

面向指-笔混合输入的交互原语和交互模型研究^{*}

仝青山^{1,2}, 康文惠^{1,2}, 付强^{1,2}, 黄进¹, 田丰¹, 戴国忠¹

¹(人机交互北京市重点实验室(中国科学院 软件研究所), 北京 100190)

²(中国科学院大学 计算机科学与技术学院, 北京 100049)

通信作者: 田丰, E-mail: tianfeng@iscas.ac.cn



摘 要: 随着触摸设备的普及, 指、笔输入已成为移动办公的一种主流输入方式. 然而现有的应用主要以手指或笔通道中的一种作为输入, 限制了用户的交互空间. 同时, 现有的指-笔混合研究主要关注串行的指-笔分工协作和特定交互任务的并行处理, 并没有系统地考虑并行协作机制和不同通道输入之间的意图关联. 为此, 提出一种面向指-笔混合输入的交互模型; 结合用户指-笔协作的行为习惯定义指-笔混合交互原语, 拓展指、笔交互空间; 继而利用部分可观察马尔可夫决策过程提出基于时序信息的指-笔混合输入意图提取方法, 用以增量式地提取多义性交互原语的交互意图. 最后通过一个用户实验说明指-笔混合输入的优势.

关键词: 笔式交互; 指-笔混合输入; 交互空间; 交互模型

中图法分类号: TP393

中文引用格式: 仝青山, 康文惠, 付强, 黄进, 田丰, 戴国忠. 面向指-笔混合输入的交互原语和交互模型研究. 软件学报, 2024, 35(4): 2022–2038. <http://www.jos.org.cn/1000-9825/6909.htm>

英文引用格式: Tong QS, Kang WH, Fu Q, Huang J, Tian F, Dai GZ. Research on Interaction Primitives and Interaction Model for Pen + Touch Inputs. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 2022–2038 (in Chinese). <http://www.jos.org.cn/1000-9825/6909.htm>

Research on Interaction Primitives and Interaction Model for Pen + Touch Inputs

TONG Qing-Shan^{1,2}, KANG Wen-Hui^{1,2}, FU Qiang^{1,2}, HUANG Jin¹, TIAN Feng¹, DAI Guo-Zhong¹

¹(Beijing Key Laboratory of Human-computer Interaction (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

²(School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the popularity of touch devices, pen + touch inputs have become mainstream input modes for mobile officing. However, existing applications mainly take one of them as input, which limits users' interaction space. In addition, existing pen + touch research mainly focuses on serial pen + touch cooperation and parallel processing of specific interactive tasks and does not systematically consider parallel cooperation mechanism and intention correlation between different inputs. This study first proposes an interaction model based on pen + touch inputs and then defines pen + touch interaction primitives according to users' behavioral habits in pen + touch cooperation, so as to extend pen + touch interaction space. Furthermore, by using a partially observable Markov decision process (POMDP), the study develops a method of extracting pen + touch input intentions based on time sequence information, so as to incrementally extract the interaction intention of polysemantic interaction primitives. Finally, the study evaluates the advantages of pen + touch inputs through a user experiment.

Key words: pen-based interaction; pen + touch input; interaction space; interaction model

手指和电子笔作为移动终端的主要交互媒介, 是自然人机交互的重要交互通道, 也是移动办公等新型应用的主流交互方式. 随着交互环境和交互任务的多样化, 以及新型传感器和输入设备的发展, 针对移动终端的输入方式

^{*} 基金项目: 国家自然科学基金 (62172397); 中国科学院前沿基础研究重点项目 (QYZDY-SSW-JSC041)

收稿时间: 2022-06-17; 修改时间: 2022-08-23, 2023-01-29; 采用时间: 2023-02-03; jos 在线出版时间: 2023-07-12

CNKI 网络首发时间: 2023-07-13

正在从单通道输入向多通道、多设备的混合输入演变^[1-3],传统的串行指-笔输入正逐渐向指-笔融合输入过渡^[4].近几年,指-笔混合输入问题已成为学术界关注的热点问题^[5-9].

目前,指-笔混合输入的研究工作主要集中于两个方面:一是探索指-笔交互协作的理论框架,二是面向具体场景的指-笔混合输入方式. Guiard^[6]提出双手动作不对称性的初步理论框架,该框架描述了人类在各种熟练的体力活动中两只手的使用偏好,指出了在活动中两只手扮演角色的差异以及彼此之间的配合; Microsoft Research 的 Hinckley 等人^[7]基于物理纸笔下用户的输入行为,提出了 Pen 书写、Touch 操作、Pen + Touch 混合的操作方式; Pfeuffer 等人^[8]与 Hinckley 等人^[9]进一步对平板上的 Thumb + Pen 进行了研究,提出了惯用手直接操作、握持设备的非惯用手间接操作等操作方式. 这些指-笔协作方式的研究成果,为利用指-笔组合操作实际问题奠定了基础. Romat 等人^[10]将 Pen + Touch 输入应用到文档编辑和审阅中,实现了特定区域的空间调整; Srinivasan 等人^[11]利用 Pen + Touch + Speech 解决了不同类型数据在平板上的一致性可视化问题; Pfeuffer 等人^[12]提出了利用 Pen + Touch 实现三维操作技术 Bi-3D. 整体来看,当前的研究工作主要集中于如何利于指-笔解决特定问题,缺乏一般情境下的指笔交互问题研究,如指-笔并行交互机制、指-笔混合交互模型、多义性交互动作意图理解等.

