

基于视觉的手势界面关键技术研究

武汇岳^{1),2)} 张凤军¹⁾ 刘玉进¹⁾ 戴国忠¹⁾

¹⁾(中国科学院软件研究所人机交互技术与智能信息处理实验室 北京 100190)

²⁾(中国科学院研究生院 北京 100049)

摘 要 针对视觉手势界面存在的问题,提出了一套行之有效的解决方案.首先,为了解决视觉手势交互中的 Midas Touch 问题,以人类注意的信息加工模型为理论依据提出了一个可扩展的视觉手势交互模型,该模型将手势交互过程分为选择性处理、分配性处理和集中处理 3 个不同阶段;然后,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阐述,其中手势检测模块和识别管理模块能够辅助系统在复杂的背景中滤除掉不相关信息而选择性地搜索人手并根据上下文信息对手势识别任务重定向,从而避免了系统时刻都处于激活状态并对所有的手势动作都进行识别分析,有效解决了 Midas Touch 问题.文中介绍了使用该方法实现的 IEToolkit 手势界面工具平台,并基于一个视觉手势交互系统进行了实验测试与评估,结果验证了文中方法的可用性.

关键词 人机交互;视觉手势界面;交互模型;识别框架

中图法分类号 TP391 **DOI号**: 10.3724/SP.J.1016.2009.02030

Research on Key Issues of Vision-Based Gesture Interfaces

WU Hui-Yue^{1),2)} ZHANG Feng-Jun¹⁾ LIU Yu-Jin¹⁾ DAI Guo-Zhong¹⁾

¹⁾(Intelligence Engineering Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100190)

²⁾(Graduate University of Chinese Academy of Sciences, Beijing 100049)

Abstract In this paper, an effective solution is described for vision-based gesture interfaces, which can help avoid what is often called the Midas Touch Problem, where everything the user does is interpreted as an interaction. First, an interactive model for hand gesture is presented based on information processing model of human attention, which divides gesture interaction into selective process, divided process and sustained process; Then a recognition framework is proposed based on the interactive model and is interpreted from the perspective of cognitive psychology. In this framework, the hand detection model and the recognition management model can help recognize hand gesture from complex environment and analyze specific context and redirect gesture recognition into the adequate system model, thus minimizing the need to activate all the different gesture recognition types simultaneously and solving the Midas Touch Problem effectively. At the end, a vision-based gesture interfaces toolkit for interactive games is presented based on the recognition framework. This paper also presents experimental results based on an interactive prototype, regarding the speed, accuracy and robustness of the implemented system, which validate the quality and usability of the proposed.

Keywords human computer interaction; vision-based gesture interface; interactive model; recognition framework

收稿日期:2009-07-15;最终修改稿收到日期:2009-08-23. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2009CB320804)、国家自然科学基金(U0735004,60673188)和国家“八六三”高技术研究发展计划项目基金(2009AA01Z337,2008AA01Z303)资助. 武汇岳,男,1979年生,博士研究生,研究方向为人机交互、VBGI等. E-mail: wuhuiyue@iel.iscas.ac.cn. 张凤军,男,1971年生,博士,副研究员,研究方向为人机交互、虚拟现实、CAD等. 刘玉进,男,1984年生,硕士研究生,研究方向为人机交互、计算机视觉. 戴国忠,男,1944年生,研究员,博士生导师,研究领域为人机交互、计算机图形学等.

1 引言

基于视觉的手势界面 (Vision-Based Gesture Interfaces, VBGI) 是 post-WIMP 时代的一种重要的界面形式,与传统的 WIMP 交互方式相比手势交互能够使用户摆脱鼠标键盘的束缚而采用一种更加自然、无约束的交互方式,从而提供给用户更大的交互空间、更多的交互自由度和更逼真的交互体验,具有较高的应用价值和良好的应用前景,因此迅速成为了国内外研究的热点,并被广泛应用于虚拟/增强现实、普适计算、智能空间以及基于计算机的互动游戏等多个领域。尽管手势交互研究已经取得了很多研究成果,但由于视觉手势交互本身具有非接触性和模糊性等特点,因此目前对手势界面的设计开发仍然是十分困难的,主要还存在以下问题:(1) 手势的理解问题。VBGI 与传统 GUI 的一个重要区别就是其非接触性,从而导致界面设计中一个非常困难的问题即“Midas Touch Problem”(点石成金问题),系统本身往往无法有效区分哪些是用户真正的交互动作哪些仅仅是用户随意的下意识动作,从而正确感知用户的交互意图;(2) 算法的鲁棒性问题。由于手势本身具有多样性、多义性、多变性以及时空差异性等特点,因此对手势的正确分析与识别仍然是十分困难的;(3) 可扩展性问题。传统的手势训练方法过程繁琐复杂,对非专家用户来说技术门槛太高,而且手势库本身也往往只能由其设计人员制定,对普通用户来说缺乏有效的界面工具方便快捷地对其扩展以及定制新的手势类型。

本文主要针对以上 3 个问题,从认知心理学的角度出发,对视觉手势界面的关键技术进行了研究并给出了一套整体的解决方案,以期为 VBGI 的设计开发提供一个统一的平台和实现方法学。本文第 2 节介绍相关研究;第 3 节对视觉交互的特点进行分析,并结合认知心理学提出一个可扩展的视觉手势交互模型,基于该模型给出一个视觉手势识别框架,有效解决视觉交互中的 Midas Touch 问题;第 4 节介绍基于本文方法的应用和评估;第 5 节总结全文。

2 相关研究

“Midas Touch”问题是 VBGI 中普遍存在的一个难题。目前,对于该问题的解决方法大致可以分为

3 类,分别是基于时间延迟的策略、基于空间接近的原则以及基于界面 widget 提供交互上下文的方法。Jacob 在文献[1]中提出了一种基于时间延迟(latency-time)的策略,通过计算手势所控制的界面跟踪符号(光标)在某一界面对象上停留的时间长短来判断该界面对象是否被选中。但该方法一方面阻碍了技术娴熟的用户加快个人交互操作过程,另一方面也增加了用户的认知负担,因为用户必须时刻注意不能使手势在某一位位置停留时间过长以免误选一些本无意选取的对象。文献[2]给出了一种基于空间接近原则的方法,通过判断用户手持的 paddle 与虚拟对象的空间接近关系来完成拣选、加载等通用的交互任务,与时间延迟策略类似的缺陷是该方法仍然无法完全避免误选问题。Kjeldsen 在手势动作与视觉界面 widget 之间建立一定的映射关系^[3],一个可视的界面实体 widget 自身具备一定的空间位置,能够提供一定的交互上下文用来约束视觉交互行为,比如用户将手放在某一个 widget 上利用上下移动来控制某一系统参数,以此来降低由于 Midas Touch 问题所引起的系统误识率。但是 Kjeldsen 的方法不够自然并且其应用范围有限,当系统需要大量的模式切换时必须产生足够多的 widget,这不仅会加大用户的认知负担,而且会占用过多的界面空间而影响交互。

