

# 人体上肢姿态的估计及多解分析\*

任海兵, 徐光祐

(清华大学 计算机科学与技术系 人机交互与媒体集成研究所,北京 100084)

E-mail: renhaibing@263.net; xgy-dcs@mail.tsinghua.edu.cn

http://www.tsinghua.edu.cn

**摘要:** 人体是复杂的变形物体(deformable objects),结构自由度多.仅仅从 2D 表观特征出发,很难分析和识别精细的动作,更谈不上理解用户的意图.因为手势是交流的主要方式,所以,以 3D 模型为基础,对人体上肢建模,重构人体上肢的 3D 姿态.从实用化和尽可能少引入误差出发,分析了重构人体上肢姿态所必须的最少条件.在此条件下,提出了端点固定的关节模型(end-determined articulate model)及其对应的方程组,以估计各个关节节点的 3D 坐标.然后,分析方程组解的最大可能数目,并给出相应的求解方法.最后,利用结果误差和位姿来检验解的合理性.

**关键词:** 姿态估计;关节模型;端点固定;3D 重建

中图法分类号: TP391 文献标识码: A

最近几年,随着智能环境(intelligent environment)、无所不在的计算(ubiquitous computing)、普遍存在的计算(pervasive computing)、智能屋(smart room)等的兴起,需要计算机系统能够感知环境中人的存在,智能地理解人的行为动作,适应人的习惯,并且作出相应的响应,实现更加自然、随意、和谐的人机交互.其中的关键问题是如何应用计算机视觉的方法感知环境中人的存在,并正确分析、理解人的表情、行为和运动,这个领域通常称为“Looking at People”.它包括:人脸识别(face recognition)、手势识别(gesture recognition)、人体的跟踪(human tracking)、姿态估计(pose estimation)以及身体语言的理解(body language understanding)等方面.

由于人是复杂的变形体(deformable object),人体四肢自由度多,运动范围很大,因此会遇到很多基本问题,例如如何对人体建模、跟踪人体各部分的运动、重建人体的 3D 姿态、解决人体中的遮挡问题、加速搜索过程、降低问题的解空间等.解决好人体姿态估计和跟踪,可以推广到其他刚体或者变形体的运动跟踪与识别、3D 重建以及遮挡情况.具有很好的普遍性.在虚拟现实、智能监控、运动员的训练、基于内容的检索、基于人体模型的高比特流的压缩算法、更高级的人机交互等方面有非常重要的应用.

对于人体运动分析和识别的研究,最早可以追溯到 70 年代.Johansson<sup>[1]</sup>从心理学的角度出发,利用贴在人体关节上亮点的运动(moving light displays,简称 MLD)研究整个人体的运动.由于 MLD(或者其他使用标记点的方法)标记和跟踪关节点非常容易,因此被广泛应用,如 Lakany<sup>[2]</sup>以此来进行步法的分析.但是,这种方法需要在人体身上贴有标记物,应用领域有很大的局限性.于是,一些研究人员尝试运用计算机视觉等非接触的方法.如文献[3]中,运动能量图像(motion energy images,简称 MEI)和运动历史图像(motion history images,简称 MHI)的方法.另外,在 MIT Media Lab 的 Pfinder 系统中<sup>[4]</sup>,Wren 也是采用基于表观的方法.这些方法没有任何人体模型,直接从 2D 表观入手,只能获取人体动作的 2D 整体特征,并不能识别人体精细的运动.

后来,越来越多的人体模型被用来估计人体头部、四肢的姿态和运动,达到捕捉和识别精细动作、理解用户意图的目的.这些模型主要分为 2D 模型和 3D 模型<sup>[5]</sup>.2D 模型是基于人体表观模型的方法,采用 2D 形状来拟

\* 收稿日期: 2001-04-12; 修改日期: 2001-08-01

基金项目: 国家自然科学基金资助项目(69873022);国家 863 高科技发展计划资助项目(863-306-ZT03-01-1)

作者简介: 任海兵(1975 - ),男,江苏宿迁人,博士生,主要研究领域为计算机视觉,人机交互;徐光祐(1940 - ),男,上海人,教授,博士生导师,主要研究领域为计算机视觉,人机交互,多媒体技术.

