

## 基于用户行为分析的手势跟踪方法\*

冯志全<sup>1,2</sup>, 杨波<sup>1,2</sup>, 郑艳伟<sup>1,2</sup>, 徐涛<sup>1,2</sup>, 唐好魁<sup>1,2</sup>

<sup>1</sup>(济南大学 信息科学与工程学院, 山东 济南 250022)

<sup>2</sup>(山东省网络环境智能计算技术重点实验室, 山东 济南 250022)

通讯作者: 冯志全, E-mail: ise\_fengzq@ujn.edu.cn

**摘要:** 人手结构的高维性而导致粒子滤波跟踪方法中采样数目非常庞大,是实现运动人手的实时性跟踪研究的主要障碍之一.以降低粒子数目为目标,以行为分析和建模为切入点,提出一种手势跟踪方法.首先分析操作者在手势操作过程中的行为特征,建立人手运动的动态模型;其次,研究动态模型的基本特征,并给出一种描述方法;然后,建立人手运动的时段模型,分析了手势状态的时间-空间关系.在此基础上,提出了状态变量微观结构的概念,重点给出了基于状态变量微观结构的手势跟踪算法;最后,设计和完成了实验,并与相关参考文献方法的实验结果进行对比.结果表明,采用该算法,用少量粒子就可以得到比较精确的跟踪结果.提出的核心算法已经用于一个基于自然手势交互的三维虚拟装配原型系统.

**关键词:** 人手跟踪;粒子滤波;状态变量微观结构;降低粒子数目;行为特征

**中图法分类号:** TP391      **文献标识码:** A

中文引用格式: 冯志全,杨波,郑艳伟,徐涛,唐好魁.基于用户行为分析的手势跟踪方法.软件学报,2013,24(9):2101-2116.  
<http://www.jos.org.cn/1000-9825/4368.htm>

英文引用格式: Feng ZQ, Yang B, Zheng YW, Xu T, Tang HK. Hand tracking based on behavioral analysis for users. Ruan Jian Xue Bao/Journal of Software, 2013, 24(9): 2101-2116 (in Chinese). <http://www.jos.org.cn/1000-9825/4368.htm>

## Hand Tracking Based on Behavioral Analysis for Users

FENG Zhi-Quan<sup>1,2</sup>, YANG Bo<sup>1,2</sup>, ZHENG Yan-Wei<sup>1,2</sup>, XU Tao<sup>1,2</sup>, TANG Hao-Kui<sup>1,2</sup>

<sup>1</sup>(School of Information Science and Engineering, University of Ji'nan, Ji'nan 250022, China)

<sup>2</sup>(Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Ji'nan 250022, China)

Corresponding author: FENG Zhi-Quan, E-mail: ise\_fengzq@ujn.edu.cn

**Abstract:** On the way forward to real time in the process of particle-filtering-based human hand tracking, one of the main obstacles is to generate a great deal of particles which are derived from high dimensionality of human hand model. Aimed at reducing the particle number, a new particle filtering approach is put forward in this paper. First, the operator's cognitive psychology features in the process of human-computer interaction are analyzed and studied and a general dynamic motion model is constituted. Second, some basic features of the dynamic motion model are studied and mathematically described. Third, a zonetime model of the moving human hand is proposed, and furthermore, the features in time-space of a hand gesture state are discussed. Based upon the abovementioned job, a new concept, microstructure of state variable, is presented, upon which the novel hand gesture tracking algorithm is put forward. Finally, experiments are implemented including some comparison experiments, and the algorithm is also compared with some referenced algorithms. The main contribution is that the study describes and models hand gesture behaviors and connect them with freehand tracking. The experimental results show that just using a small quantity of particles, compared with the referenced algorithms, the algorithm can obtain satisfactory results.

**Key words:** human hand tracking; particle filtering; microstructure of state variable; sampling reduction; cognitive psychology

\* 基金项目: 国家自然科学基金(61173079, 60973093, 61173078); 新世纪优秀人才支持计划(NCET-0863); 山东省自然科学基金(ZR2011FZ003)

收稿时间: 2011-09-09; 修改时间: 2012-02-15, 2012-07-23; 定稿时间: 2013-01-07

建立智能化的、自然的、和谐的、人性化的人机界面已经成为新一代人机交互(HCI)研究的核心目标.目前已经提出了 Post-WIMP 和 Non-WIMP 交互界面、笔式交互界面、无处不在计算(ubiquitous computing)、渗透性计算、可穿戴计算、虚拟现实(VR)、服务环境和透明计算等多种人机交互范式.基于自然人手的跟踪、识别和交互是 VR 中的关键研究内容之一.不仅如此,三维手势在手语识别<sup>[1]</sup>、手势鼠标<sup>[2]</sup>、虚拟物体控制<sup>[3]</sup>、家电遥控<sup>[4]</sup>、Windows 命令控制<sup>[5]</sup>、机器人控制<sup>[6]</sup>等领域的应用研究已经初步展开,引起业界的深度关注.

自然人手(freehand)跟踪是识别和交互的基础,实现这种跟踪的主要技术手段之一就是粒子滤波(PF)<sup>[7]</sup>.PF 虽然是从 20 世纪 90 年代中后期才发展起来的一种最优非高斯非线性滤波方法,但目前已被学术界认为是最具发展前景的状态估计方法之一,它可以有效处理由于遮挡和复杂背景所产生的非高斯分布问题和非线性问题,在每一时刻输出的是目标位置概率分布的近似,且可以保留多个假设匹配同时随时间传播,使得跟踪丢失时也能重新找到目标并继续跟踪.粒子滤波跟踪方法的另一个重要特点在于,它对于光照变化、复杂背景以及人手自遮挡等因素的影响具有较强的鲁棒性.

自然人手状态采样是基于粒子滤波方法进行自然人手跟踪研究中的关键问题之一,它直接影响到跟踪结果的实时性和鲁棒性,降维成为解决实时性问题的关键所在.沿着梯度方向进行采样<sup>[8]</sup>是降低采样数量的典型方法之一.首先,根据 Hessian 矩阵估计样本的不确定性方向;然后,根据不确定性方向进行采样.Bray<sup>[9]</sup>提出了一种基于 Stochastic Meta-Descent(SMD)的链结构跟踪方法,这是一种梯度下降方法,通过改进步长的设置而改善了跟踪的速度和精度.在文献[10]中,沿特征深度方向进行采样,迭代过程中,在模型表面随机选择少量样本点,以降低时间开销;同时,利用模型约束条件降低状态维数.在文献[11]中,进一步将 SMD 方法与 PF 滤波相结合,提出了智能粒子滤波算法,以较少的样本跟踪高维链结构.Bray 进一步对文献[5]中的方法进行改进<sup>[12]</sup>,在多个基于 SMD 的跟踪器周围设计一个粒子滤波器,经过基于简单系统动态模型的粒子传播之后,这些跟踪器可以用比原来少得多的粒子实现高维链结构体的跟踪.

通过机器学习<sup>[13-15]</sup>的方法获取高维特征集与手势之间的关系,是解决高维特征空间搜索问题的一个有效途径.例如,在文献[15]中,把人手的运动图像看成是一个高秩随机线性动态系统,通过学习获取运动图像序列的动态特性,以构造动态 Bayes 网络.把三维(3D)姿态的估计当作一个学习和推理过程的优点是算法稳定性好,速度快;缺点是只能对已经学习的运动进行跟踪,并不适合跟踪自然场景中人手的任意动作.

处理高维跟踪问题的另一类方法是先将状态空间分割为若干子空间<sup>[16-20]</sup>(search space decomposition),然后在各个子空间上分别利用粒子滤波方法进行跟踪;或者,首先对某个子空间独立地进行搜索,然后将其作为约束条件降低余下变量的搜索空间.这种方法的优点是可以改善跟踪的实时性、鲁棒性和精确性;存在的问题是,难以确定分割状态空间的方法;实验结果还表明,即使很好地状态空间分割为独立的子空间,也不能独立地对它们进行处理<sup>[21]</sup>.

