基于小样本学习的 3D 动态视觉手势个性化交互方法

武汇岳[1],王建民[1],戴国忠[2]
(1.中山大学传播与设计学院,广东广州 510006;2.中国科学院软件研究所人机交互与智能信息处理实验室,北京 100190)

摘 要: 传统的动态手势交互技术如隐马尔科夫模型、神经网络和统计分类器等都需要大量的训练样本,建模过程中需要领域专家的干预、对普通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.本文通过 WOZ 实验,分析了用户的行为特征并给出了基于手势的数字电视交互任务模型;提出了 3D 动态手势状态转移模型,解决了 Midas Touch 问题;提出了一种基于小样本学习的动态手势识别方法,解决了传统手势识别方法的缺点;构建了个性化手势设计平台,满足了用户的个性化定制需求;实验评估结果验证了本文方法的有效性.

关键词: 人机交互; 视觉手势; 小样本学习; 个性化交互
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2013) 11-2230-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.11.018

Personalized Interaction Techniques of Vision-Based 3D Dynamic Gestures Based on Small Sample Learning

WU Hui-yue[1], WANG Jian-min[1], DAI Guo-zhong[2]
(1. *School of Communication and Design, Sun Yat-Sen University, Guangzhou, Guangdong*, 510006, *China*;
2. *Intelligence Engineering Laboratory, Institute of Software, The Chinese Academy of Science, Beijing* 100190, *China*)

Abstract: There are some unresolved issues left behind for many traditional dynamic gesture recognition methods, such as Hidden Markov Model(HMM), Neural Network(NN), and statistical classifiers. For example, they require a large number of training examples and the involvement of expert users in the training process. Moreover, they are used for some specific gesture sets which are difficult to be extended. In this paper, we first build a task model and a state transition model for vision-based dynamic gestures. Then we propose a method for 3D dynamic gesture recognition based on small sample learning. Next we design a toolkit for development of user-defined gestures. Finally, we develop a gesture-based interactive television prototype. Experimental results verify the validity of our method.

Key words: human-computer interaction; vision-based gestures; small sample learning; personalized interaction

1 引言

随着人机交互技术的发展,各种新的交互手段不断涌现,使人机交互朝着更加自然、高效和智能化的方向发展.基于视觉手势的用户界面是目前主动式用户界面的主流方式之一,同时也是普适计算环境中自然人机交互的核心和热点问题之一.与传统的 WIMP 交互方式相比视觉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,被认为是当前图形用户界面基础上最自然的进化和发展方向之一[1].因此,手势交互技术受到了国内外越来越多的关注,迅速成为了人机交互领域的一个热门研究方向,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.

但是,视觉手势交互技术在给普通用户带来方便的同时,也给专业研究人员带来了很多挑战.设计鲁棒的跟踪和精准的识别算法、有效的输入/输出技术、便利的设计开发工具和有效的评估方法成为该领域中的关键问题[2~5].有的系统依赖于事先构建好的先验知识库或

收稿日期:2011-12-27;修回日期:2013-05-12;责任编辑:梅志强
基金项目:国家自然科学基金(No.61202344,No.61073132);高校基本科研业务费中山大学青年教师培育(No.1209119-17000-3161120);中山大学 958 工程"全媒体时代的新闻传播创新基地"(No.90027-3284200)

者目标用户的 3D 骨骼模型[6];有的系统易受到不稳定的光照条件或复杂背景的影响[7];有的系统在跟踪失败后难以自动初始化[8];有的系统需要繁琐复杂的训练过程,而且手势集一旦确定就难以进行个性化定制和扩展[9].上述种种约束使得多年来视觉手势技术更多地只是进行实验性研究,而难以在人们的日常生活中得以广泛应用.

近年来,新的传感器技术、生物控制理论以及无标记动作捕获技术的研究进展,极大地推动了视觉手势交互技术的发展.像索尼的 Eyetoy、微软的 XBOX360 以及 ION 的 Educational Gaming System 都已达到商业级应用.而另一方面,三网融合下的交互式数字电视(iDTV, interactive Digital Television)如今更像是一台多媒体电脑,能够上网、玩游戏以及处理一些日常生活/办公事务.相对于以前单向推送的大众传播方式,如今的数字电视可以为受众提供检索、导向和双向互动性功能,受众则变被动接受为主动参与,体现出更大的"交互性".在现代数字家庭中利用自然手势操纵数字电视能够发挥人类固有的技能,增强用户对界面的感知,诱导用户去主动完成特定的交互任务.很显然,在普通用户家庭中使用传统的方法如 HMM 等进行手势训练与识别应用是非常困难的.鉴于此,本文提出了一种基于小样本学习的三维动态视觉手势个性化交互方法,降低了学习的门槛,并应用在了 iDTV 中进行了可用性验证,取得了满意的实验结果.本文方法不仅能够应用于数字电视,还可以应用于其他数字设备的交互(例如智能空调、智能窗帘等)或者基于 PC 的虚拟现实/增强现实以及互动游戏等不同领域.