合人体各个部分,人体运动过程中跟踪这些 2D 形状的运动.如文献[6]中 Maylor 和 Yang 利用 2D 带状模型(ribbon model),从人的外轮廓中标记出各个部分.由于 2D 方法无法得到人体的 3D 姿态、运动等信息,于是研究中就引入了 3D 模型.3D 模型有 3D 关节模型(articulated model)和 3D 体积模型(volumetric model).3D 关节模型由人体中起支撑作用的骨骼组成,骨骼的连接处为关节点.由于 3D 骨架模型是最简单、实用的 3D 模型,因此应用得最为广泛,通常的方法是利用多个视图中各个关节点(或者肢体各个部分)的图像投影来获取各个关节点的 3D 位置,如文献[4,7,8].3D 体积模型,不仅包含 3D 骨架,而且包含围绕在骨骼外面的皮肤,如 3D Tapered Super-quadratics Model<sup>[9]</sup>和 3D Surface-shape Model<sup>[10]</sup>.利用 3D 模型的方法可以恢复人体 3D 姿态、跟踪人体的 3D 运动.但是,上述 3D 方法过分依赖于图像信息,没有充分利用人体自身的特点,对条件要求苛刻,鲁棒性和实用性较差.

本文从一个新视点出发,融合 2D 和 3D 的方法,根据人体结构的特点,分析恢复人体 3D 姿态所需的最少条件,并在此基础上给出具体的求解方法.具体而言,本文简化人体 3D 骨架模型,仅仅对人体上肢建模,在已知各个关节点在单幅图像中的图像坐标和人体骨架模型中各骨骼的长度比例关系的条件下,利用容易获取的其他部位的 3D 信息作为附加约束条件,建立端点固定的关节模型,估计人体的 3D 姿态.

由于本文所能得到的已知条件都是近似的,即各个关节点的图像坐标、由立体视觉计算所得到的空间 3D 坐标、像坐标、由立体视觉计算所得到的空间 3D 坐标各个骨骼的长度比例关系等,在数值计算中,这些不精确的已知条件都会引入一定的误差,对结果产生负面影响.而且多个误差的累积叠加,很容易产生灾难性的错误,例如增加或者减少解的个数.另一方面,通过视觉方法获取这些条件是相当困难的.因此,本文在能够求得解的情况下,采用尽量少的近似条件,以减少误差的影响,提高系统的实用性和鲁棒性.另外,由于 2D 到 3D 本身就是不适定问题(ill-posed),方程组很可能是多解的.本文还分析了方程组解的最大可能个数.最后,本文给出了方程组的求解方法和解合理性判据.

本文第 1 节介绍人体上肢的关节模型及其数学表示,并且分析了求解所需的最少条件.第 2 节根据已知条件得到方程组,分析解的最大可能数目.第 3 节给出求解方程组的方法,并给出解的合理性判据.第 4 节是实验结果和结论.

## 1 人体上肢模型

本文采用人体上肢的 3D 关节模型,如图 1 所示. $P_i(i=0,1,\dots,5)$ 表示人体上肢的关节点:左、右腕关节,左、右肘关节和左、右肩关节.这些关节点之间由骨骼连接在一起. $N$  是脖子与肩关节连接骨骼之间的交点.虽然人脸、双手和颈部并不属于人体的上肢模型,但这些都是人体中不可缺少的部分,而且比较容易检测,可以提供更加充分的信息.充分利用这些部位的信息,增加近似约束条件,可以减少问题的复杂度,并大幅度提高算法的鲁棒性.