时间连续性、特征多维性和意图不确定性是指-笔交互的显著特征^[1],也是构建指-笔混合交互模型的基础. 很多学者在相关领域进行了研究,如田丰等人^[13]提出了 PIBG 笔式交互范式并对笔手势进行了深入探索, Li 等人^[14,15]提出了混合自动机的概念并对触摸屏上的手势识别进行了深入研究,刘杰等人^[16]提出了三维连续交互空间下的混合手势交互模型. 虽然这些工作的研究重点不是指-笔混合输入,但对指-笔混合交互原语的定义、指-笔融合方式和指-笔状态描述都有很好的指导意义.

为此,针对传统指-笔交互过程中交互空间受限和多语义交互动作意图不确定问题,本文探索并构建了指-笔混合输入的总交互框架,提出了指-笔混合输入交互模型,将交互前期的指-笔协作与后期的交互意图理解进行了融合,并通过绩效对比实验佐证了该模型的可行性和先进性. 本文的贡献主要包括以下几点: (1) 提出了基于触摸屏的指-笔混合输入交互模型,给出了利用指-笔混合输入在触屏上进行单次交互、持续交互的方式; (2) 定义了指-笔混合输入交互原语,为规范化指-笔混合输入方式提供参考依据; (3) 提出了一种基于 POMDP 的增量式交互意图提取方法,可以提高针对多义性交互动作意图理解的准确性.

1 相关工作

笔交互与触摸交互同时具有使用自然、操作直观的特点,且二者各有优势. 笔交互可支持细粒度、高精度输入,能够避免触摸交互中存在的“胖手指”问题^[1,2],而基于手指的触摸交互不需要借助额外输入设备,并且支持多点 and 手势触控方式,能够补充和扩展单点笔式交互能力^[4]. 因此,在保持各自交互优势的前提下,融合指和笔两个通道进行输入,不仅扩展了用户交互空间,而且实现了从单通道输入向多通道的混合输入演变. 近几年,学术界和工业界开始重视指-笔混合输入相关问题. Microsoft Research 的 Hinckley 等人^[7]结合物理纸笔下的用户输入行为开展了多项 Pen + Touch 的研究,提出了 Pen 书写、Touch 操作、Pen + Touch 融合的操作方式,构建了 Microsoft Surface 指笔组合交互的原型;并在此基础上,设计了 Manual Deskterity 原型数字绘图桌^[9],从自然的人类技能和手的角色分工角度探索了笔和手指触控之间的任务分工,为指-笔混合交互输入方式的发展奠定了基础. Zhang 等人^[5]提出了一种可以用于感知用户在 Pen + Touch 交互过程中静止和微小移动状态相互转变的技术,不仅能够感知到握姿的改变,还可以为自适应用户界面提供技术支撑. Romat 等人^[10]提出了一种基于 Pen + Touch 输入的文档编辑和审阅技术 SpaceInk,用户可以在指定位置为注释腾出空间,动态地重排文档;他们的另一项相关工作是一个促进用户浏览信息和思想呈现的系统 ActiveInk^[17],该系统可以为用户提供以笔交互为核心的主动阅读服务,支持基于笔迹激活的用户活动分析方法. Srinivasan 等人^[11]构建了一个用于平板数据可视化的多通道交互系统 InChorus,它支持 Pen、Touch、Speech 这 3 种输入形式. Cami 等人^[18]使用不同握持姿势对单手的 Pen + Touch 交互进行研究,借助卷积神经网络方法能够从原始的电容传感数据实时识别出 10 个手势. 可以看到,越来越多的工作正在利用指-笔混合输入解决实际场景中的交互问题,但很少有学者关注指-笔混合交互过程中的并行机制研究.

