

# 人的三维运动的鲁棒估计

潘春洪, 马颂德

(中国科学院自动化, 模式识别国家重点实验室, 北京 100080)

**摘要:** 众所周知, 人的模型可以用多棍棒连接体来表示, 即棍棒模型. 解剖学的知识告诉我们: 人的运动有众多的约束, 如刚体约束, 静态旋转角约束等. 本文中, 我们利用人运动的刚体约束和共面约束去分析人的三维运动. 理论上, 我们得出: 在上述多约束融合下, 我们可以从单目序列图像中决定人的三维运动, 其所需要的帧数不仅与刚性连接体的个数有关, 而且与运动所在的空间平面与摄像机的相对位置也有密切关系. 为了得到鲁棒的三维运动, 我们引入了遗传算法, 实验结果令人满意的.

**关键词:** 棍棒模型; 刚体约束; 共面约束; 运动分析

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 03722112 (2003) 10151804

## Robust Estimation of 3D Human Motion

PAN Chunzhong, MA Songde

(National Laboratory of Pattern Recognition Institute of Automation, CAS, Beijing 10080, China)

**Abstract:** It is well known that human body can be approximately modeled by articulated object. The anatomical knowledge reveals that there exist multiple constraints on human motion such as rigidity constraint, static rotation constraint, and so on. In this paper, we estimate 3D human motion by rigidity and coplanar constraints. In theory, we analyze the 3D structure of multilink from motion, and investigate the relationship among the number of the frames and links and the number of equations and unknowns. By Genetic Algorithm, the robust numerical solutions can be achieved. The experiments with the simulated data and the real images are included to demonstrate the validity of the theoretic results.

**Key words:** articulated model; rigidity constraint; coplanar constraint; motion Analysis.

### 1 引言

产生真实的人的运动是虚拟现实和计算机动画的中心问题<sup>[9]</sup>. 目前, 有很多这方面的工作. 一般而言, 工作主要可以分为三类<sup>[8]</sup>: 关键帧法、过程计算方法和运动捕捉法. 用此方法所产生的动画也相应地被称作关键帧动画、过程计算动画和运动捕捉动画. 其中, 运动捕捉是一种非常流行的方法, 它能相对容易地记录人的任意运动. 事实上, 它也是到目前为止能有效地记录人的任意运动的唯一方法. 运动捕捉系统一般可分为光学运动捕捉系统和电磁波运动捕捉系统. 光学运动捕捉系统利用多个摄像机去跟踪安放在关节点上的各个特殊标志点, 然后得到人的三维运动, 这种系统能获得高精度的人的运动. 然而, 这是一个非常复杂的系统, 需要有 4~8 个被精确定标的摄像机、一些特殊的标志点和比较大的运动场合, 而且这种系统非常昂贵. 显然, 这大大限制了它的应用范围.

运动学的知识告诉我们, 脚和手的运动并不是任意的<sup>[3]</sup>, 它们有着一定的规律性. 更进一步地, 从解剖学的知识我们知道, 膝关节和肘关节仅仅有一个自由度, 所以腿的运动在一个短时间段内, 能被近似地认为是在同一平面上运动, 这种假

设在许多人的运动分析中是非常合理的, 比如, 人的行走, 由于膝关节仅仅有一个自由度, 即使不按直线行走, 但腿每一步在空中的摇摆基本上是在同一平面上, 即在短时间内可被看成在同一平面上运动, 因此, 我们可以用共面运动的假设来约束腿的运动. 许多文献中也定性地介绍过人的共面运动约束, 指出人在运动时身体的某些局部运动满足共面约束<sup>[10]</sup>. 另外, 在运动过程中, 被假设成棍棒模型的人的关节点之间的长度是不变的, 即存在刚体约束. 进一步地, 如果我们知道了共面刚体的运动, 那我们就可以仅用刚体约束求知相邻点的三维运动, 从而求得整个人的三维结构, 因此, 对棍棒模型而言, 共面刚体的运动分析是非常重要的.

