

草绘

Sketch-Based Creativity Modelling

创意建模技术

孙正兴/著



科学出版社

国家高技术研究发展(863)计划(2007AA01Z334)
国家自然科学基金(69903006、60373065、60721002)
教育部新世纪优秀人才支持计划(NCET-04-0460)
江苏省科技支撑(工业)计划(BE2010072)
江苏省自然科学基金(BK2009230, BK2010375)

草绘创意建模技术

Sketch-Based Creativity Modelling

孙正兴 著

科学出版社
北京

作者简介



孙正兴,男,汉族,博士,南京大学计算机科学与技术系教授,博士生导师,学术委员会委员,教育部“新世纪优秀人才”。兼任《计算机辅助设计与图形学学报》编委;中国图像图形学会计算机动画与数字艺术专委会委员;中国计算机学会计算机辅助设计与图形学专委会委员;中国人工智能学会人工心理与人工情感专委会委员;江苏省计算机学会计算机辅助设计与图形学专委会主任;江苏省微型电脑应用协会理事会副理事长、多媒体技术专委会主任。

曾在某国有大型企业从事计算机应用技术研发和推广工作近十年,作为技术骨干参与完成国家八五科技攻关项目一项、省部级科技攻关项目三项,先后获机械电子工业部科技进步三等奖一次(1989)、江苏省科技进步三等奖两次(1991和1995),并被授予徐州市“优秀专业人才”(1991)和徐州市“优秀青年专业技术人员”(1992)称号。进入南京大学工作以来,先后主持完成中国博士后科学基金项目一项、国家自然科学基金项目两项、国家高技术研究发展(863)计划项目一项、江苏省科技自然科学基金和江苏省科技支撑(工业)计划项目各一项;先后入选教育部高等学校骨干教师资助计划(1999)、教育部新世纪优秀人才资助计划(2004);教育部创新研究群体(2005)和国家自然科学基金委创新研究群体(2007)骨干成员;荣获南京大学航天科技奖二等奖(2002)和南京大学第七届浦发银行 SCI 论文奖(2008)。主要研究方向包括:多媒体计算、计算机视觉和智能人机交互等,重点研究面向普适计算的智能视觉环境及其交互技术,已在该领域国内外重要学术刊物上发表学术论文近百篇,编著《计算机图形学教程》、《电子政务原理与技术》等出版物四部,申请发明专利和软件著作权多项。

前 言

人类如今难以离开计算机,与计算机和平共处比以往任何时候都更显重要。尤其是以普适计算(ubiquitous computing)和环境智能(ambient intelligence)等为标志的新一代计算技术,都以构建“物理环境智能化”和“计算服务人性化”的智能环境(intelligence space)为目标,实现计算的分布、普适和透明;还需要利用环境感知、信息融合及人机交互等技术来支持环境对人类存在及其显式和隐式交互的感知和响应,实现人与环境间“现实世界就是用户界面”的普适交互(pervasive interaction),这更凸显了人机交互技术在计算机及其应用领域中的重要性。人与计算机间信息交流的自然化和流畅性需求推动着人机交互从以机器为中心的受限交互模式向以人为中心的非受限交互模式方向发展,国内外研究者已先后提出了笔迹人机交互、视觉人机交互、触觉人机交互、情感人机交互、空间人机交互和人脑计算机人机交互等多种新颖人机交互技术,以期建立能被人类自然能力或日常技能感知和理解的感知用户界面,并通过协调感知和表现间的关系来减少人类的交互认知努力,进而利用人类多个感觉和效应通道的充分而又有选择性的并行和协作来实现非精确而高带宽的多模态/多通道自然人机交互。以手绘草图和手写体为核心的笔迹人机交互是其中不可缺少的重要组成部分。

就计算机应用技术而言,设计是从抽象到具体、由模糊向精确的渐进演化和反复迭代为各种形式设计对象进行功能需求优化和物理组成选择的一种复杂思维活动。计算机辅助创意设计的困难在于解决方案或创意所具有的抽象性、模糊性与设计对象三维形体生成所要求的精确性、唯一性、实体性间的不相容性,它需要合适的媒体来表达创意构型的不确定性和模糊性,并支持创意概念的实体化和形象化过程,利于拓展思维空间、启动灵感风暴、理解创意选择和支持群体交流。