一个可用的视觉手势界面必须提供鲁棒的算法支持。目前,关于视觉手势分析与理解的算法比较多,文献[4]给了较为详细的综述介绍,本文在此就不再一一列举。VBGI 设计的关键在于如何根据具体的应用需求将已有的算法进行有效的整合,从算法的实时性、鲁棒性和准确性等方面进一步提高和改进,以适应多种多样的复杂的背景环境以及光照条件,从而满足实时交互的需要。

视觉手势界面另一个值得关注的问题是手势的灵活定制和扩展。传统的手势识别库如 HandVu^[5]等通常不具备可扩展性,此类系统大都预先定义了几种可识别的手势类型,普通用户只能按照系统既定的方式参与交互,这使得用户在使用时往往要放弃自己的一些个性化习惯,违背了自然的人机交互原则。

本文在分析国内外已有工作的基础上,首先结合人类知觉信息加工中的注意力模型提出一个基于非接触型交互的视觉手势状态转移模型,在这个模型的指导下设计并实现一个视觉手势识别框架。该框架的设计模仿了人类知觉信息加工的 3 个不同阶

段,使系统具备能够选择性地处理关键性信息而排除非关键信息干扰的能力,有效解决 Midas Touch 问题;接下来,我们对框架的每一个实现部分都有针对性地给出了一些相应的关键技术,用以解决手势识别任务本身所面临的多样性、多义性、多变性以及时空差异性等众多问题;最后,给出了一个有效的手势界面开发工具 IEToolkit,解决了传统手势工具系统复杂难用的问题,使非专家用户也能根据应用需求快速地定制新的手势类型并灵活地定义手势交互语义。

3 视觉手势识别框架

对于 Midas Touch 问题,传统的方法都是从视觉交互的角度出发去解决,普遍不能取得满意的结果. Midas Touch 问题产生的根本原因是视觉处理方法本身所造成的,由于计算机系统毫无选择地对进入摄像头捕获范围内的所有信息都进行平行加工处理,因而无法区分有意识的交互动作和下意识的其他动作,正确感知用户真正的交互意图. 而人类的视觉处理系统则不存在这个问题,因为大脑本身能够对输入的信息进行有效区分,分辨出哪些是需要进一步加工处理的关键信息而哪些是需要滤除的干扰信息. 认知心理学认为人类大脑具备一套注意机制,能够将信息加工过程分为选择性注意、分散注意和集中注意 3 个阶段^[6]. 其中,选择性注意(selective attention)是指个体在同时呈现的多种外界刺激中选择一种进行注意而忽略其他的刺激的心理过程. 由于人的信息加工系统处理容量有限,因此大脑在分析复杂的视觉信息时便采用选择性注意机制,滤除掉不相关的信息而将感兴趣区域移动到具有高分辨率的视网膜中央凹区,以便对该区域进行更精细的观察和分析. 分配性注意(divided attention)是指个体在同一时间对两种或两种以上的刺激进行注意,或将注意分配到不同的活动中. 分配注意说明了在同时操作两个或者多个任务时分配任务的能力,有时候也可以用来说明整合多个信息源的能力. 集中注意(sustained attention)是指注意在一定时间内保持在某个具体的客体或者活动上. 集中注意说明个体“屏蔽掉”了外部的非关键性信息,而将注意聚焦在关键的刺激源上. 认知心理学实验表明,正是由于人类具备这样的注意机制才能够对复杂的外部刺激进行有选择地过滤、对多个并行的任务有效地分配注意资源、对非关键信息及时屏蔽而将注意资

源聚焦于关键信息的处理上. 因此,本文认为计算机视觉系统也应该具备这样一种机制,通过模仿人类视觉系统对目标对象的识别处理机制并辅助以高层的交互语义,方能真正理解用户的交互意图,从根本上避免 Midas Touch 问题,从而有效地完成视觉手势识别交互任务. 根据这一理论依据,本文将计算机视觉系统对手势信息处理过程也分为 3 个不同处理阶段,这 3 个阶段分别对应于人类注意的信息加工模型 3 个阶段:(1)选择性处理阶段,在这个阶段中系统在复杂的背景环境里搜索关键刺激物——人手;(2)分配性处理阶段,当选取了关键目标人手之后,系统除了要实时定位人手的当前最新位置外还要不断检测是否有特定的手势事件产生,即系统需要将注意资源分配到不同的任务活动中;(3)集中处理阶段,系统将注意资源聚焦于具体的手势识别任务上,包括静态手势识别或者动态手势识别.

基于以上分析,本文将手作为一种抽象的视觉输入设备,在 Buxton 提出的接触型设备交互状态转移模型^[7]的基础上提出了一种可扩展的非接触型设备交互状态转移模型,如图 1 所示,有效解决了 VBGI 中的 Midas Touch 问题. 图 1 中,状态 0 和状态 1 均处于选择性处理阶段,其中状态 0 表示用户的手处于跟踪区域之外的 OOR(Out of Range)状态,此时手的移动不会对界面产生任何影响,状态 1 为手移入跟踪区域但尚未被系统检测到的状态,此时手的移动不会对交互产生影响,但会引起界面反馈例如实时显示手的原始图像,这一反馈能够充分利用用户的前庭感知和运动感知促进后续的交互;状态 2 和状态 3 同时处于分配性处理阶段,其中状态 2 表示系统已经成功检测出手,此时界面跟踪符号(光标)将随手的移动而移动,状态 3 表示系统需要根据一定的上下文判断当前是否存在以及存在何种类型的手势识别任务,由此来判断下一状态是静态手势识别(状态 4)还是动态手势识别(状态 5). 一旦进入某一具体的识别模块就能衍生出许多扩展状态,例如对界面交互对象的平移、旋转和缩放等,此时系统处于集中处理阶段,将对具体的手势事件进行处理和反馈. 与此相反的过程不再赘述.

针对图 1 的交互模型,本文提出一个视觉手势识别框架如图 2 所示. 与传统方法不同的是,本文的手势检测、跟踪以及识别 3 个模块相互交叠,两两之间又构成了内部循环. 其中,手势检测模块对应于认知心理学的选择性注意阶段,负责在复杂的环境中标识出第一次以某一特定姿态出现在视频流中的手

