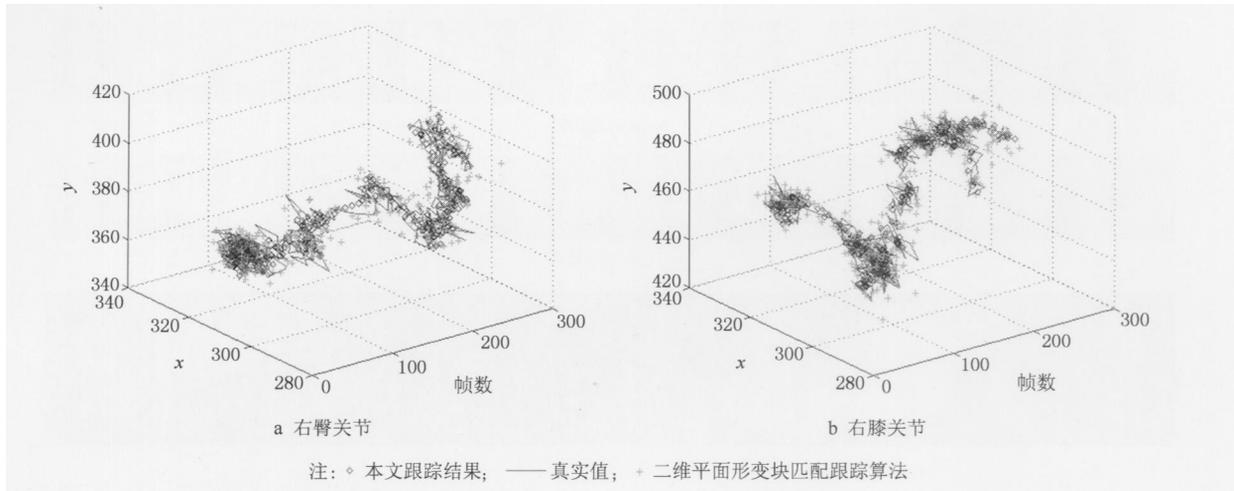


图 11 二维跟踪数据与真实值比较结果图

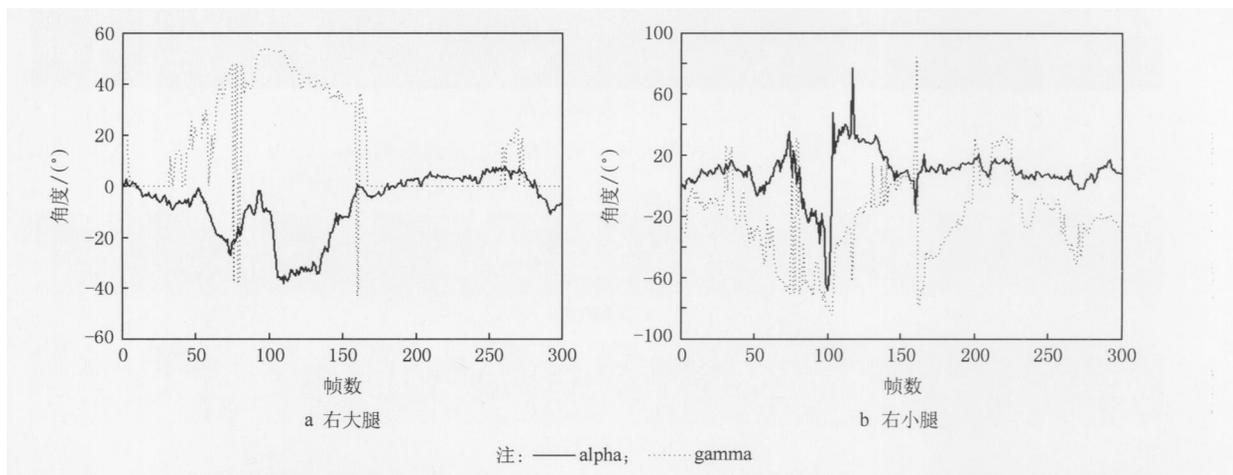


图 12 室内视频中关节运动轨迹图

6 结 论

本文提出了一种基于单目视频的人体三维运动跟踪算法, 具有使用方便、无需摄像机标定、对背景没有特殊要求等特点. 该算法按照图像的成像过程来计算形变外观模型, 整个匹配过程经历了从二维坐标到三维空间, 再投影到二维图像 2 个过程. 实验结果表明: 本文算法能够较精确地跟踪人体运动, 可以应用于人机交互和动画制作等领域.

本文算法假定待跟踪视频中的人体只沿成像面水平运动而不进行前后运动, 我们今后将研究如何提高算法的通用性, 以跟踪各种运动方向下的人体

运动; 还将研究首帧关节自动标注方法, 这对于算法的应用推广至关重要.

参 考 文 献

- [1] Poppe R. Vision-based human motion analysis: an overview [J]. Computer Vision and Image Understanding, 2007, 108 (1): 4-18
- [2] Yilmaz A, Javed O, Shah M. Object tracking: a survey [J]. ACM Computing Surveys, 2006, 38(4): Article No. 13
- [3] Wang L, Hu W M, Tan T N. Recent developments in human motion analysis [J]. Pattern Recognition, 2003, 36 (3): 585-601

- [4] Bregler C, Malik J. Tracking people with twists and exponential maps [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, 1998: 8-15
- [5] Gavrilu D M, Davis L S. 3D model-based tracking of humans in action: a multi-view approach [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 1996: 73-80
- [6] Sminchisescu C, Triggs B. Kinematic jump processes for monocular 3D human tracking [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Madison, 2003: 69-77
- [7] Toyama K, Blake A. Probabilistic tracking with exemplars in a metric space [J]. International Journal of Computer Vision, 2002, 48(1): 9-19
- [8] Chen Jian, Wang Wencheng, Wu Enhua. Markerless human motion tracking from monocular videos [J]. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(9): 2033-2039 (in Chinese)
(陈 坚, 王文成, 吴恩华. 单目视频中无标记的人体运动跟踪[J]. 计算机辅助设计与图形学学报, 2005, 17(9): 2033-2039)
- [9] Ning H Z, Tan T N, Wang L, *et al.* People tracking based on motion model and motion constraints with automatic initialization [J]. Pattern Recognition, 2004, 37(7): 1423-1440
- [10] Taylor C J T. Reconstruction of articulated objects from point correspondences in a single uncalibrated image [J]. Computer Vision and Image Understanding, 2000, 80(8): 349-363
- [11] Chen Jian, Wu Enhua. 3D human motion reconstruction from monocular videos through iterative optimization [J]. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(7): 1523-1528 (in Chinese)
(陈 坚, 吴恩华. 单目视频中人体三维运动的迭代优化估计[J]. 计算机辅助设计与图形学学报, 2005, 17(7): 1523-1528)
- [12] Luo Zhongxiang, Zhuang Yueting, Liu Feng, *et al.* Video based human body animation [J]. Journal of Computer Research and Development, 2003, 40(2): 269-276 (in Chinese)
(罗忠祥, 庄越挺, 刘 丰, 等. 基于视频的人体动画[J]. 计算机研究与发展, 2003, 40(2): 269-276)
- [13] Zhao J H, Li L, Keong K C. 3D posture reconstruction and human animation from 2D feature points [J]. Computer Graphics Forum, 2005, 24(4): 759-771
- [14] Isard M, Blake A. Condensation—conditional density propagation for visual tracking [J]. International Journal of Computer Vision, 1998, 29(1): 5-28