基于区域的手指的三维运动跟踪

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

Email: chpan@nlpr.ia.ac.cn

摘要: 在本文中，我们介绍了基于区域的多连接体（手指）的三维运动跟踪。我们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，最后利用区域跟踪的方法获取它的三维运动，并在真实的手指序列图像中实现了我们的算法；实验结果证实了该方法的有效性。

关键词: 三维运动跟踪, 约束融合, 刚性多连接体, 图像序列

Region-Based 3D Motion Tracking of Hand

Abstract: In this paper, we propose a region-based 3D motion tracking on multi-linkage. Firstly by constraint fusion and special properties on fingers, we obtain the initial 3D structure of finger from the image, and then by the motion properties of rigidity multi-linkage and corresponding gesture constraints，we give the general descriptions of 3D region tracking of this linkage, finally we apply the techniques to 3D motion tracking of hand. The experiment with real images is included to demonstrate the validity of the theoretic results.

Key Words: 3D motion tracking, Constraint fusion, Rigidity multi-linkage, Image sequence

引言

手势识别和手的运动跟踪在人机交互中技术中起着重要的作用，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。这方面存在的主要困难有：

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。另外由于人的观察力敏锐，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。再者，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，其某一部分经常会遮挡另一部分，由于遮挡，这部分信息就会在图像序列中丢失，从而不可能跟踪或估计出这部分结构的三维运动。

目前，主要有三种三维运动估计的方法：

基于特征点：这种方法的最大困难在于如何从人的序列运动图像中获取高精度的特征点匹配 [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，以便在序列图像中准确跟踪这些特殊的标志 [3]。这种系统由于需要安放特殊的标志和穿戴特殊的设备，

因而极大地约束了它的应用范围。

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，这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像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。

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 [13]， 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。

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。在本文中，我们详细地讨论了基于区域特征和约束融合的运动跟踪方法，并以手指的运动为例来具体实现我们的算法。我们首先分析刚性多连接体的运动特殊性，给出了这一特殊运动模型一般的三维运动估计算法，为了得到初始帧手指的三维结构，我们利用了多约束融合的方法和手指运动的特殊性，最后，我们给出了真实图像的实验结果。

一：刚性多连接体的运动分析

1: 单区域的运动分析

考虑到图像在象素 $\mathbf{m}=(x,y)^T$ 时刻 t 的灰度值为 $I(x,y,t)$，点 m 的图像运动速度为 $\mathbf{V}_m=\dot{\mathbf{m}}=(u_x,u_y)^T$，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，我们有：

$$I(x+u_x(x,y,\phi),y+u_y(x,y,\phi),t+1)=I(x,y,t) \tag{1}$$

这里：$u(x,y,\phi)$ 是运动模型，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：

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

方程（1）的一阶台劳展开可以写成：

$$I_t(x,y)+[I_x(x,y),I_y(x,y)]\cdot\mathbf{u}(x,y,\phi)=0 \tag{3}$$

进一步可以写成：

$$\phi^T=-(H^T\cdot H)^{-1}\cdot H^T Z \tag{4}$$

其中: $H=[(\nabla I_1)^T\cdot\mathbf{X}_1,\ (\nabla I_2)^T\cdot\mathbf{X}_2,\ \ldots.\ (\nabla I_N)^T\cdot\mathbf{X}_N]^T$； $Z=[\Delta I_1,\Delta I_2,\ldots.,\Delta I_N]^T$； 这是仿射模型的运动估计, 详细见 [6][7]。 I_t， I_x， I_y 分别是两幅图像之间时域和空域的灰度变化，可以用两幅图像间的灰度差来求得。

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：

$$q_c = G_0 q_0 \tag{5}$$

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, 详细见[8]。

为简单起见，我们假设摄像机模型是带有尺度的平行投影，那么点投影到图像上的图像坐标可以写成：$[x_{im}, y_{im}] = \lambda[x_c, y_c]$, 这里 λ 是尺度, $[x_{im}, y_{im}]$ 是图像坐标。单刚体运动有六个自由度（三个旋转，三个平移）。因此要估计单刚体的运动，需要决定的参量有 $[\lambda, v_1, v_2, v_3, w_1, w_2, w_3]$,其中, $[v_1, v_2, v_3]$ 为平移矢量，$[w_1, w_2, w_3]$ 为旋转矢量。文章[13]给出了单刚体的三维运动估计的公式：

$$I_{\mathbf{t}} + H_{\mathbf{i}} \cdot [\lambda, v_1', v_2', \omega_x', \omega_y', \omega_z']^T = 0 \tag{6}$$