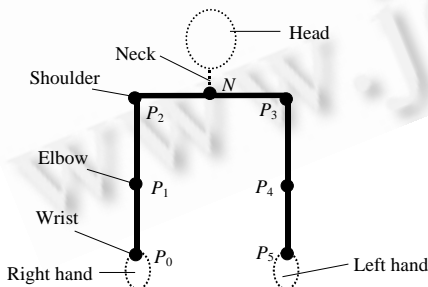


Fig.1 The human upper-limb model  
图 1 人体上肢模型

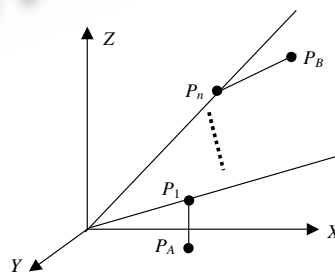


Fig.2 The end-determined articulated model  
图 2 端点固定的关节模型

### 1.1 可行性分析

本文需要计算 6 个关节点的 3D 空间坐标  $P_i(X_i, Y_i, Z_i)(i=0, \dots, 5)$ , 所以总共有 18 个未知参数.已知  $P_i$  的图像坐标  $p_i(x_i, y_i)$ , 在摄像机坐标系中,根据摄像机的透视模型,关节点的空间坐标  $P_i(X_i, Y_i, Z_i)$  与图像投影  $p_i(x_i, y_i)$  的关系如下:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{s}{Z_i} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \frac{s}{Z_i} \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \quad i=0,1,\dots,5, \quad (1)$$

其中  $s$  为比例因子,由摄像机的焦距等内参数决定,可以通过摄像机的标定而获得.

另外,已知各个骨骼的长度比例关系如下:

$$\frac{|P_0P_1|}{m_0} = \frac{|P_1P_2|}{m_1} = \frac{|P_2P_3|}{m_2} = \frac{|P_3P_4|}{m_3} = \frac{|P_4P_5|}{m_4}, \quad (2)$$

其中  $|P_iP_{i+1}|$  ( $i=0,\dots,4$ ) 表示  $P_iP_{i+1}$  的 3D 长度,  $m_i$  表示长度的比例因子,所以可得到 4 个等式.加上式(1),总共有 16 个约束条件、18 个未知参数,所以无法求解,至少还需要增加两个约束条件.

## 1.2 附加约束信息

本文采用立体视觉方法获取 3D 信息,增加约束信息,解得未知参数.在计算机视觉领域,图像中的人手和眼睛是人体中最容易识别的部分,据此本文提出一种比较新颖、鲁棒性更好、效率更高的方法.

(1) 在两幅图像中,利用人体皮肤的色度信息<sup>[11]</sup>检测双手区域,计算区域中心点的坐标.利用立体信息,计算双手的深度信息.由于人是面对摄像机运动的,因此双手区域是很容易检测到的.

(2) 以双手的深度去近似双腕的深度信息.这个近似是合理的,因为无论人体如何运动,这个近似误差相对于人体各个骨骼长度都是很小的,相对于人体与摄像机之间的深度就更加微乎其微.

可以得到如下两个约束条件:

$$Z_0 \approx Z_{r\_hand}, \quad (3)$$

$$Z_5 \approx Z_{l\_hand}, \quad (4)$$

其中  $Z_{l\_hand}, Z_{r\_hand}$  分别是左、右手的深度.综合方程组(1)~(4),联立得到 18 个方程,有 18 个未知数,理论上是可以求解的.

## 1.3 端点固定的关节模型

由式(3)和式(4)以及两个腕关节的图像坐标可以直接得到双腕的 3D 坐标.所以,人体上肢模型可以看成是两端固定的关节模型( $P_A P_1 \dots P_i \dots P_n P_B$ ) (end-determined articulate model),如图 2 所示.已知开始端点  $P_A$  的 3D 坐标( $X_A, Y_A, Z_A$ )、最后端点  $P_B$  的 3D 坐标( $X_B, Y_B, Z_B$ )、中间各个关节点的图像坐标  $p_i(x_i, y_i)$ ,  $i=1, \dots, n$ , 另外,各个骨骼的长度成一定比例关系:

$$\frac{|P_A P_1|}{m_A} = \frac{|P_1 P_2|}{m_1} = \dots = \frac{|P_i P_{i+1}|}{m_i} = \dots = \frac{|P_n P_B|}{m_n} = L, \quad (5)$$

其中  $m_A, m_1, \dots, m_i, \dots, m_n$  为长度比例系数,  $L$  为单位长度.