2　相关研究

手是人类最灵敏的身体部分,在物理世界中能够被用来完成各种操作任务,而具有高效运动/操作技能的双手也可以很方便地被训练用来执行人机交互上下文中的各种虚拟控制任务.例如,将手映射为一个虚拟鼠标,来完成虚拟场景中各种指点和勾画任务[10];将手势应用在虚拟/增强现实环境下,用来驱动漫游或者完成对虚拟对象的抓取、平移、旋转和缩放等各种操作[9];将手势应用在交互桌面系统中,使得用户能够像在物理桌面上操作真实物体一样操作交互桌面系统中的数字物体[7].

除了双手之外,更大幅度的肢体/全身运动也被很多研究人员利用来设计更多的交互系统.MIT 的研究人员开发了一系列基于肢体手势的互动娱乐系统[11];Tollmar 等人开发了一个无约束的 3D 手势漫游系统[6];另外,Sony、Microsoft 和 GestureTek 等国际大公司也都开发了商业级的视觉手势应用系统像 Eyetoy, XBOX360 和 GestureFX 等.

但是上述手势交互技术大都利用了一些较为简单的身体动作来控制基于 PC 的各种游戏类应用,很少有用来控制精确的交互任务,例如菜单漫游或者导航等.Freeman[12]等人设计了一个概念原型,用户能够利用双手控制屏幕光标来调节电视的音量、对比度和亮度等;Bretzner 等人[13]设计了一套静态手势,用来触发相应的电视操作命令,例如打开/关闭电视,切换到上/下一个频道等.但是 Freeman 和 Bretzner 等人只是提出了一个概念原型,系统离真正实用还有一定的距离,而且他们的系统无法支持手势的个性化定制,这是本文重点解决的问题之一.

3　视觉手势交互模型

3.1　数字电视任务模型

在开发一个基于视觉手势的 iDTV 原型系统之前,首先要从用户的视角来定义系统的功能和任务模型,以此来帮助开发人员在系统设计之初就准确地把握用户的需求,从而提高系统的可用性.为此,我们设计了一个实验来收集用户的需求反馈.12 名来自不同专业的在校大学生参加了本次实验,这 12 名大学生的年龄介于 16~23 岁之间,以前从未接触过视觉手势交互技术.

实验被安排在一个配有数字电视的可用性实验室中进行,用来模拟用户看电视的场景.整个过程分为两个部分:(1)我们对被试进行 1 对 1 的面对面访谈,要求每个被试在纸上列出基于视觉手势的电视系统所需要具备的最基本的功能,接下来我们对所有用户的数据进行汇总统计,得出一个通用的功能集合;(2)我们要求所有被试独立地设计出一套适合于完成这些功能的手势集合.为了不受当前手势识别算法和交互技术的限制和影响,以便收集到用户最自然和最习惯的手势动作,我们使用了"Think-aloud"技术来确定用户的动作语义,即用户必须大声说出他们正在从事的任务以便于实验人员记录;同时,我们使用了 WOZ 方法[14]来实时呈现用户的动作效果,用户被告知他们正在使用一个手势识别系统与电视交互,但实际上身后的一个实验人员(Wizard)使用一个遥控器控制电视来产生该动作的预期效果.为了保证手势集的原始创新性,实验过程中我们并不给用户提供任何语言/动作提示以及手势的识别反馈.

实验使用了 5 个监控摄像头从不同角度对 12 名被试的所有的交互行为进行录制并进行后续数据分析和提取,对 12 名被试的实验结果进行定量分析和统计,结果如表 1 所示.

