

¹运动图象序列的人体三维运动骨架重建

庄越挺 刘小明 潘云鹤 杨骏

(浙江大学人工智能研究所 杭州 310027)

摘要 在计算机视觉领域,人体运动分析的研究具有相当广泛的应用前景。由于人体运动的复杂性,已有的研究方法都对研究对象—人体,加上了许多限制条件。本文中提出了一种新的方法来获得各类人体运动的信息。本文主要论述了该方法中对人体三维运动骨架的重建部分,其基本思想是在摄像机定标的基础上,应用三维人体模型知识和运动连续性依次建立每个图象的人体运动骨架。最后给出了实验结果。

关键词 图象序列 运动骨架 摄像机定标

中图法分类号 TP391.4

Human Three Dimension Motion Skeleton Reconstruction of Motion Image Sequence

ZHUANG Yue-Ting LIU Xiao-Ming PAN Yun-He YANG Jun

(Institute of AI, Zhejiang University, Hangzhou 310027)

Abstract In the domain of computer vision, there is a very wide application for the research of human motion analysis. For the complexity of human motion, the existing research methods lay much limitation on human. This paper proposes a new approach to obtain the information of human motion. It focuses on the human three dimension motion skeleton reconstruction. The main scheme is to establish the human motion skeleton of every image in turn using three dimension human model knowledge and motion continuity on the basis of camera calibration. The final result is presented at the end of the paper.

Keywords Image sequence, Motion skeleton, Camera calibration

1.引言

人体运动分析,就是从图象序列或视频中获取人体的运动信息,并对它进行分析,识别等工作。由于这一研究具有相当广泛的应用前景,它正越来越受到研究者的重视。如在图象中分割出人体的各个部分,并在图象序列中跟踪关节的运动,以至恢复相应的三维人体结构,在运动员的动作分析,医学分析,动画制作等方面非常有用。另外,自动化地识别,理解人体运动能够广泛地应用于监视,人机交互,视频检索等领域。

传统的人体运动分析方法一般分为两种。一是在人体的各个关节部位分别加上传感器。在人体运动过程中,传感器会不断地将各关节在空间中的位置返回给计算机,这样计算机就可以精确获得人体在各个时刻的运动信息。二是针对图象序列进行分析。图象序列可以是单个相机拍下的,也可以是多个相机同时从各个视角拍下的。这一般分三步顺序完成。首先从图象帧中提取低层特征,如各关节的位置,并对之进行识别;然后是在各个帧的特征间建立对应关系;最后是从特征对应中恢复三维人体结构与运动信息。在这方面, O'Rourke 与

¹ 本文得到国家自然科学基金资助。庄越挺,男,1965年6月生,博士,副教授,主要研究方向为多媒体数据库,智能CAD。刘小明,男,1975年3月生,硕士研究生,主要研究方向为多媒体技术,视频检索。潘云鹤,男,1946年11月生,教授,博士生导师,院士,主要研究方向为计算机美术,形象思维,智能CAD系统,GIS,计算机动画,多媒体技术等。杨骏,男,1978年9月生,本科生,主要研究方向为计算机视觉。

Badler[1]采用将图象与一个方块模型进行映射的方法进行研究。Hogg[2]，Rohr[3]的系统仅针对自由度为一的人体行走模型。他们从图象中提取了边特征与角特征，并与一个三维圆柱人体模型进行匹配。Chen 与 Lee[4]用 17 条线段和 14 个关节组成了一个骨架模型来表示人体，并在分析过程中加入了多种约束。Bharatkumar et al[5]也用骨架模型描述人体下肢，但他们试图建立一个通用模型，以用于人体行走的步态分析。由于人体运动的特殊复杂性，几乎所有这些方法都加上了许多限制[6]。如要求背景单一甚至不动，人体运动方向与投影平面平行，人体穿紧身衣裤等；另外，加上传感器后人体运动不可避免地变得不自然。