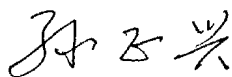
草绘创意建模(sketch-based creativity modelling)技术正是顺应了上述两个方面的技术需要和发展趋势而提出的,它将草绘交互的自然性、模糊性和高效性与视觉媒体的保真直观性、语义可演化性和操作灵活性有机地结合起来,使得用户可以在交互指令、图形绘制和三维建模等多个层次上运用日常技能以形象思维模式来快速而流畅地表达、记录和交流设计方案和创意,以期实现对计算机辅助创意设计、计算机辅助创作/设计灵感捕捉的支持,同时,可推动笔迹人机交互技术和计算机辅助设计技术的进一步深入研究和发

展。自1999年以来,作者领导的项目组从事草绘交互技术及其应用的研究,是国内较早开展这项技术研究的单位之一,先后完成了国家自然科学基金项目“基于数

据仓库的方案设计智能化方法及其工具研究(69903006)”和“基于手绘草图的智能化构思环境及其关键技术研究(60373065)”、教育部新世纪优秀人才支持计划项目“面向普适计算的智能化构思环境及其交互技术研究(NCET-04-0460)”、国家高技术研究发展(863)计划项目“基于手绘草图的三维动画模型创作技术研究(2007AA01Z334)”,参与完成了国家自然科学基金创新研究群体项目“面向Internet的软件方法与技术研究(60721002)”,取得了一系列与国际同类研究水平相当的研究成果,创立并形成了草绘创意建模的技术理念和方法体系,研发了多项具有自主知识产权的技术专利和软件工具。目前正以承担的江苏省科技支撑(工业)计划项目“三维动漫创意设计网络化协作软件系统(BE2010072)”及江苏省自然科学基金项目“交互式纹理合成技术研究(BK2009230)”和“面向普适交互的分布式视觉计算技术研究(BK2010375)”为平台实施该项技术在动漫游戏企业的应用推广工作。

本书是对项目组前期研究工作和成果积累的系统总结,是项目组全体人员集体智慧的结晶,每个从本项目组毕业和目前在项目组工作的研究生都为此付出了心血和汗水:王惠锋、金翔宇、朱华宇、徐晓刚和彭彬彬等先期开创了项目组草绘交互技术研究基础;邱庆华、周若鸿、孙建勇、张斌、李曼舞、孙怀志、尹建峰、王强、李彬、王卫、谢强、刘婧、远博和徐文晖等拓宽了项目组草绘交互技术的研究和应用范围;张莉莎、梁爽、冯贵焕和蒋维的创新性成果提升了项目组草绘交互技术的研究水平;张尧焯、陈松乐、刘凯、李毅、王斌、宋沫飞、章菲倩等为项目组创立了草绘三维建模的技术研究平台,并为草绘创意建模技术体系的建立作出了贡献;张岩副教授、陈圣国副教授、项建华副教授等在项目组技术研究和推广应用中提出了许多建设性意见,并给予了大力协助。同时,项目组所在的南京大学计算机科学技术系和计算机软件新技术国家重点实验室和谐而开放的研究环境和高品质技术研发平台为我们提供了强力保障。

草绘创意建模的技术理念和方法体系尚有许多不足,有待于我们后续的完善和深化并进行应用验证。本书旨在抛砖引玉,期待同行的指点和纠正,以共同推动该领域的发展和应用。书中如有不妥之处,望读者不吝指教。



2011年1月于南京大学蒙民伟楼

目 录

前言

第 1 章 绪论	1
1.1 人机交互技术机理	1
1.2 新颖人机交互技术	6
1.3 草绘交互技术	10
1.3.1 草绘交互的原理特点	11
1.3.2 草绘交互的处理过程	16
1.3.3 草绘交互的应用特性	20
1.4 草绘交互的研究课题	24
1.4.1 图形识别	25
1.4.2 语义解释	28
1.4.3 领域知识	30
1.4.4 用户习惯	31
1.4.5 交互界面	33
1.5 本章小结	36
参考文献	37
第 2 章 草绘创意建模技术原理	41
2.1 草绘创意建模原理	42
2.1.1 问题的提出	42
2.1.2 基本技术机理	47
2.1.3 相关研究状况	50
2.2 草绘创意建模的研究课题	53
2.2.1 深度恢复方法	54
2.2.2 模型表示方法	62
2.3 草绘创意建模的应用	64
2.3.1 概念设计与数字内容创建	64
2.3.2 三维模型检索和编辑	66
2.3.3 三维动画创作与生成	67
2.4 本章小结	69
参考文献	69