表 1 iDTV 交互任务映射模型

手势	语义	功能
Push	水平前推	确认/选择/进入;播放/暂停;取消
Swipe left	水平向左	下一页/上一页
Swipe right	水平向右	上一页/下一页
Swipe up	水平向上	增大音量/对比度/亮度;取消;上一页;关闭电视
Swipe down	水平向下	减小音量/对比度/亮度;下一页
Clockwise circular motion	顺时针画圆	进入/选择/确认;下一页;播放
Double-push	双击	停止
Anticlockwise circular motion	逆时针画圆	返回主菜单;取消;上一页;停止
Waving goodbye	挥手再见	关闭电视
Double-swipe right	重复水平向右	快进
Double-swipe left	重复水平向左	快退

实验发现 1 手势的个性化需求.为了完成同一个交互任务,不同的用户设计了不同的手势动作.例如,对于一个切换到下一页(turn to next page)的命令,有 7 个用户设计了 Swipe left,2 个用户设计了 Swipe right,2 个用户设计了 Swipe down,还有 1 个用户设计了 Clockwise circular motion 手势.

实验发现 2 手势的可重用性需求.例如,用户使用了一个 Push 手势来完成一个播放电影的命令,紧接着又重用 Push 手势来停止电影播放.

实验发现 3 上下文需求.在视觉手势交互中上下文是一个非常重要的因素.设计合理的上下文有助于手势的重用,并能有效避免所谓的 Midas Touch 问题[15].我们发现,用户在每一个手势动作的开始之前和结束之后都会有意识地明显停顿,以此来区分不同的手势语义.

基于上述实验发现,接下来本文提出了一个基于视觉的 3D 动态手势状态转移模型,用以指导视觉手势交互技术的设计和实现,在实践中有效避免 Midas Touch 问题;提出了一种基于小样本学习的 3D 动态手势识别方法并在此基础上开发了一个手势设计工具,用以满足不同用户的手势个性化定制需求.

3.2 交互状态转移模型

摄像头属于非接触性输入设备,只是用来捕获人体运动以及感知周围环境变化,并没有实质性地参与交互.在传统 WIMP 界面中应用广泛的设备状态转移模型[16]已经不适合描述视觉手势的交互特征,因为用户无法像操作鼠标或手写笔那样利用额外的按键来完成不同状态的转换.Midas 问题产生的本质就是因为系统无法自动进行手势的时空分割.结合 WOZ 的实验发现,本文将人手作为一种抽象的输入设备,并结合了用户心理模型,提出了一种新的视觉手势交互状态转移模型(图 1),用于指导视觉手势交互技术的设计与实现.将认知心理学与手势交互技术相结合,能够从更高层次上有效地描述用户在自然环境下的各种交互行为特征,符合普适人机界面的设计策略[17].

图 1 中,状态 1 表示用户的手处于摄像头视野范围之外,此时,用户无论做什么动作都不会对交互产生影响,界面也不提供任何形式的反馈.状态 2 表示用户的手进入摄像头的视野范围之内但尚未被系统识别出来.此时,用户的动作对交互也没有影响,但是界面应提供反馈,例如实时显示摄像头所捕获的手的深度图像,这一反馈能够充分利用用户的前庭感知和运动感知(Vestibular and Kinesthetic Senses,负责感知手在空间中的位置及位置变化,并通知中枢神经及时调节保持人体平衡),从而有效地促进交互.状态 3 为用户的手被成功检测出来之后的状态,称之为调整状态,此时界面自动产生一个跟踪符号(例如一个手形的图标),该跟踪符号将随着手的移动而实时变化状态(通过不断变换自身的 2D 平面坐标以及缩放比例来可视化人手在 3D 物理空间中的位置及深度信息变化).我们以条件阈值 T_1 和 D 作为两个不同手势之间的状态转移条件,当用户手的调整时间(每一个手势动作开始之前,用户有一个时间差来调整以得到一个理想的起始位置)$t > T_1$ 并且两帧之间手的移动距离 $d > D$ 时(手在物理空间中的实际移动距离,主要用于除噪和降低误识别率),系统进入状态 4,此时界面除了显示跟踪符号外还显示手势的运动轨迹.从 WOZ 实验结果可以看出,用户所设计的手势大都是较为简单的单笔画手势,通过对 12 名用户的手势运动时间进行统计分析,我们定义了时间阈值 T_2 来设定每个手势的最大运动时间,当满足 $t > T_2$ 并且两帧之间手的运动距离 $d > D$ 时,系统切换到状态 5 进行手势识别.在状态 5 系统一方面将识别结果实时

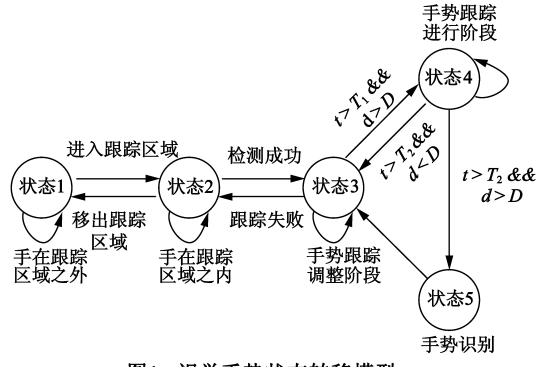


图1 视觉手势状态转移模型

反馈给用户,另一方面自动跳转到状态 3,此时用户可以调整手的空间位置或方向以准备下一个手势动作;反之,如果 $t > T_2$ 但是 $d > D$ 时,系统重新切换至调整阶段,从而减少了无意义的手势识别.