不确定性问题是触控交互研究的热点和难点^[15,19],指-笔混合交互过程中的不确定性也是指-笔混合交互的核

心问题之一。正如 Shi 等人^[19]所述,“与传统的依赖于鼠标、键盘的精确交互不同,自然用户界面支持触控、语音、手势、书写和视觉等几种具有一定不确定性的交互方式。关于语音、手写和视觉的研究已经取得了巨大进展,而与交互任务息息相关的触控和手势的研究依旧相对薄弱。对具有不确定性的交互方式研究的难点在于如何从连续的随机数据中识别用户的真实输入意图”,而指笔输入意图的识别和理解严重依赖于输入笔迹信息的识别和意图理解方法。常见的用户意图理解方法可以分为基于规则和基于统计两大类^[20]。其中,基于规则的推理和决策通常是以有限状态机为模型的问答控制流,它以系统为主导(system-initiative)^[21];而基于统计的推理和决策将信息处理过程视为一个决策过程,主要根据会话的整体成功情况来优化系统动作的选择过程^[22]。随着机器学习以及人工智能的发展,基于统计的方法逐渐成为主导。例如易鑫等人^[23]尝试利用用户自然输入的行为数据来解释用户输入意图,主要通过贝叶斯定理将交互意图的后验概率转化为交互意图的先验概率与用户输入行为的条件概率的乘积,再计算建模。Bai 等人^[24]采用混合整数线性规划方法优化驾驶场景中的输入信号,进而识别其他车辆在交叉口和高速公路上换道时的意图。Nguyen 等人^[25]通过马尔可夫决策过程推断用户正在执行的简单任务或子任务,将意图识别应用于游戏场景。吴涛等人^[26]将部分可观测马尔可夫决策过程(POMDP)应用于网络安全领域,用以识别网络入侵意图。Dong 等人^[27]针对单模态输入时交互性不足和低准确率问题,设计了一种基于决策级融合的算法,用以解决虚拟化学实验中用户操作意图的感知问题。

指-笔混合输入能够弥补触摸交互和笔式交互各自交互能力不足的问题,但目前缺乏相应的指-笔混合模型和交互原语来定义混合输入空间、规范指-笔并行操作,从而有效指导指-笔混合交互的设计和开发过程,如混合交互原语设计、用户意图提取等。目前已有有一些笔式交互和触摸交互的研究能够为指-笔混合输入交互模型的研究提供理论支撑。如田丰等人^[1]和戴国忠等人^[2]提出的笔式交互原语生成模型,可以很好地描述 4 种基本原语的状态变化和生成过程,为混合交互原语的提出奠定了基础。刘杰等人^[16]基于连续交互空间的概念,将混合交互手势、空中手势和表面触控手势进行了统一,提出了三维连续交互空间下的混合手势交互模型,并给出了多种交互空间的融合方案,虽然其核心在于三维空间中的手指交互,但也为指-笔混合交互原语的归一化提供了参考。Google Research 的 Li 等人^[15]针对触摸交互中的不确定性导致开发人员难以开发自定义手势的问题,分析了人的触控行为并建立了动作模型,开发了手势设计工具 Touch 并提供了 API,可以促进和加速开发人员描述和设计想要的触摸手势。虽然该研究主要关注了触摸手势,但对本文并行交互原语的状态描述提供了重要依据。

通过上述文献调研可以看出,虽然很多研究成果对本文的研究有很好的借鉴意义,但针对指-笔混合交互中的一些核心问题,如指-笔并行交互机制、指-笔混合交互模型和交互原语、用户意图提取等并没有进行深入研究。基于此,本文从指-笔混合的并行协作机制和意图关联出发,提出了一种面向指-笔混合输入的交互模型、定义了指-笔混合交互原语、提出了基于时序信息的指-笔混合输入意图提取方法,并通过一个用户实验说明了指-笔混合输入的优势。

2 指-笔混合输入交互模型

指-笔混合输入支持串行和并行的指、笔混合输入。作为一种自然交互方式,指-笔混合输入扩展了指笔交互信道,扩充了指笔交互空间,具有交互通道多、学习成本低、交互效率高、用户体验好等优点。但与此同时,指-笔混合输入也带来了输入时更多的不确定性。

指-笔混合输入的交互模型是基于指-笔混合输入过程中的时序特征提出的一种解决交互过程中不确定性的解决方案,如图 1 所示,该方案共包括 5 个部分和 2 个过程。其中,5 个部分是指交互信息、原语生成器、交互原语、意图提取器和意图;2 个过程是指交互原语产生过程和交互任务产生过程,分别与原语生成器和意图提取器相对应。原语生成器在一定程度上解决了交互手势识别中的不确定性问题,意图提取器解决了多义性交互原语的意图推断不确定性问题。

- 交互信息:在指、笔与触摸屏的接触过程中,系统通过不同的传感器可以获得指点设备的位置、压力、倾角等信息。基于电子笔传感器种类和数量的差异以及手指的交互特点,本文中的交互信息主要采用指-笔输入的位

置信息, 结合指-笔混合交互的双设备融合特征以及时序特征, 我们将指-笔交互信息 (pen-touch information, PTI) 定义为 $PTI = \langle Position, Time, Device \rangle$, 表示某时间、某种指点设备在触摸屏上的位置. 其中, $Position$ 为平面坐标系中的 X 、 Y 坐标, $Device$ 是指 Pen 或 Touch 两种指点设备. 指点设备在持续交互过程中的轨迹信息可以表示为交互信息的集合 (pen-touch information set, $PTIS$), 即 $PTIS = \langle PTI_1, PTI_2, \dots, PTI_n \rangle$, 可以描述某时间段内所有指点设备的状态变化.