采用搜索策略可以缩小采样范围.清华大学的崔锦实<sup>[22]</sup>博士提出一种基于回归-优化方法的关节式物体的姿态估计方法.该方法把回归分析与全局优化搜索相结合,保证了估计的精度和连续性;针对现有滤波器在高维非线性多峰跟踪问题上的困难,将粒子滤波器与全局搜索算法的演化粒子滤波器方法相结合,提高了高维跟踪的精确度.

PCA(principal components analysis)是降低状态向量维数的典型方法之一,其基本思想是:将原来的一组变量变换到另一组变量,再选择其中一部分重要成分作为自变量,最后利用最小二乘方法确定主成分,从而得到低维向量.在文献[23]中,采用 PCA 方法将 28 种手势维数降为 7 维.其他所有手势被表示为这 28 种手势的线性组合.ICA(independent components analysis)算法是一种基于高阶统计量的去相关多元数据处理方法,其基本思想是:用一些基函数来表示一系列随机变量,这些基函数的各成分之间是统计独立的,它是降低状态向量维数的另一种重要途径之一.在文献[24]中,采用 ICA 方法将手势状态空间的维数降为 5 维,并证明其效果比 PCA 要好.

2007 年,德国绍姆堡(Schaumburg)摩托罗拉(Motorola)实验室的 Qu 等人分别提出对链状物体的两种跟踪方法<sup>[25]</sup>:分散跟踪(decentralized articulated object tracking)方法和分层跟踪(hierarchical articulated object tracking)方法.在分散跟踪方法中,首先采用隐式状态层和观测层所组成的图来描述被跟踪体,然后再对图进行分解.将

这种方法进行扩展,得到分层跟踪方法,它可以处理遮挡问题.中国科学院软件研究所王西颖等人<sup>[26]</sup>利用一组 2D 手势模型替代高维度的 3D 手模型,跟踪计算量大为减少.Raskin 等人<sup>[27]</sup>将退火粒子滤波跟踪器与高斯过程动态模型相结合,降低高维状态向量之维数.Stefanov 等人<sup>[28]</sup>把粒子滤波与退火算法相结合,用变长马尔可夫模型对手势进行描述.

2008 年,清华大学的姚安邦等人融合尺度估计和目标模型的更新,提出一种基于核的跟踪方法<sup>[29]</sup>.他们首先用固定尺度的基于核的方法估计物体中心位置,然后用基于核密度估计策略刷新目标模型.北京理工大学的贾云得教授及其研究团队结合图像模型、人手模型、CAMSHIFT 算法以及非参数化置信传播算法,采用三维人手图模型描述人手结构、运动学、动力学及自遮挡特性,将人手高维跟踪问题转化为并行跟踪 16 个 6 维变量的问题<sup>[30]</sup>.在 Nayak 等人<sup>[31]</sup>从手势方向、颜色等低层特征的统计分布出发,建立运动帧变化的分布函数,从而将高维问题转换为低维空间进行处理.

2009 年,Robert 等人在其 SIGGRAPH 论文中<sup>[32]</sup>,让操作者戴上彩色手套,他们对彩色区域块进行编号,建立手势数据库及其采样搜索机制,实现对手势的三维实时跟踪.Wu 等人<sup>[33]</sup>通过学习低维内部运动流形,实现降维.

2010 年,Mohr 等人<sup>[34]</sup>利用分层模板匹配,得到多个候选手势模型.这种模板实际上是一个很大的 3D 手势库.Aristidou<sup>[35]</sup>利用光学运动跟踪系统 3D 人手跟踪,但要求在手的关键部位打上标记.

近年来,研究者逐渐意识到,不同手势跟踪方法往往只能在一定条件下解决局部问题,与手势交互应用的要求仍然有较大差距.为此,他们开始把认知理论引入到手势的跟踪研究中,开始对“人”进行研究.武汇岳等人<sup>[36]</sup>以人类注意的信息加工模型为理论依据,提出了一种可扩展的视觉手势交互模型,并基于该模型提出了一个视觉手势识别框架,并结合认知心理学从手势检测、跟踪和识别 3 个方面对该框架的各个组成模块的关键技术进行了研究.

综上所述,在单目视觉条件下实现对自然人手进行实时性三维跟踪,是目前基于三维人手人机交互系统中公认的研究热点和研究难点之一<sup>[37,38]</sup>.从目前的主流降维技术来看,需要首先产生大量样本作为这些算法的输入.如果采用机器学习或事先统计方法,则影响跟踪算法的适应能力;如果在跟踪过程中生成大量样本,则时间开销足以影响跟踪的实时性;而沿着梯度方向进行采样又容易陷入局部极值;缺乏对“人”进行分析和研究,是目前该研究领域的一个共性问题.

针对上述缺陷,本文从新的角度把人机交互环境和人机交互心理纳入到手势跟踪系统的讨论中,重点对粒子滤波方法在高维人手跟踪中如何降低样本数量问题展开研究.

## 1 人机交互中的行为特征分析

### 1.1 人机交互实验

首先,对一个数字虚拟样机系统进行调查和分析,从中分解出几个基本操作:(1) 抓取一个物件;(2) 把螺帽套在螺杆上;(3) 旋转一个物件;(4) 移动一个物件;(5) 选中一个物件.为了使得实验结果客观、真实、可靠,在实验前不告诉实验者本实验的基本目的,只要求实验者(操作者)按照自己喜欢的方式自然地完成操作目标和操作任务即可(不暗示、更不要求操作人员按照某种特定的运动规律改变运动手势或运动路线).为了使得实验结果及其结论具有一定的统计意义,将实验者分为若干组,每组由若干人组成,同组操作者完成同一个基本操作,并考虑到实验者的性别、年龄和学历分布的合理性.

在实验过程中,需要详细记录两部分数据:一是操作者手指关节角度变化数据,以便进行定量分析;二是实验过程的录像数据,以便进行定性或辅助性分析.为了采集到第 1 部分数据,要求操作者戴上数字手套,其手势变化数据由数字手套传感器实时存入计算机数据库中;为了得到第 2 部分数据,对整个实验过程进行实时录像.

图 1 给出了本文人机交互实验场景图.图 2 给出了在一次特定的人机交互过程中,中指 3 个关节的变化曲线,该曲线是由 10 名学生、每人将同一交互过程重复 15 次,然后取其平均值而得到的(图中给出在抓取操作过程中操作者中指 3 个关节角度变化图(18 000 帧)).



Fig.1 The operator is interacting with the computer with data glove and position tracker on her hand