4　动态视觉手势识别方法

为了识别表 1 中的 3D 动态手势,本文在文献[18]的基础上提出了一种基于小样本学习的 3D 动态手势识别方法,大大降低了视觉手势的设计开发门槛.

4.1　手势模板及手势库

我们基于 Kinect 搭建了一个实验平台,并基于 OpenNI SDK 设计实现了手势识别引擎.利用 Kinect 提供的 SDK 可以获取由 20 个关节所构成的人体骨骼模型,我们在每一帧中直接提取人体左右手的关节点三维空间坐标 $P_t(X,Y,Z)$作为手的质心位置.记录每个手势动作过程中产生的一个连续的帧序列,并连接每一帧中所包含的手的质心点便形成了一个三维手势运动路径.下面给出 3D 动态手势识别问题的形式化描述:

假定 L 为一个既定的手势类库,$L = \{G_1, G_2, \cdots G_n\}$,其中 $G_i(\forall i \in \{1,\cdots,n\})$称为一个手势类,它代表一组相似的手势模版,$G_i = \{T_{i1}, T_{i2}, \cdots T_{im}\}$.每一个手势模板 T_{ij} 都是由一组空间三维轨迹点组成,$T_{ij} = \{P_{x1,y1,z1}, P_{x2,y2,z2}, \cdots P_{xn,yn,zn}\}$,其中$(x_i, y_i, z_i)$为空间中某三维轨迹点 P_i 的物理坐标.那么,对一个待识别的手势样本 X(我们同样定义 X 由一组空间三维轨迹点组成),其分类问题可转化为 X 与已定义模板 T_{ij}之间的最佳匹配度问题(即两组空间三维轨迹点之间的最佳匹配).

与传统的 HMM 等方法不同的是,为了满足 WOZ 实验结果中的用户个性化手势设计这一需求,在实际应用时,手势库的建立可根据具体的功能需求以及用户的个性偏好而实时创建和修改,每个用户创建自己的手势库,而非面向所有用户事先创建好一个通用的手势库.一个手势库对应一个用户的个性化设计空间,其中保存用户自己的手势特征参数.在线识别时,由于当前待识别的手势样本是与用户自己库中的训练模板匹配,因此相对于与一个通用的手势库中的模板匹配来讲,能在一定程度上提高识别率.

4.2　算法实现

为了消除手势之间的时空差异性,需要先对样本进行一定的预处理,确保手势样本中的轨迹点等数目等距离分布,并且使其具备平移与旋转不变性,然后对其进行匹配和分类.下面给出具体的算法流程:

Step 1　重采样,消除时间及速度噪音.由于受不同的视频捕获设备以及用户动作速度等因素的影响,手势运动轨迹的采样率会不尽相同.为了优化分类,需要对待识别手势样本 X_N 进行重采样(N 为重采样前手势路径中轨迹点的个数),从而得到与手势库中模板具有相同采样点数目的新样本 X_n(n 表示重采样后手势路径轨迹点的数目,n 是一个经验值,可通过实验统计方法获得).采样过程如下:首先计算采样前 X_N 中由 N 个轨迹点构成的路径总长度 L,然后将整个路径划分为 $n-1$ 等份;接下来从路径的起始点开始,以 $\text{step} = L/(n-1)$为步长,对原始样本 X_N 重采样.采样过程中如果两个点之间的距离超过了步长 step,则使用线性插值法构建一个新的采样点.

Step 2　消除位置噪音.将 X_n 的质心点平移到空间坐标系原点(0,0,0)(图 2),消除不同手势空间位置的差异性.文献[18]将所有手势样本非均匀缩放到一个统一的基准正方形区域中进行了归一化处理,这样处理完之后的手势样本将被严重扭曲,将无法区分依赖于特定的方向角或长宽比等特征信息的手势,例如长方形与正方形,圆与椭圆,以及本文 WOZ 实验结果中的 Swipe up 与 Swipe down, Swipe left 与 Swipe right 等手势.因此,本文并不对手势样本进行缩放归一化处理,从而保持了原始手势的长宽比等特征信息.