第3章 草绘图形的表示和处理	76
3.1 草绘图形的表示框架	76
3.1.1 草图的模糊几何模型	76
3.1.2 草图的层次化表示	81
3.1.3 草图的操作性描述	85
3.1.4 草图的结构文档定义	92
3.2 草绘笔画的表示和处理	95
3.2.1 笔画的预处理	96
3.2.2 笔画的特征表示	99
3.2.3 笔画的分割处理	105
3.2.4 笔画的拟合处理	125
3.3 草绘对象的表示与处理	127
3.3.1 草绘对象的特征表示	127
3.3.2 草绘对象的提取	136
3.3.3 草绘图形的检索	158
3.4 本章小结	170
参考文献	171
第4章 草绘图形的识别与解析	175
4.1 图符识别的图匹配方法	176
4.1.1 草绘图符的模式描述	176
4.1.2 图匹配的约束部分枚举算法	184
4.1.3 基于图匹配的草绘图符识别	195
4.2 草图识别的动态规划方法	197
4.2.1 动态规划方法概述	197
4.2.2 图符识别的动态规划方法	199
4.2.3 草图解析的动态规划方法	208
4.3 图符识别的机器学习方法	218
4.3.1 SVM	219
4.3.2 HMM	222
4.3.3 BN	229
4.3.4 实验比较与分析	235
4.4 图符识别的人工免疫方法	239
4.4.1 手写签名认证	241
4.4.2 样本克隆选择	250
4.4.3 实验结果及分析	258

4.5 草绘交互的相关反馈	262
4.5.1 查询点位移方法	263
4.5.2 线性规划方法	265
4.5.3 主动有偏式 SVM 学习	269
4.5.4 实验结果与分析	277
4.6 本章小结	280
参考文献	282
第5章 草绘图形的三维建模	287
5.1 草绘三维建模的场计算方法	288
5.1.1 草绘三维发型建模	289
5.1.2 草绘三维服饰建模	304
5.2 草绘三维建模的模板形变方法	314
5.2.1 草绘三维人体建模	315
5.2.2 草绘三维人脸建模	328
5.2.3 草绘三维模型编辑	341
5.3 草绘三维建模的规则驱动方法	349
5.3.1 草绘三维树木建模	351
5.3.2 草绘多面体建模	361
5.4 草绘三维建模的曲面重建方法	368
5.4.1 草绘膨胀体三维建模	370
5.4.2 草绘隐式曲面建模	375
5.4.3 草绘自由曲面建模	385
5.5 本章小结	395
参考文献	396
第6章 草绘创意建模技术的应用	402
6.1 草绘建模软件模型	402
6.2 图表绘制应用	410
6.2.1 PDA 图形绘制	410
6.2.2 电路图绘制	412
6.2.3 UML 图绘制	415
6.2.4 流程图绘制	418
6.3 图形检索应用	421
6.3.1 草绘图形检索	424
6.3.2 结构图形检索	426
6.3.3 三维模型检索	431

6.4 三维建模应用	434
6.4.1 智能画布系统设计	434
6.4.2 智能画布的建模功能	437
6.5 运动建模应用	442
6.5.1 草绘运动生成	443
6.5.2 草绘运动编辑	448
6.6 本章小结	450
参考文献	451

第 1 章 绪 论

自人类使用机器设备以来,人类的大量活动是借助人机接口(human-machine interface)与机器共同完成的,人机交互技术的研究目标就是协调和消融人机界限,实现自然而通畅的人机协作和交流。人类如今难以离开计算机,人与计算机这两个智能系统间信息交流的便捷和通畅就成为人机交互(human-computer interaction, HCI)技术的核心问题,尤其是 20 世纪 90 年代以来,计算机正逐渐成为“人类智力的放大器”,人类借助计算机不仅能从形式化的定量计算为主的结果中得到启发来完成计算任务或工作,而且可从定性和定量综合集成的多维化信息环境中得到(形式化和非形式化的)感性和理性认识,并实现对人类创造性工作的协助,人机交互也从单纯的计算机输入/输出技术逐渐形成将人类基本技能与计算机的感知和推理能力结合的多学科综合技术。

1.1 人机交互技术机理

计算机系统的人机交互是以用户界面(user interface, UI)为中介的人与计算机间信息传递和转换的过程,如图 1-1 所示。人向计算机传达交互意图,即用户通过其侧重于多媒体信息接受的感觉通道(包括视觉、听觉、触觉和嗅觉等)对用户界面中诸如语音、图形和图像等多媒体信息的感知,先利用逻辑思维和形象思维能力产生对交互意图的心理描述,再按用户界面对交互的输入要求使用其侧重于过程控制与信息输入的效应通道(包括手、嘴、眼、头、足及身体等)来实施交互行。计算机向人反馈计算结果,即它利用其感知机制从精确的用户界面框架结构中获取用户的交互信息,并完成获取交互行为信息的感知处理、判断用户交互意图的认知处理和回馈交互结果的响应处理等一系列计算过程,最终运用其输出机制以用户界面所定义的特定形式输出其计算结果。