这里：$H_{\mathbf{i}} = [\, X_c \cdot I_x + Y_c \cdot I_y,\ I_x,\ I_y,\ -Z_c \cdot I_y,\ Z_c \cdot I_x,\ -Y_c \cdot I_x + X_c \cdot I_y \,]$； 对于每一个象素都有上述方程，这样当有 N 个象素时，就有 N 个上述的方程，因此，上式能被写成矩阵的形式:

$$H \cdot \phi + Z = 0 \tag{7}$$

$\phi = [\lambda, v_1', v_2', \omega_x', \omega_y', \omega_z']^T$ 为所求的变量，I_t， I_x， I_y 可以从两幅图像的灰度变化中求出，见公式（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，我们利用[12]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。详细的分析见后面。事实上，我们只要知道了初始帧物体的三维坐标，就可以利用上述迭代公式计算出相应的旋转、平移和尺度，从而进一步地求出下一帧的 q_c。

进一步地，对于刚性多连接体，我们可以把它看成一个共面的旋转链，$\theta_i \ (i = 1, 2, \cdots k)$ 表示相邻两连接体之间的夹角，其三维运动估计的一般式子为：

$$I_t + H_i \cdot [\lambda', v_1', v_2', w_x', w_y', w_z']^T + J_i \cdot [\dot{\theta}_1, \dot{\theta}_2, \cdots, \dot{\theta}_k]^T = 0 \tag{8}$$

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

$$J_{ik} = [I_x, I_y] \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \widehat{\xi}_k \cdot q_c \tag{9}$$

上式可以写成矩阵形式：

$$[H, J]\phi + Z = 0 \tag{10}$$

其中: $\phi = [\lambda', v_1', v_2', \omega_x', \omega_y', \omega_z', \dot{\theta}_1, \dot{\theta}_2, \cdots \dot{\theta}_k]$，式中 λ'， v_1'， v_2'， ω_x'， ω_y'， ω_z' 和，$\dot{\theta}_1$， $\dot{\theta}_2$， …，$\dot{\theta}_k$，是未知变量。一旦求出这些变量，我们就求出了这种刚性多连接体的三维运动。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[13]。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。

二：手指的三维模型和其约束关系

下面我们根据解剖学的知识，分析手指的特殊性，并给出相应的静态约束和动态约束，然后把这些约束用来估计手指的三维运动。

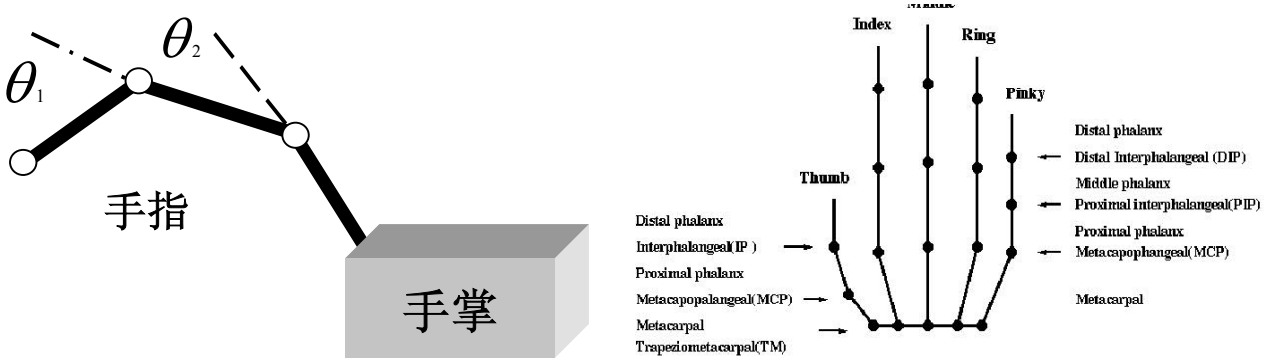

图1：手指模型

一般而言，手指可以被看成是棍棒模型[9]。每根手指都由三节组成（拇指除外），因而有三个关节点，但每个关节点的自由度是不一样的，图1（右）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----伸张和弯曲，因而我们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相对运动范围，即如果把每一个手指的第一节（与手掌相连）定义成MCP（见图1左），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$[0, 90^\circ]$。而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，其它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，左栏针对其它各手指，表中相应的符号见图1）。例如每一个手指第二节的运动必定会带来第三节的运动，而且两指节的旋转角之间满足关系式：$\theta_{PIP}^{y} = \frac{3}{2}\theta_{DIP}^{y}$，见图1所示$\theta_1$和$\theta_2$的关系。另外，表中左栏相应的第二项是关于各手指的第一节在X方向运动的动态约束关系。由于我们可以通过三节手指(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[10]。事实上,这里我们用到了三类手指约束：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。我们可以利用这些约束关系来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。

