

基于区域的手指三维运动跟踪

潘春洪 马颂德

(中国科学院自动化研究所模式识别国家重点实验室, 北京 100080)

摘要 提出了基于区域的多连接体(手指)的三维运动跟踪算法。该算法首先用多约束融合的方法以及手指的运动特性, 得到初始帧手指的三维结构; 然后根据刚性多连接体的运动模型以及相应的姿势约束模型, 给出了这一特殊运动模型三维运动估计的优化算法, 此算法能够鲁棒地估计手指的三维运动; 最后利用区域跟踪的方法获取多连接体三维运动, 并在真实的手指序列图象中实现了该算法, 实验结果证实了该算法的有效性。

关键词 模式识别(520·2040) 三维运动跟踪 约束融合 刚性多连接体 图象序列
中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2003)10-1205-06

Region-Based 3D Motion Tracking of Hand

PAN Chun-hong, MA Song-de

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080)

Abstract In this paper, a algorithm of region-based 3D motion tracking on multi-linkage is proposed. Firstly based on the fact that the length of fingers is same in motion, and each of them lies in the same virtual plane because of only one degree of freedom at their joints, which is defined as the rigidity constraint and coplanar constraints respectively, we obtain the initial 3D structure of finger from the image by exploiting these constraints and special properties on fingers. Then by the motion properties of rigidity multi-linkage and corresponding gesture constraints, we give the general descriptions of 3D region tracking of the linkage. Finally we apply the techniques to 3D motion tracking of hand. Here genetic algorithm is employed to our optimal iteration. In order to verify the validity of our approach, we use 3D motion of fingers, which is obtained from the real image sequence, to drive our virtual model of fingers. The results demonstrate the validity of the method.

Keywords Pattern recognition, 3D motion tracking, Constraint fusion, Rigidity multi-linkage, Image sequence

0 引言

手势识别和手的运动跟踪在人机交互技术中起着重要的作用, 但由于各种原因, 手的三维运动跟踪显得非常困难。迄今为止, 除了运动捕捉系统(光学、电磁波)和数据手套以外, 几乎没有特别鲁棒的手的三维运动跟踪方法, 更谈不上对手的三维运动进行实时跟踪。其主要原因是:

(1) 手指的三维运动是一种既非形变也非刚体的多自由度的运动, 因此在跟踪过程中需要同时估计很多个运动参数;

(2) 人的观察力敏锐, 任何微小的错误运动都很容易被看出, 因此在实际的运动跟踪中需要获取非

常精确的三维运动;

(3) 手运动时, 仅仅通过一个摄像机是不可能始终跟踪手的所有部分, 因其某一部分经常会遮挡另一部分, 由于遮挡, 被遮挡部分的信息就会在图象序列中丢失, 从而不可能跟踪或估计出被遮挡部分结构的三维运动。

目前, 主要有以下3种三维运动估计的方法:

(1) 基于特征点 这种方法的最大困难在于如何从人的运动序列图象中获取高精度的匹配特征点^[1,2]。虽然此类方法中的运动捕捉系统和数据手套比较有效, 但该系统需要在关节上安放特殊的标志和穿戴特殊的设备, 以便在序列图象中准确跟踪这些特殊的标志^[3]。正是因为需要安放特殊的标志和穿戴特殊的设备, 因而极大地约束了该系统的应用范围。

(2) 基于边缘特征 这种方法通常需要比较简单的背景,以便能够从背景中分割出运动物体的边缘^[4,5]。但通常情况下,即使有简单的背景,也很难做到有效的分割。因为衣服的许多皱褶和皮肤的相似性足以使许多图象分割的方法不可能从背景中分割出有语义的身体和手指的各个部分。

(3) 基于区域特征 这种方法绝大部分是通过基于光流的匹配模板来实现的^[6],其跟踪误差是逐步积累的,因而随着跟踪帧数的增加,误差也越来越大,从而产生错误的匹配,而一旦产生错误匹配,由于没有反馈过程,就不可能回到正确的匹配中去。

事实上,手指的运动有着一定的动态和静态约束关系,用这种约束关系可以修正在跟踪过程中所造成的错误匹配,从而形成一个有效的反馈过程,实现鲁棒的三维运动跟踪。

1 刚性多连接体的运动分析

1.1 单区域的运动分析

考虑到图象像素 $m = (x, y)^T$ 在时刻 t 的灰度值为 $I(x, y, t)$, 点 m 的图象运动速度为 $v_m = \dot{m} = (u_x, u_y)^T$, 假设在很短的时间间隔内(即两帧之间)点的灰度值保持不变,有:

$$I(x + u_x(x, y, \varphi), y + u_y(x, y, \varphi), t + 1) = I(x, y, t) \quad (1)$$