图1 操作者戴上数据手套和位置跟踪器进行人机交互实验

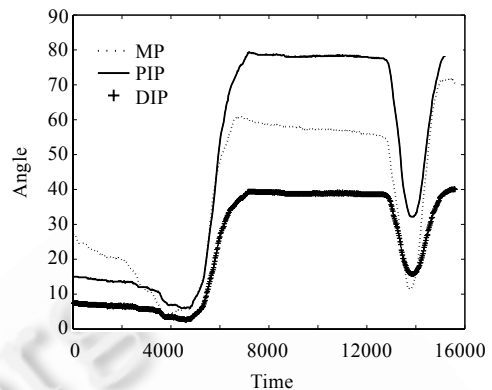


Fig.2 Data analysis for the process of interacting

图2 人机交互实验数据分析

## 1.2 人手运动的动态模型

对所有实验操作者的调查表明,在特定的 HCI 过程中,操作者均凸显出注意(attention)的心理特征.所谓注意<sup>[39]</sup>,是指人的心理活动对一定心理对象的指向与集中,其核心在于,操作者对输入的刺激信息进行有选择地加工分析而忽视其他刺激信息的心理活动.根据行为特征学理论,注意具有持续性,即根据人的目的与意愿,对某刺激信息在一段时间内保持不变,并且不受其他刺激信息的干扰而分心.据此,在操作者的手势运动过程中,其手势运动参数具有保持不变的倾向.这在第 1.1 节的实验中可以得到印证.如果说注意这种心理特征仅仅暗示一种心理趋势的话,那么行为心理学可以从实验和观察到的行为来解释和描述操作者心理活动的结果.根据 Shneiderman 的行为心理学理论<sup>[40]</sup>,人类行为一般可以分为 7 个阶段,即:(1) 确定目标;(2) 确定意图;(3) 确定明确的行动内容;(4) 执行;(5) 感知外部状况;(6) 解释外部状况;(7) 评估行动结果.可见,阶段性是人类行为的一个普遍特征,每个阶段又可以分为若干子阶段.在正常的人机交互过程中(即,不是刻意的乱动或漫无目的的盲动),仅就执行阶段进行分析,我们发现,无论采用什么运动模型来刻画人手的动态模型,有关参数的变化呈现出阶段性特点,每个阶段往往具有明显的特征和规律.调研和观察发现,操作者往往选择最简单最容易的运动形式、以最少的认知负荷去实现认知任务——从本质上讲,操作者企图利用与生俱来或后天学习所获取的生活经验或知识完成交互任务——这正是和谐人机交互的本质所在.基于人机交互实验和人手运动的行为分析,我们得到下述基本结论:在正常的人机交互过程中,人手运动模型具有以下基本特征:

- (1) 人手运动的过程具有比较明显的阶段性,这样,操作者不容易感到疲惫,以较少的认知努力完成认知任务;
- (2) 在同一个阶段内,其参数变化往往具有相似的数学描述形式,即手势变量的改变往往具有同步性——它们共同为同一个认知任务而服务;
- (3) 人手运动的过程是一个由连续形变和突变交替进行的过程;
- (4) 连续形变是主要的运动形式,占据整个运动过程的绝大部分时间.

用一个  $n$  维向量  $X=(\alpha_0, \alpha_1, \dots, \alpha_{n-1})$  描述手势状态,如果某状态变量在某个阶段内的加速度变化量满足:

$$\Delta a = \text{常数},$$

则称该状态变量在该阶段内进行连续形变,并把相邻两次连续形变之间的运动形式称为突变.人手运动的动态模型如图 3 所示.

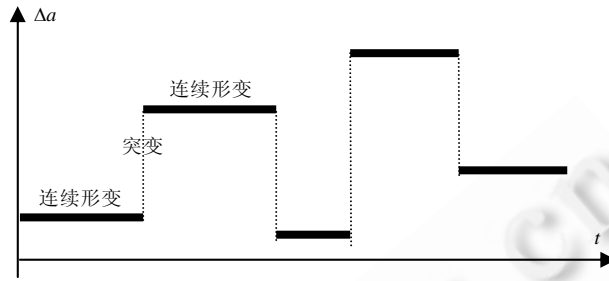


Fig.3 The schema of dynamic model of gesture motion

图3 人手运动的动态模型示意图

## 2 动态模型的基本特征及其概率描述

在人手运动的动态模型中,怎样判断一个连续形变阶段的结束?该问题等价于:怎样判断一个突变的开始?也就是说,一次连续形变的持续时间究竟有什么规律?如何对其进行有效描述?本文通过研究连续形变的时空特征和时间特征,对这些问题进行初步探讨和研究.为叙述方便,设当前时刻  $k$  的预测手势模型为  $G$ ,它在像平面上的投影为图像  $I_g$ ,当前帧观测图像特征为  $I$ .

### 2.1 连续形变的时空特征

为了增加对图像变化的描述能力,我们把图像特征分为点特征集和轮廓特征,所谓点特征是指由指尖、指根、关节以及手势图像中的其他凹点、凸点等所构成的基本信息,所谓轮廓特征是指图像的边沿、轮廓等边界信息.把  $I_g$  和  $I$  之间的点特征误差和轮廓特征误差分别记为  $E_p$  和  $E_e$ ,定义  $G$  的连续形变概率为

$$P_C = \eta e^{-\mu E_p^2} + (1 - \eta) e^{-\nu E_e^2} \quad (1)$$

$G$  的总体突变概率为:

$$P_{total} = 1 - P_C \quad (2)$$

其中,  $\mu$  和  $\nu$  是调节总体突变概率曲线平滑度的经验常数,  $\eta$  是调节这两类误差对突变概率影响能力的常数因子.点特征和轮廓特征二者在图像变化描述中的重要程度不一样,公式(1)给出了一种将二者融合起来的有效途径,这种重要程度由参数  $\eta$  进行调节.总体突变概率从总体上刻画手势空间中部分或全体状态变量的突变能力,反映了连形变的时空特征.

### 2.2 单状态变量特征

总体突变概率反映了所有状态变量的总体空间特性,并不能描述某一个状态变量的变化趋势.问题等价于:如何估计某一个状态变量的改变而产生的观测改变量?由于提取手势模型观测值的复杂性和困难性,即使要获取一个指尖特征也是一件很不容易的事情.为此,本文通过特征点的区域特征来近似代替点特征,引入单状态变量关联集的概念.

如果像平面上的某观测区域仅由一组状态变量集  $(\alpha, \beta_1, \beta_2, \dots, \beta_L)$  所对应的  $G$  中部分或全体模型所决定,而与手势状态向量中其余状态变量无关,则称  $(\alpha, \beta_1, \beta_2, \dots, \beta_L)$  为状态变量  $\alpha$  的一个单状态变量关联集,简记为  $\zeta_\alpha$ ,由它所确定的观测区域简记为  $\pi(\alpha)$ .显然,不同变量的单状态变量关联集之间可能存在交集.

如果单状态变量  $\alpha$  是手势的局部变量,定义单状态变量  $\alpha$  的突变误差概率为

$$P_\alpha = 1 - e^{-\lambda E_\alpha^2} \quad (3)$$

其中,  $E_\alpha$  表示观测区域  $\pi(\alpha)$  与  $I$  之间的误差,称为单状态变量  $\alpha$  的突变误差;  $\lambda$  是一个常数.

如果单状态变量  $\alpha$  是手势的全局变量,定义单状态变量  $\alpha$  的突变误差概率为

$$P_\alpha = P_{total} \quad (4)$$

$P_\alpha$  越小,则 $\alpha$ 发生突变的概率也越小.

### 2.3 连续形变的时间特征

通过研究人机交互过程中人手的动态模型我们可以发现,一次连续形变持续时间的分布呈现出一定的规律性,据此可以建立连续形变的时间模型(如图 4 所示).

用

$$P_d(\alpha, t) = e^{-\mu t^2} \quad (5)$$

来描述变量 $\alpha$ 的这种时间模型.其中, $\mu$ 是参数,通过统计或学习训练等方法可以对其进行估计; $t$  是一次连续形变的持续时间.

我们让一名操作者反复将虚拟装配系统实验 50 次,按照式(6)

$$P(t_1 \leq t \leq t_2) = \frac{1}{T} \sum_{t_1 \leq t \leq t_2} t \quad (6)$$

对手势各状态变量连续形变的概率分布进行统计,发现统计特征具有极大的相似性,图 5 给出了中指中间关节的统计特征及逼近曲线.图 5 中,利用 PSO(particle swarm optimization)算法经过 100 次迭代对人机交互统计数据曲线进行逼近,得到统计参数 $\mu$ (实心圆标识的曲线代表概率分布曲线,空心圆标识的曲线代表逼近曲线).