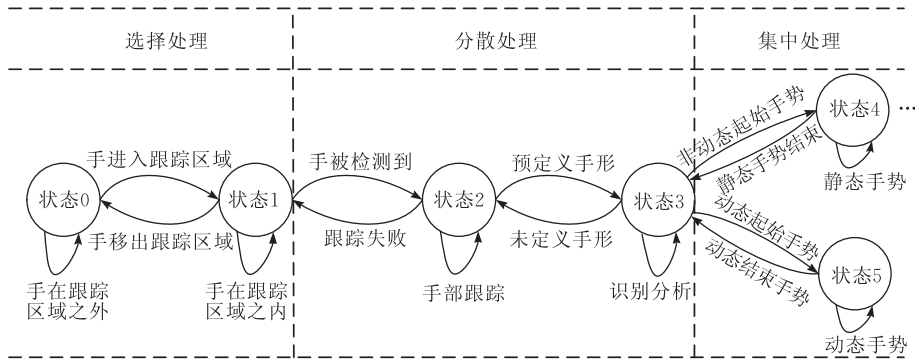


图 1 可扩展的视觉手势交互模型

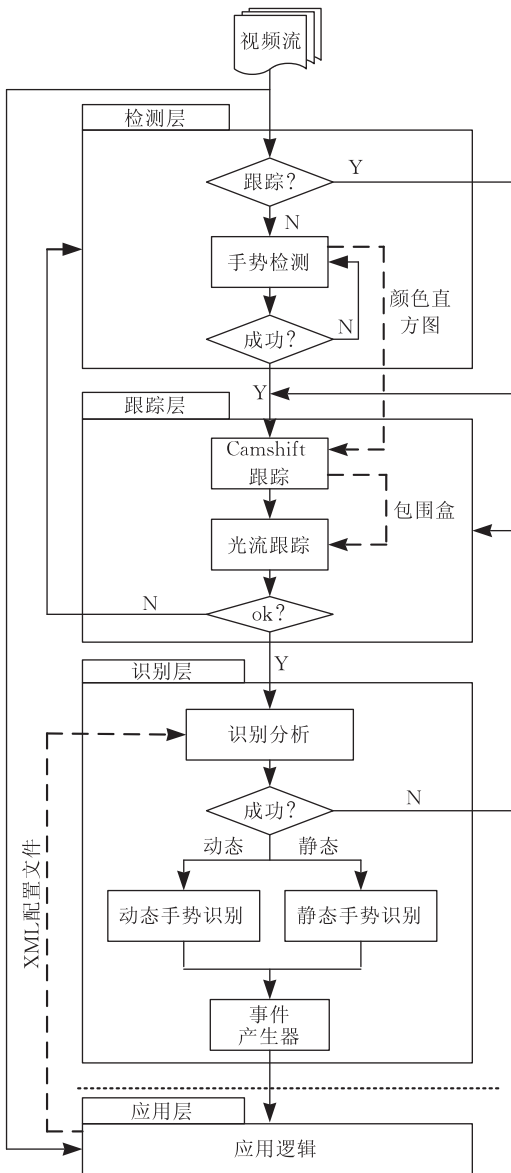


图 2 视觉手势识别框架

性注意阶段,负责解决如何对检测成功的目标手进行后续处理的问题,在这一阶段中系统将把注意资源分配到几个子任务上,它根据一定的准则判断当前跟踪是否有效,如果跟踪失效则重新返回检测模块进行检测,如果跟踪有效则在每一次成功跟踪之后都根据上下文信息判断是否有特定的手势事件产生,如果有则启动相应的识别模块.具体该启动静态识别模块还是启动动态识别模块将由识别分析模块完成,这一模块实际上又对应于选择性注意阶段.识别模块对应了注意的集中处理阶段,在手势识别过程中,系统在某一特定时刻只启动静态识别或动态识别模块之一,而并非时刻都并行地运行这两个识别模块,这样做的好处是一方面降低了系统的计算负载,另一方面由于系统聚焦于某一特定类型的识别任务,因而能够排除其他无意识手势的干扰,降低了系统误识率.最终,整个识别框架向高层应用的输出包括手的数目、手势类别名称、质心运动轨迹、手形包围盒、倾斜角、手部区域的面积等多个属性.下面给出框架中各个组成部分的关键实现技术.

3.1 手势检测

本文在分析视频流中被分割对象——手的特征信息以及运动信息的基础上,使用了一种基于模糊集理论的运动手势分割方法^[8].它利用模糊集合分别描述视频流中空域和时域上的背景、肤色和运动等信息,并定义了相关的模糊计算规则.通过对模糊集合执行模糊计算,得到感兴趣区域(Region of Interest, ROI)的模糊表示,然后由清晰化计算得到精确的手区域,能够在实时条件下对复杂背景中的手实现区域分割.一旦分割出手势区域,我们便计算它所包含的肤色像素的数量并在 HSV 空间中建立基于直方图的统计模型.

手势的成功分割实现了对手部区域和轮廓的提取,接下来描述手势轮廓的形状特征.成对几何直方图(Pairwise Geometric Histogram, PGH)^[9]是一种

势,如果检测成功则激活后续的跟踪、识别等几个模块.在 VBGI 中这一模块是十分重要的,因为它能够有效避免整个系统时刻处于激活状态而对所有的手势动作都进行识别分析;手势跟踪模块对应于分配

应用广泛的多边形形状描述子,其编码方式基于多边形线段间的距离和角度信息,具有很好的平移、对称和尺度不变性,即便是在对象发生部分遮挡的情况下也能获得较好的效果.该方法计算简单,不仅能够克服传统的匹配算法计算量大、对形状的平滑和非平滑性不敏感、对复杂轮廓匹配实时性不高等缺点,而且只需要训练少量的手势模板就能达到很好的匹配效果.因此,本文引入 PGH 方法对手势形状进行描述.由于人手轮廓属于不规则形状,因此我们首先采用 D-P 算法^[10]对手势轮廓进行多边形近似,然后利用 PGH 对近似多边形的几何特征进行描述,从而得到手势特征向量:

$$f_{\text{PGH}} = [E_r(1)E_r(2)\cdots E_r(N)E_c(1)E_c(2)\cdots E_c(M)]^T,$$

其中 N 为 PGH 的行数, $E_r(i) = \frac{\sum_j p(i,j)}{\sum_j p(i,j)}$ 为第 i

行的期望, M 为 PGH 的列数, $E_c(j) = \frac{\sum_i p(i,j)}{\sum_i p(i,j)}$

为第 j 列的期望, $p(i,j)$ 为点 (i,j) 的 PGH 值.

3.2 手势跟踪

对于手势跟踪,我们将 CamShift^[11]与金字塔光流^[12]两种算法结合,综合利用了手势图像的颜色、区域和特征信息,具有较好的鲁棒性. CamShift 是基于颜色概率模型的跟踪方法,是 Mean Shift 算法的推广,它使目标点能够“漂移”到密度函数的局部极大值点.但 CamShift 算法的一个缺点是需要跟踪开始前手工选择 ROI 来提取跟踪对象的颜色分布.本文框架由于在跟踪开始前增加了分割检测模块,手部区域已经被分割出来,所以手的肤色直方图计算是自动完成的,不再需要用户手工标定 ROI,从而实现了跟踪的自动初始化.在建立被跟踪目标的颜色直方图模型后,可将视频图像转化为基于颜色的概率分布图,跟踪的每一个迭代过程是一个反复搜索的过程,根据已有的肤色直方图模型,在计算区域内搜索窗口所覆盖部分的每个像素的肤色概率被计算出来.搜索窗口根据当前帧的肤色概率图进行位置搜索,在跟踪过程中对每一帧图像中搜索窗口的位置和尺寸实时更新,根据搜索窗口来定位手的中心和大小.