从信息传递的角度来看,用户界面的任务是为人类的感觉和效应通道与计算机的感知和输出机制提供通信接口,计算机系统的人机交互包含互不可分的两类感知过程:一类是人对计算机提供信息的感知过程。在这个过程中,计算机利用其计算能力和各种媒体处理技术将计算机可供交互的输入要求或交互处理结果的反馈以适当形式在用户界面上呈现给用户;而人类使用感觉通道(如视觉和听觉)接受和理解用户界面所提供的各种信息。这种感知过程的关键是用户界面呈现信息的形式和丰富程度,并部分与用户的心理因素及其认知能力相关,这决定了人机交

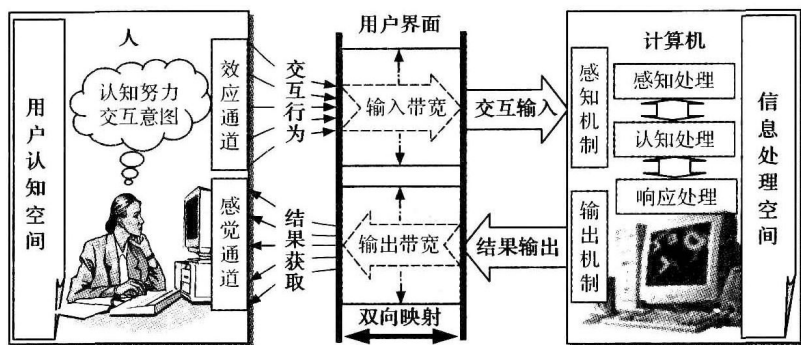


图 1-1 人机交互中人与计算机间的关系

互中计算机与人间通信的输出带宽。另一类是计算机对人及其所处环境的感知过程。在这个过程中,人类按照计算机所要求的输入事件使用效应通道(如语音和动作等)在用户界面中输入交互操作和指令;计算机则通过对来自用户界面的用户交互媒介信号的处理和分析来获取和识别用户交互意图,并以适当形式在用户界面上呈现给用户。这种感知过程不仅取决于计算机感知和识别人和物及其检测用户情绪变化或用户个性等信息处理能力,而且取决于用户有意识地将交互意图“分解”和“组织”为计算机所特有输入事件过程中所付出的“认知努力”,这决定了人机交互中计算机与人间通信的输入带宽。

此外,如果计算机是可以感知外部世界的机器,那么还存在两种特殊的感知:一种是计算机感知并传递人及其环境信息,人借助于计算机传递感知信息,它基本上不属于人机交互的感知,但可以作为人机交互感知的基础。例如,在视频会议、远程教学和远程医疗等系统中,计算机并不一定需要识别人的交互行为和信息,而只成为两人或多人之间相互交流的媒介;另一种是人作为虚拟世界中的成员而从计算机中获得的感知,这种感知是人机交互的感知,而且人和计算机的感知地位是等同的,其感知信息的形态不仅内容丰富,而且可以是现实世界里所看不到的(如虚拟的场景、文本和语音等)。

从信息转换的角度来看,用户界面的作用是实现“用户认知空间”与“计算机信息处理空间”之间信息的双向映射。由于人类的交互意图通常是一个涉及人的认知和行为因素的高维复杂信息,且这些信息是非精确或模糊的;而计算机遵循线性信息表示和存储模式来表示和处理交互信息,所能捕获和处理的必须是形式化的精确信息,且其感知和处理结果不一定能够依从人的感知规则。两者间的差异决定了其双向映射实现的技术复杂性和方式多样性,因此,人机交互技术的任务就是要选择合适的映射方式来解决其技术实现的复杂性与用户交互的易用性问题。技术实现的复杂性主要体现在如何提升计算机感知和识别用户交互行为和状态,并

进而理解其交互意图;用户交互的易用性则表现为如何充分利用和协调人类的多个感觉和效应通道,并进而降低其完成交互行为所需付出的“认知努力”。显然,一个好的映射方式将会极大地提高人机交互的自然性和高效性。