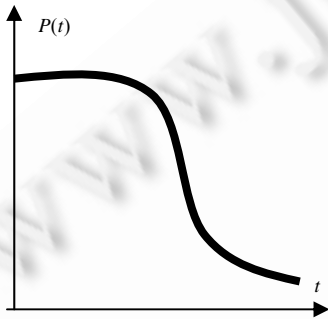


Fig.4 Time characteristics of continuous deformation

图 4 连续形变的时间特征

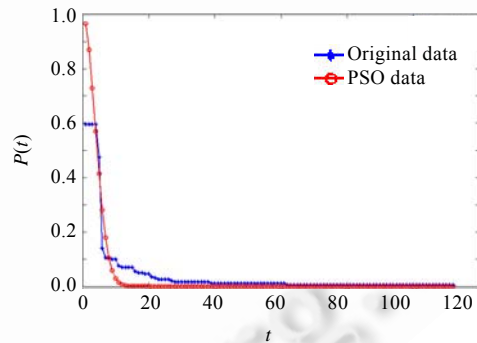


Fig.5 The probability distribution and its approximation of duration of continuous deformation for middle finger articulation in a virtual assembling system

图 5 在一个虚拟装配系统中,中指中间关节连续形变之持续时间概率分布及其逼近曲线, $\mu=0.03522095$

## 3 人手运动的时段模型

结合连续形变的时间特征和运动动态模型,进一步提出并建立基于周期概念的人手运动的时段模型.

一个跟踪周期由连续形变时段和突变时段所构成.在变量的连续形变时段,其动态模型往往具有确定性和稳定性.在变量的突变时段,匀加速度运动状态被打破,系统进入振动并重新建立新的平衡.进一步把连续形变时段分为安全时段和风险时段.大量统计实验表明,系统进入连续形变时段以后,在一定的时间内以大概率持续保持匀加速度运动状态;但是随着时间的推移,这种连续形变状态被打破的可能性越来越大,我们将这个时段称为风险时段(如图 6 所示).图 6 中, $P(t)$ 是变量之持续时间为  $t$  时仍为连续形变的概率.本文采样统计方法确定概率意义上的安全时段上限值  $K_1$  和风险时段上限值  $K_2$ .这种基于周期概念的人手运动时段模型进一步揭示了人机交互过程中人手运动的新特征,为设计跟踪算法建立了基础.

要实现这种时段模型的划分具有一定难度,为此,我们不是单纯根据第 1.2 节中的连续形变的定义和图像投影的变化来估计手势加速度的变化,而是结合连续形变的时间模型和突变概率,计算突变强度(见第 5.2 节),

据此实现突变时段的划分;安全时段和风险时段则分别由参数  $K_1$  和  $K_2$  实现划分.可见,为了实现阶段划分,需要提取以下特征:

- (a) 连续形变的时空特征(公式(1)).这里需要计算两个数据集合:一是当前图像帧中的手势轮廓数据集合,二是当前时刻三维手势预测模型在对应帧图像上的投影数据结合;
- (b) 连续形变的时间统计特征(公式(5));
- (c)  $K_1$  和  $K_2$  的统计模型.

在数字手套和位置跟踪器的帮助下,很容易得到特征(b)、特征(c)中的统计变量.

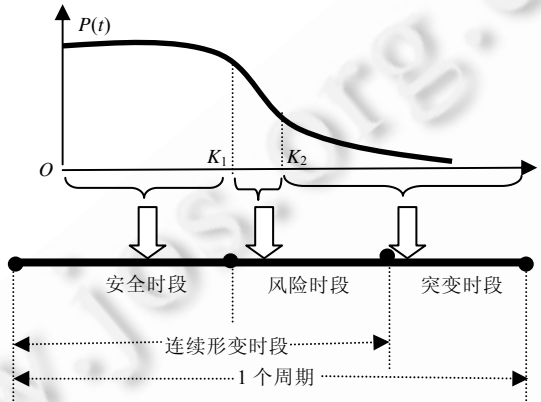


Fig.6 The periodic-based time interval model of gesture motion  
图 6 基于周期概念的人手运动的时段模型

### 4 手势状态的时间-空间关系

在时段模型的基础上,进一步探讨手势状态的时间-空间关系.在  $n$  维手势姿态向量  $X=(\alpha_0, \alpha_1, \dots, \alpha_{n-1})$ 中,根据人手运动的时段模型,每个状态变量  $\alpha_i(0 \leq i \leq n-1)$ 的运动形式在安全时段、风险时段和突变时段这3个时段上交替进行,图7示意了这种时间-空间关系.在时刻  $k$ ,其手势状态由处于不同时段的状态变量所构成,这种手势状态的时间-空间关系从微观角度描述了手势运动中时间和空间之间联系的新形式.

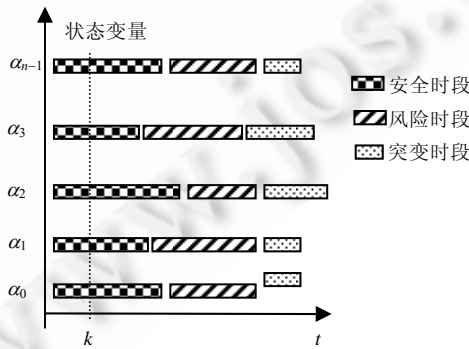


Fig.7 The spatio-temporal representation for gesture states  
图 7 手势状态的时间-空间关系示意图

## 5 基于行为分析的手势跟踪算法

### 5.1 状态变量的微观结构

现有的采样方法主要建立在状态先验分布的基础之上,这种先验分布只能从单一层面描述状态变化的特征.如何设计一种新的状态描述方法,使之可以从不同层面、不同角度描述状态变化的真实面貌,使得采样过程可以建立在更加有效、高效和统一的数据结构的基础之上,为降低采样数目、方便程序设计提供一种现实基础,是本文要解决的关键问题之一.本文提出状态变量微观结构的概念和相关方法探索该问题的解决途径.本文将状态变量的一些内在特征、变化趋势、数据结构以及状态变量之间的联系形式统称为状态变量的微观结构.实际上,第1节~第4节提出的数据和模型已经形成一个状态变量微观结构之实例(如图8所示),这种微观结构是建立在对操作者的认知行为模型分析的基础之上的.图8中,自遮挡特征描述手势的自遮挡部分所对应的状态变量的运动特性,该状态变量保持其被遮挡前的运动状态值,并据此进行预测.例如,如果在被遮挡前手势某部分作匀速运动,则在被遮挡期间仍然按照匀速运动规律对该部分对应的变量进行预测.空间特征根据手势图像的改变量描述手势发生连续形变的概率.单状态变量特征旨在描述某一个状态变量的突变误差概率.时间特征主要描述状态变量进行连续形变沿时间方向的改变规律.时段特征是对时间特征进行进一步建模,更加深入地描述状态变量的变化趋势.时空关系同时从时间和空间两个维度上描述状态变量的时段特征.

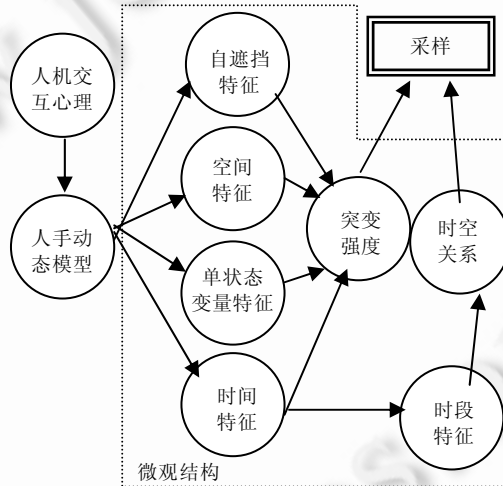


Fig.8 The representation of microstructure of state variable

图8 状态变量微观结构示意图

### 5.2 算法描述

#### 5.2.1 基本思想

算法设计的总体原则是:在 PF 算法设计中,既要减少手势状态样本的数目,又要使得这些样本可以较好地反映手势的真实分布;既要考虑跟踪时间开销,又要考虑跟踪精度.为此,根据预测手势中各变量所处的时段,把手势分为以下 3 类分别进行研究:安全型连续形变手势(G-1)、风险型连续形变手势(G-2)、突变型手势(G-3).所谓 G-1,是指预测手势模型中每个状态变量均处于安全时段条件下跟踪得到的手势模型;G-2 是指预测手势模型中部分状态变量处于安全时段、部分状态变量处于风险时段条件下跟踪得到的手势模型;G-3 是指预测手势模型中有状态变量处于突变时段条件下跟踪得到的手势模型.算法总体框架如图 9 所示.