我们提出了一种新方法来获取人体运动信息。该方法分为两个部分。第一部分是二维图象序列的特征点跟踪[7]：给定一个包含被分析人体的图象帧序列，用户先在第一帧上标注人体各个关节点，然后系统跟踪出后续帧中人体各个关节点的位置。这样就获得了人体经投影后的二维运动骨架序列。第二部分是三维人体运动骨架重建：针对这一骨架序列，我们应用计算机视觉的方法进行摄像机定标，最后恢复了相应的透视投影下三维人体运动骨架序列。也即在三维空间中，人体各关节点在各个时刻的空间坐标。这样，该人体模型就能用于人体动画，虚拟现实等各类需建立三维场景的应用中。该过程的整体框架如图 1 所示：

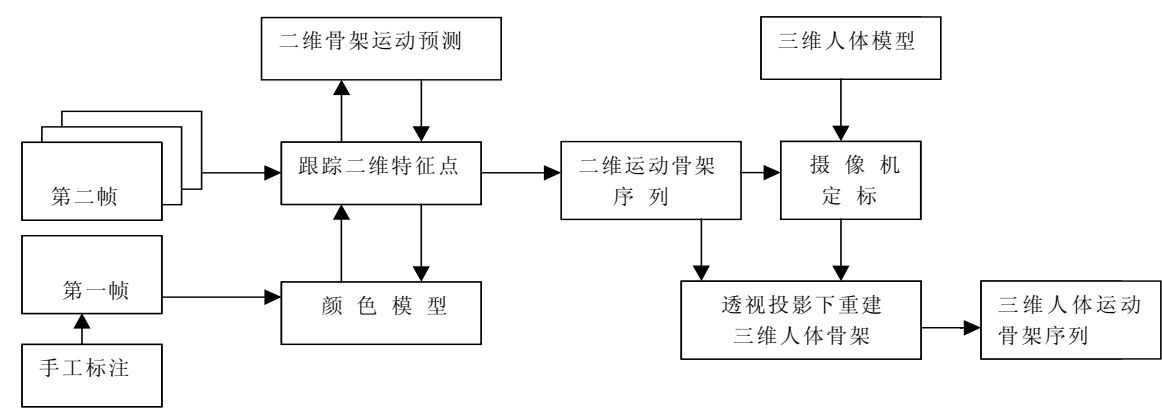


图 1. 系统整体框架

本文将专门介绍第二部分的具体实现。第二节将介绍本文所用的三维人体骨架模型，第三节介绍三维人体运动骨架重建的具体算法，第四节给出实验的结果，最后对文章进行总结并探讨未来的工作方向。

2. 三维人体骨架模型

我们将人体看成一个由关节点连接的刚体的集合[4]。如上肢是由肘关节连接的上下臂两个刚体组成的，上臂与躯干是由肩关节连接的。我们以一条线段来表示一个刚体，将人体运动简化为人体骨架的运动，这样就得到了一个三维人体骨架模型。如图 2 所示。该人体模型共包含 16 个人体的关节点，本文中称它们为三维特征点。

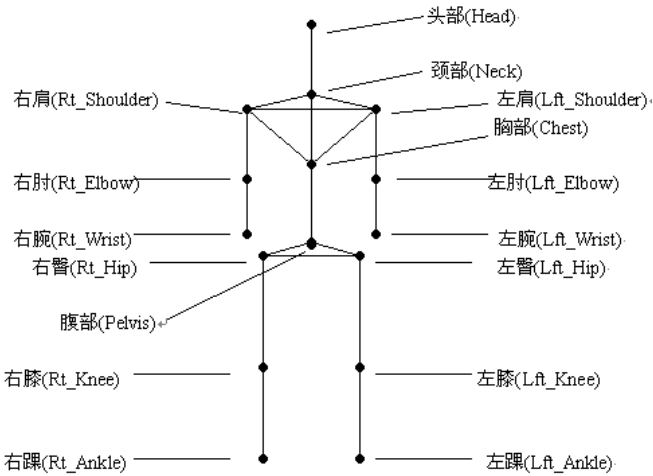


图 2. 人体三维骨架模型

由于特征点之间的线段是代表刚体，因此它们在任何时刻都不会形变。同时我们假设各

个线段（即人体骨架）之间的长短比例关系并不因人而异，这一比例可以从解剖学中获得。本文把三维特征点在投影平面中的投影点称为二维特征点。在前面的工作——二维图象序列的特征点跟踪[7]中，已经得到了每帧中人体各个二维特征点的位置。下面我们就开始讨论在此前提下，如何运用计算机视觉的方法来恢复相应的三维人体骨架的运动信息。