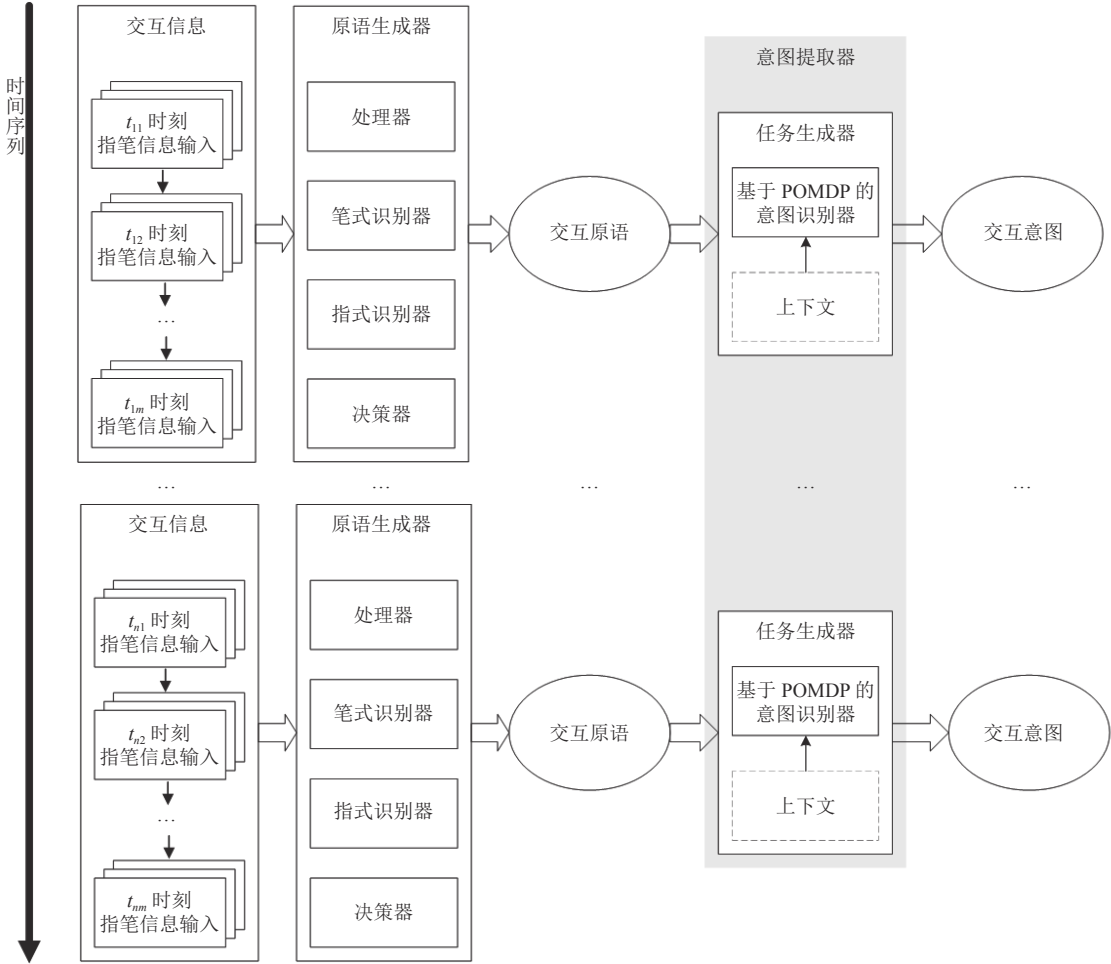


图 1 指-笔混合输入交互模型

- 交互原语: 交互原语是用户施加到计算机上的独立的、不可分割的最小操作^[2], 反映了人的交互动作的运动状态和方式. 根据指点设备在触摸屏上的运动轨迹信息, 利用二维空间上的时间、空间约束, 可将其划分为多种交互原语, 以完成更多的交互任务, 如单击、长击、平移、绘制等交互原语. 本文中的交互原语不仅包括归一化的独立或串行的指、笔交互原语, 还包括并行的混合交互原语, 如表征“笔长击同时双指旋转”的原语.
- 原语生成器: 原语生成器由处理器、笔式原语识别器、指式原语识别器和决策器构成, 它们一起协作对接收到的交互信息进行预处理和标准化、轨迹识别、混合原语决策等操作, 最终得到当前输入的交互原语.
- 意图: 在人机交互领域, 用户意图主要是指计算机对用户交互动作的理解和处理, 主要表现为从多种可能性中识别出用户的真实意图. 不同上下文中的同一交互动作可能有多种交互意图, 最终目标是借助计算机来完成用户期望的交互任务. 本文中交互意图主要指用户期待的计算机操作, 尤其指需要结合上下文才能确定的复杂交互

任务,如左对齐、垂直分布、高亮文本等.