已知关节点  $P_i$  在图像中的投影坐标,根据摄像机透视模型可知,物体  $P_i$  在 3D 空间中一条过光心的直线上.则  $P$  可以表示为

$$P_i = P'_i + k_i \vec{p}_i, \quad (6)$$

其中  $P'_i(X'_i, Y'_i, Z'_i)$  为光心,  $\vec{p}_i(\bar{x}_i, \bar{y}_i, \bar{z}_i)$  为直线的方向向量,  $k_i$  为表示比例关系的系数.在摄像机坐标系中,光心为坐标系原点(0,0,0),则直线方程可以简化为  $P_i = k_i \vec{p}_i$ .  $P_i$  的 3D 坐标( $X_i, Y_i, Z_i$ )可以表示为( $k_i \bar{x}_i, k_i \bar{y}_i, k_i \bar{z}_i$ ).

## 1.4 迭代法求解方程组

综合式(5)、式(6),可以得到问题的公式化描述:

$$\begin{cases} k_1^2(\bar{x}_1^2 + \bar{y}_1^2 + \bar{z}_1^2) - 2k_1(Y_A \bar{y} + Z_A \bar{z} + Z_A \bar{z}) + X_A^2 + Y_A^2 + Z_A^2 - Lm_A^2 = 0 \\ k_i^2(\bar{x}_i^2 + \bar{y}_i^2 + \bar{z}_i^2) - 2k_i k_{i-1}(\bar{x}_{i-1} \bar{x}_i + \bar{y}_{i-1} \bar{y}_i + \bar{z}_{i-1} \bar{z}_i) + (k_{i-1} \bar{x}_{i-1})^2 + (k_{i-1} \bar{y}_{i-1})^2 + (k_{i-1} \bar{z}_{i-1})^2 - Lm_{i-1}^2 = 0, \quad i=2,3,\dots,n, \quad (7) \\ k_n^2(\bar{x}_n^2 + \bar{y}_n^2 + \bar{z}_n^2) - 2k_n(X_B + Y_B + Z_B) + X_B^2 + Y_B^2 + Z_B^2 - Lm_n^2 = 0 \end{cases}$$

其中  $k_i$  ( $i=1, \dots, n$ ) 和  $L$  是未知参数.一种简单直观的求解方法是利用 Newton 迭代法、SOR 迭代法等普通的多元非线性方程组解法求得未知数的代数解.

然而,本文并不是一个纯粹的数学问题,而是计算机视觉中的实际应用,其中无论是骨骼的长度比例关系,还是各个关节的坐标,都是近似值,因此,实际问题的解满足上述方程组的概率为 0.利用迭代的方法则有如下缺点:

(1) 本文是一个实际的问题,其中的  $L$  具有明显的意义和一定的取值范围,迭代的方法得到的结果可能会超出这个范围.而本文的目的是得到允许取值范围内的最优解,这是迭代方法所不能达到的.

(2) 从方程组本身而言,其解不一定是惟一解.迭代的方法只能得到局部最优解,并不一定是实际问题的解.只有当迭代初始值选得比较好时,才能够收敛到所需要的解.

### 1.5 遮挡情况

在某些情况下,人体各部分之间可能存在着遮挡,导致一只手不可见.这时,可以在立体图像中检测实验者的双眼,计算双眼的深度信息,以其均值近似双肩中点  $N$  的深度值.由于  $N$  是两个肩关节的中点,因此,图像坐标近似为两个肩关节坐标的均值.由于实验者面对摄像机,所以双肩深度值之差相对而言极其微小,误差  $N$  点坐标值误差可以忽略不计,

$$Z_2 = (Z_2 + Z_3) / 2 \approx (Z_{l\_eye} + Z_{r\_eye}) / 2, \quad (8)$$

其中  $Z_{l\_eye}, Z_{r\_eye}$  分别为左、右眼睛的深度信息.此时为两个  $n=2$  的端点固定的关节模型.