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

表 1：手指的静态和动态约束

三：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，由于图像噪声的影响，以及公式推导过程中的许多近似假设，方程（10）的关系不可能得到满足，因此该方程可以写成：

$$\zeta_1 = [H, J] \cdot \phi + Z \tag{11}$$

这里，我们称之为运动约束。另外，上述的静态和动态约束关系也是近似的，可以表示成：

$$\zeta_2 = h(\theta^{p}_{s}),\ p = x, y;\ s = MCP, PIP, DIP \tag{12}$$

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，我们可以表示成如下刚体约束方程，详见[14]：

$$\zeta_3 = g(q_c, \vec{V}) \tag{13}$$

其中：$\vec{V} = [\nu_1, \nu_2, \nu_3, w_1, w_2, w_3]$ 为单刚体的旋转和平移矢量，q_c 为刚体初始帧的三维姿势。上述的二种约束我们称之为姿势约束。因此，一般地我们可以写成：

$$\zeta = \alpha \cdot \zeta_1 + \beta \cdot \zeta_2 + \gamma \cdot \zeta_3 \tag{14}$$

这里，α， β， γ 为权重因子，$\alpha, \beta, \gamma \subset (0, 1)$，且 $\alpha + \beta + \gamma = 1$，误差 ζ 是参数 $\lambda, \vec{V}$ 和 $\theta_1, \theta_2 \ldots, \theta_k$ 的函数，因此，上述的三维运动估计就转化为如何求得一组 $\lambda, \vec{V}$ 和 $\theta_1, \theta_2 \ldots, \theta_k$ 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进行，这里，我们应用遗传算法求解该问题。

上述的误差准则 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 $\lambda, \vec{V}$ 和 $\theta_1, \theta_2 \ldots, \theta_k$ 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：$\lambda \in (\lambda_{\min}, \lambda_{\max})$，$\vec{V} \in (\vec{V}_{\min}, \vec{V}_{\max})$，$\theta_1, \theta_2 \ldots, \theta_k \in (0, 90^0)$。在我们的算法中，交叉概率为 15%，变异概率为 0.5%，详见[14][15]。

四：算法流程及相应的实验结果

这里，我们简述一下我们的算法，并给出相应的计算结果。

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，选择出合适的图像跟踪区域，利用上述的优化方法跟踪它的三维运动，需要说明的是:

(1): 在文章[12][14]中，我们论述过，当有三个连接体且其相互间的长度关系已知时，我们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。因此我们可以用该方法得到手指初始帧的三维结构，图 2(左)是所获到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，我们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像的区域不能超过图像中手指的区域。通过手指图像某区域的投影线相交于给定圆柱体，一般有两个交点，我们取距离图像最近的那一点作为 q。图 2(右)给出了带有圆柱体的手指的三维初始结构，图 3(左)给出了反投回相应图像的区域， 图 3(左)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(左)相对应的圆柱体的半径。上述分析表明，我们可以通过增大圆柱体的半径来扩大图像中的相应区域，从而增加跟踪区域的信息量。

