

普适计算中多通道交互建模研究*

马翠霞⁺, 戴国忠, 王宏安

(中国科学院 软件研究所, 北京 100080)

Research on Multi-Modal Interaction Model in Pervasive Computing

MA Cui-Xia⁺, DAI Guo-Zhong, WANG Hong-An

(Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-10-62559307, Fax: +86-10-62629509, E-mail: cuixia@ios.cn

Ma CX, Dai GZ, Wang HA. Research on multi-modal interaction model in pervasive computing. *Journal of Software*, 2006,17(Suppl.):243-248. <http://www.jos.org.cn/1000-9825/17/s243.htm>

Abstract: As technologies are rapidly evolving in the area of storage, recognition and displays, the vision of multi-modal interaction is positioning the human needs central to pervasive access to services. Users expect better looking interfaces that are also easy to use to improve high efficiency and productivity and to focus on task itself instead of tools as providing the invisibility in pervasive computing. However, it is hard to obtain the right service at the right time naturally in the multimedia environment based on current interaction approaches. In this paper, a framework is proposed for multi-modal interaction which offers the sketch and speech interactive models without limitation of the diversity and richness of the provided services in the pervasive computing environment. Furthermore, it is applied to describe sample application and shows how it is used to create a new interface with high performance.

Key words: multi-modal interaction; pen and speech; sketch; context aware; pervasive computing

摘要: 计算机处理速度和性能的迅猛提高并没有相应提高用户使用计算机交互的能力,其中一个重要原因就是缺少一个与之相适应的高效、自然的人机交互界面。普适计算的发展为人们提供了一个无处不在的获取服务的环境,而多通道用户界面允许用户使用自然的交互方式,如语音、手势、表情等在此环境中与系统进行协同工作,改变了传统交互手段的局限性。重点讨论了基于笔和语音的多通道交互,给出了普适环境中对多通道交互建模的框架,进一步研究了在图形信息应用中的多通道交互实例,论证了创建自然多通道交互方式的优势。

关键词: 多通道交互;笔和语音;草图;上下文感知;普适计算

普适计算的发展改变了传统人机交互的模式,它描述了具有丰富计算资源和通信能力的人与环境之间关系的场景,能够在需要的何时何地提供信息和服务,环境与人们逐渐地融合在一起^[1,2]。传统计算模式是以计算机为中心的,其使用方式不是按照用户的习惯,而主要是由计算机规定如何做,用户有太多的注意力是在工具上,而不是在解决的问题上。目前的用户界面是基于鼠标和键盘的 WIMP(windows,icon,menu,pointing devices)界

* Supported by the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312103 (国家重点基础研究发展规划(973))

Received 2006-03-30; Accepted 2006-10-08

面,它以桌面为隐喻,频繁的菜单选择、按钮操作和键盘命令输入不符合人们日常主要依靠语音和视觉的交互习惯,操作与内容分离、界面布局复杂,加重了认知负担,用户需花费大量时间去学习,记忆大量的命令符号,对于非熟练的用户更是效率低下.多通道的交互方式主要是利用人们的日常习惯和技能来与机器进行交互,可以简化用户命令,缩短完成任务的时间,用户易于接受,操作流程自然、连续并具有灵活性(适用于个体间的空间能力差异与意愿差别).

1 自然交互

人类社会正在向信息社会迈进,人机之间的交互理念也发生了巨大变化,追求一种自然、和谐以及不可见的交互方式逐渐成为下一代交互技术研究的目标.充分利用人体丰富多彩的感知和动作器官,以及人们与日常物理世界打交道时所形成的自然交互技能来获得计算机提供的服务越来越成为新一代用户界面的主要交互方式^[3].自然用户界面和笔式用户界面的使用越来越广泛.在基于桌面隐喻的交互方式下,人们习惯于各种窗口间的切换,习惯于在桌面上去寻找合适命令来完成任务,这种交互方式至今广泛存在,灵巧的鼠标使得很多工作在几下点击之后就可以轻松完成.但在很多情况下这种方式剥夺了人们的自然交互技能,使用过程中需要较为繁重地培训和去熟悉操作过程,并且对操作步骤的精确要求也增加了交互难度,影响了使用上的自由.