这里, $u(x, y, \varphi)$ 是运动模型, φ 为运动参数, 如果运动模型是线形的, 当为二维仿射运动时, 参数可以被定义为

$$u(x, y, \varphi) = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \quad (2)$$

式(1)的一阶泰勒展开可以写成

$$I_t(x, y) + [I_x(x, y), I_y(x, y)] \cdot u(x, y, \varphi) = 0 \quad (3)$$

进一步可以写成

$$\varphi^T = - (H^T \cdot H)^{-1} \cdot H^T Z \quad (4)$$

其中, $H = [(\nabla I_1)^T \cdot X_1, (\nabla I_2)^T \cdot X_2, \dots, (\nabla I_N)^T \cdot X_N]^T$; $Z = [\Delta I_1, \Delta I_2, \dots, \Delta I_N]^T$, 这是仿射模型的运动估计, 可详见文献[7]、[8]。 I_t, I_x, I_y 分别是两幅图象之间时域和空域的灰度变化, 可以用两幅图象间的灰度差来求得。

1.2 多刚体的三维运动描述

物体在摄像机坐标系下和世界坐标系下的坐标变换关系可以写成

$$q_c = G_0 q_0 \quad (5)$$

其中, $q_0 = [x_0, y_0, z_0, 1]^T$ 是摄像机坐标系下的齐次

坐标, $q_c = [x_c, y_c, z_c, 1]^T$ 是世界坐标系下的齐次坐标, G_0 是两者间的变换关系, 详见文献[9]。

为简单起见, 假设摄像机模型是带有尺度的平行投影, 那么点投影到图象上的图象坐标可以写成 $[x_m, y_m] = \lambda [x_c, y_c]$, 这里 λ 是尺度, $[x_m, y_m]$ 是图象坐标。单刚体运动有 6 个自由度(3 个旋转, 3 个平移), 因此要估计单刚体的运动, 需要决定的参量有 $[\lambda, v_1, v_2, v_3, \omega_x, \omega_y, \omega_z]$, 其中, $[v_1, v_2, v_3]$ 为平移矢量, $W = [\omega_x, \omega_y, \omega_z]$ 为旋转矢量。文献[6]给出了单刚体的三维运动估计, 即

$$I_t + H_t \cdot [\lambda, v'_1, v'_2, v'_3, \omega'_x, \omega'_y, \omega'_z]^T = 0 \quad (6)$$

这里, $H_t = [x_t I_x + y_t I_y, I_x I_y, -z_t I_x, z_t I_x - y_t I_y + x_t I_y]$ 。对于每一个像素都有上述方程, 这样当有 N 个像素时, 就有 N 个如式(6)的方程, 因此, 上式能被写成矩阵的形式

$$H \cdot \varphi + Z = 0 \quad (7)$$

$\varphi = [\lambda, v'_1, v'_2, v'_3, \omega'_x, \omega'_y, \omega'_z]^T$ 为所求的变量, I_t, I_x, I_y 可以从两幅图象的灰度变化中求出, 见式(3)。 $q_c = [x_c, y_c, z_c, 1]^T$ 是摄像机坐标系下物体的三维坐标, 这里, 利用文献[10]中提出的方法对多连接体进行初始化分析, 然后用得到的三维结构去构造一圆柱体的多连接体模型, 这样, 就可以得到 q_c 。事实上, 只要知道了初始帧物体的三维坐标, 就可以利用式(4)计算出相应的旋转、平移和尺度, 从而进一步求出一帧的 q_c 。

进一步地, 可以把刚性多连接体看成一个共面的旋转链, $\theta_i (i = 1, 2, \dots, k)$ 表示相邻两连接体之间的夹角, 其三维运动估计一般为

$$I_t + H_t \cdot [\lambda', v'_1, v'_2, v'_3, \omega'_x, \omega'_y, \omega'_z]^T +$$

$$J_i \cdot [\dot{\theta}_1, \dot{\theta}_2, \dots, \dot{\theta}_k]^T = 0 \quad (8)$$

这里, $J_i = [J_{i1}, J_{i2}, \dots, J_{ik}]$, 而

$$J_{ik} = [I_x, I_y] \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \xi_k \cdot q_c \quad (9)$$