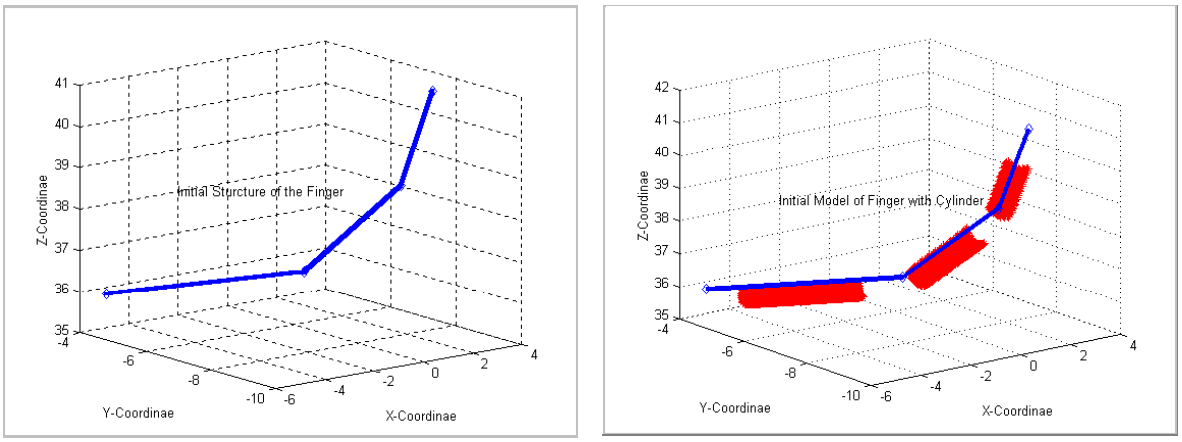

图 2：初始帧手指的三维结构：左：不带有圆柱体，右：带有圆柱体

(2): 每节手指的跟踪区域是不同的，见图 3(左)，第一节的跟踪区域最大，相应的象素点个数也最多，三个区域的象素点个数分别为 1523、856、353。在每帧的跟踪过程中，其每个区域的象素点个数是不变的；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，因而不需要进行图像的分割，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。