人能被简化成棍棒模型(Articulated Model)<sup>[4]</sup>, 这种棍棒模型是由一些棍棒通过关节点连接而成的, 它既非形变模型(Deformable Model), 也非刚体模型(Rigid Model), 虽然其中的每一部分是刚体, 然而就整体而言, 它是非刚体的. 由于在棍棒模型中存在着多运动, 因此, 从单目序列图像中, 仅仅通过关节点的对应, 是不可能从单目序列图像中恢复这种棍棒模型的三维结构的. 必须对物体的运动或投影方式进行建模. 特殊的运动: 如共面运动<sup>[1]</sup>, 至少一个点的三维运动已知<sup>[5]</sup>等,

特殊的投影方式:如平行投影<sup>[2]</sup>等,只有在这些特殊情况下,才可能恢复它的三维结构.这里,为了从单目序列图像中获取相应棍棒模型的三维运动信息,除了刚体约束外,我们增加了共面运动约束.

文[10,7]中详细地分析了共面多刚体的三维运动,然而在实际的图像序列中,几乎不可能鲁棒地找到特征点(关节点)的对应,而且,很难找到同时满足共面约束和刚体约束的数学解.本文中,我们引入了基于约束融合的鲁棒的人的三维运动估计,为此,我们利用了遗传算法去优化估计它的三维运动.带有模拟数目和真实图像的实验结果有效地证实了该方法的有效性.

### 2 棍棒模型运动分析

这里,我们用针孔摄像机模型获取不同时刻运动物体的图像序列,  $A^{(j)}$ 表示三维空间坐标系中,在时刻  $t_j$  时刚体端点的三维坐标,  $a^{(j)}$  是  $OA^{(j)}$  方向的单位向量,如果特征点  $A^{(j)}$  相应的图像坐标为  $(x_a^{(j)}, y_a^{(j)})$ , 那么其空间坐标点的单位向量可表示成:

$$a^{(j)} = \frac{1}{\sqrt{x_a^{2(j)} + y_a^{2(j)} + f^2}} \# (x_a^{(j)}, y_a^{(j)}, f) \quad (1)$$

刚体约束 对于所有单刚体,其长度在所有图像序列中应保持相应不变,因此对于所有的  $j = 0, 1, 2, \dots, n$  和  $i = 1, 2, \dots, k$ , 我们可以表示成如下刚体约束方程:

$$K_{A_i^{(j)}}^2 + K_{A_{i+1}^{(j)}}^2 - 2\#K_{A_i^{(j)}}\#K_{A_{i+1}^{(j)}}\#\cos A_{A_i^{(j)}}A_{A_{i+1}^{(j)}} = l^2 \quad (2)$$

这里:  $f$  是摄像机的焦距,  $K_{A_i^{(j)}}$  是摄像机光心到空间坐标点  $A^{(j)}$  的距离,  $n$  为图像帧数,  $k$  为连接体的个数.

共面约束 因为刚体是在同一平面上运动,即刚体的所有端点在不同时刻总在同一平面上,这样我们就可以得到如下共面约束方程:

$$k_a^{(j)}\#K_{A_i^{(j)}}\#(x_{a_i^{(j)}}\#n_x + y_{a_i^{(j)}}\#n_y + f) = D \quad (3)$$

这里,  $D$  是摄像机光心到空间平面的距离,  $n = \frac{1}{\sqrt{n_x^2 + n_y^2 + 1}}$

$(n_x, n_y, 1)$  是空间平面的单位法向量,  $k_a^{(j)} = \frac{1}{\sqrt{n_x^2 + n_y^2 + 1}}$

$\frac{1}{\sqrt{x_{a_i^{(j)}}^2 + y_{a_i^{(j)}}^2 + f^2}}$ ; 对所有的  $i = 1, 2, \dots, k$ , 和  $j = 1, 2, \dots, n$ ,

$K_{A_i^{(j)}}$ ,  $n_x, n_y$  是未知变量.