光流法是利用光流场运动信息完成目标跟踪的方法,基于光流的跟踪方法不需要预先获取图像背景,其计算结果仅仅依靠相邻两帧中手部的相对运动,因此不受复杂环境的影响比较适用于动态手势

跟踪过程.但传统的计算方法要求相邻两帧图像的亮度恒定并且运动速度不能过大,否则光流计算会产生较大的误差,难以得到精确的手部运动光流场. Bouguet 提出使用图像金字塔方法对传统的光流法进行改进^[12],从图像金字塔的最高层(即分辨率最低的一层)开始逐渐向下,在每一层中迭代地运行传统的光流算法,通过将高层的处理结果不断向下传播最后得到总的误差累计为 $d = \sum_{L=0}^{L_m} 2^L d^L$,其中 L_m 是金字塔的高度.假定最高层可以处理的图像速度场最大为 d_{max} ,则整个金字塔可以处理的速度场最大能达到 $d_{\text{max-final}} = (2^{L_m+1} - 1)d_{\text{max}}$,从而有效解决了传统方法对像素移动速度限制的问题.

光流跟踪之前首先要初始化,即在图像中选择具有较大特征值易于跟踪的特征角点. KLT^[13]是一种较为典型的算法,但是该方法所跟踪的目标仅仅是特征角点,无法提供全局的对象级信息,算法无法将手与上臂或场景中其他的运动对象有效区分.并且, KLT 比较适用于跟踪移动的刚体对象,而人手属于非刚性物体,跟踪过程中其外形会发生改变从而引起不同帧之间计算得到的特征匹配相关度降低,导致跟踪特征点丢失或逐渐转移到背景中具有更加突出的灰度梯度值的区域中,从而引起跟踪特征不一致问题.为了跟踪稳定的特征点,本文在文献[5]的基础上将 CamShift 算法与金字塔光流相结合,充分考虑了手势图像的皮肤、区域和特征信息.为了选取好的特征点达到更好的跟踪效果,在光流跟踪之前首先利用手的肤色信息进行整体跟踪,实现对手部区域粗定位,即确定当前手势区域的外包矩形 bounding_box,然后在 bounding_box 所限定的范围之内使用金字塔光流算法有针对性地跟踪同时具有肤色和距离约束的 KLT 特征点在时空上的运动情况(见图 3),算法描述如下:

(1) 初始化部分

运行 KLT 算法找到 M 个彼此之间最小距离不小于 ξ_{min} 的特征点.基于颜色概率分布模型,从 M 个特征点中选取 N 个最靠近 bounding_box 中心的特征点形成特征点集 \mathcal{R} , bounding_box 的中心坐标 P_{center} 由式(1)计算得出:

$$X_c = \frac{M_{10}}{M_{00}}, \quad Y_c = \frac{M_{01}}{M_{00}} \quad (1)$$

其中 $M_{00} = \sum_x \sum_y I(x,y)$, $M_{10} = \sum_x \sum_y xI(x,y)$, $M_{01} = \sum_x \sum_y yI(x,y)$.

(2) 跟踪部分

1. 利用图像金字塔模型更新特征点集 \mathcal{R} ;

2. 计算 \mathcal{R} 的重心 P_{median} , 在计算过程中将距离 P_{median} 较远的特征点剔除;
3. 对 \mathcal{R} 中的特征点, 如果不能同时满足下列 3 个约束条件则转步 4, 否则转步 5;
 - (a) 与其他特征点之间的最小距离小于阈值 ξ_{min} ;

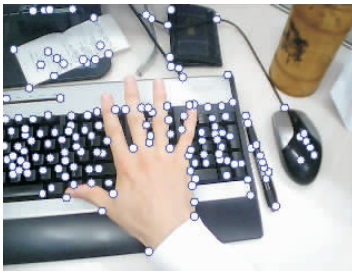
(a) M 个 KLT 特征角点(b) 由 bounding_box 所约束的 N 个特征角点

图 3 光流跟踪初始化

由于 CamShift 算法仅仅基于肤色模型跟踪, 因此手势运动过程中极易受到具有肤色信息的背景干扰, 造成搜索窗口定位不准确从而无法准确判断手部的正确位置, 而光流跟踪计算结果仅仅依靠相邻两帧中手部的相对运动, 不受复杂环境的影响能够很好地弥补 CamShift 的缺点. 但光流跟踪的 KLT 特征点无法提供对象级信息, 在跟踪过程中极易造成特征点丢失引起跟踪不一致问题. 本文将这两种方法有效结合起来, 跟踪初始化时利用 CamShift 的搜索窗口限定光流跟踪特征点的传播范围, 使其局限在目标手上, 在跟踪过程中实时更新 bounding_box 范围内的特征点. 系统使用权值法对两种方法进行有效的组合: 如果跟踪过程中 KLT 特征点丢失或者逐渐转移到了背景中那些具有更加突出的灰度梯度值的区域中, 其重要性权值将不断下降, 当低于一个特定的阈值时系统将使用颜色概率分布模型利用 bounding_box 对特征点进行重定位; 如果 CamShift 的跟踪窗口“飘移”到了与肤色相近的背景物体上, 即 P_{center} 与 P_{mean} 的距离超过一定的阈值时, 系统将根据 P_{mean} 对 bounding_box 重新调整.

3.3 手势识别

3.3.1 静态手势识别

静态手势是仅依靠手的外部形状与轮廓传递信息的方式, 在人机交互中不仅可以表达某些概念或意图, 还能作为某些特殊的状态控制系统转移, 因此对手形的识别成为了手势应用中的主要研究内容之一. 要进行静态手势识别, 首先要训练手势模型库, 识别时将提取的手势特征向量代入模型库中的各个模型, 取概率值最大者为待识别手势. 本文首先使用

(b) 位于外包矩形 bounding_box 之内;

(c) 具有较高的特征匹配相关度.

4. 重定位特征点, 使其满足上述 3 个约束条件并且位于具有高肤色概率的位置;
5. 计算 \mathcal{R} 的平均位置 P_{mean} .

3.1 节基于 PGH 的描述方法计算得到静态手势的特征向量 f_{PGH} , 然后采用贝叶斯分类器对手势进行分类. 贝叶斯分类器是基于贝叶斯决策理论的分类器, 有着成熟完善的数学基础, 它通过选择最小化条件风险的方式来使预期的损失最小化, 对所有的 $i=1, 2, \dots, m$, 计算条件风险 $R(a_i | X) = \sum_{j=1}^c \lambda(a_i | \omega_j) p(\omega_j | X)$, 其中 $\lambda(a_i | \omega_j)$ 为风险函数, 描述了类别状态为 ω_j 采取行动 a_i 的风险. 使总风险最小的类别划分 ω_j 就是 X 所属的类. $p(\omega_j | X)$ 通常被认为符合高斯分布模型, 这种模型运算简单, 而且现实世界中的很多事件都与高斯分布有极大的相似性. 高斯型贝叶斯分类器的模型参数(均值矢量与协方差矩阵)是通过训练样本的学习得到的, 经过训练后的高斯型贝叶斯分类器根据手势的特征向量 f_{PGH} 进行分类, 按照最大似然假设, 最终得到手势的所属类别.