在人机交互技术发展历程中,人机交互的界面形式经历了手工作业、命令接口(作业控制语言及交互命令语言批处理)、(文本)菜单到多通道、多媒体人机交互和虚拟现实系统;人机交互的信息载体经历了以文本和符号为主的字符用户界面(character UI)、以二维图形为主的图形用户界面(graphical UI, GUI)、兼顾视听感知的多媒体用户界面(multimedia UI)及综合运用多种感觉(包括触觉等)的虚拟现实系统(virtual reality, VR);人机交互的信息维度从一维(主要指文本流)、二维(主要是利用色彩、形状、纹理等维度信息的二维图形技术)、三维(主要是三维图形技术,但图形显示仍以二维平面为主)向多维空间(利用人的多个感觉通道和效应通道的信息)发展,计算机与用户间的通信带宽不断提高。就其技术机理本质而言,这种发展蕴涵着对人机交互中两个主体不同侧重点的映射方式演化。

一种是以机器为中心的受限方式。这种方式强调将计算机的信息处理需求有效地呈现给用户,为用户提供一个形式化、半双工、串行的低维度信息展现和操作界面,主要表现在两个方面:一是人的交互需基于离散事件,用户界面等同于一个解释器,它接受事件输入并反馈计算机处理的结果,在此过程中,多维信息矢量被压缩为一个低维矢量,形成一个串行的输入流;二是人的交互必须基于精确交互,用户在交互中需要把意图按模型要求“分解”并“组织”为计算机所特有的输入事件,以完全说明用户交互目的和意度。用户必须面对机器操作定义的限制并付出较多的认知努力来使用计算机。从用户角度来看,这种方式支持的是受限人机交互(formal HCI),用户需饱受命令记忆和新功能学习之苦。这种方式的关键在于计算机处理能力展示的有效性和合理性及用户操作的简便性和易记忆程度,具体表现为两个方面的问题:一是如何充分而有效地展示计算机及其软件的处理能力,即如何定义操作及其布局,如弹出菜单、下拉菜单、分层菜单等;二是如何让用户方便地使用和记忆,即用户采用何种方式进行操作,如命令语言、文字菜单和图标点击,等等。显然这种方式限制了人机间的通信带宽,使得既不能有效地利用当今计算机的强大计算和处理能力,也给用户使用电脑带来诸多不便。

传统用户界面都采用了这种映射方式,它主要包括三大类:

(1) 命令语言人机交互(command HCI)。这种人机交互始于联机终端的出现,用户与计算机间借助一种双方都能理解的语言进行对话。根据语言特点可分为:①形式语言,这是一种人工语言,特点是简洁、严密、高效,它不仅是操纵计算机的语言,而且是处理语言的语言;②自然语言,特点是具有多义性、微妙、丰富;③类自然语言,这是计算机语言的一种特例,命令语言的典型形式是“动词+宾语”结构,二者都可带有限定词或量词。命令语言可以具有非常简单的形式,也可有非常

复杂的语法,它要求惊人的记忆和大量的训练,且容易出错,但具有灵活和高效的特点,适合于专业人员使用。

(2) 图形人机交互(graphic HCI)。图形用户界面是当前人机交互的主流,成熟的商品化系统有 Apple 的 Macintosh、IBM 的 PM(presentation manager)、Microsoft 的 Windows 和运行于 UNIX 环境的 X-Window、OpenLook 和 OSF/Motif 等。这种界面包含三个重要思想:① 桌面隐喻(desktop metaphor),即在用户界面中用人们所熟悉的桌面上的物品来清楚地表现计算机可处理的能力;② WIMP (windows, icons, menu and pointer)是组成图形用户界面的基本单元,以可重叠多窗口管理的窗口管理系统为核心,使用键盘和鼠标作为输入设备,并广泛采用事件驱动(event-driven)技术;③ 直接操作及所见即所得的界面,交互过程极大地依赖用户视觉和手动控制的参与,具有强烈的直接操作特点。菜单(menu)本身与图形人机交互并没有必然联系,在图形用户界面出现前,字符型菜单已经在用户界面中得到了广泛应用。但图形人机交互中菜单的表现形式比字符人机交互更为丰富,在菜单项中可以显示不同的字体、图标甚至产生三维效果,同时,还可以诸如层叠式(hierarchy)、弹出式(popup)、下拉式(pulldown)和滚动式(scrolling)及其组合等多种形式出现,用户只需确认而不需回忆系统命令,大大降低了记忆负荷。图形人机交互具有一定的文化和语言独立性,并可提高视觉目标搜索的效率;其主要缺点是需要占用较多的屏幕空间,并且难以表达和支持非空间性的抽象信息的交互,且灵活性和效率较差,可能不适合于专家用户。