● 意图提取器: 意图提取器主要是根据用户当前输入的交互原语、上一个交互原语及历史记录, 利用 POMDP 构建的模型来推断当前交互场景下的用户意图, 以期完成用户的交互任务. 因此, 任务生成器是对意图提取器的更为具体的描述. 如图 1 所示, 基于 POMDP 的意图识别器利用上下文 (context) 中的信息进行意图推断. 用虚线框表示的 context 是对当前交互环境的抽象表示, 它不仅包括前一个交互原语和交互任务, 还包含历史信息和交互对象等. 图中深色背景表达了用户的序列交互动作之间可能存在着隐式关联, 为复杂交互原语的交互意图提取提供了更为丰富的语义信息.

独立或并行的指-笔交互信息是用户与计算机进行交互的最小输入, 是实现其交互目的的最小交互动作, 表现为交互原语. 人们施加在计算机上的交互动作在计算机中被识别为交互原语, 通常具有明确的交互意图, 通过有限状态机可以实现交互原语与交互任务的绑定, 它们之间是一一对应的. 但人们的交互动作在不同环境下可能具有不同语义, 这催生了多义性交互原语. 借助用户的历史交互信息和当前交互上下文来提取用户的交互意图, 是本文解决这种意图不确定的策略, 表现为利用 POMDP 对用户的交互意图进行推断.

指-笔混合输入交互模型的核心是混合交互原语的生成和识别过程, 多义性交互原语在交互原语中的占比较小, 交互意图或交互任务通常是确定的, 不需利用 POMDP 模型对交互原语的语义进行再次推断, 故意图提取器是可选的. 图 1 中灰色背景的意图提取器和意图部分, 一方面表示当前交互意图与上下文有关, 受交互原语序列和各自的交互意图影响, 另一方面表示该部分是可选的, 受交互原语的交互意图数量制约.

3 指-笔混合输入交互原语

3.1 混合输入交互原语的形式化表达

指、笔交互原语的归一化表达是实现指、笔融合的基础, 特征分析和抽象是实现归一化的主要手段.

指-笔混合交互信息是组成交互原语的基本元素, 它是在指、笔交互过程中与屏幕接触时产生的. 指、笔在触摸屏上的运动轨迹表现为连续的点的集合, 它们除了输入精度上存在差异外, 采集到的数据具有一致的表达形式. 因此, 我们将每个轨迹点信息用前文所述的指笔交互信息 PTI 来表达.

交互原语是用户通过交互设备施加到计算机上的一个独立的、最小的、不可分割的操作. 指-笔混合交互的基本原语是一段连续轨迹信息所代表的操作, 它代表了用户在完成交互任务过程中输入的独立的、最小的、不可分割的一段指-笔交互信息的集合. 由此, 我们可以将指-笔混合交互原语表示为 $IP = \langle PTI_1, PTI_2, \dots, PTI_n \rangle = PTIS$.

虽然上述表达形式可以用来表示所有的混合交互原语, 但仅有一种交互原语是很难完成复杂交互任务的. 指-笔混合交互原语是两种交互原语的混合体, 因此, 我们有必要将指式原语和笔式原语的定义囊括进来. 为了更清晰地体现该原语的产生设备和定义更多的交互原语, 我们将交互原语进一步优化为 $IP = \langle Gesture, PTIS, Constraint, Device \rangle$. 其中, *Gesture* 为交互原语的名称, 可以是单击、双击、绘制和自定义原语; *Constraint* 为约束条件, 表示该原语所具有的特征, 如时间约束、图形约束等; *Device* 为交互原语的产生设备, 主要包括 *Pen*、*Touch* 和 *PenTouch*.

前面用元组的形式定义了指-笔混合输入交互信息和混合输入交互原语, 其精确的 BNF 描述如下所示.

```
< IP > ::= < Gesture > < PTIS > < Constraint > < Device >,
< Gesture > ::= Tap|Hold|Pan|Drag|CustomizedPrimitive,
< PTIS > ::= {PTI},
< PTI > ::= < Position > < Time > < Device >,
< Device > ::= Pen | Touch | PenTouch.
```

3.2 混合输入交互原语空间

指-笔混合输入交互原语空间不是指、笔交互空间的简单合并, 它是由指、笔交互原语有机融合在一起所形成的交互空间, 主要包括归一化的单设备交互原语和并行的指-笔混合交互原语.

单设备 (指或笔) 交互原语是并行混合交互原语产生的基础, 它们拥有通用的表示形式, 其主要差异在于约束条件, 主要包括明确了时间和空间的约束以及具有相似特征的图形约束. 如用当前时间与起始时刻的时间间隔 Δt 、当前位置与起始点的距离 Δd 以及给定的两个非常小的阈值 T 和 D 可以定义 4 种基本交互原语: 单击 Tap、长击 Hold、移动 Pan、绘制 Drag^[1,2]. 利用有向图约束可以定义更多的交互原语, 如左滑 SwipeLeft、左向圆弧 ArcLeft、有向矩形 Rectangle、波浪线 WaveLine、往复折线 ReciprocateLine 等, 可以完成翻页、撤销、选择、强调、删除等操作任务. 手指作为一种特殊的输入设备, 不仅可以完成上述交互原语, 还可以实现符合人们交互习惯的多手指交互原语, 如双指缩放 Pinch、双指旋转 Rotate、4 指捏合 Pinch4 等.