算法设计的基本思想是:从概率意义上讲,使得最后确认的处于突变时段的变量个数最小化,因为粒子数目与这部分变量个数成指数级倍增关系;使得最后确认处于安全时段的变量个数最大化,因为在允许正常误差的前提下,对这部分变量处理的时间开销最小.



手势姿态预测是本文算法的基础.首先,根据当前时刻的前  $W$  帧数据(本文取  $W=3$ ),动态建立运动模型,然后采用 UKF<sup>[41]</sup>方法,对当前帧的手势姿态进行预测.分类器主要根据状态变量的时段模型对 G-1、G-2 和 G-3 进行初步分类,算法将进一步根据状态变量的空间特征对初步分类结果进行调整和优化.

图 9 中的分类器是这样工作的:在一个新的时变阶段,如果  $t \in [0, K_1]$ ,则选择 G-1;如果  $t \in [K_1, K_2]$ ,则选择 G-2;如果  $t \in [K_2, \infty]$ ,则选择 G-3.其中,  $K_1$  和  $K_2$  通过统计分析得到.

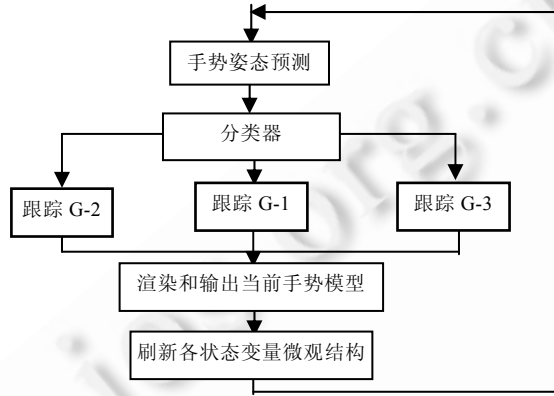


Fig.9 The global framework of the proposed algorithm

图 9 本文算法总体框架图

### 5.2.2 G-1 跟踪算法

跟踪 G-1 的基本方法是直接把手势预测值近似地作为当前帧手势模型值.这样处理的基本依据在于:其一,以绝对精确为目标,在理论上是不可行的,在实践中是不必要的;其二,本文第 1 节对人手运动动态模型的研究结论可以支持这种近似处理方法.

### 5.2.3 G-2 跟踪算法

G-2 跟踪算法的处理重点是处于风险时段的状态变量:首先,确认验证.确认验证的主要目的是确认状态变量是否处于连续形变状态;然后,根据一定的概率模型对经确认后处于突变状态的状态变量进行随机采样. G-2 跟踪算法描述如下:

确认验证.

(1.1) 初始化:刷新状态微观结构.

(1.2) 计算预测手势的总体突变概率  $P_{total}$ ;

(1.3) 如果  $P_{total} < \eta_1$ ,则返回预测手势;否则,转下一步;

调用 G-3 跟踪算法.

其中,  $0 < \eta_1 < 1$ .

### 5.2.4 G-3 跟踪算法

在有全局变量处于突变时段的条件下,需要分别处理安全时段、风险时段和突变时段的状态变量,其基本途径是把所有状态变量分为两类:一类采用近似的确定性方法估计其值;另一类采用贝叶斯方法估计其值.为了利用贝叶斯方法,算法融合预测手势的总体突变概率和变量的连续形变时间持续概率等先验信息,建立采样概率模型.G-3 跟踪算法描述如下:

(1) 初始化:把处于风险时段和突变时段的所有状态变量置入集合  $\phi$  中;刷新状态微观结构.

(2) 把预测模型中所有被遮挡部分有关的状态变量从  $\phi$  中取出来,将其放入安全时段状态变量集;

(3) 处理处于安全时段的状态变量集中的状态变量:

$$\alpha^{(k)} = \hat{\alpha}^{(k)} \quad (7)$$

(4) 建立 $\varphi$ 中各变量的概率采样模型

## (4.1) 计算突变强度:

对于风险时段变量 $\alpha$ ,定义其突变强度为

$$s_{\alpha} = \delta P_d(\alpha, t) P_{\alpha} P_{total} \quad (8)$$

对于突变时段变量,定义其突变强度为

$$s_{\alpha} = (1 - \delta) P_d(\alpha, t) P_{\alpha} P_{total} \quad (9)$$

## (4.2) 把所有突变强度归一化;

(4.3) 把突变强度小于阈值 $\eta_2$ 的所有状态变量从 $\varphi$ 中取出来,将其放入安全时段的状态变量集中,并按步骤(3)对其进行处理;(4.4) 将 $\varphi$ 中状态变量的突变强度重新归一化;

## (5) 随机采样

根据 $\varphi$ 中变量突变强度归一化值,沿变量变化的梯度方向进行随机采样,结合已经近似确定其真实值的状态变量,生成 $N$ 个样本;

## (6) 计算样本权值

计算各样本对应手势的总体突变概率 $P_{total}$ ,然后归一化,得到各样本权值;

(7) 计算 $N$ 个粒子的加权和,得到当前帧的跟踪模型.

其中, $\delta$ 是经验参数,用于调节风险时段和突变时段对突变强度的影响程度.一般来讲,风险时段对突变强度的影响要相对小一些,所以一般取 $0 < \delta < 0.5, 0 < \eta_2 < 1$ .

## 5.3 算法分析

## 5.3.1 算法的时间开销

本文所提出的模型和算法旨在降低手势样本空间的维度——本文的维度是指手势状态变量中需要用概率性方法采集大量样本进而得到其值的独立变量的个数 $\|\varphi\|$ .

首先,基于人机交互过程和心理分析,人手运动过程是连续形变时段和突变时段交替进行的过程,且前者往往表现为人手运动的主要特征之一.为此,避免对这些时段每帧进行大规模采样,成为降低平均时间开销的基本途径之一.例如,对于G-1手势,其时间复杂度为 $O(1)$ .在大部分正常人机交互过程中,运动人手往往保持刚性匀速或匀加速运动状态(尤其是整体手势的平移或旋转),就是比较典型的G-1手势.

其次,对于G-2或G-3手势,算法的时间开销主要取决于集合 $\varphi$ 中元素的个数 $\|\varphi\|$ .集合 $\varphi$ 中的元素由部分风险时段和部分突变时段的状态变量所组成.本文算法融合连续形变的时间特征、单状态变量特征以及总体突变特征等信息,最小化 $\|\varphi\|$ ——降维机理与时段模型的关系如下:

- (1) 在安全时段,状态变量的值是用确定性方法直接计算出近似值而无需采样估计,这时手势向量维度最低.在正常的人机交互过程中,手势变量运动具有同步性,即手势向量中的各变量往往同时处于同一个阶段,因此大部分情况下 $\|\varphi\|=0$ ;
- (2) 在风险时段,状态变量有发生突变的可能性,通过判断状态变量是否处于安全时段,对依然处于安全时段的变量(假设其数目为 $n_1$ )用确定性算法估计其值,仅对处于突变状态的变量进行随机采样,则 $\|\varphi\|=n-n_1-n_2$ .如果某一个变量处于安全时段,则其他变量往往也处于安全时段,因此 $n_1$ 往往比较大;
- (3) 在突变时段,需要用随机采样的概率性方法通过大量粒子刻画手势状态的后验概率分布,在最坏情况下手势维度为 $\|\varphi\|=n-n_2$ .其中, $n_2$ 表示自遮挡部分手势所对应的状态变量数目;另外,一般情况下 $n=26$ .