我们假设棍棒模型有  $k$  个连接体, 即有  $k + 1$  个节点, 在某一时间间隔内, 它们是在同一平面上运动的. 对空间中的任一固定平面当给定  $n$  帧图像时, 有未知变量  $K_{A_i^{(j)}}$ ,  $n_x, n_y$ , 在尺度意义下其未知变量的数目是  $n(k + 1) + 1$ , 由等式(2)我们可以得到  $k(n - 1)$  个刚体约束方程, 由等式(3)可得到  $n(k + 1) - 1$  个共面约束方程, 因此总的方程数为  $2nk + n + k - 1$ . 对于单刚体的情况, 即  $k = 1$  时, 方程数为  $2n + 1$ , 而未知变量有  $3n - 2$  个. 显然, 当提供三幅图像时未知数的个数和方程的个数均为 7, 因此当提供三帧图像时, 利用上述方程有可能恢复单刚体的三维运动, 详细的推导过程见文[7,10].

### 3 鲁棒性估计

鲁棒解的分析 在真实的人的三维运动估计中, 由于图像噪声的影响, 不同帧之间刚体端点的匹配误差, 以及棍棒模型的近似假设, 刚体约束方程以及共面约束方程不可能得到同时满足, 必须要用优化估计的方法去求解该方程.

为了求解, 我们可以假定最后的解仅满足共面约束方程, 那么刚体约束方程可以写成:

$$N_i = \sum_{i \times 1, i, i=1}^n \sum_{i \times 1, i, i=1}^k \left| |A_{1i}^{(j)} - A_{2i}^{(j)}| - |A_{1i}^{(l)} - A_{2i}^{(l)}| \right| \quad (4)$$

同样地, 我们假设所求解仅仅满足刚体约束方程而不真正满足共面约束方程, 我们可以假设存在一空间平面, 那么连接体端点到该平面的距离可以写成:

$$d_{ij} = \frac{|K_{A_i^{(j)}}\#a_i^{(j)}\#n|}{\sqrt{n_x^2 + n_y^2 + 1}}$$

这里, 我们的目的是找出一空间平面, 使得连接体端点到该平面的距离为最小, 因此我们有如下共面误差准则:

$$N_B = \sum_{i=0}^n \sum_{j=1}^k d_{ij} \quad (5)$$

以上是两种极端的情况, 对于一般情况我们可以假设存在一虚拟平面, 使得共面约束误差和刚体约束误差之和最小, 这样, 我们可以将最终的误差准则写成:

$$N = A\#N_1 + B\#N_2 \quad (6)$$

这里,  $A, B$  为权重因子,  $A, B \in (0, 1)$  且  $A + B = 1$ , 误差  $N$  是参数  $K_{A_i^{(j)}}$ ,  $H_x$  和  $H_y$  的函数, 因此, 上述的三维运动分析就转化为如何求得一组  $K_{A_i^{(j)}}$ ,  $H_x$  和  $H_y$  的值, 使得  $N$  最小, 这是一个典型的优化问题, 可以用动力学编程的方法进行优化, 这里, 我们应用遗传算法求解该问题.

遗传算法 遗传算法模拟了生物遗传、变异, 由低级到高级的发展进化过程<sup>[6]</sup>. 生物进化遵循达尔文的/优胜劣汰, 适者生存0的思想. 1965年, Holland 将这种思想应用到实际优化搜索算法中, 由此形成了遗传算法. 将遗传算法应用于各种优化问题时, 实际上是将进化操作应用到一群对搜索空间编码的基因串中, 通过反复应用选择、复制、杂交和变异等遗传算子, 择优淘劣, 最终得到全局最优解.

本文以上述的误差准则  $N$  作为目标函数, 通过下面的等式建立目标函数与适应度函数  $f$  之间的关系:  $f = \frac{N}{1 + N}$ . 对待优化参数  $K_{A_i^{(j)}}$ ,  $H_x$  和  $H_y$  ( $j = 0, 1, 2, \dots, n, i = 1, 2, \dots, k$ ) 进行二进制编码, 每个参数编码的长度为  $l_i = 10$ , 各参数的寻优区间为:  $K_{A_i^{(j)}} \in [K_{\min}, K_{\max}]$ ,  $H_x$  和  $H_y \in \{0, 180\}$ .

依法构造  $N$  个这样的个体组成初始群体, 计算各个体的适应度, 根据一定的选择概率及个体适应度, 对初始群体进行操作, 选出  $M$  个具有较高适应度的个体组成新群体, 并将这些个体放入匹配池中, 根据交叉和变异概率对它们进行相应的遗传操作. 重复该过程, 使群体中具有较高适应度的个体被保留到下一代, 直到满足终止条件(预先定义的误差大小), 迭代过程结束. 在我们的算法中, 交叉概率为 15%, 变异概率为 0.5%.