### 1.6 模型的推广应用

本模型可以推广更加广泛的情况.各个关节的图像坐标可以采用不同视图中的坐标,此时,公式(6)需要变化为

$$P_i = T_i + R_i k_i \vec{p}_i. \quad (9)$$

其中  $(T_i, R_i)$  为图像坐标  $p_i$  所在视图的摄像机外参数,  $T_i$  为 3D 平移分量,  $R_i$  为 3D 旋转分量,可以通过摄像机的标定来获得.在多个视图的情况下,只要关节在一个视图中可见,就可以恢复其 3D 坐标,对于去除遮挡具有重要意义.

## 2 方程组求解

本文将方程组(7)转化成多个简单的方程组,针对方程组自身的特点,把方程组降维成一元多次方程,利用搜索的方法求得最优解,然后判别解的合理性.

### 2.1 解数目的分析

由方程组(7)可得

$$\begin{cases} k_1 = a_1 \pm \sqrt{b_1 + c_1 L} \\ k_i = a_i k_{i-1} \pm \sqrt{b_i k_{i-1}^2 + c_i L} \\ L = \frac{(k_n \bar{x}_n - X_B)^2 + (k_n \bar{y}_n - Y_B)^2 + (k_n \bar{z}_n - Z_B)^2}{m_n^2} \end{cases}, \quad i=2,3,\dots,n, \quad (10)$$

其中  $a_i, b_i, c_i$  是与  $k_i, L (i=1, \dots, n)$  无关的参数.方程组(10)实际上是  $2^n$  个方程组,在有意义的解空间中,每个方程组最多有一个最优解,则最多可能有  $2^n$  组解.方程组(10)最多可能有  $2^n$  组解,但并不是每组解都是合理的.

### 2.2 方程组的求解

对方程组(10)中任意一组而言,总共有  $n+1$  个未知参数  $k_i (i=1, \dots, n), L$ .实际上,从公式中可以看出,只要确定  $k_n$  或者  $L$ ,其他参数可以依次计算获得.所以,本文将方程组看成是一元多次方程:

$$L = \frac{(k_n \bar{x}_n - X_B)^2 + (k_n \bar{y}_n - Y_B)^2 + (k_n \bar{z}_n - Z_B)^2}{m_n^2}. \quad (11)$$

其中,只有  $L$  是未知参数,  $k_i (i=1, \dots, n)$  是中间变量,可以用  $L$  表示.计算使式(11)误差  $E(L)$  最小的  $L$  作为最优解.

$$E(L) = (L - \frac{(k_n \bar{x}_n - X_B)^2 + (k_n \bar{y}_n - Y_B)^2 + (k_n \bar{z}_n - Z_B)^2}{m_n^2})^2. \quad (12)$$

由于人体的身高有一个合理的范围,如 150cm~200cm,所以相对而言,未知参数  $L$  也有一个合理范围,离散化后搜索取得误差最小的  $L\_Resolution$  即为方程的最优解.离散化的精度以远大于初始条件的精度为原则,具体由下面的方法确定:

- 人体距离摄像机的最近距离可以定为  $d$ ;
- 由于各个关节点的图像坐标都是离散的,其误差为 0.5 个像素.已知摄像机的比例参数为  $s$ ,所有这 0.5 个像素对应于人体长度为  $0.5 \times d/s$ .未知参数  $L$  的搜索间隔只需要远小于  $0.5 \times d/s$ ,就可以保证得到结果的精确性.
- 在本文的实验中,一般只需要搜索 200 个采样点.

### 2.3 解的合理性判据

在区间  $[L\_low, L\_high]$  上,在各个方程组有解的情况下,获取各个方程组的最优解  $L\_Resolution$ ,然后从结果误差和位姿两方面来判定解的合理性.只有当两个条件都满足时,  $L\_Resolution$  才是所求的解.由于本文是逆过程求解,理论上至少有一组解满足实际问题.