Step 3　消除方向噪音.由于交互习惯的不同,动态手势输入过程中存在方向差异性,我们接下来对输入样本 X_n 进行旋转,使其与模板 T 达到最佳拟合角度.文献[18]将所有手势样本的方向角(手势路径的质心点到起始点的连线方向)均旋转至 0 度角位置,使其具备方向无关性.而在实际应用中,手势的方向信息可能是有用的(例如 WOZ 实验中的 Swipe up 和 Swipe down 就是两个不同的手势),因此我们令向量 $\boldsymbol{V} = P_0 - P_c$(其中 P_c 为 X_n 的质心,P_0 为 X_n 的起点),然后分别作如下处理:

(1)如果样本 X_n 是方向无关的,则将它沿着其质心旋转 ω_1 度,使得方向角为 0.ω_1 的计算方法如式(1)所示:

$$\omega_1 = \arccos \frac{\boldsymbol{V} \cdot \boldsymbol{i}}{|\boldsymbol{V}|\,|\boldsymbol{i}|} \tag{1}$$

其中,$\boldsymbol{i} = (1,0,0)$为单位向量.

(2)如果样本 X_n 是方向相关的,则将它沿着其质心旋转一个最小的角度 ω_2,使得 X_n 的方向角与某一基准轴对齐(X 轴正/负向、Y 轴正/负向、Z 轴正/负向),ω_2 的计算方法如式(2)所示:

$$\omega_2 = \arccos \frac{\boldsymbol{V} \cdot \boldsymbol{P}_m}{|\boldsymbol{V}|\,|\boldsymbol{P}_m|} \tag{2}$$

其中,$\boldsymbol{P}_m$ 为与 X_n 方向角对齐的某基准轴的单位向量.

Step 4　计算 X_n 与模板 T 之间的最大相似度 $\text{score}(X_n, T)$用于实时在线模板匹配:

$$\text{score}(X_n, T) = \frac{1}{\arccos \dfrac{X_n \cdot T}{|X_n| \, \| \, T|}} \tag{3}$$

从式(3)可以看出,相似度的计算只与 n 维特征空间中两个向量的余弦距离有关,而与手势大小无关,因此满足尺度缩放的不变性.

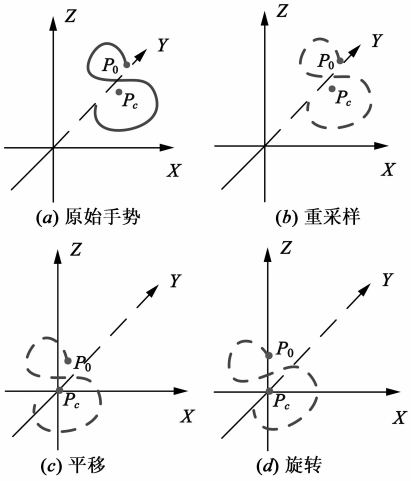


图2 原始手势运动轨迹及预处理示意图

Step 5 文献[18]的方法主要应用在支持手写输入的移动设备中,系统处理的是基于触控界面的精确的2D运动轨迹,而本文研究对象是物理空间中基于非接触界面的模糊的3D动态视觉手势,其运动轨迹不像2D笔手势那样精确和规范,因此我们对每一类手势分别收集5个样本作为模板并添加到手势库中.为了降低可能出现的假阳性(false positive)概率,本文并非直接取Step4中计算得到的最大值 $\text{score}(X_n, T)_{\max}$ 所对应的模板所在的手势类作为输入样本 X_n 的最终识别结果,而是按照得分多少进行二次排序:如果某一手势类中的模板 T 与 X_n 的匹配度超过一定的经验阈值 ξ_1,则直接取 T 所在的手势类作为样本 X_n 的识别结果;否则,在得分最高的三个候选模板中,如果存在两个模板是属于同一手势类,并且得分均超过了经验阈值 ξ_2($\xi_2 < \xi_1$),则返回这两个模板所属的手势类别;否则,返回手势未被识别.

4.3 性能评估

接下来,我们将本文方法与文献[18]的方法进行了对比.由于文献[18]只能处理二维平面手势,为了便于二者比较,我们对文献[18]进行扩展,增加了 Z 轴上的深度信息,使其能够处理三维手势.接下来使用2.1节WOZ实验结果中的11个手势作为测试集.实验配置为一台2.4GHz CPU、8G内存 、2T硬盘的主机,实验结果如图3所示.