### 4 实验结果

#### 4.1 模拟数据的鲁棒估计

我们首先用带噪声的模拟数据分析上述方法的有效性,然后再用该方法去估计人的三维运动.这里,我们给出四帧单刚体端点的图像坐标:

$$(a_1^{(1)}:(2.0, 2.0)), (a_2^{(1)}:(0.403, 0.62));$$

$$(a_1^{(2)}:(5.0, 5.0)), (a_2^{(2)}:(1.39, 0.788))$$

由上述分析可知,对于单刚体而言,给出3帧时,运动的解不唯一,当给出4帧时,有唯一解,此时,共有10个参数,其中8个端点深度值,2个空间平面的法向量,把这10个参数用如上所介绍的规则级联在一起形成一个个体,通过遗传算法,我们可以求出其空间平面的法向量和刚体的结构.对于理想情况(不带噪声),平面约束方程和刚体约束方程均可以得到满足,在这种情况下,我们利用上述的遗传算法,求出刚体所在平面的法向量  $n_x = 0.4986$ ,  $n_y = 0.4986$  以及相应的法向角  $H_x = 261.5b$ ,  $H_y = 261.5b$  此时,如果再给定单刚体的长度,我们就可以算出理想情况下单刚体的三维运动轨迹.

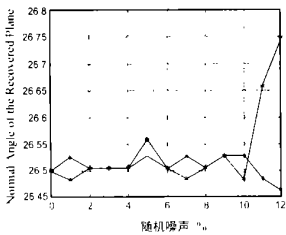


图1 被求出的平面法向量与随机噪声的关系

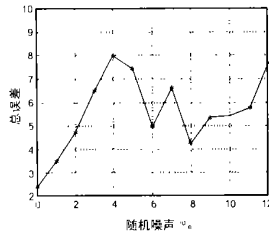


图2 总误差与随机噪声的关系(一个端点加噪声)

当我们在上述单刚体端点的二维序列图像坐标上增加一些随机噪声时,上述的平面约束和刚体约束方程显然不可能完全得到满足,但我们可以利用上述所提出的优化方法来求出一虚拟平面,使得平面约束和刚体约束的误差和最小.具体过程:首先,我们在单刚体某端点所有的二维图像坐标上加一些随机噪声,该端点上所加的随机噪声是按该端点图像坐标的百分比来增加的,比如,某端点图像坐标为(100, 100),那么增加10%的噪声,意味着端点的图像坐标将在(90, 90), (110, 110)之间随机地取.应用上述遗传算法,我们分别求出不同噪声下的平面法向量,图像坐标加有噪声的该端点的三维坐标值,以及共面约束和刚体约束的总误差值,虽然共面约束方程和刚体约束方程不能同时满足,但存在一虚拟平面,使得被恢复的三维运动满足一定的误差准则.图1给出了被求出的平面法向量与随机噪声的关系,从图中我们可以看出,由于噪声是随机加上的,求出的平面法向量总是在理想点附近变化,但当随机噪声增加时,就会远离理想的情况.图2给出了计算总误差与随机噪声的关系,图5中是图像坐标加有噪声的该端点的三维坐标值与噪声的关系,理想情况应是:(2.0, 2.0, 1.0).

进一步地,我们在单刚体两个端点的所有图像坐标上都增加一些噪声,如前,我们也能找到满足一定误差准则的唯一

的虚拟平面,图3说明了不同的噪声将有不同的虚拟平面,但都是在理想情况附近变化,而总误差则随着噪声的增加而增加.

比较图2、图4以及图1、图3,我们可以看出,虽然加有相同的噪声,但由于前者是在一个端点处加有噪声,而后者是在两个端点处,因此,所带来的误差,前一种情况要明显地小于后一种情况,而且,其估计的平面法向量的偏差值也是后一种要大.