人类双手协作的行为习惯^[4]和指笔协作的相关研究工作^[5-12]为我们对并行混合交互原语的定义提供了指导. 加入新指点设备前长按当前指点设备, 描述了并行指-笔混合交互的状态, 可以用 Hold+Action 来抽象表示, 其 BNF 形式为<并行交互原语> ::= <笔长击><指原语> | <指长击><笔原语>. 按照这种构成规则并结合人类的交互习惯, 可以定义出多种并行混合交互原语. 表 1 列出了常用的指-笔并行混合交互原语, 并通过文字和图形两种形式对其进行了描述. 并行混合交互原语的命名, 借用了原有指、笔交互原语的名称, 约定了笔先、指后的顺序, 制定了数字代表手指数量的规约, 同时遵循骆驼命名法则, 如 TapHold4 表示“笔单击同时 4 指长击”, 换为更符合人们习惯的中文描述为“4 指长击同时笔单击”.

表 1 指-笔并行混合交互原语

混合原语	示意图	混合原语	示意图	混合原语	示意图
HoldTap 笔长击同时单指点击		HoldPinch 笔长击同时双指缩放		PanHold 单指长击同时笔移动	
HoldTap2 笔长击同时双指点击		HoldPinch3 笔长击同时 3 指捏合		PanHold2 双指长击同时笔移动	
HoldTap3 笔长击同时 3 指点击		HoldPinch4 笔长击同时 4 指捏合		PanHold3 3 指长击同时笔移动	
HoldTap4 笔长击同时 4 指点击		HoldPinch5 笔长击同时 5 指捏合		PanHold4 4 指长击同时笔移动	
HoldTap5 笔长击同时 5 指点击		HoldUp2 笔长击同时 双指上滑		PanHold5 5 指长击同时笔移动	
HoldUp 笔长击同时单指上滑		HoldDown2 笔长击同时 双指下滑		TapHold 单指长击同时笔单击	
HoldDown 笔长击同时单指下滑		DragHold 单指长击同时笔绘制		TapHold2 双指长击同时笔单击	
HoldRight 笔长击同时单指右滑		CHold 单指长击同时笔写字母C		TapHold3 3 指长击同时笔单击	
HoldLeft 笔长击同时单指左滑		VHold 单指长击同时笔写字母V		TapHold4 4 指长击同时笔单击	
HoldRotate 笔长击同时双指旋转		ZHold 单指长击同时笔写字母Z		TapHold5 5 指长击同时笔单击	

3.3 混合输入交互原语的产生过程

指-笔混合输入交互原语产生过程是指从指点设备按下至其抬起的全过程, 既包括笔或手指单独交互, 也包括指、笔交叉按下并行协作至其全抬起的过程. 隐式交互是指-笔混合交互最核心的特征之一, 交互原语是实现隐式交互的主要手段. 在使用指、笔与触摸屏进行交互时, 计算机并不能直接识别用户的交互动作, 系统需要从其轨迹信息中识别出与之对应的交互原语, 进而完成相应的交互任务.

交互原语是现实世界中用户的交互动作在计算机中的映射, 图 2 详细地描述了指-笔混合输入交互原语的产生过程. 随着用户不断输入交互信息, 接收器不断地接收数据, 并利用时间器为它们添加时间标志; 之后, 接收器将这些时空信息送到处理器, 处理器根据输入信息的设备特征将其分发给笔式原语识别器或指式原语识别器, 待识别器识别完成后再将结果反馈给处理器; 处理器再将两个原语识别器的结果送入决策器; 决策器再次识别汇总后, 将最终结果反馈给处理器; 处理器将最终的交互原语以概率分布的形式输出; 最后意图分布通过映射函数 π 获得最终的用户意图.