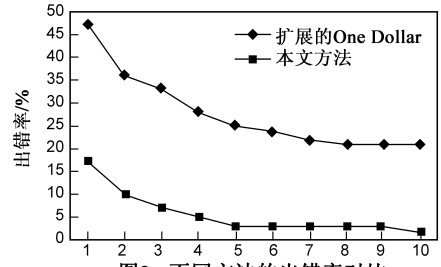


图3 不同方法的出错率对比

图3可以看出,随着样本数量的增加,两种方法的出错率都会随之降低.但整体来看,本文的方法要优于扩展的One Dollar方法.后者出错率较高的主要原因正是由于本文在4.2节所述的那样,它对于所有的样本都统一地非均匀缩放到一个统一的基准正方形区域中进行了归一化处理使其具备大小不变性,并且把所有样本的方向角都旋转到了0度,使其具备方向不变性,而其匹配度计算公式则直接受这两个因素决定,因此无法有效区分WOZ实验中的Swipe left与Swipe right,Swipe right与Swipe down,以及Clockwise circular motion与Anti-clockwise circular motion这些成对手势.而本文方法则很好地解决了这些问题,从而提高了识别率.另外,从图3中我们也可以看出,本文方法在模板数量增加到5个的时候,就能将出错率控制在5%以内,而随着模板数量的继续增加,出错率并没有显著降低,相反增加模板数量则会消耗更多的系统计算资源和计算时间.

5 个性化手势开发平台及实验评估

从WOZ实验可以看出,不同用户使用手势完成交互任务的个体差异性很大,为了遵循自然人机交互的原则、尊重用户的个性化交互习惯、减少用户的认知负

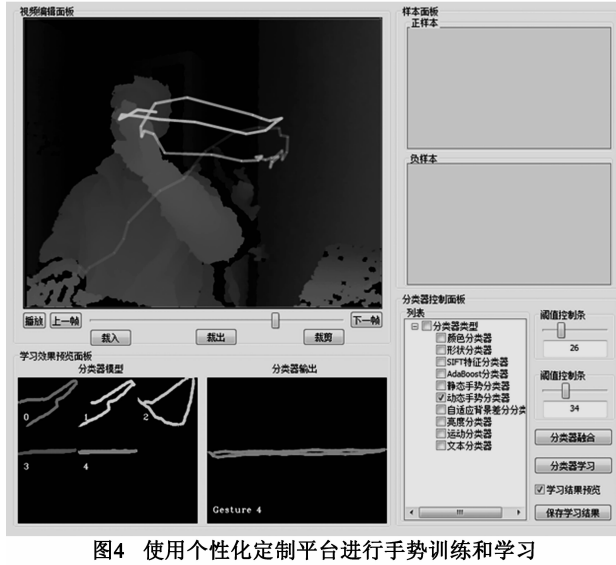
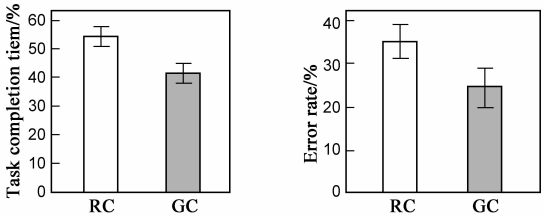


图4 使用个性化定制平台进行手势训练和学习

担并提高系统的识别率和可用性,我们设计开发了 3D 动态视觉手势个性化定制平台[19].图 4 所示为用户正在使用该平台进行手势训练和学习的过程.每个用户在手势训练学习过程中所产生的数据将被系统自动保存下来,生成个性化手势设计空间.

为了满足用户个性化需求,我们在工具箱中实现了可视化界面方便用户灵活地建立高层的手势语义映射模型,工具箱在后台将自动地根据映射模型而生成基于 XML 的手势-语义映射文件,在系统实时运行时将根据该语义映射文件并结合上下文信息模型完成用户交互意图的自动解析和自动感知.为了完成既定的交互任务,不同用户先使用工具箱定制一套符合自己交互习惯的手势模板,然后基于工具箱建立手势-语义映射模型,整个过程不需要用户深入底层进行复杂的程序编码,而只需要利用工具箱提供的可视化界面进行简单的配置就可以了,从而大大降低视觉手势开发门槛、提高了原型开发的效率.