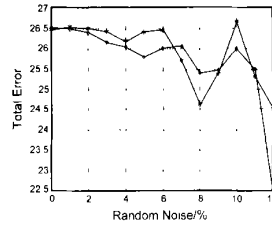


图3 被求出的平面法向量与随机噪声的关系(两个端点加噪声)

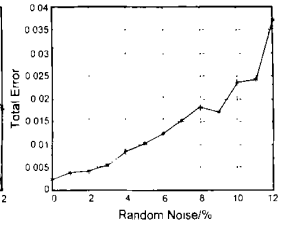


图4 总误差与随机噪声之间的关系(两个端点加噪声)

不同的随机噪声,将会有不同的虚拟平面,而且随着噪声的增加,越来越偏离理想的情况,当噪声增加到12%时,误差就会比较大,而且就遗传算法而言,当误差较大时,较难收敛.仅在单端点加噪声的情况下,当噪声为1%时,其迭代次数为10次左右即可收敛,而且,误差很小;当噪声增加到12%时,其迭代次数就要增加到60次左右;对于两个端点均加有噪声的情况,当噪声为1%时,迭代次数为25次左右;当噪声为12%时,其迭代次数就要增加到110次左右,明显难以收敛,而每次迭代时间和每个参数编码的长度有关.这里我们的参数编码长度为10,这种情况下,其每次迭代时间是1.2秒左右,这些都与我们分析的结论相符.由上述结果,我们可以看出,该方法是非常有效的.

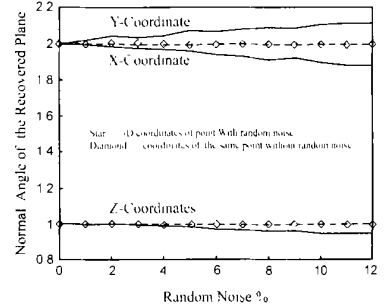


图5 图像坐标加有噪声的该端点的三维坐标值与噪声的关系

#### 4.2 真实图像

以下,我们应用上述的方法来估计真实的人的三维运动.这里,我们主要跟踪腿的运动,我们知道当人行走时,即便不走在一条直线上,在某一很短的时间内,其关节点仍可以近似看成在同一平面上,我们用摄像机(已定标)去跟踪人的行走(采样速率为30帧/秒),然后我们在图像序列中跟踪这些标志点,得到它们的图像坐标,图6给出了带有标志点的人行走的序列图像.最后,我们用上述的优化算法去估计关节点的三维运动.

需要说明的是:

(1) 摄像机是定过标的,我们是通过定标块而实现定标

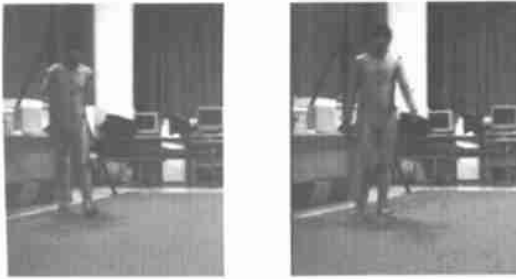


图 6 带有标志点的人行走的序列图像

的,常用定标软件有: Tsai 方法(从网上下载). 该程序非常鲁棒,是目前用定标决定标中最好的软件;这里,我们摄像机的内参数为:  $u_0 = 322.5155$ ;  $v_0 = 232.47737$ ;  $d_x = 4.9/640$ ;  $d_y = 3.71/480$ ;  $f_a = 7.0607$ ;

(2) 膝关节只有一个自由度,因而腿可以看成是由两个连接体连接而成的棍棒模型,理想情况下,虽然在相邻两帧内,可以假设腿是在同一平面上运动,但事实上,由于标志点的安放位置不可能在相应腿部的关节处,因而很难保证这些标志点在同一平面上运动. 这里我们把两部分分开考虑,因为当摄像机以 30 帧/秒的速率采样时,人的一步要多于 4 帧,而由 4 帧我们就能恢复它的三维结构;在具体计算中,我们用测试的方法来实现. 先手工选出最初 4 帧,求出相应的虚拟平面,当增加一帧时,重新计算相应的误差,当误差小于我们给定的域值时,认为它们在同一平面内,否则它们属于下一个平面. 通过以上的方法,我们获取了腿上三个关节的三维结构,图 7 分别给出了左右脚归一化后踝关节,膝关节和髋关节的三维坐标,其结果是令人满意的.