3. 三维人体运动骨架重建的实现

为了在透视投影下重建三维人体运动骨架，我们必须先获得摄像机参数。通过三维模型和二维图象上的线对应关系，本文采用牛顿方法来进行摄像机定标。然后，利用透视投影关系和人体骨骼的比例关系，我们求解了人体模型中各个三维特征点的坐标。在帧序列中，我们利用运动连续性来有效地消除了三维运动信息的二义性。

3.1 基于线性模型的摄像机定标

我们知道，世界坐标系中的任何一点经过以下两个变换都可以转化为投影平面上的点 $P(u, v)$ [8]: 首先是世界坐标系中的点 $P_w(X_w, Y_w, Z_w)$ 通过旋转和平移变换得到该点在摄像机坐标下的坐标 $P_c(X_c, Y_c, Z_c)$ ，用矩阵表示如(1)。其中 R 为旋转变换矩阵， t 为三维平移矢量。其次是通过透视投影将摄像机坐标下的点 $P_c(X_c, Y_c, Z_c)$ 投影为投影平面上的点 $P(u, v)$ ，如(2)所示，其中 f 为焦距。这两次变换如图 3:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} - t \quad (1)$$

$$(u, v) = \left(f \cdot X_c / Z_c, f \cdot Y_c / Z_c \right) \quad (2)$$

我们要解决的问题是：给定一些三维人体模型上的特征对象和与其对应的二维图象上的特征对象，求得投影参数 R ， t 以及 f 。在我们的方法中， f 为已知。在上两式中，计算 u, v 相对于未知参数的偏导数较困难，因此将(1)和(2)进行变化得到如下形式：

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \quad (3)$$

$$(u, v) = \left(\frac{f \cdot X'}{Z' + Dz} + Dx, \frac{f \cdot Y'}{Z' + Dz} + Dy \right) \quad (4)$$

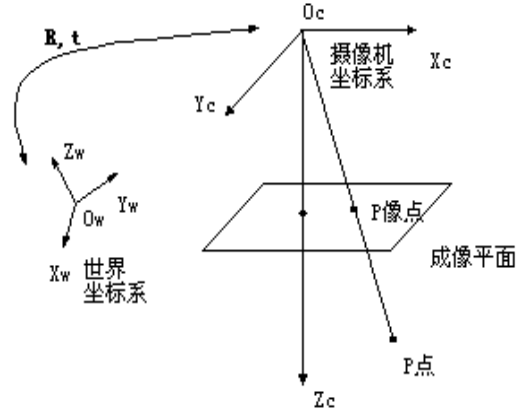


图 3. 投影变换示意图

此两式中参数 R 和 f 的意义与式(1)和式(2)相同，而将平移矢量 t 用参数 Dx, Dy, Dz 代替，其中 $P' = (X', Y', Z')$ 为三维点。当下式成立时上述两种表示方法等同：

$$t = R^{-1} \cdot [-Dx(Z' + Dz)/f, -Dy(Z' + Dz)/f, -Dz]^T \quad (5)$$

本文用旋转向量表示法来表示旋转参数 R 。即由一个三维向量 $r = (W_x, W_y, W_z)^T$ 来定义一个旋转，它的方向就是旋转轴的方向，它的模等于旋转角。至此，投影参数可以用矢量 $(Dx, Dy, Dz, W_x, W_y, W_z)$ 来表示，可以方便地求得 u 和 v 对于 $Dx, Dy, Dz, W_x, W_y, W_z$ 这六个投影参数的偏导数。在我们计算投影参数时，用于对应的特征对象是图像与三维模型之间的对应