利用该平台我们开发了一个基于视觉手势的 iDTV 系统原型,并设计了一组对比实验,重新邀请了参与前面 WOZ 实验的 12 名被试分别使用视觉手势(GC)和遥控器(RC)两种不同的交互方式完成事先指定的一组页面导航和节目选取任务.在实验开始前,这 12 名被试首先使用工具箱进行手势训练(根据 4.3 的实验结果,我们对每一类手势均收集 5 个样本作为模板).经过一段时间的训练之后,12 名用户的手势的平均识别率均达到了 93% 以上.接下来,用户利用手势工具箱进行手势语义建模并最终完成所规定的任务.两种不同技术的对比结果如图 5 所示.



(*a*) 任务完成时间对比　(*b*) 失误率对比

图5　定量评估结果

配对样本 T 检验(matched-pair t test)结果表明两种不同的交互技术存在显著性差异.在平均任务完成时间方面,手势要快于遥控器($t_{23}=7.389, p<.001$).主要原因是遥控器属于离散设备,用户经常需要连续多次按键才能控制光标从界面的一端漫游到另一端来切换目录或选取菜单项或激活命令;而手势是连续输入,用户只需要一个简单的直线运动就能控制光标从界面的一端到另一端,从而大大缩短了交互的时间(mean = 12.5s);在失误率方面,手势也低于遥控器($t_{23}=3.745$, $p<.001$).遥控器出错率较高主要是由于交互过程中的手眼分离造成的,尤其是为了完成某些交互任务而需要使用组合按键的时候更容易出错,比如连续快速地多次按箭头键再按确认键会走错目录而不得不再次按返回键回到起点,所以当需要使用组合按键的时候用户不得不经常低下头来看按键;而使用手势则不会产生这一问题,因为手与屏幕都是在视野前方.

接下来,我们从易学习性、易使用性、舒适性和交互性四个方面对这两种交互技术进行主观评估(其中 1 为最坏,7 为最好),结果如图 6 所示.

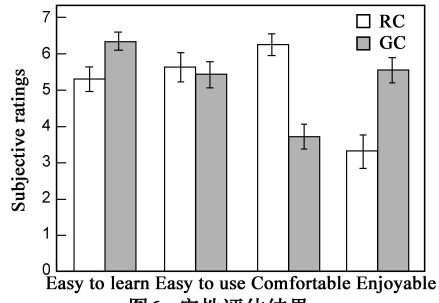


图6　定性评估结果

威氏符号秩检验(Wilcoxon signed ranks test)结果表明用户认为手势比遥控器更容易学习($Z_{GC\ vs.\ RC}=3.65$, $p=.0001$),并且更具有互动性($Z_{GC\ vs.\ RC}=4.28$, $p<.0001$).这主要是因为实验所使用的手势是基于 WOZ 实验结果而定义的,并且都是由用户本人个性化定制的,手势的简单性、原始性和易用性对用户的交互起到了关键的作用;但在舒适性方面遥控器则具有显著优势($Z_{RC\ vs.\ GC}=4.28, p<.0001$).这是因为交互过程中用户可以一手抓着遥控器进行按键操作,而同时手则可以放到腿上或沙发上,不易产生疲劳;而手势交互过程中所有动作都在空中完成并且需要很多肌肉群组的参与,手势过程中用户无法找到一个支点,长时间操作就很容易产生疲劳.在可用性方面,二者没有显著性差异($Z_{RC\ vs.\ GC}=0.7, p=.242$),这一结果充分表明了用户对手势交互技术的认可程度.

6　结论和展望

视觉手势以其非接触的直接操纵方式受到了国内外研究者的青睐,成为目前人机交互的研究热点.本文提出了一种基于小样本学习的 3D 动态视觉手势个性化交互方法.为了验证本文方法的有效性,我们开发了支持用户个性化定制的手势设计工具,并在此基础上构建了一个原型系统并进行了可用性评估.实验结果表明,用户对手势交互给予了较大肯定.本文下一步的工作重点是设计更为详细和深入的实验,研究分析不同手势相互之间的可理解性、可记忆性、可检测性以及

可区分性.

参考文献

[1] Porta, M. Vision-based user interfaces: Methods and applications[J]. International Journal of Human-Computer Studies, 2002(57):27 – 73.

[2] 杜友田,陈峰,徐文立,等.基于视觉的人的运动识别综述[J].电子学报,2007,35(1):84 – 90.
Du Y T, Chen F, Xu W L, et al. A survey on the vision-based human motion recognition[J]. Acta Electronica Sinica, 2007, 35(1):84 – 90. (in Chinese)

[3] Moeslund, T B, Hilton, A , Kruger, V. A survey of advances in vision-based human motion capture and analysis[J]. Computer Vision and Image Understanding, 2006(104):90 – 126.