3.3.2 动态手势识别

动态手势是具有时空模型概念的手势, 可以表达更加丰富和准确的信息, 也是人们日常生活中最常用的交流方式之一, 它主要由姿势信息(形状、位置和方向)和运动信息(一系列跟踪手势质心点的像素坐标)来表示. 由于手的三维姿态信息在没有其他辅助设施的条件下是很难实时恢复的, 因此本文将动态手势建模为参数空间里的一条轨迹而不考虑手的姿势信息, 轨迹中的点数取决于视频捕获的帧率以及手部运动速度. 传统的动态轨迹识别方法如 HMMs、神经网络和统计分类器等都需要大量的训练样本, 建模过程中需要很多人工干预, 对很多普通 UI 用户来说使用起来较为困难, 并且它们针对的是

特定的手势集合,很难对其进行扩展.本文提出了一种基于小样本学习的模板匹配方法,可以在少量样本情况下取得较高的识别率.下面给出一个动态手势分类问题的描述:

对于一个有 M 个数据类的分类问题,其中 ϕ_i , $\forall i \in \{1, \dots, M\}$ 称为一个类,每一类别代表一组相似的手势模板.令 T_{ij} 表示类别 ϕ_i 所对应的第 j 个手势模板,它是由许多个二维轨迹点构成.一个分类任务就是判断待识别样本 X 的类别所属,分类问题可转化为样本 X 与已定义的模板 T_{ij} 的最佳匹配度问题.

由于不同的用户做手势时运动的速度、空间、幅度和完整性等方面都是有非常明显的差异的,即使是同一用户每次做同一种手势时在这些方面也会有很大的不同.为了消除同类别不同动态手势之间的时空差异性,在分类之前我们首先对其进行预处理,

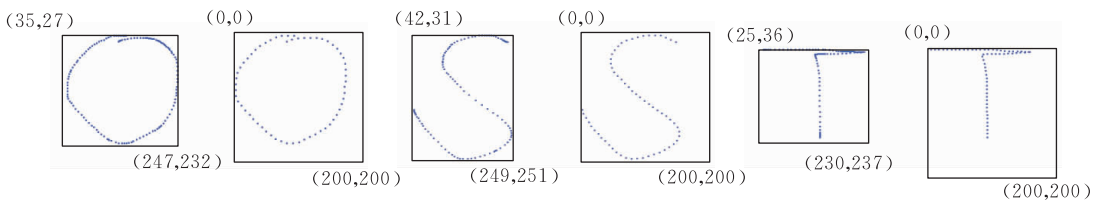


图 4 原始手势运动轨迹以及相应的处理结果示例

3.3.3 识别管理模块

识别管理模块是识别框架的智能核心,它能够根据一定的上下文对当前的手势识别任务进行适当的调度.上下文是计算系统运行环境中的一组状态或变量,这些状态和变量可以直接改变系统的行为.上下文感知是指系统自动地对上下文、上下文变化以及上下文历史进行感知和应用,根据它来调整自身的行为^[14],是提高计算智能性的重要途径.本文在识别管理模块中引入了上下文的概念,系统运行时将通过上下文信息对当前的手势识别任务进行分析,将手势识别重定向到合适的识别模块中,比如当系统检测到一个“fist”的手势后开始启动动态手势识别模块,用户的手部运动轨迹将被识别成某种特定的语义符号,从而使系统具有一定的“智能性”,这样系统在某一特定时刻只运行一个识别模块而不必同时激活所有的手势识别模块,不仅降低了系统的计算负载而且能有效避免由于 Midas Touch 引起的误识别问题.系统提供了可视化的界面,设计开发人员可以通过界面对特定手势识别任务的上下文语义信息进行灵活的设置,系统内部则实现了以 XML 为主的应用上下文管理机制,对用户所有的操作或更改都生成一个 XML 配置文件,该配置文件在识别

使其具备平移以及尺度变换的不变性.下面给出算法流程:

1. 使用线性插值法对样本重采样,使所有样本具有相同数目的采样点;
2. 将样本缩放到一个既定大小的正方形区域内,并将左上角顶点平移到(0,0);
3. 使用式(2)计算待识别样本 X 与模板 T_{ij} 的平均欧式距离用于在线模板匹配.

$$d_{ij} = \frac{\sum_{k=1}^N \sqrt{(X[k]_x - T_{ij}[k]_x)^2 + (X[k]_y - T_{ij}[k]_y)^2}}{N} \quad (2)$$

其中,待识别样本 X 与模板 T_{ij} 均具有 N 个采样点, $(X[k]_x, T_{ij}[k]_x)$ 和 $(X[k]_y, T_{ij}[k]_y)$ 分别为 X 和 T_{ij} 中第 k 个采样点的横纵坐标.取最小值 d_{ij}^* 所对应的标号 i ($1 \leq i \leq M$) 作为样本 X 的类别标号,此时样本 X 属于类 ϕ_i .图 4 给出了几组动态轨迹手势预处理结果的实例,其中每一组的前一个为原始轨迹,后一个为预处理后的结果.

时由应用层提供给识别管理模块,如图 2 箭头所示.

3.4 视觉手势事件

VBGI 主要是通过视觉手势事件驱动的,手势事件是形成交互的最终数据形态,事件的语义信息由界面设计人员定义,通常与任务直接相关,手势事件模块则是整个手势识别框架与高层应用的接口.下面给出一个通用的动态手势事件(Dynamic Gesture Event, DGE)的形式化描述:

$$DGE = \langle ID, Type, "posture", start_Gesture, end_Gesture, t, r, data, sample_Rate, mean_Time, variance_Time \rangle,$$

其中, ID 是事件的唯一标识符; $Type$ 指定该事件的类别; $posture$ 是字符串,标识手势事件的语义信息,例如“挥手”表示关闭一个文档; $start_Gesture$ 为一个预定义的静态手形,用来触发一个动态手势的开始; $end_Gesture$ 为一个可选的预定义静态手形,用来结束一个动态手势; t 为标志位,如果目标手正在被跟踪则为“1”,否则为“0”; r 为标志位,如果手势识别成功则为“1”,否则为“0”; $data$ 为浮点数组,用来储存归一化的运动轨迹点坐标; $sample_Rate$ 表示数组中点的采样率; $mean_Time$ 和 $variance_Time$ 分别表示从检测到目标手的第一帧起持续时间的均

值和方差,单位为 s.

由于 XML 具有平台无关性、自描述性、易于标准化等特点,因此本文选择 XML 语言来表达视觉交互应用中的各种事件,用户引发的各种手势事件被封装成 XML 节,每一个 XML 节表示一个单一的事件,应用端通过解析 XML 字符流还原事件,然后根据会话标志符和事件参数将事件发送给上层的逻辑处理模块.使用 XML 描述手势事件有两大优点:一方面,XML 的可扩展性使得界面设计人员可以针对具体的应用定义自己的标签,以统一的方式描述各种手势事件,作为信息共享与数据交换的基础,从而极大地方便了系统的扩展;另一方面,XML 作为数据表示的一个开放的国际化标准,将数据表示独立于机器平台、供应商以及编程语言,使得它能够较为容易地与传统的界面设计工具有效集成.