线而不是对应点。我们将通过点(u, v)的直线方程用如下形式表示:

$$\frac{-m}{\sqrt{m^2+1}}u + \frac{1}{\sqrt{m^2+1}}v = d \quad (6)$$

其中 d 为原点到该直线的垂直距离, m 为该直线的斜率。由该式可得 d 对 u, v 的偏导数, 再结合前面的计算, 可以得到 d 对于 Dx, Dy, Dz, Wx, Wy, Wz 的偏导数。这样, 我们就可以使用牛顿迭代法来计算投影参数。具体的做法为: 首先从一组投影参数 (Dx, Dy, Dz, Wx, Wy, Wz) 的初值出发(f 作为已知值设为 1), 按目前的参数估计将三维模型上的线段投影到二维投影平面上, 然后计算投影直线与实际的二维特征线段之间的误差偏离, 于是有以下的方程:

$$\frac{\partial d}{\partial Dx} \Delta Dx + \frac{\partial d}{\partial Dy} \Delta Dy + \frac{\partial d}{\partial Dz} \Delta Dz + \frac{\partial d}{\partial Wx} \Delta Wx + \frac{\partial d}{\partial Wy} \Delta Wy + \frac{\partial d}{\partial Wz} \Delta Wz = Ed \quad (7)$$

方程中的 Ed 为二维特征线段的端点到投影直线的垂直距离。由于一条线段有两个端点, 因此对于任何一对对应线, 都能列出两个象 (7) 这样的方程。所以只要有三对不共线的对应线就可以有六个方程组成线性方程组, 求解该方程组得到修正矢量 $h=[\Delta Dx, \Delta Dy, \Delta Dz, \Delta Wx, \Delta Wy, \Delta Wz]$, 然后将 h 的各个分量分别与当前的投影参数相加。修正投影参数后, 再重新计算线性方程组。这样经过有限次的迭代, 方程组中所有的 Ed 值都将小于预设的一个阈值。这样就求得了六个投影参数。

应用牛顿方法至少需要三对对应线。在三维模型中, 我们选择了人体左右肩部之间的一条线段, 左右肩部与胸部之间的两条线段。这三条线段组成了人体上身的一个稳固的等边三角形。这种选择是基于在人体运动中该三角形最不易发生形变的考虑。对于该三角形的各个特征对象, 本文在以后的表述中分别称为关键节点, 关键线段与关键三角形。

由于首帧的二维特征点是由用户标注的, 因此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位置是由系统来指定的。因为只要关键三角形各个边的比例符合人体解剖学, 那么我们总是能在世界坐标系中确定摄像机的位置和方向, 使得关键三角形的透视投影与二维图象帧上躯干上部三角形重合。由于人体其他部分的骨架长度是通过关键线段的长度按比例得到的, 所以关键三角形的大小表示了整个人体模型的大小。事实上, 我们也可以使用多于三对的对应线段。这样就能用最小二乘法减少定标的误差。

3.2 首帧图像对应三维人体模型的建立

从 3.1 的工作中, 已经求得了六个投影参数。但此时除了三个关键节点之外, 人体模型的三维特征点中都不可知, 下一步就是从已知的一个二维特征点坐标 (U, V) 出发, 求出与该点对应的人体模型上三维特征点的空间坐标 $P_c(X_c, Y_c, Z_c)$ 。但从透视投影的针孔模型可以看到, 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长度。

假设右肩部点 P_c' 已知, 要求出未知的右肘部点 P_c , 以它为例来说明本文的算法。设 $P_c(X_c, Y_c, Z_c)$ 为右肘部点在摄像机坐标系中的坐标, (u, v) 为 P_c 在投影平面中的坐标。则有如下关系:

$$d(P_c', P_c) = Len \quad (8)$$

其中 $d(P_c', P_c)$ 表示求点 P_c' 与 P_c 之间的距离。Len 为人体模型中上臂的长度。结合上式与 (2), 可以得到仅含有未知量 Z_c 的方程。该方程可以形象地表示为求投影直线与以 P_c' 为球心, Len 为半径的球面的交点。根据空间直线与球面可能出现的相交, 相切, 相离三种可能, 此方程的解也有下列三种可能情况:

1) 有两个解, 即对应于直线与球面相交, 有两个交点的情况。这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。本文通过两种方法来消除二义性。首先是利用人体解剖学中的各种人体约束。如当上臂前伸时, 小臂不可能向后伸等。其次是利用颜色信息。通过本文的方法解出的两个点总是一个离摄像机光心较近, 一个较远。本文假设离摄像机光心较近的图象区域亮度较高, 反之则较低。因此, 先在肩部与肘部分别以相应特征点为中心选择一小块区域。然后将两个区域中点的 RGB 颜色模型转化为 HLS 颜色模型, 分别计算出两区域中点的 L 分量的均值。最后, 如果肘部区域的 L 分量均值大于肩部区域的相应值, 则选择离摄像机光心较近的点为最终解, 反之则选择离摄像机光心较远的点。实验表明, 结合这两种方法能有效地消除首帧骨架的二义性。但当有些情况下, 由于光照等因素而使亮度假设不成立时, 我们也提供了用户辅助的手段来消除二义性。

2) 只有一个解, 对应于直线和球面相切的情况。此时该切点就是所求的三维右肘部节点。

3) 无解, 对应于直线和球面相离的情况。产生的原因有两个: 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, 系统会适当调整骨骼长度之后再行计算。

至此, 通过上述的从已知二维点计算未知三维特征点的方法, 按图 4 所示的扩展顺序就可以求出所有的三维特征点。

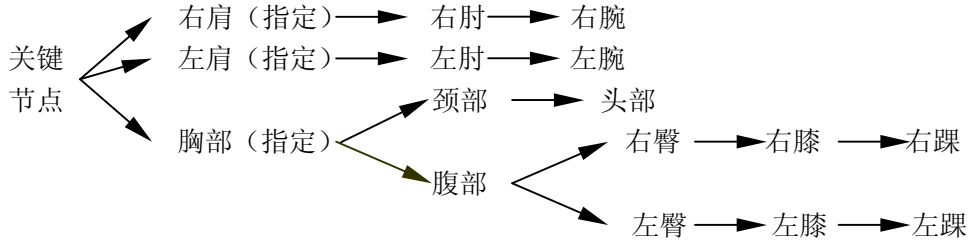


图 4. 人体运动模型三维特征点求解顺序

3.3 各后续帧中三维人体运动模型的对应

在 3.2 中, 我们已经对第一个图像帧建立了相应的三维人体模型。事实上, 对任一帧只要知道三个人体关键节点的三维坐标, 就可以用 3.2 中所述方法求出该帧所对应的三维人体模型。三个关键节点的坐标在首帧中是由系统确定的, 下面讨论在后续帧中如何确定它们的坐标[9]。

设第 n 帧中三个关键节点在摄像机坐标系中的坐标为 $P_i' (X_i', Y_i', Z_i')$ ($i=1\sim3$) 是已知的, 要计算第 $n+1$ 帧中三个相应的三维关节点的摄像机坐标 $P_i (X_i, Y_i, Z_i)$ ($i=1\sim3$)。 P_i 在投影平面中对应的二维点坐标为 (U_i, V_i) ($i=1\sim3$), 则由式 (2) 可知 P_i 与 (U_i, V_i) 的换算关系为:

$$P_i = \left(\frac{U_i \cdot Z_i}{f}, \frac{V_i \cdot Z_i}{f}, Z_i \right) \quad (i=1\sim3) \quad (9)$$

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, 则有下式:

$$\begin{cases} d(P_1', P_2') = d(P_1, P_2) \\ d(P_2', P_3') = d(P_2, P_3) \\ d(P_3', P_1') = d(P_3, P_1) \end{cases} \quad (10)$$

将 P_i' 和 (9) 代入 (10) 后可得一个非线性方程组。它有三个未知数, 用上一帧关键节点的 Z 值做初值, 通过梯度法来迭代求出一组解。这样就得到了本帧中三个关键节点的摄像机坐标 P_c 。

此时再利用 3.2 中所述的扩展方法, 我们就能求出该帧对应的人体模型中的全部三维特征点的坐标。本文假设人体运动具有连续性, 在出现二义性时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模型中充分体现了运动的连贯性与真实性。如图 5 中, 加粗的三角形为人体的关键三角形在投影平面中的位置, 当匹配右肘部节点出现 P1 和 P2 两个可能的位置时, 利用图中虚线所示的上一帧右肘部节点 P' 位置, 选取离 P' 较近的点 P1 为此帧中右肘部的位置。