事实上,在正常的人机交互过程中,手势不变量和手势自遮挡所对应的变量占很大比例.爱荷华州立大学的Vaswani研究结果表明<sup>[42]</sup>:在任何情况下,大多数状态的改变都仅仅发生在极少数维度上.例如,对于一个动态的三指抓手势,必然有两个手指所对应的变量保持不变.由于充分考虑到了操作者的这些行为模型特征,以一种简单、有效、高效的方法最小化 $\|\varphi\|$ ,使得基于时段模型的手势跟踪算法可以达到比较理想的降维效果.

### 5.3.2 算法的精度

本文算法允许一定的误差,但这种误差必须得到有效控制.如果所有状态变量都处于安全时段(G-1),直接用预测值代替真实值而无需确认验证,通过优化连续形变持续时间概率模型和阈值的选取,这种免检方法以大概率保证跟踪精度;只有在突变风险(本文用突变强度定义)很大或者正处于突变状态的状态变量才进行比较大规模的采样过程,该过程也是上一帧源自免检或其他误差源而产生的累积误差的一个纠正过程;UKF 状态估计方法可以使得非线性估计值以较高精度接近于真实值,且将这种估计方法建立在当前时刻状态观测值的基础上,进一步优化了预测精度.

其次,对手势进行分类处理,并采用由粗到精的方法,逐步确定 $\phi$ 中元素,系统化研究方法在客观上控制了跟踪精度.

## 6 实验

我们在 PC 机上进行实验,基本配置为: Intel(R)Core (TM) 2, 1.66 GHz CPU, 1.00G Memory.

实验 1. 首先,采用本文提出的跟踪算法对一段真实的录像序列进行实验.用校正过的数码相机拍摄一段运动人手录像序列,然后以 60ms 间隔将其分解为帧序列,然后用手工方式去掉背景.图 10 中给出部分实验结果.每部分实验结果由两部分内容组成:上部分是跟踪得到的三维人手模型,下部分是该三维模型在对应帧图像上的投影.实验结果表明,本文算法可以实现目标跟踪,跟踪得到的人手模型与录像中手手的真实运动状态基本一致. $\|\phi\|$ 的平均数目为 4.

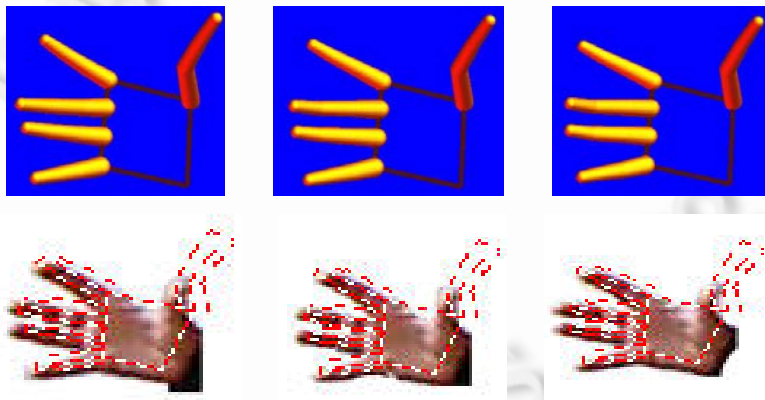


Fig.10 Some of experimental results of the proposed method

图 10 采用本文算法得到的部分实验结果

实验 2. 我们对不同的手势序列分别采用本文方法和 PF 算法<sup>[7]</sup>进行跟踪,截取了部分跟踪结果.图 11 和图 13 是本文方法得到的实验结果,图 12 和图 14 是采用 PF 算法得到的实验结果(分别与图 11 是和图 13 相对应);在每个图中,(a)图是跟踪得到的三维模型,(b)图是三维模型在对应帧图像平面上的投影.

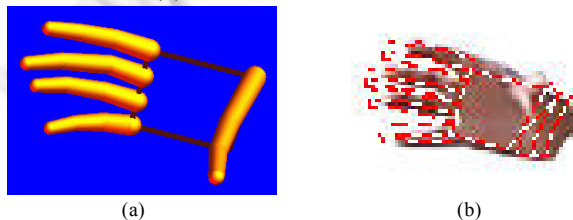


Fig.11 The tracked posture using the proposed method

图 11 本文方法得到的跟踪结果

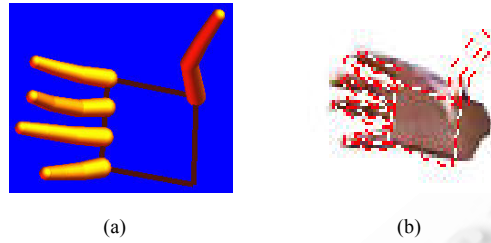
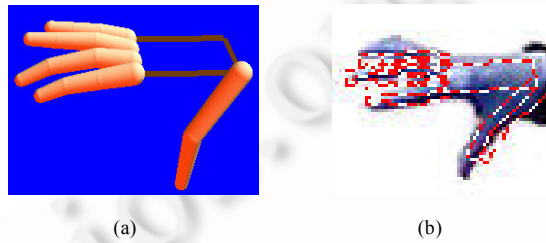
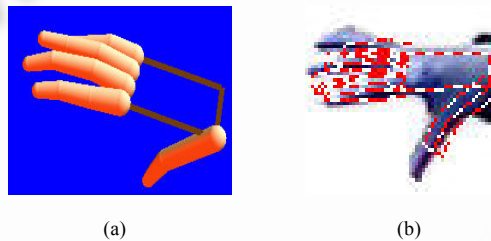
Fig.12 The tracked posture using PF algorithm<sup>[7]</sup>图 12 用 PF 算法<sup>[7]</sup>得到的跟踪结果

Fig.13 The tracked posture using the proposed method

图 13 本文方法得到的跟踪结果

Fig.14 The tracked posture using PF algorithm<sup>[7]</sup>图 14 用 PF 算法<sup>[7]</sup>得到的跟踪结果

其中,  $\eta_1=0.8, \eta_2=0.85, K_1=5, K_2=10, \delta=0.4$ , 公式(1)中的  $\mu=\nu=0.01, \eta=0.5$ , 公式(5)中的  $\mu$  统一取为 0.03. 综合手势实际姿态和投影误差情况, 本文算法有更好的跟踪精度. 图 11 和图 13 中都属于 G-1 手势, 避免了粒子采样; 在图 12 和图 14 中, 采样粒子数目为 40.

实验 3. 本实验旨在进一步比较算法性能. 在一组一般性的手势运动跟踪中, 本文算法和参考文献算法<sup>[40]</sup>实际所使用的粒子数目的分布情况如图 15 所示(其中, 横坐标单位为帧(Frame); 纵坐标代表粒子数目),  $\|\phi\|$  的平均数目为 3. 从图 15 所示的一段跟踪序列可以看出, 在第 1 帧~第 5 帧, 粒子数目为 0, 说明系统运行在 G-1 跟踪算法或 G-2 跟踪算法; 在第 6 帧~第 8 帧、第 18 帧, 粒子数目为 20, 说明系统切换到了 G-3 算法. 在图 15 所示的片段中, 第 1 帧以前处于安全时段, 在第 1 帧~第 5 帧处于风险时段, 在第 5 帧、第 6 帧处于突变时段.

本文用总体误差

$$E=E_p+E_e \quad (10)$$

来评价跟踪结果, 其中,  $E_p$  和  $E_e$  在第 2.1 节中已给出定义. 图 16 给出了本文算法和 PF 算法<sup>[7]</sup>总体误差比较曲线.

很明显, 在两种不同算法的跟踪精度大致相同的条件下, 在 18 帧中, 本文算法共用 120 个粒子, 但参考文献中的方法共用 3 600 个粒子.