## 3 实验结果和结论

原始图像的分辨率为  $384 \times 288, 24\text{bits}$ .在本文的人体上肢模型中  $n=4$ ,最多可能有 16 组解.人体骨骼  $P_i P_{i+1}$  ( $i=0, \dots, 4$ ) 长度比例关系分别为 28.26, 36.9, 47.52, 36.9, 28.26. 当以 cm 为单位时,上述比例数字相当于身高 180cm 的人的骨骼长度.

采用一组模拟数据,见表 1,其中的骨骼长度采用身高为 175cm ( $L=0.9722$ ) 的人的值.

Table 1 The simulated data

表 1 实验模拟数据

$P_i$	3D coordinates ( $X_i, Y_i, Z_i$ )	Image coordinates ( $x_i, y_i$ )
0	(0.0000, -20.0000, 600.0000)	
1	(10.6163, -6.6217, 621.5219)	(203.2393, 151.0104)
2	(22.0245, -0.3894, 654.9588)	(214.1267, 144.3912)
3	(68.2245, -0.3894, 654.9588)	(260.5413, 144.3912)
4	(92.6618, -17.1078, 675.2153)	(282.2992, 160.6717)
5	(103.2781, -30.4861, 696.7372)	

3D 空间坐标, 图像坐标.

人体身高范围:150cm~200cm,每 0.25cm 一个采样点,总共 200 个采样点.分别计算 16 组解,判别其合理性,以 PIII933 的计算机处理,算法没有任何优化,总共需要 2.688 毫秒,远远超过实时.最后,只得到一组合理的解.见表 2.

Table 2 Experimental result

表 2 实验结果

$P_i$	3D coordinates ( $X_i, Y_i, Z_i$ )
1	(10.6165, -6.6218, 621.5333)
2	(22.0253, -0.3894, 654.9825)
3	(68.2402, -0.3895, 655.1096)
4	(92.6838, -17.1119, 675.3759)

3D 空间坐标.

在结果中  $L=0.9725$ ,骨骼长度相当于身高为 175.057 2cm,误差为 0.032 68%,  $P_1 \sim P_4$  与模拟数据的 3D 空间坐标误差平均为 0.031cm,这是一个相当微小的误差.

图 3 为一组实际应用结果,其中图 3(a)是主视图图像,“+”处为双手和各个关节点图像坐标(关节点的图像坐标为手工标注),图 3(b)为辅助视图,图 3(c)为重构上肢姿态的正视图,图 3(d)为重构上肢姿态的俯视图.将各个关节点投影回图像中,计算关节点  $P_1 \sim P_4$  实际投影点与已知条件之间的平均误差小于 0.000 1 个像素.