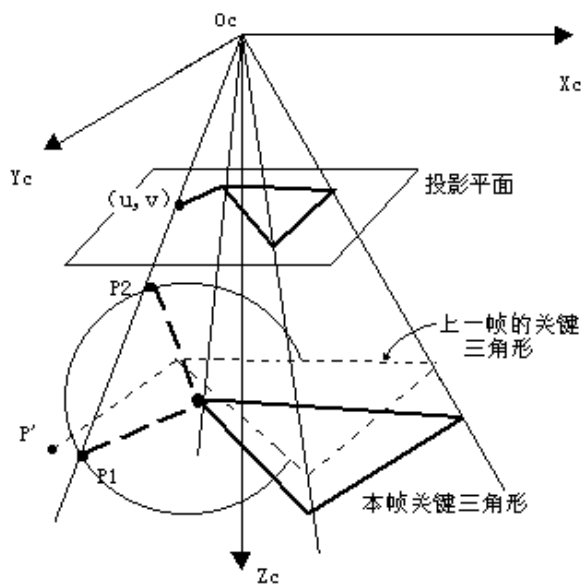
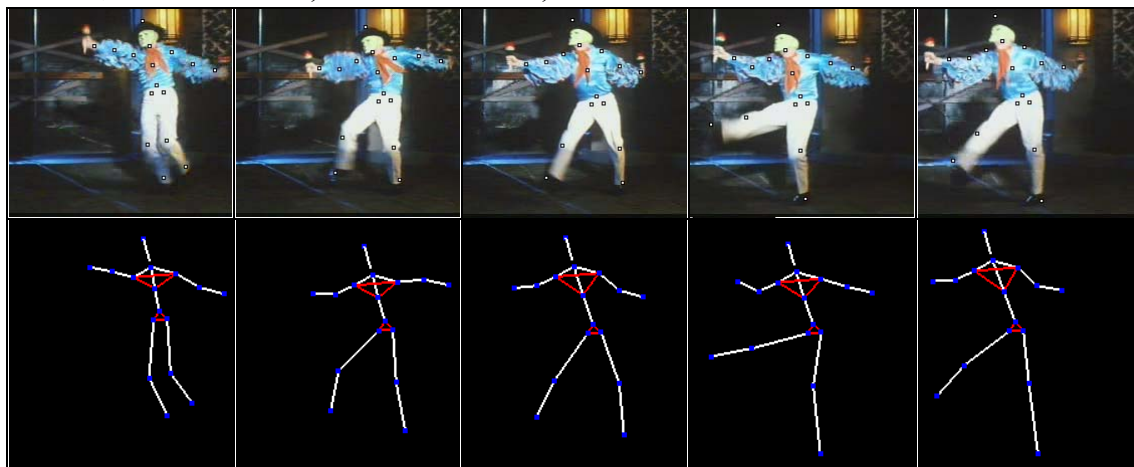


图 5. 后续帧中二义性的自动消除

4. 实验结果

按上文所述算法, 我们用 VC++6.0 与 OpenGL 实现了能从一个图像序列中生成三维人体运动模型的原型程序 Human Animation, 在 PENTIUM II 微机上运行通过。我们从电影《MASK》中截取了一段长为 2 秒共 48 帧的图像序列, 图 6 显示的就是该系统对其作用并生成三维人体运动模型的结果。最上面的一行分别显示了图像序列中的第 1, 5, 17, 39, 48 帧图象。其中, 用户手工标注了第 1 帧图象中的 16 个特征点, 另外四帧图像上的特征点是由程序[7]跟踪得到的。我们还定义了一个虚拟相机以模拟拍摄图像序列时所用的相机, 中间一行是虚拟相机拍摄的透视投影图, 它们与原相机采用同样的视角。最下面的一行显示了从不同角度面对人体三维模型得到的透视投影图。它们分别是将相机右转 30 度, 左转 30 度, 右转 20 度, 左转 150 度, 右转 15 度得到的投影图。从图中可以看到, 下面两行生动地反映了从各个角度看到的人体形态。除了《MASK》之外, 我们还进行了许多组实验, 结果发现有两个因素会影响最终三维模型的真实程度。首先, 如果二维图像序列的特征点跟踪有较大误差的话, 重建的三维模型会受很大影响。其次, 某些人骨架的长短比例与解剖学的数据有一定差异。因此当我们以一个短跑运动员作为分析对象时, 对骨架比例作了调整后, 才获得了很好的效果。但相对前者, 这一因素较少发生, 影响也小。