其中基于笔和语音的多通道界面是目前应用最广泛、面向广大普通用户的主流应用界面^[4,5].笔和语音作为最主要的交互通道,具有很强的互补性,极大地提高了操作效率.笔式手势指点、圈点等标记形式和勾画等操作方式比较适合于对空间或几何信息的操作,使得脱离鼠标指点和菜单选择的交互方式,代之以笔式手势操作,可以自然、高效地完成图形操纵,通过用户徒手勾画方式来完成一系列的操作命令符合人们的自然习惯.而语音信息可以弥补笔式操作的不足,比如对超出屏幕显示区域对象的操作,采用语音方式可以更方便、快速地完成.两种交互方式可以相辅相成,改善单通道交互识别效果差的问题.我们在强调两种交互方式的同时,应当将两者有机地整合起来,笔输入和语音输入有其各自的特点,既可以同时使用,也可以在适合的时候单独使用.

就交互模型上来看,系统和用户作为两个独立而又相互联系、相互制约的实体,在信息传递和表达上存在两种方式:

(1) 系统本身具有一定的干扰性,因为系统信息是复杂的,是多种多样的,而这种复杂性使得用户应去利用更自然、简便的交互方式来辅助操作,并且去帮助系统完成复杂性的处理.

(2) 用户具有的干扰性,这时候就需要系统提供一些手段来帮助用户,需要系统能够准确地明白用户意图,在合适的时间和合适的任务下给出合适的建议,而在其余的情况下还是秉承自然输入的形式,两个方面是相辅相成的.

基于笔和语音的交互技术是自然用户界面的一个主要研究领域,用户免除了被频繁的菜单等交互操作打断思路的烦恼,设计思路保持了过程连续性以及流畅性^[6,7].

2 基于笔和语音的多通道交互模型

2.1 信息表征

当在传统 WIMP 界面下利用菜单、按钮等交互组件的方式,执行对内容的操作时,用户所关注的焦点由内容转向交互组件,当操作完成后,焦点再转回内容.关注焦点的转移使得用户在执行一个交互任务时,过多地关注于交互的执行过程,并没有完全关注于任务本身,这种关注焦点的来回切换会增加用户的认知负担,给加工过程带来难度.而基于笔和语音的多通道交互方式是一种直接面向内容的交互方式,用户所关注的是当前执行的操作和内容本身,用户的笔和语音输入直接作用在内容上,笔输入主要提供所需的空间信息,语音输入主要给出操作命令信息,操作过程与内容并不分离,系统会自动地将用户的交互动作转变为任务执行的命令,这种方式不需要用户过多地关注任务的执行方式^[8,9].可基于对象、任务、操作过程 3 个方面研究多通道系统中的交互行为,采用多个通道同时表现来完成系统功能,对输入条件给出合适的表示方式和描述方法.我们通过对空间信息

的分析和抽取来研究多通道信息的表征结构,其中包括笔输入的草图信息表征和语音信息的表征.草图信息表征描述用户通过笔式手势输入的对象信息以及对象间的拓扑结构、相对方位和相对距离等属性信息^[10,13];语音表征结构描述用户通过语音输入的空间对象的信息.通过对空间对象、空间对象的属性以及对象间的关系或关系属性的描述来给出表征的定义,如图 1 所示.