其中, $\xi_k = \begin{bmatrix} -W_k \times q_c \\ W_k \end{bmatrix}$

式(8)可以写成矩阵形式

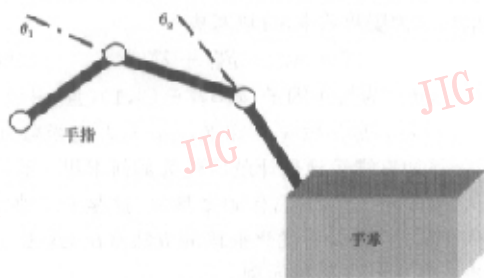
$$[H, J] \varphi + Z = 0 \quad (10)$$

其中, $\varphi = [\lambda', v'_1, v'_2, v'_3, \omega'_x, \omega'_y, \omega'_z, \dot{\theta}_1, \dot{\theta}_2, \dots, \dot{\theta}_k]$, $\lambda', v'_1, v'_2, v'_3, \omega'_x, \omega'_y, \omega'_z$ 和 $\dot{\theta}_1, \dot{\theta}_2, \dots, \dot{\theta}_k$ 是未知变量。一旦求出这些变量, 就获得了这种刚性多连接体的三维运动描述。虽然刚体与刚体之间的相对运动有着特殊的关系, 但其整体结构的运动是任意的, 以上详细

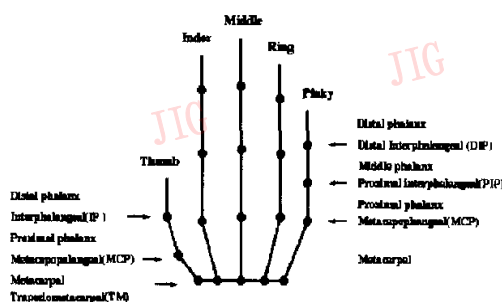
的推导过程见文献[6]。虽然该算法不需要优化迭代,但它的误差是积累的,而且在整个算法实现中没有反馈过程,所以,随着被处理帧数的增加,其跟踪的区域将逐渐偏离实际跟踪的目标,从而得出错误的结果,而且一旦产生错误匹配,由于没有反馈过程,就不可能回到正确的匹配中去,因此必须在该算法的基础上增加约束。

2 手指的三维模型和其约束关系

一般而言,手指可以被看成是棍棒模型^[11]。每根手指都由 3 节组成(拇指除外),因而有 3 个关节点,但每个关节点的自由度是不一样的,图 1(a)给出了手指的简单模型。对于所有的指关节,除与手掌相连的关节外,其余关节都只有一个自由度(见图 1(a)中的 θ_1 或 θ_2),因而可以认为它们是在同一个平面上运动,即共面约束。进一步地,不同指节间都有相互约束,这种约束可以分为静态约束和动态约束两种。静态约束是指手指有着一定的相对运动范围,即如果把每一个手指的第 1 节(与手掌相连)定义成 MCP(见图 1(b)),Y 方向表示手指的伸缩,而 X 方向表示手指的局部拐动,那么 MCP 相对手掌的运动范围就可以表示成:在 X 方向的活动范围为 $[-15^\circ, 15^\circ]$,而在 Y 方向的活动范围为 $[0, 90^\circ]$ 。动态约束是指某一指节运动时,必定会带动另外一节运动,如表 1 所示动态约束的最后一项表示各节手指骨之间在 X 方向的运动约束关系,其他项表示各节手指骨之间在 Y 方向的运动约束关系(右栏仅对大拇指而言,左栏针对其他各手指,表中相应的符号见图 1)。例如每一个手指第 2 指节的运动必定会带来第 3 指节的运动,而且两指节的旋转角之间满足关系式: $\theta_{PIP}^y = \frac{3}{2}\theta_{DIP}^y$, θ_1 和 θ_2 的关系如图 1 所示。另外,表中左栏相应的第 2 项是关于各手指的第 1 指节在 X 方向运动的动态约束关系,由于可以通过 3 节手指(已知相互间的长度关系)来获取其三维结构,同时利用上面的角度动态约束来消除它的歧义性,因此,这一项就不再作详细介绍,各种约束关系如表 1 所示^[12]。事实上,这里用到了 3 类手指约束:刚体约束、共面约束,以及在 X/Y 方向的动态依赖关系的约束。利用这些约束关系,可以估计手指的初始三维结构,以及在后面的区域估计中修正所估计出的三维运动。



(a) 手指的简单模型



(b)