[4] 任海兵,祝远新,徐光佑,等.基于视觉手势识别的研究——综述[J].电子学报,2000,28(2):118 – 121.
Ren H B, Zhu Y X, Xu G Y, et al. Vision-based recognition of hand gestures: A survey[J]. Acta Electronica Sinica, 2000, 28(2):118 – 121. (in Chinese)

[5] 徐一华,李善青,贾云得.一种基于视觉的手指屏幕交互方法[J].电子学报,2007,35(11):2236 – 2240.
Xu Y H, Li S Q, Jia Y D. A vision-based method for finger-screen interaction[J]. Acta Electronica Sinica, 2007, 35(11): 2236 – 2240. (in Chinese)

[6] Tollmar K, Demirdjian D, Darrell T. Navigating in virtual environments using a vision-based interface[A]. Proc of the 3th Nordic Conference on Human – Computer Interaction[C]. New York: ACM Press, 2004. 113 – 120.

[7] Hilliges O, Izadi S, Wilson A D, et al. Interactions in the air: Adding further depth to interactive tabletops[A]. Proc of the 22th Annual ACM Symp on User Interface Software and Technology[C]. New York: ACM Press, 2009. 139 – 148.

[8] Sidenbladh H, Black M, Fleet D. Stochastic tracking of 3D human figures using 2D image motion[A]. Proc of the 6th European Conf on Computer Vision[C]. London: Springer-Verlag, 2000. 702 – 718.

[9] Kölsch M, et al. Vision-Based Hand Gesture Interfaces for Wearable Computing and Virtual Environments[D]. Santa Barbara: University of California, 2004.

[10] Mo Z Y, Lewis J P, Neumann U. SmartCanvas: A gesture-driven intelligent drawing desk system[A]. Proc of the 10th International Conference on Intelligent User Interfaces[C]. New York: ACM Press, 2005. 239 – 243.

[11] Wren C, Azarbayejani A, Darrell T, et al. PFinder: Real-time tracking of the human body[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):780 – 785.

[12] FreemanW T, Weissman C. Television control by hand gestures[A]. Proc of IEEE Intl Workshop on Automatic Face and Gesture Recognition[C]. Zurich: IEEE, 1995. 1 – 5.

[13] Bretzner L, Laptev I, Lindeberg T, et al. A Prototype System for Computer Vision Based Human Computer Interaction[R]. Stockholm: KTH (Royal Institute of Technology), 2001.

[14] HÖysniemi J, Hämäläinen P, Turkki L, et al. Children′s intuitive gestures in vision-based action games[J]. Communications of the ACM, 2005, 48(1):45 – 52.

[15] Jacob R J K. What you look is what you get: Eye movement-based interaction techniques[A]. Proc of the SIGCHI Conf on Human Factors in Computing Systems Computer[C]. New York: ACM Press, 1991, 11 – 18.

[16] Buxton W. A three-state model of graphical input[A]. Proc of the Human-Computer Interaction (INTERACT′90)[C]. Amsterdam: Elsevier Science Publishers B V, 1990. 449 – 456.

[17] 岳玮宁,董士海,王悦,等.普适计算的人机交互框架研究[J].计算机学报,2004,27 (12):1657 – 1664.
Yue W N, Dong S H, Wang Y, et al. Study on human-computer interaction framework of pervasive computing[J]. Chinese Journal of Computers, 2004, 27(12): 1657 – 1664. (in Chinese).

[18] Wobbrock J O, Wilson A D, Li Y. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes[A]. Proceedings of the 20th Annual ACM Symp on User Interface Software and Technology[C]. New York: ACM Press, 2007. 159 – 168.

[19] 武汇岳,张凤军,刘玉进,等.基于视觉的互动游戏手势界面工具箱[J].软件学报,2011,22(5):1067 – 1081.
Wu H Y, Zhang F J, Liu Y J, et al. Vision-based gesture interfaces toolkit for interactive games[J]. Journal of Software, 2011, 22(5):1067 – 1081. (in Chinese).

作者简介

武汇岳 男,博士,讲师,1979 年生于山东烟台.研究方向为人机交互、用户界面.
E-mail: wuhuiyue@gmail.com

王建民 男,博士,教授,1973 年生于内蒙古包头.研究方向为交互设计、社会媒体计算.