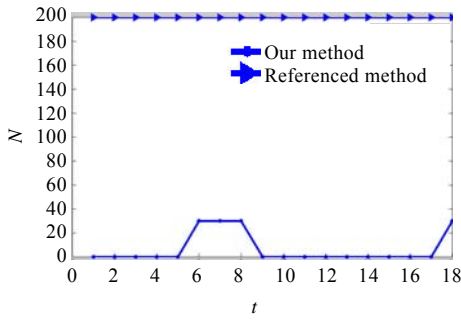


Fig.15 The comparison of particle numbers in each frame to the proposed algorithm and PF<sup>[7]</sup> algorithm

图 15 本文算法与 PF 算法<sup>[7]</sup>进行比较,各帧所使用粒子数目对比图

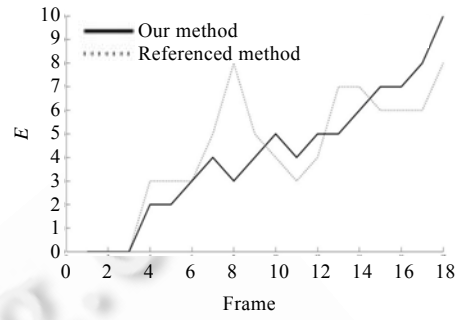
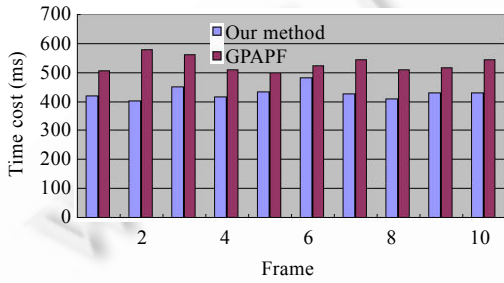


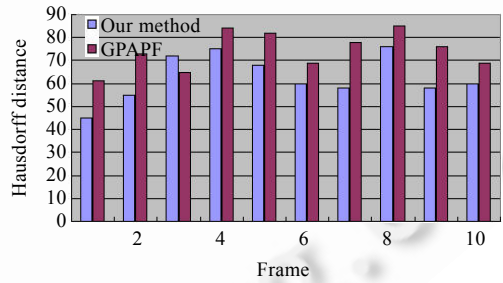
Fig.16 Error curves of the proposed method and PF algorithm

图 16 不同算法的误差比较曲线

实验 4. 我们进一步将本文算法与 Raskin 等人<sup>[27]</sup>提出的 Gaussian Process Annealed Particle Filter(GPAPF)方法从时间开销和跟踪精度两个方面进行对比分析,结果在图 17 中给出.



(a) 每帧的运行时间



(b) 每帧的 Hausdorff 距离

Fig.17 Comparison of the proposed method and GPAPF to arbitrary ten frames out of the whole tracking process  
图 17 任取 10 帧,本文算法与 GPAPF 算法的比较

由于在性能评价系统中,直接获取真实数据(ground-truth data)是一件非常困难的事情<sup>[43]</sup>,一种国际上通行的做法是:通过跟踪得到的三维人手模型在帧图像上的投影与对应帧图像之间的 Hausdorff 距离<sup>[44]</sup>作为跟踪精度的一个评价标准,Hausdorff 距离越大,则跟踪精度越低.实验结果表明,本文算法无论从时间开销方面考察,还是从跟踪精度方面分析,本文算法都得到了改善.

最后,结合复杂背景下的手势特征提取<sup>[45]</sup>以及三维人手模型自动初始化系统<sup>[46]</sup>,我们已经实现了一个基于自然人手的三维虚拟样机装配系统(如图 18 所示,图中操作者正在利用自然手装配一个虚拟样机).



Fig.18 A human-computer interactive system based on the proposed algorithm

图 18 根据本文算法实现一个基于自然人手的人机交互系统

在该系统中,我们用一个摄像机完成用自然人手(不戴数字手套,手上也不加任何标记)机器部件的装配,初步提供了一种自然、和谐、方便的三维人机交互界面。

## 7 结 论

操作者的认知过程、行为模型与手势维度之间存在着深刻的联系,本文通过研究人机交互过程中操作者的行为特征,提出了人手动态模型的基本假设,进一步研究了连续形变的空间特征、单状态变量特征以及连续形变的时间特征,并将其融合以刻画状态变量的突变强度;根据连续形变的时间特征和人手动态模型,研究了人手运动的时段模型以及建立在该模型基础之上的手势状态时间-空间关系;最后,在手势状态时间-空间中根据突变强度实现粒子采样.这种方法可以较好解决连续形变或突变的判断问题——这正是本文要着力解决的重点问题.本文的研究还表明,层次化方法是处理高维链式体跟踪问题中跟踪时间和跟踪精度这一共性矛盾的途径之一.一方面,把手势分为无需免检、需要确认检测以及需要粒子滤波处理这3个不同层次;另一方面,分别从时间特征和空间特征两个不同层次实现手势分类.

在继续探索认知过程、行为模型与手势降维之间关系的基础之上,采用智能化、自适应方法优化微观结构中主要模型参数( $\eta_1, \eta_2, K_1, K_2, \delta, \mu$ ),进一步从理论上探讨 $\|\varphi\|$ 的优化方法,将是下一个阶段的研究重点.

## References:

- [1] Wang CL, Gao W, Shan SG. An approach based on phonemes to large vocabulary Chinese sign language recognition. In: Proc. of the 5th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Washington: IEEE Press, 2002. 393-398.
- [2] von Hardenberg C, Brard F. Béare-Hand human-computer interaction. In: Proc. of the ACM Workshop on Perceptive User Interfaces. New York: ACM Press, 2001. 1-8.
- [3] Oka K, Sato YC, Koike H. Real-Time tracking of multiple fingertips and gesture recognition for augmented desk interface systems. In: Proc. of the 5th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. 2002. 411-416.
- [4] Lenman S, Bretzner L, Thuresson B. Computer vision based hand gesture interfaces for human-computer interaction. Technical Report, TRITA-NA-D0209, CID, Centre for User Oriented IT Design, 2002.
- [5] Lee MS, Weinshall D, Cohen-Solal E, Colmenarez A, Lyons D. A computer vision system for on-screen item selection by finger pointing. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2001). Puerto Rico: IEEE Press, 1997. 1027-1033.
- [6] Cipolla R, Ollinghurst NJ. Human-Robot interface by pointing with uncalibrated stereo vision. Image and Vision Computing, 1996, 14(3):171-178.
- [7] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing, 2000,10:197-208.
- [8] Chen R, Liu GY, Deng Y, Li H. Combining particle filter with local optimization for human body tracking. Journal of Computer-Aided Design & Computer Graphics, 2006,18(2):276-282 (in Chinese with English abstract).
- [9] Bray M, Koller-Meier E, Schraudolph NN, Van Gool L. Stochastic meta-descent for tracking articulated structures. In: Proc. of the Computer Vision and Pattern Recognition Workshop. 2004. 7-7.
- [10] Bray M, Koller-Meier E, Müller P, Van Gool L, Schraudolph NN. 3D hand tracking by rapid stochastic gradient descent using a skinning model. In: Proc. of the 1st European Conf. on Visual Media Production (CVMP). 2004. 231-237.
- [11] Bray M, Koller-Meier E, Van Gool L. Smart particle filtering for 3D hand tracking. In: Proc. of the IEEE Int'l Conf. on Automatic Face & Gesture Recognition. 2004. 675-680.
- [12] Bray M, Koller-Meier E, Van Gool L. Smart particle filtering for high-dimensional tracking. Computer Vision and Image Understanding, 2007,106(1):116-129.
- [13] Wachs J, Kartoun U, Stem H, Edan Y. Real-Time hand gesture telerobotic system using the fuzzy c-means clustering. In: Proc. of the World Automation Congress (WAC). 2002. 403-410.
- [14] Rosales-Del-Moral RE. Specialized mappings architecture with applications to vision-based estimation of articulated body pose [Ph.D. Thesis]. Boston University Graduate School of Arts and Sciences, 2002.
- [15] Zhou HN, Huang TS. Tracking articulated hand motion with eigen dynamics analysis. In: Proc. of the 9th Int'l Conf. on Computer Vision. Nice: IEEE Computer Society Press, 2003.