图 1 手指模型

表 1 手指的静态和动态约束

	手指	拇指
静态约束	$0 \leq \theta_{MCP}^y \leq 90^\circ, -15^\circ \leq \theta_{MCP}^x \leq 15^\circ$	
动态约束	$\theta_{PIP}^y = \frac{3}{2}\theta_{DIP}^y$	$\theta_{IP}^y = \theta_{MCP}^y$
	$\theta_{MCP}^x = \frac{\theta_{MCP}^y}{90}(\theta_{MCP}^x, \text{converge} - \theta_{MCP}^y) + \theta_{MCP}^x$	$\theta_{TM}^y = \frac{1}{3}\theta_{MCP}^y$ $\theta_{TM}^x = \frac{1}{2}\theta_{MCP}^x$

3 手指三维运动的优化估计算法

事实上,在真实的三维运动估计中,由于图象噪声的影响,以及公式推导过程中的许多近似假设,式(10)的关系不可能得到满足,因此该方程可写为

$$\zeta_1 = [H, J] \cdot \varphi + Z \quad (11)$$

这里, ζ_1 为运动约束。另外,上述的静态和动态约束关系也是近似的,其误差可以表示成

$$\zeta_2 = h(\theta_s^y), p = x, y; s = MCP, PIP, DIP \quad (12)$$

再则,对于多连接体的所有刚体,其长度在三维运动中应相应地保持不变,刚体约束方程用误差的形式可以表示成^[13]

$$\zeta_3 = g(q_c, V) \quad (13)$$

其中, $V = [v_1, v_2, v_3, \omega_1, \omega_2, \omega_3]$ 为单刚体的旋转和

平移矢量, q_i 为刚体初始帧的三维姿势. 上述的二种约束称之为姿势约束, 可以写成

$$\zeta = \alpha\zeta_1 + \beta\zeta_2 + \gamma\zeta_3 \quad (14)$$

这里, α, β, γ 为权重因子, $\alpha, \beta, \gamma \in (0, 1)$, 且 $\alpha + \beta + \gamma = 1$, 误差 ζ 是参数 λ, V 和 $\theta_1, \theta_2, \dots, \theta_k$ 的函数, 因此, 上述的三维运动估计就转化为如何求得一组 λ, V 和 $\theta_1, \theta_2, \dots, \theta_k$ 的值, 使得 ζ 最小, 这是一个典型的优化问题, 可以用优化编程的方法解决, 这里, 应用遗传算法^[14]求解该问题.

将误差准则 ζ 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: $f =$

$$\frac{\zeta}{1+\zeta}$$

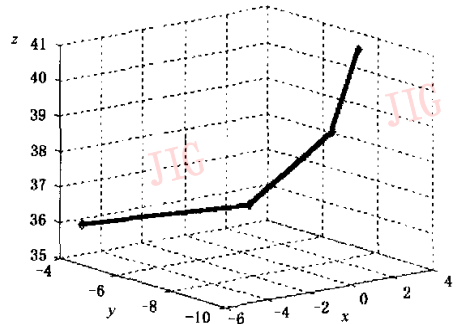
. 对待优化参数 λ, V 和 $\theta_1, \theta_2, \dots, \theta_k$ 进行二进制编码, 每个参数编码的长度为 $l_i = 10$, 各参数的寻优区间为 $\lambda \in (\lambda_{\min}, \lambda_{\max}), V \in (V_{\min}, V_{\max}), \theta_1, \theta_2, \dots, \theta_k \in (0, 90^\circ)$. 在文献^[15]算法中, 交叉概率为 15%, 变异概率为 0.5%.

4 算法流程及相应的实验结果

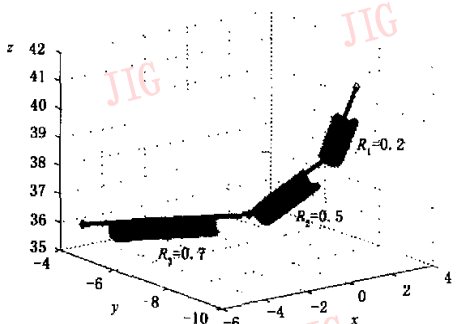
算法的基本思路是: 首先利用共面约束、刚体约束, 以及运动的动态约束, 对手指的三维结构进行初始化分析, 同时, 选择出合适的图象跟踪区域, 利用上述的优化方法跟踪它的三维运动, 需要说明的是:

(1) 在文献^[12]、^[14]中, 论述了当有 3 个连接体且其相互间的长度关系已知时, 就可以用共面约束和刚体约束的方法计算出其相应的三维结构. 因此可以用该方法得到手指初始帧的三维结构, 图 2(a) 是所获得的初始帧手指的三维结构, 此时的三维结构是摄像机坐标系下的. 得到了初始三维结构后, 就可以用圆柱体来模拟手指的结构, 圆柱体的半径大小必须保证该圆柱体反投到手指图象的区域不能超过图象中手指的区域. 通过手指图象某区域的投影线相交于给定圆柱体, 一般有两个交点, 取距离图象最近的那一点作为 q . 图 2(b) 给出了带有圆柱体的手指的三维初始结构, 图 3 给出了反投回相应图象的区域, 图 4 说明经计算得到的初始三维结构是正确的, 并给出了与图 3 相对应的圆柱体的半径. 上述分析表明, 可以通过增大圆柱体的半径来扩大图象中的相应区域, 从而增加跟踪区域的信息量.

(2) 每节手指的跟踪区域是不同的(图 3, 图 4), 第一节手指的跟踪区域最大, 相应的像素点个数也最多, 三个区域的像素点个数分别为 1523、856、353.



(a) 不带有圆柱体



(b) 带有圆柱体

图 2 初始帧手指的三维结构

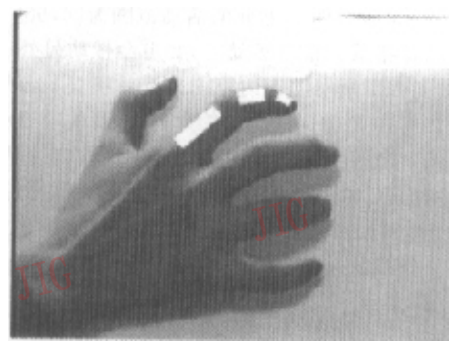


图 3 反投回相应图象的区域

在每帧的跟踪过程中, 其每个区域的像素点个数是不变的; 另外, 在跟踪过程中不需要已知摄像机的内参数, 也即摄像机的内参数在拍摄过程中是可以变化的. 最后, 由于选中的区域并没有用到边缘信息, 因而不需要进行图象的分割, 同时也不需要特征点的对应, 因此该方法在这方面显示了极大的优越性. 同时, 由于在跟踪过程应用了刚体约束和手指相互之间的静态和动态约束, 因此又极大地提高了跟踪的鲁棒性.