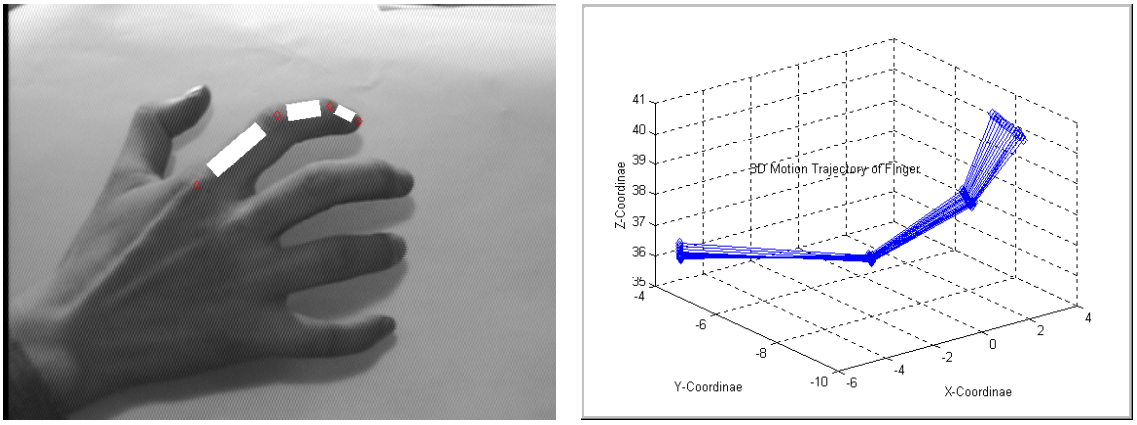

图 3：左：反投回相应图像的区域，右：跟踪的手指的三维运动轨迹

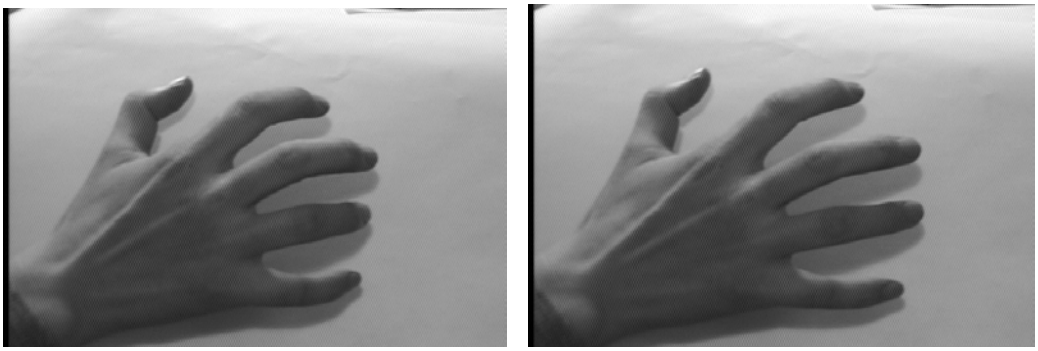

图 4：人手的运动图像

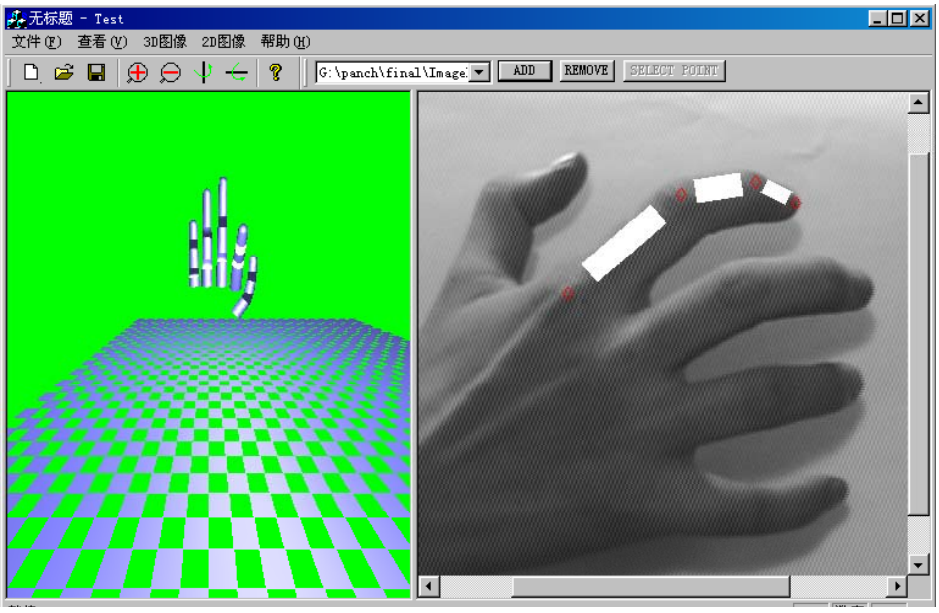

图5： 程序的基本框架，左边框中的示图是简单的三维虚拟手指，右边的是显示的序列图像

我们用 CCD 摄像机来拍摄人手的运动图像，图 4 给出了其中的两帧；然后，我们用上述方法来求取人手的三

维运动，图 3(右)给出了我们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，我们用该方法获取的三维数据去驱动人手的三维模型。如图 5 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像。

本文中，我们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，我们用真实的序列图像实现了我们的算法；实验结果证实了该方法的有效性。

参考文献

1. Q.Cai and J.K. Aggarwal. *"Automatic Tracking of Human Motion in Indoor Scenes across Multiple Synchronized Video Stream"*， In Intl. Conf. On Computer Vision, Bombay. India, 1998.

2. E. Huber，*"3D Real-time Gesture Recognition using Proximity Space. In Proc. of Intl. Conf. On Pattern Recognition"*， Vienna, Austria, August 1996, pages 136-141.

3. G. Cameron, A. Bustanoby, K.Cope, S. Greenberg, C. Hayes, and O.ozox，*" Panel on Motion Capture and Character Animation"*， SIGGRAPH'97, pages 442-445, 1997.

4. H.A. Rowley and J.M. Regh，*"Analyzing Articulated Motion Using Expectation-maximization",* In Proc. of Intl. Conf. On Pattern Recognition, Puerto Rico, 1997, pages 935-941.

5. M.K. Leung and Y.H. Yang, *"First sight: A Human Body Outline Labeling System",* IEEE Trans. PAMI, vol.17 (4), pages 359-377, 1995.

6. B.K.P. Horn and B.G. Schunck, *"Determining Optical Flow, Artificial Intelligence",* Vol.17, pages 185-204, 1981.

7. A. Verri and T. Poggi, *"Motion Field and Optical Flow: Qualitative Properties",* IEEE Trans. PAMI. Vol.11, pages 490-498, 1989.

8. 马颂德， 张正友； 《计算机视觉》，科学出版界。

9. J. Lee and T.L. Kuni, *"Model-based Analysis of Hand Posture",* IEEE Computer Graphics and Applications. Pages 77-86, Sept. 1995.

10. J.J. Kuch. *"Vision-based Hand Modeling and Gesture Recognition for Human Computer Interaction"*, Master's Thesis, Univ. Of Illinois at Urbane-Champaign, 1994.

11. R.M. Murray, Z.Li, and S.S. Sastry, *"A Mathematical Introduction to Robotic Manipulation"*, CRC Press, 1994.

12. Chunhong PAN and Songde MA; *"3D Motion Analysis Based on Coplanar Constraints"*, In proceeding of International Conference on Pattern Recognition 2000, Spain, pages 159-163.

13. Christoph Bregler and Jitendra Malik, *"Tracking People with Twists and Exponential Maps"*， In Proceeding of International Conference on Computer Vision and Pattern Recognition, 1998.

14. 潘春洪， *"*单目序列图像中人的三维运动分析*"*, 博士论文，2000，7。

15. D.E. Goldberg, *"Genetic Algorithms in Search, Optimization and Machine Learning, Reading"*, Addison Wesley,

1989.

潘春洪: 1987年毕业于清华大学自动化系，获工学学士学位，2000年毕业于中国科学院北京自动化研究所，获工学博士学位，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。现工作于中国科学院自动化研究所，模式识别实验室。 研究方向为: 计算机视觉，图像处理，和模式识别。

马颂德: 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，1986 年获该校国家博士学位。研究员，博士生导师，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，图像处理，和模式识别。