4 应用实例与评估

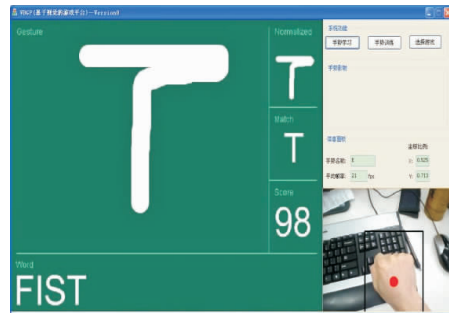
4.1 应用实例

与传统交互方式不同,基于视觉的交互必须要

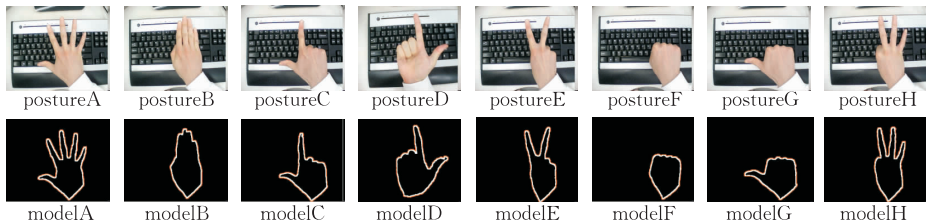
从一个含噪声的输入中获得用户的真实意图,同时每个用户又有着不同的交互习惯和个人特点.为了解决个人特性问题,对交互系统进行训练是很有必要的,界面设计者应该允许用户根据个人的喜好调整界面元素和系统设置.鉴于目前缺乏此类有效的手势界面开发工具问题,作者基于前面所述的手势识别框架设计开发了一个面向互动娱乐的视觉手势界面工具 IEToolkit,其目的是使得非计算机视觉专业的游戏开发人员不必局限于复杂的图像处理、机器学习等底层的技术细节,从而有效地加速游戏开发过程.该平台从底层到高层完全基于视觉交互的特征进行设计,能够支持灵活的视觉交互信息的处理,支持高层的手势语义事件的定义和管理,支持不同手势类型的灵活定制,支持快速的原型系统开发,简化 VBGI 的开发过程.限于篇幅本文只对其简单介绍,具体的实现过程将在另一篇文章中详细描述.图 5(a)所示为平台运行界面,其中左上角为样本编辑区,用户可以利用鼠标在视频流中采集不同的静态/动态手势样本;左下角为提取的手势模板;中间面板为训练学习区,上半部分显示样本缩略图,下半



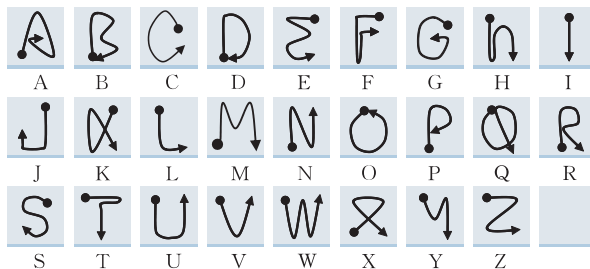
(a) IEToolkit 工具箱界面



(b) 基于视觉手势的写字板系统



(c) 标准静态手势集



(d) 标准动态手势集

部分显示分类器列表,用户可以选择不同的分类器或分类器组合进行手势模板训练;右下角为应用配置面板,系统提供了可视化界面,开发人员可以根据应用需求对输出结果进行配置,并设定上下文约束条件或自定义高层交互语义,为手势与系统命令或场景行为之间建立起一定的映射关系;右上角为游戏测试区,可以对手势训练情况进行实时测试与评估.目前基于该平台已经构造出多个原型系统,如基于视觉的虚拟家居展示系统、城市漫游系统、吞食鱼游戏等,实践表明它能够有效支持 VBGI 原型的便捷开发.

4.2 实验评估

我们邀请了 6 名人机交互专业的研究生分别从手势识别率、手势跟踪性能以及交互任务完成情况等方面对本文所提框架的可用性进行了初步的评估,这 6 名研究生虽然都来自计算机专业但以前均未接触过视觉交互,由他们来模仿非视觉专业的普通系统设计开发人员具有一定的代表性.为了便于比较,文献[5]中所使用的 8 种手势类型被作为本文的静态手势评估数据集,而动态手势评估方面则选择了 26 个英文字母作为测试集,原因是这 26 个英文字母涵盖了多种多样的动态手势变化,能够较好地验证本文识别算法的通用性及可用性.下面给出本文的实验条件:

(1)所有用户首先使用 IEToolkit 工具箱平台对 8 类静态手势和 26 类动态手势进行样本的采集、训练和测试.评估结果中的识别率采用交叉验证的方法计算得到,即从数据集中分别选择不同的样本作为训练样本,剩余的样本作为测试样本而得到的识别率的平均值.实验中每种手势类型都将在用户相关的手势库(即手势库中所存的手势模板均取自当前被试用户)中测试 10 次,为了最大限度验证本文方法的识别性能,我们对手势库中每种手势只保存 1 个模板,因此实验中用户共需完成 $6 \times (1+10) \times 8 = 528$ 个静态手势样本以及 $6 \times (1+10) \times 26 = 1716$ 个动态手势样本.

(2)实验场景中包含了各种不同颜色和形状的物体,这些物体具有多种多样的色度值,其中包括一些和人手色度值非常接近的物体,例如图 5(b)中的线框框出的盒子.

(3)本文实验硬件配置为 Intel 酷睿四核处理器,时钟频率为 2.66GHz,内存为 4GB,ATI Radeon HD 3650 集成显卡,容量为 1GB,输入视频分辨率大小为 320×240 .

在上述实验条件下,我们对 8 类静态手势和

26 类动态手势进行了测试,其识别率分别如表 1 和图 6 所示.其中,表 1 给出了 8 种静态手势的识别率与误识率.其中识别率是指手势被识别出来且能与模板正确匹配的比率,误识率是指手势被识别出来但被误匹配成了其他手势类型.表 1 中 postureE 的识别率最高,达到了 91%,postureH 的识别率最低但也超过了 60%;postureC 和 postureD 之间以及 postureF 和 postureG 之间由于手形相近比较容易混淆,具有较高的误识率;postureH 容易被识别为 postureA 或 postureE.本文实验结果并没有达到很多文献中所提方法 90%以上甚至高达 100%的识别率,主要是由于 PGH 算法本身的特点以及本文所使用的手势集中部分手势之间的相似度较高所造成的.PGH 算法的核心思想是分别以手势轮廓多边形线段间的最大、最小距离差以及相对夹角为行和列所构成的直方图匹配为基础的,而本文 6 名用户以前从未接触过视觉手势,在手势测试过程中部分手形不太规范,对于比较接近的几种手势经轮廓提取后计算得到的 PGH 直方图之间相似性较高,产生了较高的误识率.为了进一步验证本文方法,我们测试了上述手势集的几个子集,其中具有高度相关性的几个手势被分配到不同的子集中,平均识别率达到了 92%以上.