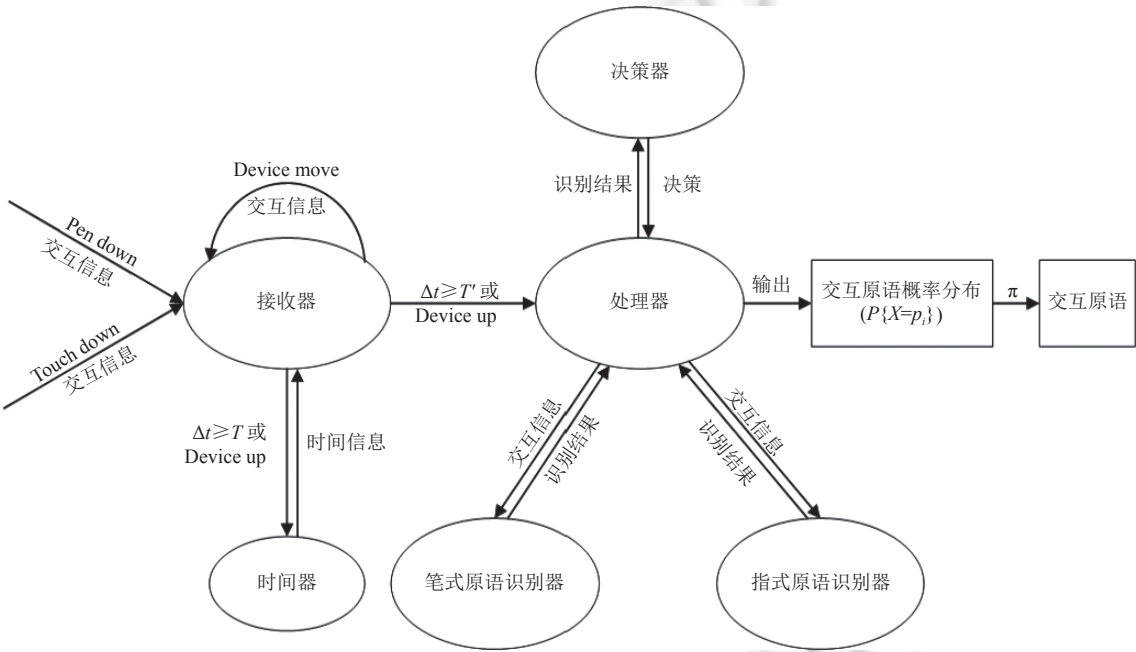


图 2 指-笔混合输入交互原语产生过程

在指-笔混合交互原语的产生过程中, 其核心是数据处理, 主要包括接收器的数据预处理和处理器的交互原语识别. 数据预处理主要工作是将数据归一化和去噪, 而交互原语识别是对数据进行的专业化、业务化处理. 图 3 通过数据流程图的形式描述了处理器识别指笔混合输入交互原语的过程. 从流程图中可以看出, 本文所述模型不仅可以识别指式和笔式交互原语, 还可以识别指-笔并行混合交互原语.

指-笔混合输入交互原语产生过程与以往交互原语的一个重要区别是支持过程中的增量识别. 交互信息较少时, 识别器很难做出精准的原语识别, 只能得到几种交互原语的一个概率分布, 此时概率分布近似均匀分布, 不确定性较强; 随着交互信息的不断输入, 某种或某几种交互原语的概率明显增加. 如图 2 所示, 通过时钟可以将用户的连续输入信息进行离散化, 时间间隔 Δt 的引入为交互原语的实时识别提供了可能, 但系统向用户反馈的时机可以依据识别到的原语类型和概率分布情况而有所差异, 如原语概率分布中的最大概率值大于阈值直接作为最终识别结果, 而小于阈值不进行反馈或者在用户界面中以弹出菜单等形式向用户展示最有可能的几种交互原语.

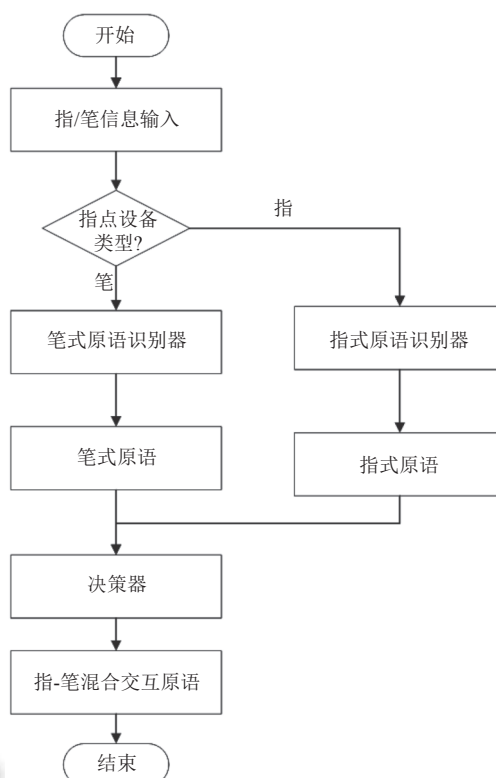


图3 指-笔混合输入信息数据处理流程

3.4 指-笔并行混合交互原语示例

指-笔并行混合输入交互原语拓展了指笔交互的带宽,是指-笔混合输入交互模型的重要组成部分。根据指-笔混合输入交互原语产生过程,本文用按下 (device down)、移动 (device move)、抬起 (device up) 来抽象表示用户与系统的交互。利用这3个交互动作,可以将交互原语进行初分类,然后再根据初分类进行细分,主要表现为对 device move 数据的处理和识别。接下来,我们将结合示例来探索人的交互动作对并行混合交互原语的影响。