(3) 直接操纵人机交互(direct-manipulation HCI)。直接操纵人机交互更多地借助物理的、空间的或形象的表示,而不是单纯的文字或数字表示。前者已为心理学证明有利于“问题解决”和“学习”。视觉的、形象的(艺术的、右脑的、整体的、直觉的)人机交互对于逻辑的、直接性的、面向文本的、左脑的、强迫性的、推理的人机交互是一个挑战。直接操纵人机交互的操纵模式采用“宾语+动词”结构,Windows 95 设计者称之为“以文档为中心”,用户最终关心的是其欲控制和操作的对象,即只关心任务语义,而不用过多为计算机语义和句法而分心。对于大量物理、几何空间及形象的任务,直接操纵人机交互表现出巨大的优越性,但它在抽象而复杂的应用中具有局限性。对于用户界面设计者来说,图形的设计比较困难,需大量测试和实验;而复杂和抽象语义的表示也比较困难,不容易实现交互与应用程序的独立设计。

另一种是以用户为中心的非受限方式。它强调将人类自然能力(尤其交流、运动和感知能力)与计算设备及其感知和推理结合起来,通过采用多种模式(multimodal)感知人类的自然行为,并以易理解的多媒体(multimedia)形式实现多通道(multichannel)通信,建立“以人为中心”的感知用户界面(perceptive UI, PUI)^[1]。这种界面旨在充分使用人类诸如语音、手势、视线及头部跟踪等多个感

觉和效应通道,并使之可选择地并行和协作来实现多通道、非精确、高带宽、高效及不限制地点的人机通信^[2],以拓展人机间信息通信的“带宽”,减少人类的交互认知努力,提高人机交互的自然性和高效性。从用户角度来看,这种方式支持非受限人机交互(informal HCI)^[3,4],其主要出发点是借助于媒体感知技术获取并理解用户动作、行为、习惯、偏好及其他相关交互信息,来缩小或消除计算机和用户交互意图的认识和描述存在的精确信息(人机系统的内部表示)和非精确信息(人机系统的外部表示)间的“鸿沟”;其实现的关键在于计算设备对人的动作和行为识别的有效性,因此,其用户界面又称为基于识别的用户界面(recognition-based UI)。

这种方式的界面则代表着人机交互技术的发展方向,体现了对人的因素的重视,标志着人机交互技术从“人适应计算机”向“计算机不断地适应人”的方向发展,就是要给用户“充分的自由”,使得人类不再需要按机器的要求扭曲自己最自然的思维和行为方式,而可以像在电脑出现前人类那样自由地工作和生活。传统的界面事实上成为隔离物质世界和信息世界之间的屏障,虚拟现实、网络计算、智能代理和软件机器人、计算机支持的协同工作等技术的应用,将实现“虚物实化”和“实物虚化”,消除物理对象和抽象对象、输入装置和输出装置在交互空间中的差别,并为人提供多感觉通道的自然临境体验;语音及文字识别和自然语言理解等言语计算、手写体和草绘识别等笔式计算及手势和表情识别、视觉-目标拾取认知技术等视觉计算等技术的不断发展和完善,将不断提高人机交互的智能化程度,使机器能够根据上下文及使用者的特点主动识别人的身体姿态、手势、语音和表情等各种自然行为,进而判断出人的意图。同时,机器的功能齐全、高效,既适于人的操作使用,也能满足人的审美和认知需要,机器的操控和使用将更加符合人的习惯,并具备听、看和说的能力,人可以用语言、文字、图像、手势、表情等自然方式与机器打交道,恰如与他人交流一样自然。因此,其用户界面是能模拟多种智能和真实环境的虚拟空间,人们能够在任何时间、任何地点与任何需要的环境(包括人)以最自然和“身临其境”的方式来完成所需要完成的工作;实现“无形而又无处不在,有形而又自然和谐”的普适人机交互(ubiquitous/pervasive HCI)模式,这种方式的特性体现在以下三个方面:

(1) 以用户为中心(human-centered)。以用户的需求变化为出发点,使人机交互的外在形式和内部机制能符合不同用户的需要。人类的交互行为是自然的,用户将可利用语音、手势、笔画等自然方式,不受地点限制地与计算机进行交互,既能满足用户个性化的需要,又使得用户不脱离自然社会关系(包括社会经济环境和人类沟通交流)。

(2) 多模态人机交互。充分利用人类多种感觉和效应通道的互补特性,并使之可选择地、充分地并行和协作来捕捉用户的交互意图,从而增进用户交互的自然性。模态(modal)和通道(channel)间的联系和区别主要在于,人类通过视觉、听