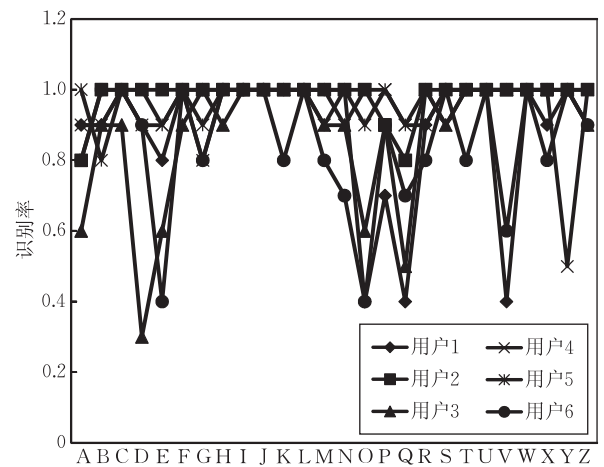


图 6 26 种动态手势的识别率

表 1 8 种静态手势的识别率

手势类型	识别率/%	误识率/%
postureA	83	4
postureB	75.5	5.5
postureC	62	20
postureD	63.5	18
postureE	91	2
postureF	85.5	8
postureG	83	10.5
postureH	61	21

动态手势识别结果可以从图 6 中体现出来,可

以看出大部分手势的识别率均超过 80%，部分手势保持了 100% 的识别率，但也有部分手势识别率很低，经分析发现主要是这部分手势如 D、O、P、Q 之间以及手势 U、V 之间具有较高的相关度，产生了较高的误识率，影响了整体识别率。本文实验结果有助于为相关开发人员提供有益的借鉴，在动态手势设计时可以考虑避开相关度较高的手势，而选择相关度较低的手势以提高整体识别率。

为了验证本文方法在交互任务完成方面的可用性以及能否有效解决 Midas Touch 问题，我们基于 IEToolkit 设计开发了一个基于视觉的写字板系统如图 5(b) 所示^①。6 名被试人员分别使用上述 26 种动态视觉手势完成指定数量的单词拼写，系统使用本文 3.3.2 节提出的模板匹配方法对动态手势轨迹进行识别。由表 1 可见 postureE 具有最高的识别率和最低的误识率，因此系统使用 postureE 作为系统初始化的检测手势，一旦检测到 postureE 则启动后面的跟踪识别模块。交互任务完成过程中，涉及到空格、删除和撤销等 3 种基本的系统操作命令，这 3 种交互命令分别由 3 种不同的静态手势来完成，用户可以在图 5(c) 的静态手势集中挑选 3 种相关度较低且符合自己交互习惯的手势类型，并通过 IEToolkit 的应用配置面板对所选的几种手势添加高层交互语义从而为手势与系统命令之间建立起一对一的映射关系。

VBGI 的交互质量和可用性主要取决于实时性、准确性和鲁棒性等几个因素^[5]，因此我们从上述几个指标出发对本文方法进行定量评估。

(1) 准确性方面。6 名被试者均选择使用 postureF 即保持握拳姿势作为动态勾画时的开始手势，而负责删除、空格和撤销等系统控制的 3 种静态手势则根据每个用户的喜好和选择各有不同。由于每个用户选择的 3 种静态手势之间的相关度较低，因此手势的整体识别率有了显著提高，平均识别率达到了 96%。我们将 postureF 手势的质心即 P_{mean} 映射为界面光标完成字母轨迹的勾画，通过目测发现光标的移动能够与手部运动较好地匹配。

(2) 实时性方面。系统在使用 postureF 进行动态轨迹勾画时，跟踪速度最快达到了 32fps，而删除、空格和撤销等其他 3 种静态手势的平均跟踪速度为

25fps，实时性方面满足人机交互的需求。

(3) 鲁棒性方面。实验结果表明本文方法能够有效适应各种不同的复杂背景与光照条件，图 5(b) 中盒子和人手色调值非常接近，利用本文方法也可以非常精确地排除它的干扰，有效跟踪时间长度占总时间的 97% 以上，从而验证了本文方法的鲁棒性。

在完成上述定量评估基础上，我们采用了调查问卷的形式收集用户反馈意见，请 6 名被试者从易学性、易用性、自然性、满意度等几个方面对本文方法进行了定性评估，结果如图 7 所示。

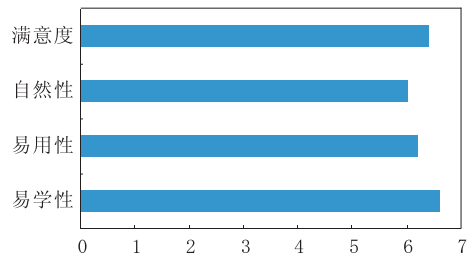


图 7 用户主观评分

从图 7 可以看出，用户对本文方法给予了较大肯定。在易学性方面，IEToolkit 实现了前文所述的关键技术并提供了灵活的训练、学习机制和可视化的用户界面，降低了学习门槛，开发过程中不需要用户从底层的代码入手而只需要关注于高层的交互任务，即便是新手学习起来也较为容易，对用户来说普遍容易掌握；在易用性方面，工具箱为用户提供了个性化定制和扩展功能，用户能够根据喜好选择并使用符合自己习惯的交互手势，交互过程中不需要过多的认知负担而将精力集中于具体的任务本身，从而大大提高了交互效率；自然性方面，手势交互一方面具有灵活、自然和高自由度等优点，另一方面又受非接触性输入影响而不可避免地产生 Midas Touch 问题。本文将前面所述的关键技术应用到系统设计中，智能的识别框架与鲁棒的算法支持成功地滤除了用户无意识的手势动作而提取出真正的交互意图，实验中用户能够灵活地在不同的静态手势和动态手势之间切换而不会产生歧义，解决了 Midas Touch 问题并有效完成了任务。

最后，我们给出了本文方法与 Kolsch 方法^[5] 的比较结果如表 2 所示。

表 2 与 Kolsch 方法的比较

	平均跟踪速度/fps	是否支持动态手势识别	训练过程	模板数量	受样本相关度的影响
本文方法	25	是	简单	少量	高
Kolsch 的方法	12	否	复杂	大量	低

① 本系统是基于 <http://www.dncompute.com/blog/2006/05/22/gesture-recognition-in-flash.html> 的 VBGI 应用。

(1) 在处理速度方面. Kolsch 扩展了 Viola 的对象检测方法^[15]用于在跟踪过程中实时识别静态手势,在本文上述机器配置条件下对各种手势的平均跟踪速度为 12fps;而本文方法在跟踪过程中采用 PGH 算法对手势轮廓进行描述并利用 Bayesian 分类器进行手势分类,对各种手势的平均跟踪速度为 25fps.

(2) 在训练方法方面. Kolsch 的方法手势训练过程复杂繁琐,为了训练图 5(c)的静态手势集共收集了 2300 多张不同性别、不同背景和光照的手势图片作为正样本,23000 多张不含手部区域的背景图片作为负样本,然后使用交叉验证的方法进行训练,即将样本集一分为二,其中一半作为训练集另一半作为测试集;而为了简化训练过程方便非视觉专业的普通用户使用,我们开发了 IEToolkit 工具箱实现了本文所提的方法,基于 IEToolkit 平台手势训练过程非常简单,用户只需要在界面上的样本编辑窗口中利用鼠标圈选就能在视频流中实现对不同静态或者动态样本的采集,样本训练过程对用户是可见的,用户只要选择了“学习结果预览”选项就能实时查看样本的训练结果,如果对学习结果不满意还可以通过界面提供的阈值控制条对分类器学习参数的阈值进行调节,总的来说这一训练过程是一个“取样—检查反馈—修正”迭代渐进的过程.训练过程透明化对用户而言可以在短时间内迅速掌握选取与应用领域最为相关的样本以及分类器训练技巧,通过这种训练策略有助于用户迅速了解分类器的功能以及内部工作机制而不再把分类器当作一个黑盒子,从而帮助他们建立起一种本能的直觉从而在以后的开发过程中大大缩减训练时间,提高开发效率.最后,通过本文上面的实验结果可以看出,不管是静态手势识别还是动态手势识别,本文方法都不需要太多的模板就能达到较高的识别率,取得满意的结果.