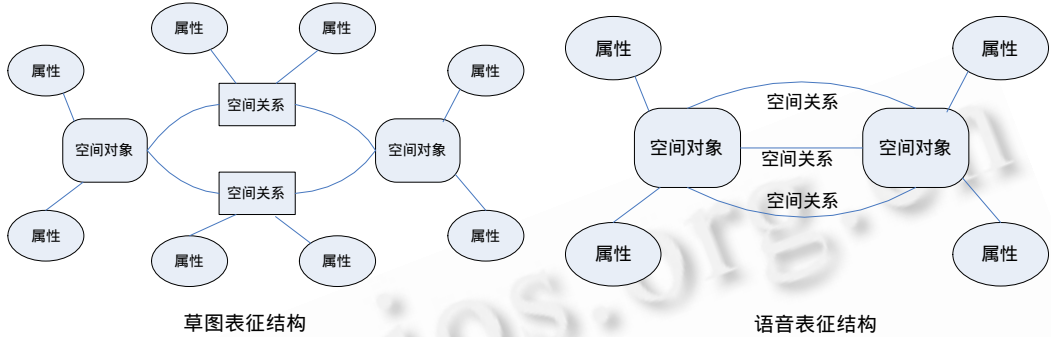


图 1 信息表征

2.2 基于层次概念模型的多层求解机制

来自多个通道的信息需要整合起来共同完成指定的任务,具体来说,多通道整合就是要通过确定用户所选择的交互任务,根据相应任务的结构信息、时间相关性和来自多个通道信息的表征结构,利用多通道输入的语义互补性来达到对用户交互输入的理解.这一过程需要确定用户所要完成的任务并确定任务的每一个参数,采用多通道交互方式这两者可以同时进行,即操作和内容不分离.我们采用基于层次概念模型描述笔画信息、语音信息、文字信息、上下文信息以及历史信息等交互信息之间的关系,给出了一个关于以笔输入和语音输入为主的多层求解机制,用来平滑多通道技术和应用之间的鸿沟.如图 2 所示,其中包含了 4 个层次,分别是设备层、输入层、整合层以及应用层.最底层是设备层,包含笔式输入、语音输入、文字和按钮输入等.第 2 层是输入层,得到的是用户输入的原始笔划信息、语音、文字或者按钮控制信息;第 3 层是整合层,根据当前的操作对象、任务特征、输入通道信息、上下文以及历史信息进行信息整合;第 4 层是应用层,根据不同的应用映射为不同的功能.

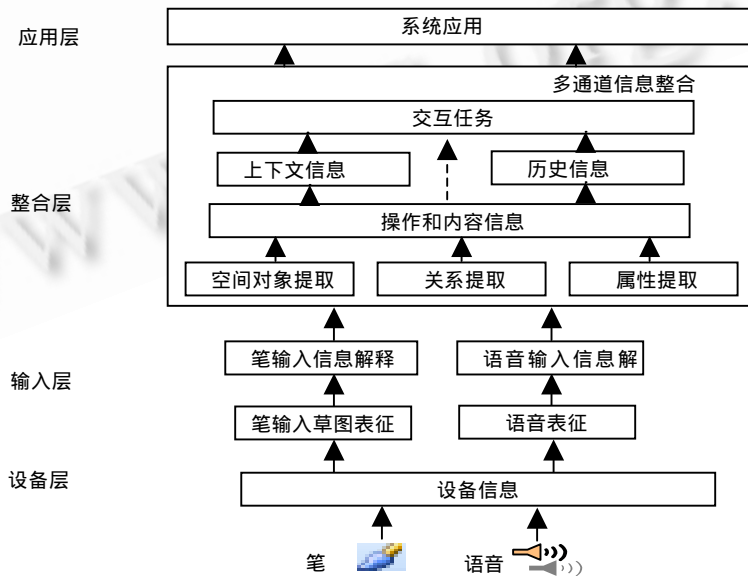


图 2 层次结构

尽管基于笔和语音的交互方式很灵活、很方便,符合人们的传统认知习惯,具有很大的吸引力,并在一些应用中产生了很好的效果,使用户体会到了一种自然的操作方式,但它也给设计者和用户带来了一些问题和限制.手势由自由勾画的一系列有意义的线条组合,大部分的任务利用手势可以很自然地实现,提供一种可视化的交互方式.但是对某些命令还是难以用手势的形式来表达,使得手势和目标语义表达之间存在差距,并且语义表达是否真正符合用户的意图也不容易判断,设计不好的手势往往会误导用户,影响了功能和可用性^[11,12].语音是一种更加自然的交互方式,但其识别率的高低直接影响了使用的性能,特别是对自然语言的识别上还存在很大的难度,对使用环境的要求较高,目前大多数采用受限语言集合的方式,在一定程度上提高了识别效率.