- [16] MacCormick J, Isard M. Partitioned sampling, articulated objects, and interface-quality hand tracking. In: Proc. of the ECCV. Dublin, 2000. 13–19.
- [17] Deutscher J, Davidson A, Reid I. Articulated partitioning of high dimensional search spaces associated with articulated body motion capture. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. 2001. 669–676.
- [18] Zhao QJ, Zhang BJ, Hu HS. Novel two-step filtering scheme for a logging-while-drilling system. *Computer Physics Communications*, 2009,180:1566–1571.
- [19] Wu Y, Huang TS. Capturing articulated human motion: A divide-and-conquer approach. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 1999. 606–611.
- [20] Zhao QJ, Zhang BJ, Wang W. Data processing techniques for a wireless data transmission application via mud. *EURASIP Journal on Advances in Signal Processing*, 2011,1:1–8.
- [21] Deutscher J, Davison A, Reid I. Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. 2001. 187–193.
- [22] Cui JS. Studies on three-dimensional model based posture estimation and tracking of articulated objects [Ph.D. Thesis]. Beijing: Tsinghua University, 2004 (in Chinese with English abstract).
- [23] Wu Y, Lin JY, Huang TS. Capturing natural hand articulation. In: Proc. of the Int'l Conf. on Computer Vision (ICCV). Vancouver: IEEE Press, 2001.
- [24] Kato M, Chen YW, Xu G. Articulated hand tracking by PCA-ICA approach. In: Proc. of the IEEE Conf. on Automatic Face and Gesture Recognition. 2006.
- [25] Qu W, Schonfeld D. Real-Time decentralized articulated motion analysis and object tracking from videos. *IEEE Trans. on Image Processing*, 2007,16(8):2129–2138.
- [26] Wang XY, Zhang XW, Dai GZ. An approach to tracking deformable hand gesture for real-time interaction. *Ruan Jian Xue Bao/Journal of Software*, 2007,18(10):2423–2433 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/2423.html> [doi: 10.1360/jos/18/2423]
- [27] Raskin L, Rivlin L, Rudzsky E. Dimensionality reduction for articulated body tracking. In: Proc. of the 3DTV Conf. 2007. 1–4.
- [28] Stefanov N, Galata A, Hubbold R. A real-time hand tracker using variable-length Markov models of behaviour. *Computer Vision and Image Understanding*, 2007,108(1-2):98–115.
- [29] Yao AB, Wang GJ, Lin XG, Wang H. Kernel based articulated object tracking with scale adaptation and model update. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2008). 2008. 945–948.
- [30] Liu TL, Wu XX, Liang W, Jia YD. 3D articulated hand tracking by nonparametric belief propagation on feasible configuration space. *Journal of Computer-Aided Design & Computer Graphics*, 2008,20(4):476–481 (in Chinese with English abstract).
- [31] Nayak S, Sarkar S, Loeding B. Distribution-Based dimensionality reduction applied to articulated motion recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008,99:1–8.
- [32] Wang RY, Popović J. Real-Time hand-tracking with a color glove. *ACM Trans. on Graphics*, 2009,28,3:1–8.
- [33] Wu XX, Liang W, Jia YD. Tracking articulated objects by learning intrinsic structure of motion. *Pattern Recognition Letters*, 2009, 30(3):267–274.
- [34] Mohr D, Zachmann G. Real-Time hand tracking for natural and direct interaction. In: Proc. of the CHI 2010. Atlanta: ACM Press, 2010. [doi: 978-1-60558-930-5/10/04]
- [35] Aristidou A. Motion capture with constrained inverse kinematics for real-time hand tracking. In: Proc. of the ISCCSP 2010. 2010.
- [36] Wu HY, Zhang FJ, Liu YJ, Dai GZ. Research on key issues of vision-based gesture interfaces. *Chinese Journal of Computers*, 2009,32(10):1102–1120 (in Chinese with English abstract).
- [37] Jiang YY, Tian F, Zhang XL, Liu W, Dai GZ, Wang HA. Unistroke gestures on multi-touch interaction: Supporting flexible touches with key stroke extraction. In: Proc. of the ACM Conf. on Intelligent User Interfaces (IUI). 2012. 85–88.
- [38] Erol A, Bebis G, Nicolescu M, Boyle RD, Twombly X. Vision-Based hand pose estimation: A review. *Computer Vision and Image Understanding*, 2007,108:52–73.
- [39] Liang NJ. *Contemporary Cognitive Psychology*. Shanghai: Shanghai Education Press, 2006 (in Chinese).
- [40] Shneiderman B. *Design the User Interface-Strategies for Effective Human-Computer Interaction*. Reading: Addison-Wesley Publishers, 2004.
- [41] Julier SJ, Uhlmann JK. A new extension of the kalman filter to nonlinear systems. In: Proc. of the SPIE, Vol.3068. 1997. 182–193.
- [42] Vaswani N. Particle filtering for large dimensional state spaces with multimodal observation likelihoods. *IEEE Trans. on Signal Processing*, 2008,56(10):4583–4597.

- [43] Erol A, Bebis G, Nicolescu M, Boyle RD, Twombly X. A review on vision-based full DOF hand motion estimation. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005). 2005. 15–22.
- [44] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1993,15(9):850–863.
- [45] Feng ZQ, Yang B, Chen YH, Zheng YW, Xu T, Li Y, Xu T, Zhu DL. Features extraction from hand images based on new detection operators. Pattern Recognition, 2011,44(5):1089–1105.
- [46] Feng ZQ, Zhang MM, Pan ZG, Yang B, Xu T, Tang HK, Li Y. 3D-Freehand-Pose initialization based on operator's cognitive behavior models. The Visual Computer, 2010,26(6-8):607–617.

#### 附中文参考文献:

- [8] 陈睿, 刘国翌. 结合粒子滤波和局部优化方法的人体运动跟踪. 计算机辅助设计与图形学学报, 2006, 18(2): 276–282.
- [22] 崔锦实. 基于三维模型的关节式物体姿态估计与跟踪方法研究[博士学位论文]. 北京: 清华大学, 2004.
- [26] 王西颖, 张习文, 戴国忠. 一种面向实时交互的变形手势跟踪方法. 软件学报, 2007, 18(10): 2423–2433. <http://www.jos.org.cn/1000-9825/18/2423.html> [doi: 10.1360/jos/18/2423]
- [30] 刘棠丽, 吴心筱, 梁玮, 贾云得. 基于非参数信念传播的可行 C-空间关节人手跟踪方法. 计算机辅助设计与图形学学报, 2008, 20(4): 476–481.
- [36] 武汇岳, 张凤军, 刘玉进, 戴国忠. 基于视觉的手势界面关键技术研究. 计算机学报, 2009, 32(10): 1102–1120.
- [39] 梁宁建. 当代行为特征学. 上海: 上海教育出版社, 2006.



冯志全(1964—),男,四川绵阳人,博士,教授,CCF 会员,主要研究领域为计算机图形图像处理,运动人手跟踪,人机交互技术.  
E-mail: ise\_fengzq@ujn.edu.cn



杨波(1965—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络智能信息处理.  
E-mail: yangbo@ujn.edu.cn



郑艳伟(1977—),男,讲师,主要研究领域为人机交互.  
E-mail: Ise\_zhengyw@ujn.edu.cn



徐涛(1979—),男,讲师,主要研究领域为人机交互.  
E-mail: Ise\_xt@ujn.edu.cn



唐好魁(1972—),男,讲师,主要研究领域为人机交互.  
E-mail: Ise\_tanghk@ujn.edu.cn