(3) 对动态手势的支持方面. Kolsch 的方法不支持动态手势识别;而本文方法同时支持静态手势识别和动态手势识别,并能够根据上下文在两种不同的识别模块之间灵活地切换,有效地完成人机交互任务.

(4) 在样本相关度影响方面. Kolsch 的方法受的影响比较低,本文方法在具有更多手势类型的情况下,由于受相似手势形状的影响在一定程度上降低了识别效率.

5 总结和展望

VBGI 以其自然、无约束的交互方式受到了广

泛关注并被应用到了多个领域,但是相比传统的 GUI 来说,VBGI 的设计开发更加复杂.本文针对 VBGI 存在的 3 个问题,以认知心理学中注意的信息加工模型为理论依据,从交互模型、识别框架和开发工具 3 个方面对视觉手势界面的关键技术进行了讨论,提出了一套整体的解决方案,对人机交互及 VBGI 领域的研究和应用具有一定的借鉴意义.本文提出的手势识别框架以及在此基础上开发的工具箱平台 IEToolkit 具有广阔的应用前景.我们在一个基于视觉手势的写字板系统中,针对具体的交互任务对本文所提出的解决方案进行了实验测试与评估,验证了本文方法的有效性.

本文接下来的工作包括:在算法实现方面,目前该框架基于 PGH 算法进行静态手势识别,在实验中发现 PGH 算法对用户相关的训练和应用具有较高的识别率,但对用户无关的训练应用识别率并不高.因此,下一步的工作是研究 PGH 与简单形状描述子(Simple Shape Descriptors, SSD)的有效组合,增强算法的通用性;在应用开发方面,目前已设计实现了手势界面工具 IEToolkit 并开发了面向互动娱乐的几个典型的应用系统,下一步我们将在该平台基础上开发面向其他领域的手势界面系统,从而支持更多的手势交互应用开发.

致 谢 感谢香港科技大学高元在演示系统开发中所做的工作;同时,作者所在实验室许多研究人员包括任磊、姜映映、徐礼爽、吕菲、朱继玉等人对本文提出了很多有益的建议,在此一并表示感谢!

参 考 文 献

- [1] Jacob R J K. Eye-movement-based human-computer interaction techniques: Toward non-command interfaces//Proceedings of the Advances in Human-Computer Interaction, Ablex Publishing Corporation. Norwood, New Jersey, 1993: 151-190
- [2] Kato H, Billinghurst M, Poupyrev I. Virtual object manipulation on a table-top AR environment//Proceedings of the ISAR2000. Munich, 2000: 111-119
- [3] Kjeldsen R, Levas A, Pinhanez C. Dynamically reconfigurable vision-based user interfaces. Machine Vision and Applications, 2004, 16(1): 6-12
- [4] Wu Y, Huang T S. Vision-based gesture recognition: A review//Proceedings of the Gesture Workshop. Gifsur Yvette, France, 1999: 103-115
- [5] Kolsch M. Vision based hand gesture interfaces for wearable

- computing and virtual environments [Ph. D. dissertation]. University of California, Santa Barbara, 2004
- [6] Wichens C D, Hollands J. Engineering Psychology and Human Performance. New Jersey: Prentice Hall, Inc., 2003: 82-133
- [7] Buxton W. A three-state model of graphical input//Proceedings of the Human-Computer Interaction-INTERACT'90. Amsterdam, North-Holland, 1990: 449-456
- [8] Zhu Ji-Yu, Wang Xi-Ying, Wang Wei-Xin, Dai Guo-Zhong. Hand gesture recognition based on structure analysis. Chinese Journal of Computers, 2006, 29(12): 2130-2137 (in Chinese)
(朱继玉, 王西颖, 王威信, 戴国忠. 基于结构分析的手势识别. 计算机学报, 2006, 29(12): 2130-2137)
- [9] Ashbrook A P, Thacker N A, Rockett P I. Pairwise geometric histograms: A scaleable solution for the recognition of 2D rigid shape//Proceedings of the 9th Scandinavian Conference on Image Analysis. Uppsala, Sweden, 1995, (1): 271-278
- [10] Douglas D H, Peucker T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Canadian Cartographer, 1973, 10(2): 112-122
- [11] Bradski G R. Real time face and object tracking as a component of a perceptual user interface//Proceedings of the IEEE 4th Workshop on Application of Computer Vision. New Jersey, 1998: 214-219
- [12] Bouguet J Y. Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm. Intel Corporation, Microprocessor Research Labs, OpenCV Documents, 2001
- [13] Shi J, Tomasi C. Good features to track//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, 1994: 593-600
- [14] Dong Shi-Hai. Intelligent user interfaces, its primary techniques, and instance//Progress of Artificial Intelligence in China. Beijing: Beijing University of Posts and Telecommunications Press, 2003: 632-637 (in Chinese)
(董士海. 智能用户界面、主要技术及实例//中国人工智能学会 2003 全国学术年会(CAAI-10)论文集(上册). 北京: 北京邮电大学出版社, 2003: 632-637)
- [15] Viola P, Jones M. Robust real-time object detection//Proceedings of the 2th International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, 2001: 1-25



WU Hui-Yue, born in 1979, Ph. D. candidate. His major research interests include human computer interaction, vision-based interfaces.

ZHANG Feng-Jun, born in 1971, Ph. D., associate professor. His major research interests include human computer

interaction, CAD, Virtual Reality.

LIU Yu-Jin, born in 1984, M. S. His major research interest include human computer interaction, vision-based interfaces.

DAI Guo-Zhong, born in 1944, professor, Ph. D. supervisor. His major research interests include human computer interaction, software engineering and computer graphics.

Background

This research is supported by the National Basic Research Program of China (973 Program) of China under grant No. 2009CB320804, the National Natural Science Foundation of China under grant Nos. U0735004, 60673188, and the National High Technology Research and Development Program (863 Program) of China under grant Nos. 2009AA01Z337, 2008AA01Z303.

Vision-based gesture interfaces provide more natural interactions for many interactive applications and have been widely used in many areas such as VR/MR, ubiquitous computing and vision-based computer games. But the development of vision-based hand gesture interfaces is still difficult, because it faces the following three problems: First, it is difficult for system to interpret gestures correctly because of the

Midas Touch Problem, where everything the user does is interpreted as an interaction. Second, it is hard for system to analyze and recognize hand gesture thanks to the complex environments and variable light conditions. At the end, it is embarrassed for traditional HCI practitioners to customize and expand new types of gestures.

In this paper, an effective solution is presented for vision-based gesture interfaces from the viewpoint of Cognitive Psychology, which includes an interactive model for hand gesture, a recognition framework following with many key technologies, and a vision-based gesture interfaces toolkit. Evaluation shows that our solution can effectively address the above three issues.