3 实例

3.1 功能概要与技术特点

本文给出了基于笔和语音的多通道几何软件系统,它提供了一个灵活、自然、动态的几何环境.通过基于笔的自由勾画方式来进行板书和绘图,同时辅助以语音输入方式,提高了交互的自然性和高效性.完全摒弃多余的按钮和菜单,使得用户不必在繁多的菜单和按钮项之间切换和花费太多的时间,只提供具有明确交互的按钮来代替手势和语音不太方便表达的命令,吸取已有人机交互技术的成果,与传统的用户界面,特别是广泛流行的 WIMP/GUI 兼容.系统中的主要交互任务分为 3 类:草图任务、语音任务、上下文感知任务;草图任务是指通过手势完成的各类操作;语音任务是指通过语音命令完成的各类操作;此处的上下文感知任务是指融合草图和语音等上下文的各类操作.如图 3 给出的交互界面的例子.

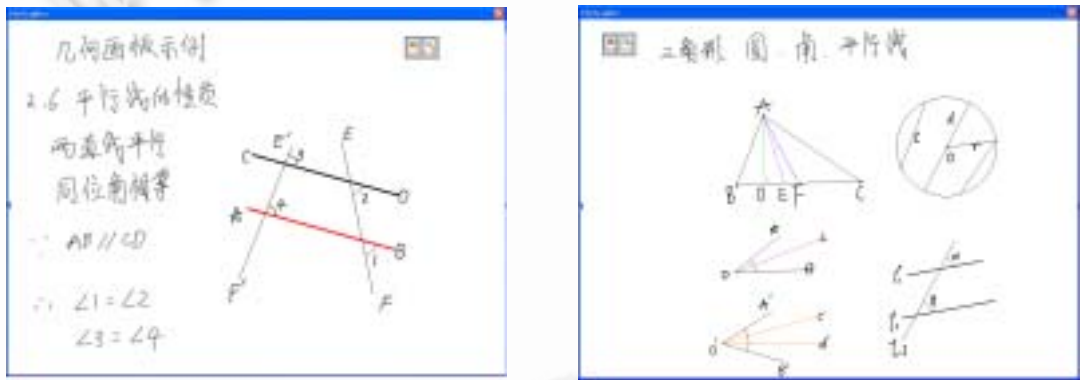


图 3 几何系统界面

多通道几何教学系统的开发支持了自动捕捉用户意图的功能,提供了动态显示方式,脱离了大量的鼠标指点和菜单选择的交互方式,提供手势编辑方式,可以更方便地辅助教师操作,同时让学生更容易理解它们所表示的涵义.

3.2 交互技术

基于多通道交互技术的支持,可以主要从以下几个方面来介绍:

(1) 用户意图捕捉,即对约束关系的主动捕捉,在对几何图形的编辑过程中,系统自动对当前操作的几何图形以及存在的其他图形间的约束关系进行判断,并给用户实时反馈.

(2) 信息融合,融合来自多个通道的信息去完成一个完整的交互任务,与空间图形操作有关的主要通过笔交互完成,而图形属性信息的获取和操作则主要通过语音命令来完成.如,绘制一条红色标识线,绘制过程中或者绘制完成的一定时限内,通过语音输入“红色”命令,可以完成此操作,操作效率要高于通过菜单和按钮方式的多步搜索来完成的属性修改和设定操作.

(3) 上下文感知交互,根据当前操作的对象,操作手势和操作位置为下一步的具体操作类型作出正确的判

断.如绘制完一个三角形,对其内部属性线的绘制根据落笔的起点以及相应的捕捉判断机制来获取,由顶点起笔向对边绘制线,根据与边或者顶角的关系来判断绘制的线是角平分线(给定一个阈值判断是否平分角)还是高线(给定阈值判断绘制线是否和边垂直),由对边起笔向顶点绘制线,根据落笔点在边上的位置判断绘制的是否为中线(给定一个阈值判断是否是在边中点附近).对圆来讲,可绘制弦、直径、半径,直接按照自然属性进行绘制,系统自动根据位置信息捕捉到用户要绘制的意图.比如,在圆内部绘制过圆心的线为直径,一端点为圆心的线为半径,两端点与圆相交且不过圆心的为弦.对角来讲,对于任意相交的直线,直接绘制弧形角度符号即可得到两条线组成的角约束,可以给出角平分线、三分线.同样支持对多线的约束捕捉操作.