觉、触觉、味觉和嗅觉这五种感官的信息发送和接收来实现与世界交互的,一种模式对应一种官能;而通道是指信息传递的过程或途径。在人机交互应用中,通道描述一种交互技术,这种技术可基于某种特殊设备,如键盘或鼠标;也可基于某种特殊行为,如口语、书面语或手势。例如,命令语言用户界面中的每个命令都会使用视觉和触觉等多种模式;而图形用户交互既使用了视觉和触觉等多种模式,又用到了诸如键盘、鼠标和声音及图像等多种通道。多模式人机交互是使用同等方式将两种或两种以上输入模式相结合的技术,其重点是将基于诸如语音识别、笔形手势识别、计算机视觉等传感识别的输入技术集成到用户界面中。也就是说,多模式人机交互使用多种模式来实现多种通道通信。使用多种模式生成单独的通道,如利用视觉和声音进行三维用户定位,属多传感器融合(multisensor fusion);类似地,使用一种模式生成多种通道,如左手鼠标指示、右手鼠标选择,则属于多通道或多设备(multichannel or multidevice)人机交互。

(3) 多媒体感知(multimedia perceptive)。机器利用其感知及推理能力对来自用户感觉和效应通道的交互信号进行识别、集成和协调,并获取用户动作和行为习惯、偏好及其他相关信息,并以人类易理解的多媒体信息方式为用户提供输出信息,从而提供不受时空限制而又效能最大化的个性化计算服务。这种人机交互方式的双向信息流动是以多媒体感知和处理为核心的。用户通过其感觉和效应通道传递的交互意图在计算机内表示为文本、语音、图形和图像等多媒体信息,人到机(human to compute)的信息流动是多媒体信息的获取及识别过程;计算机经过处理的信息需要以文本、语音、图形和图像等用户理解概念所需的多媒体信息形式展现出来,机到人(computer to human)的信息流动是多媒体信息的合成和呈现过程。

1.2 新颖人机交互技术

自 Weiser 博士提出普适计算以来,人类向未来计算探索的进程,也催生了以网格计算(grid computing)和云计算(cloud computing)为代表的计算分布化(distribution)、以嵌入式计算(embedded computing)为代表的计算透明化(transparent)、以虚拟现实和可穿戴计算机(wearable computer)为代表的计算机系统拟人化(personate)和有形化(tangible)、以个人数字助理(personal digital assistant, PDA)和智能手机(smart phone)为代表的计算机系统微型化(micromation)和随身化(mobile),及以智能空间(smart space)和环境智能(ambient intelligence)为代表的计算机系统嵌入化(embedded)和无形化(invisible)等众多新型计算技术。这些技术推动着“机箱时代”迈向一个崭新时代:计算设备将从人类视野中“消失”,那些装在机箱中的声音、视频和通信等将“游荡”于周围环境中而变得

“无处不在”。这个时代具有两个重要特征：一方面，小型化嵌入式计算设备将使得人类生活的自然环境“具有”计算能力，它能自主感知人的存在，并在人类需要时出现和对其行为作出反应；另一方面，计算将日常生活化，能帮助人类安排、构造和控制日常生活，人类将能享受计算服务，人类能以最自然的方式、在任何时间、任何地点、通过任何设备从环境中“无意识地”获取“有感觉”的计算资源和计算服务。概括地讲，在这个时代，人类将片刻离不开计算机，人与计算机间的交流更加自然和没有阻碍，人和计算机间将是非侵入且不可见的交互，现实世界就成为用户界面，用户只需运用其日常技能来操作和使用计算设备，而不再关注交互媒体形式，也无需特定的用户界面(idea interface is no interface)。这既对人机交互技术提出了更高的要求 and 新的挑战，也为人机交互技术的发展注入了新的动力，因此，自然、高效、无障碍的新型人机交互技术成为该领域发展方向和研究热点。