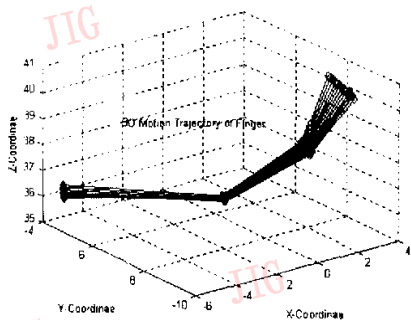


图 4 跟踪的手指的三维运动轨迹

用 CCD 摄像机来拍摄人手的运动图象,图 5 给

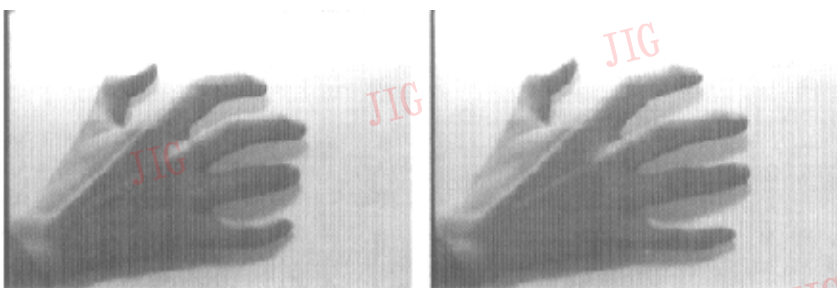


图 5 人手的运动图象

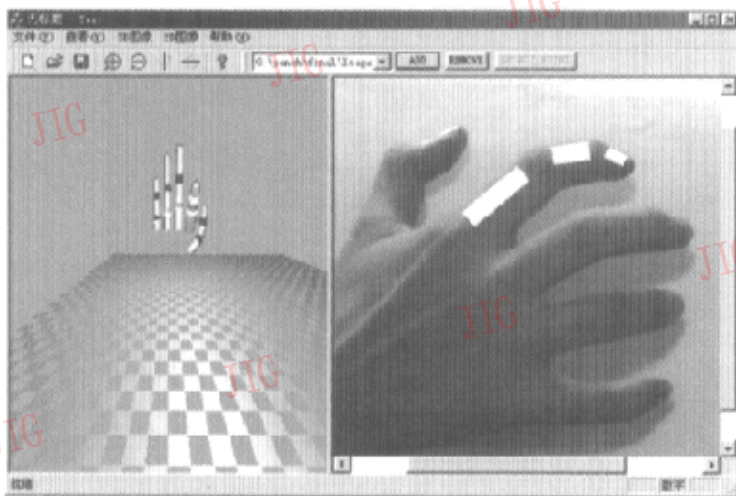


图 6 程序的基本框架,左边框中的示图是简单的三维虚拟手指,右边的是显示的序列图象

出了其中的 2 帧;然后,用上述方法来求取人手的三维运动,图 4 给出了跟踪的手指的三维运动轨迹,图中右边上翘的部分为指尖端,相对应的另一端为与手掌相联的指末端;最后,用该方法获取的三维数据去驱动人手的三维模型.如图 6 所示是程序的基本框架,左边框中的示图是简单的三维虚拟手指,右边的是显示的序列图象.

本文利用基于区域的多约束融合的方法实现了手指的三维运动跟踪,由于该方法利用了手指的运动特性以及相应的姿势约束模型,所以它能鲁棒地估计手指的三维运动,用真实的序列图象实现了该算法;实验结果证实了该方法的有效性.

参考文献

- 1 Caia, Aggarwal J K. Automatic tracking of human motion in indoor scenes across multiple synchronized video stream[A]. In: Proceeding of International Conference on Computer Vision[C], Bombay, India, 1998:156~162.
- 2 Huber E. 3D Real time gesture recognition using proximity space[A]. In: Proceeding of International Conference on Pattern Recognition[C], Vienna, Austria, 1996:136~141.
- 3 Cameron G, Bustanoby A, Cope K et al. Panel on motion

- capture and character animation [A]. In: SIGGRAPH'97 [C], Los Angeles, California, USA, 1997: 442~445.
- 4 Rowley H A, Rehg J M. Analyzing articulated motion using expectation-maximization [A]. In: Proceeding of International Conference on Pattern Recognition [C], Puerto Rico, 1997: 935~941.
- 5 Leung M K, Yang Y H. First sight: A human body outline labeling system [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(4): 359~377.
- 6 Horn B K P, Schunck B G. Determining optical flow [J]. Artificial Intelligence, 1981, 17: 185~204.
- 7 Verri A, Poggi T. Motion field and optical flow: Qualitative properties [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(5): 490~498.
- 8 马颂德, 张正友. 计算机视觉 [M]. 北京: 科学出版社, 1998.
- 9 Lee J, Kuni T L. Model-based analysis of hand posture [J]. IEEE Computer Graphics and Applications, 1995, 15(9): 77~86.
- 10 Kuch J J. Vision based hand modeling and gesture recognition for human computer interaction [D]. Urbane-Champaign, Illinois, USA, University of Illinois at Urbane-Champaign, 1994.
- 11 Murray R M, Li Z, Sastry S S. A mathematical introduction to robotic manipulation [M]. Florida, USA, CRC Press, 1994.
- 12 PAN C, MA S. 3D motion analysis based on coplanar constraints [A]. In: Proceeding of International Conference on Pattern Recognition [C], Barcelona, Spain, 2000: 159~163.
- 13 Bregler C, Malik J. Tracking people with twists and exponential maps [A]. In: Proceeding of International Conference on Computer Vision and Pattern Recognition [C], Santa Barbara, California, USA, 1998: 276~282.
- 14 潘春洪. 单目序列图像中人的三维运动分析 [D]. 北京: 中国科学院自动化研究所, 2000.
- 15 Goldberg D E. Genetic algorithms in search, optimization and machine learning, reading [M]. Boston, Massachusetts, USA: Addison Wesley, 1989.



潘春洪 1964年生, 1987年获清华大学自动化系工学学士学位, 2000年获中国科学院北京自动化研究所工学博士学位, 2000年7月至2001年12月, 作为访问学者, 工作于美国南加州大学计算机系. 研究方向为计算机视觉、图象处理和模式识别.



马颂德 1946年生, 研究员, 博士生导师, 1968年毕业于清华大学自控系, 1983年获法国第六大学工程博士学位, 1985年获该校国家博士学位, IEEE高级会员, 现任国家科技部副部长. 研究方向为计算机视觉、图象处理和模式识别.