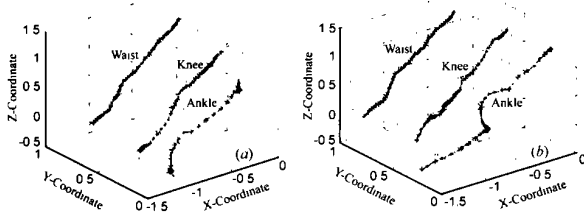


图 7 归一化后的踝关节,膝关节和髋关节的三维运动数据 (a)左脚;(b)右脚

最后,我们用获取的三维运动数据来驱动虚拟人的行走,图 8 给出了某一视点下的一些关键帧.

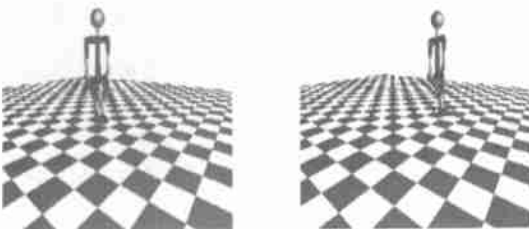


图 8 某一视点下的一些关键帧

## 5 小结

本文中,我们首先从理论上探讨了多约束融合的共面连接体的三维运动估计方法,基于共面约束和局部刚体约束,我们可以唯一地确定共面连接体的三维运动,然后,利用这种多约束融合的方法去估计人的三维运动,并利用遗传算法获取了鲁棒的三维运动. 最后,我们给出了模拟数据和真实图像的实验结果.

## 参考文献:

- [1] J K Roach, J K Aggarwal. Determining the Movement of Objects from a Sequence [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, 2: 554- 562.
- [2] J A Webb, J K Aggarwal. Structure from Motion of Rigid and Jointed Objects [J]. Artificial Intelligence, 1982, 19: 107- 130.
- [3] D D Hoffman, B E Flinchbaugh. The Interpretation of Biological Motion [J]. Biological Cybernetics, 1982, 42: 195- 204.
- [4] D Terzopoulos, D Metaxas. Dynamic 3D Models with Local and Global Deformable Superquadrics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13: 703- 714.
- [5] D T Lawton. Constraint-based Inference from Image Motion [A]. In Proc. 1<sup>st</sup> Annu. Nat. Conf. Artificial Intelligence [C]. USA: Stanford Univ., 1980.
- [6] D E Goldberg. Genetic Algorithms in Search, Optimization and Machine Learning. Reading [M]. USA: Addison Wesley, 1989.
- [7] Chunhong PAN, Songde MA. 3D Motion Estimation of Human by Genetic Algorithm [A]. In proceeding of International Conference on Pattern Recognition 2000 [C]. Spain: ICPR, 2000. 159- 163.
- [8] J K Hodgins, J F O'Brien. Computer Animation [M]. The Encyclopedia of Computer Science, International Thomson Business Press, 1998.
- [9] V I Pavlovic, R Sharma, T S Huang. Visual Interpretation of Hand gesture from Human-Computer Interaction, A Review [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19: 607- 617.
- [10] 潘春洪. 单目序列图像中人的三维运动分析 [D]. 北京: 中科院自动化所, 2000.

## 作者简介:

潘春洪 1987年毕业于清华大学自动化系,1990年毕业于中国科学院上海光机所,获工学硕士学位,2000年毕业于中国科学院自动化研究所模式识别国家重点实验室并获工学博士学位,2000年7月至2001年10月在美国南加州大学做博士后研究工作,现工作于中国科学院自动化所模式识别国家重点实验室,主要研究方向为图像处理,计算机视觉和计算机图形学.

马颂德 男,1969年毕业于清华大学自动化系,1983年和1986年先后于法国巴黎第六大学获硕士和博士学位,1983年至1984年为Maryland大学计算机视觉实验室访问学者,1984年至1986年为法国INRIA机器人视觉组访问学者,曾任中国科学院自动化研究所所长,现为中国科学技术部副部长,模式识别国家重点实验室特聘教授,博士生导师,IEEE高级会员,主要研究方向为图像处理,计算机视觉,计算机图形学和模式识别.