新型人机交互技术的最主要特性就在于用户交互的“非受限性(informality)”：机器给人以最小的限制并对人的各种动作做出反应，人是主动参与者，可以最大的自由度操作机器，如日常生活中人与人间的交流一样自然、高效和无障碍。这种人机交互技术强调两个关键特征：一是交互隐含性。理想的人机交互应当使用户把所有注意力集中于完成任务而无需为交互操作分心，且应允许使用模糊表达手段来避免不必要的认识负荷，有利于提高交互活动的自然性和高效性。传统用户界面都是等待用户输入命令才开始采取行动，而新颖人机交互技术的用户界面则应是积极地感觉和理解世界，并让用户在各个层次上根据目标和已有知识采取行动，并以充分性代替精确性。理想地，这是一个使用“被动”或“非侵入性”感知的“主动”交互方式，无需用户显式说明交互成分，仅在交互过程中隐含地表现而允许非精确的交互，如用户视线自然地落在所感兴趣的对象上；用户的手自然地握住被操纵的目标，等等。二是交互多模态性和双向性。允许使用多种感知模态(如视觉、听觉、触觉等)或多个感觉和效应通道，突破了模态和鼠标键盘显示器通信通道的限制，使用连续形式和并行形式在内的多种模态。尽管感觉通道侧重于多媒体信息的接受，效应通道侧重于交互控制与信息输入，但两者是密不可分、相互配合的。只用一种通道不能充分表达意图，就需要其他通道信息辅助。有时，使用辅助通道仅为增强表达力，否则，就必须允许充分地并行和协作的通道配合关系。此外，人的感觉和效应通道通常具有双向性特点，如视觉可看又可注视、手可控制又可触及等，新颖人机交互技术让用户避免生硬、频繁或耗时的通道切换，从而提高自然性和效率。如视线跟踪系统可促成视觉交互双向性、听觉通道在利用三维听觉定位器实现交互双向性等。

新型人机交互技术的核心问题在于如何确保计算设备对人的动作和行为识别的有效性，并实现人机的和谐共处。计算设备及其传感器是透明和被动的，但可以主动感知人类相关通信通道，并输出人类可理解的结果。这就需要不同类型和层

面技术的集成,并随着关键维度或特征发生变化,包括输入模态的数目和类型,通信通道的数目和类型,使用并行模式、串行模式或同时使用并行和串行模式的能力,识别模式的大小和类型,传感器和通道的集成方法及支持的应用种类,等等。依据充分利用人类感觉和效应通道并使得人机交互更加自然的目标,新型人机交互的关键技术主要集中在以下几个方面。

1. 支持语音人机交互(speech-based HCI)的言语计算(speech computing)

语音是人类一种重要而灵活的通信模态,言语交互的核心是语音识别,其任务就是利用语音学和语言学知识,先对语音信号进行基于信号特征的模式分类(这是语音信号处理的范畴)得到拼音串,再利用语言学知识对拼音串进一步处理,得到一个符合语法和语义的句子。简单地说,语音识别就是让计算机能听懂人说话,将人说的话转换成计算机文本。

2. 支持笔迹人机交互(pen-based/calligraphic HCI)的笔迹计算(calligraphic computing)

笔迹人机交互是通过计算机软硬件技术和相关领域的研究,模拟人类“笔录纸现”这一日常技能的一种人机交互方式。在这种用户界面中,用户借助鼠标、笔迹交互器及触摸屏等设备用手自由地书写或绘制各种文字和图形,计算机通过对这些输入对象的识别和理解获得执行某种任务所需要的信息。它充分利用书写的自然性和墨水丰富的表达能力,使人们通过笔迹交互自然地使用计算机的高性能计算能力:从计算机处理角度看,笔迹交互可以在一个或多个连续的笔画中自然地提交计算任务所需要的命令和参数;从人机交互角度来看,它通过将用户的交互意图映射为由压力、方向、位置和旋转等信息共同构成的多维矢量序列^[5,6];从认知心理学角度看,用户的交互意图通过几个效应和感知通道提交。与言语交互相比,笔迹交互以视觉形象表达和传递概念,既有抽象、隐喻等特点,还具有形象、直观等特征,易于理解和记忆,更适于推论和构思;与传统图形图像工具相比,它具有自然、简便,可表达不完备模糊概念,拥有强大的信息表达能力,有利于创造性思想的快速表达、抽象思维的外化和自然交流。

3. 支持视觉人机交互(vision-based HCI)的视觉计算(vision computing)

在人类日常面对面交互中,除使用语音和文字外,还可利用身体各部位的姿态和动作(即所谓身体语言)来表达自己的意思。视觉和言语是联合表达、相互补充的通道,诸如脸部表情、注视焦点、肢体动作及其他活动等行为线索能为交流提供相关上下文信息,这些信息不仅能加强语音的表达能力,还能起到语音交互所不能起到的作用,并影响交谈的内容和进程,如头部动作可能是“否定”或“肯定”的信号,脸部表情可能是“无奈”或“赞许”的表示,注视角度可能可消除“这边”或“那边”