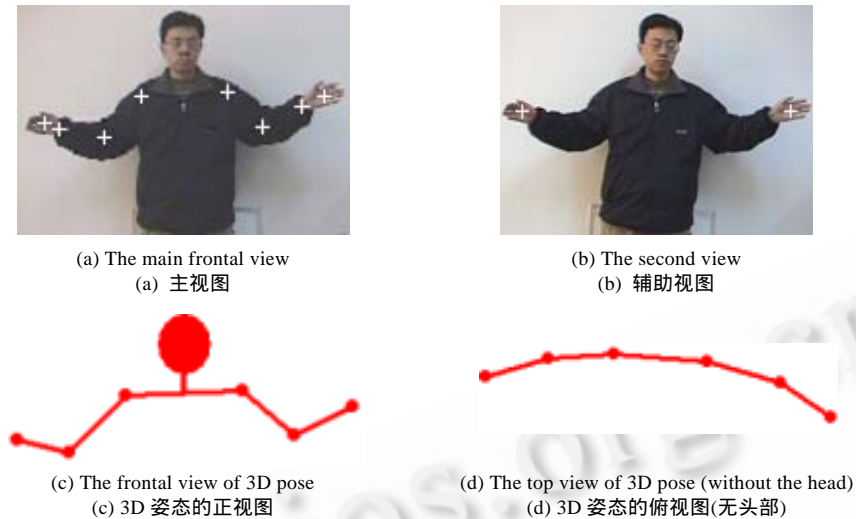


Fig.3 The experimental result of real data

图3 真实图像的实验结果

在计算机视觉领域,由于手和瞳孔的检测是比较成熟的,所以本文利用这些已有成果,在立体图像中检测双手等的深度信息,建立端点固定的关节模型,以此来估计上肢 3D 姿态.由此增加约束条件,准确地获取人体上肢各个关节的 3D 坐标,恢复姿态信息.另外,在实际应用中,并不是每个约束条件都可以准确获取的.例如,一只手被挡住,仍然可以通过双眼两个约束条件求得未知参数,所以具有很好的鲁棒性.

#### References:

- [1] Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 1973,14(2): 201~211.
- [2] Lakany, H.M., Haycs, G.M., Hazlewood, M.E., *et al.* Human walking: tracking and analysis. *IEE Colloquium on Motion Analysis and Tracking* (Ref. No.1999/103), 1999. 1~14. <http://ieeexplore.ieee.org>.
- [3] Davis, J.W., Bobick, A.F. The representation and recognition of action using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001,23(3):257~267.
- [4] Wren, C., Azarbajani, A., Darrell, T., *et al.* Pfunder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997,19(7):780~785.
- [5] Gavrilu, D.M. The visual analysis of human movement: a survey. *Computer Vision and Image Understanding*, 1999,73(1):82~98.
- [6] Leung, M.K., Yee, Hong-yang. A model based approach to labelling human body outlines. In: *Proceeding of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects*. 1994. 57~62. <http://ieeexplore.ieee.org>.
- [7] Ren, Hai-bing, Xu, Guang-you. Articulated-Model based upper-limb pose estimation. In: *Proceedings of the 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation*. 2001. 450~454. <http://ieeexplore.ieee.org>.
- [8] Hel-Or, Y., Werman, M. Constraint fusion for recognition and localization of articulated objects. *International Journal of Computer Vision*, 1996,19(1):2~28.
- [9] Gavrilu, D.M., Davis, L.S. 3-D model-based tracking of humans in action: a multi-view approach. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1996. 73 ~80. <http://ieeexplore.ieee.org>.
- [10] Kakadiaris, I., Metaxas, D. Model-Based estimation of 3D human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000,22(12):1453~1459.
- [11] Ren, Hai-bing, Xu, Guang-you, Zhu, Yuan-xin, *et al.* Motion-and-Color based hand segmentation and hand gesture recognition. In: *Proceedings of the 1st International Conference on Image and Graphics*. 2000. 384~388.

## Human Upper-Limb Pose Estimation and Multi-Resolution Analysis\*

REN Hai-bing, XU Guang-you

*(Institute of Human-Computer Interaction and Media Integration, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)*

E-mail: renhaibing@263.net; xgy-dcs@mail.tsinghua.edu.cn

<http://www.tsinghua.edu.cn>

**Abstract:** In physics, human body is a very complex deformable object with many degrees of freedom. Only with 2D appearance features, there exist a lot of difficulties in subtle motion analysis and recognition, and it is even more difficult to understand the user purpose. Since the gesture is an important approach for communication, based on 3D model, 3D pose of upper-limbs is reconstructed in this paper. For practical application and error reduction, the minimal conditions for pose estimation are analyzed, under which an end-fixed articulate model and the corresponding equations are proposed for modeling the upper-limb and estimating the 3D coordinates of each joints. Then, maximal number of the resolutions is analyzed and an approach is given to obtain these ones. At last, two principles on the result error and the pose correctness are utilized for verification.

**Key words:** pose estimation; articulated-model; end-fixed; 3D reconstruction

---

\* Received April 12, 2001; accepted August 1, 2001

Supported by the National Natural Science Foundation of China under Grant No.69873022; the National High Technology Development 863 Program of China under Grant No.863-306-ZT03-01-1