4 结束语

随着计算技术的飞速发展以及人们交互观念的改变,用户自由的交互方式越来越成为人们所关注的问题之一,而自然、高效的多通道用户界面系统是其中的一个重要的发展方向.设计早期的概念想法有很高的价值,一种好的表达方式既能够给早期设计提供框架结构,又能够更好地去辅助理解要解决的问题.基于笔和语音的多通道交互方式,以草图作为信息载体,采用笔和语音交互,使得设计者间的交流也容易实现,与人们几千年来纸笔交互方式相适应,使得交互与回归自然的人类本性相结合,避免了妨碍设计者灵活构思的嵌套菜单、成列按钮和系列对话框的操作.多通道交互更适合人们的设计理念,将自然、和谐的交互方式与计算机的强大计算功能相结合,实现人机交互的理想模式,即用户自由的模式,真正让软件更适用于用户^[14].

References:

- [1] Robertson P, Laddaga R, van Kleek M. Virtual mouse vision based interface. In: Proc. of the IUI 2004. Madeira, 2004.
- [2] Abowd GD, Mynatt ED. Charting past, present, and future research in ubiquitous computing. ACM Trans. on Computer-Human Interaction, 2000,7(1):29-58.
- [3] Oviatt SL. Multimodal interfaces. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. 2003. 286-304.
- [4] Oviatt SL. Multimodal interactive maps: Designing for human performance. Human-Computer Interaction, 1997,12(1-2):93-129.
- [5] Chu C-CP, Dani TH, Gadh R. Multimodal interface for a virtual reality based computer aided design system. In: Proc. of the IEEE Int'l Conf. on Robotics and Automation, Vol 2. Albuquerque, 1997. 1329-1334.
- [6] Oviatt SL, Cohen PR, Wu L, Vergo J, Duncan L, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J, Ferro D. Designing the user interface for multimodal speech and gesture applications: State-of-the-Art systems and research directions. Human Computer Interaction, 2000,15(4):263-322.
- [7] Cohen PR, Johnston M, McGee D, Oviatt S, Pittman J, Smith I, Chen L, Clow J. Quickset: Multimodal interaction for distributed applications. In: Proc. of the 5th ACM Int'l Multimedia Conf. New York: ACM Press, 1997. 31-40.
- [8] MedSpeak: Report creation with continuous speech recognition. In: Proc. of the Conf. on Human Factors in Computing (CHI'97). ACM Press, 1997. 431-438.
- [9] Duncan L, Brown W, Esposito C, Holmback H, Xue P. Enhancing maintenance environments with speech understanding. In: Boeing M&CT TechNet. 1999.
- [10] Ma CX, Zhang FJ, Chen YD, Dai GZ. Feature-Gestures modeling of sketch in conceptual design. Journal of Computer-Aided Design & Computer Graphics, 2004,16(4):559-565 (in Chinese with English abstract).
- [11] Holzman TG. Computer-Human interface solutions for emergency medical care. Interactions, 1999,6(3):13-24.
- [12] Leclercq P. Invisible sketch interface in architectural engineering. LNCS 3088, 2004. 353-363.
- [13] Ma CX, Dai GZ, Teng DX, Chen YD. Research of interaction computing based on pen gesture in conceptual design. Journal of Software, 2005,16(2):303-308 (in English with Chinese abstract). <http://www.jos.org.cn/1000-9825/16/303.htm>
- [14] Oviatt SL, DeAngeli A, Kuhn K. Integration and synchronization of input modes during multimodal human-computer interaction. In: Proc. of the CHI'97. ACM Press, 1997. 415-422.

附中文参考文献:

- [10] 马翠霞,张凤军,陈由迪,戴国忠.支持概念设计的特征手势建模.计算机辅助设计与图形学学报,2004,16(4):559-565.
- [13] 马翠霞,戴国忠,滕东兴,陈由迪.概念设计中基于笔式手势的交互计算研究.软件学报,2005,16(2):303-308.
<http://www.jos.org.cn/1000-9825/16/303.htm>



马翠霞(1975 -),女,山东高唐人,博士,主要研究领域为多通道人机交互,概念设计,笔式计算,草图,普适计算.



王宏安(1963 -),男,博士,研究员,博士生导师,主要研究领域为实时数据库,实时调度.



戴国忠(1944 -),男,研究员,博士生导师,主要研究领域为人机交互技术,计算机图形学.

www.jos.org.cn

www.jos.org.cn