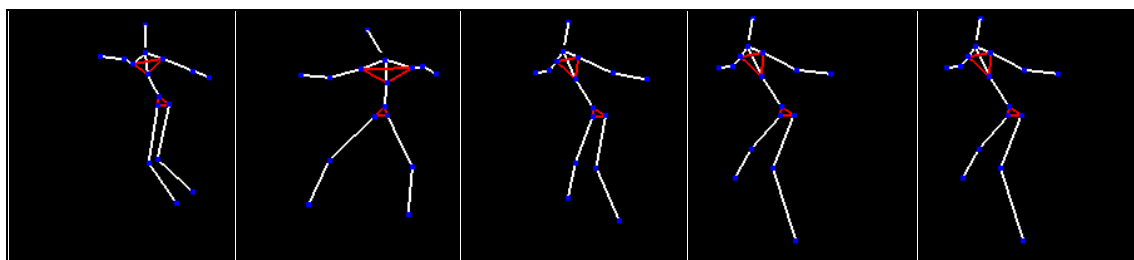


图 6. 《MASK》的实验结果

5. 结论

本文主要介绍了从二维人体骨架序列出发,对人体三维运动骨架的重建。通过三维模型和二维图象上的线对应关系,我们用牛顿方法求解了摄像机参数。然后,利用透视投影关系和人体骨骼的比例关系,我们求解了人体模型上各个三维特征点的坐标。在帧序列中,我们利用运动连续性来有效地消除了三维运动信息的二义性。从实验结果来看,最后得到的三维运动骨架的效果是很好的。今后我们将给三维人体骨架加上肌肉块,再贴上纹理[10],这样从各个角度看到的运动人体将更加动人。显然,计算机动画,虚拟现实等领域都可以应用我们的人体运动模型。

致谢: 感谢与童一颖,周昆进行的许多有益的讨论。

参考文献

1. J.O'Rourke, N. I. Badler. Model-based image analysis of human motion using constraint propagation. IEEE Trans. Pattern Anal. Mach. Intell., November 1980, 2(6):522-536.
2. D.Hogg. A program to see a walking person. Image Vision Computing, 1983, 5(20):87-96
3. K. Rohr. Incremental recognition of pedestrians from image sequences. In: Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. New York City: June 1993.8-13.
4. Z. Chen, H. J. Lee. Knowledge-guided visual perception of 3D human gait from a single image sequence. IEEE Trans. On Systems, Man, and Cybernetics, 1992, 22(2):336-342.
5. A. G. Bharatkumar, K. E. Daigle, M. G. Pandy, Q. Cai, J. K. Aggarwal. Low limb kinematics of human walking with the medial axis transformation. In: Proc. Of IEEE computer Society Workshop on Motion of No-Rigid and Articulated Objects. Austin,TX: 1994.70-76.
6. J. K. Aggarwal, Q. Cai. Human Motion Analysis: A Review. In: Proceedings of the IEEE Nonrigid and Articulated Motion Workshop 1997. Piscataway, NJ, USA: IEEE. 1997.90-102.
7. Liu XiaoMing, Zhuang YueTing, Pan YunHe. submitted to Study On Motion Image Sequence Based Human Skeleton Tracking
(刘小明, 庄越挺, 潘云鹤. 基于运动图象序列的人体骨架跟踪的研究,已投《模式识别与人工智能》)
8. Ma SongDe, Zhang ZhenYou. Computer Vision—Compute Theory and Arithmetic Foundation. Beijing: Scientific Press, January 1998. 52-59.
(马颂德, 张正友《计算机视觉—计算理论与算法基础》科学出版社, 1998年1月。52-59)
9. Thomas S. Huang, Arun N. Netravali. Motion and Structure From Feature Correspondences: A Review. Proceedings of The IEEE, February 1994, Vol.82 No.2: 252-268
10. Jia-Ching Cheng, Jos M. F. Moura. Capture and Synthesis of Human Motion in Video Sequences. In: Proceedings of IEEE Signal Processing Society 1998 Workshop on Multimedia Signal Processing. Los Angeles, California, USA: December 7—9 1998.