Device down、device move、device up 是触控交互系统中的基本事件,是人的交互动作在交互系统的直接映射,是触控 (操作) 系统不可或缺的事件或方法。这3个交互动作是交互系统获取交互信息的主要途径,每个指-笔混合交互原语都可以用它们来表示。Device down 或 device up 通常表示交互原语的开始、结束或原语内的状态转换。Device move 用来接收用户指点设备的运动轨迹信息,通常表示系统状态的变迁;但由于用户难以将笔或手指稳定在光滑的触摸表面上,即使没有滑动的意图,往往也会发生 device move 事件,如图4所示的 HoldTap 混合交互原语时序动作序列。该图描述了指-笔混合交互过程中所有可能的 device move 操作,显然它们是无意识下的指点设备的微小移动,全为噪声,在进行交互原语识别时需要将其删除。从交互动作序列中删除这些噪声,得到了如图5所示的 HoldTap 交互原语状态迁移图,它不仅明确了交互原语的语义信息,规范了交互原语的约束条件,还使交互原语更具有可识别性。同时,状态迁移图也隐含了状态转移概率的信息,如 $P(S_1|S_2)=1$ 和 $P(S_2|\text{TouchDown})=1$ 。

从图5所示的状态迁移图中可以观测到,新指点设备加入和退出必将引起系统状态的变化。此外,指点设备的显著运动 device move (位移大于阈值) 也会产生状态的变化。图6展示了去除噪声后的标准的 HoldPinch 混合交互原语的交互动作序列,其中 touch move 为 HoldPinch 原语的交互动作,产生了大量的交互数据,为原语识别提供了数据支持。图7为 HoldPinch 混合交互原语的状态迁移图,它不仅描述了标准输入的状态转移信息,还在核心区域添加了两种可能的状态迁移 $S'_2 \rightarrow S_3$ 与 $S_4 \rightarrow S'_5$ 。其中,状态 S'_2 与 S'_5 一般表示用户输入动作错误或更改输入意图的情形;用户的这种操作行为,为交互原语识别增加了难度。

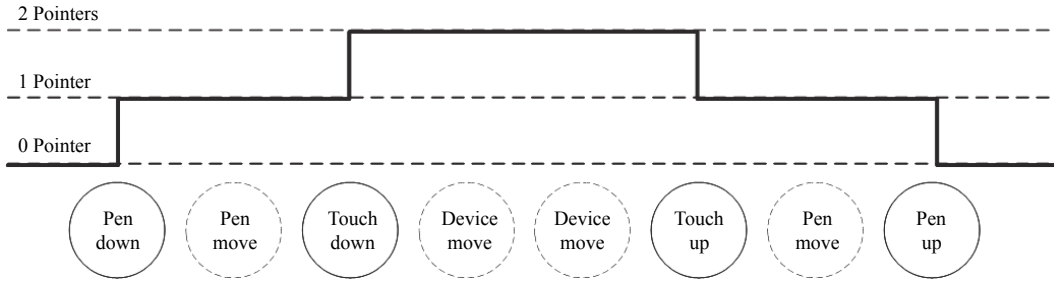


图 4 HoldTap 混合交互原语时序动作示意图

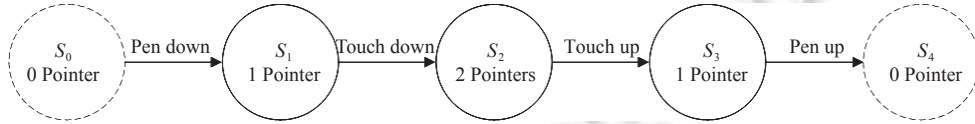


图 5 HoldTap 混合交互原语状态迁移图

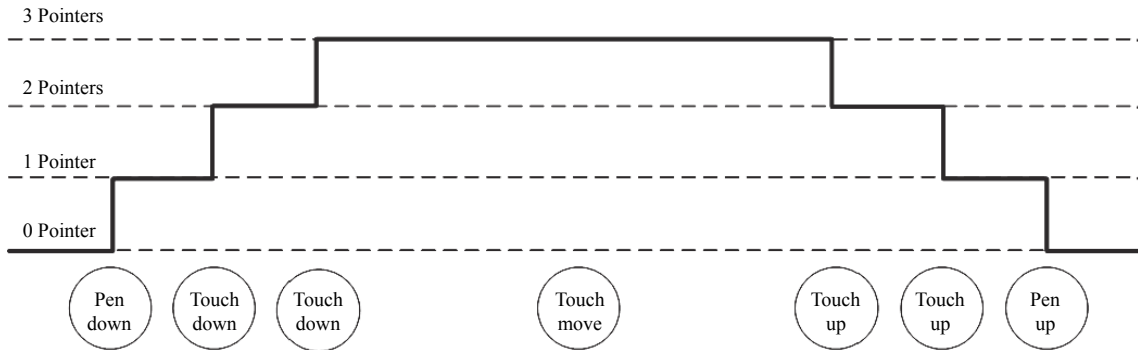


图 6 HoldPinch 混合交互原语时序动作示意图

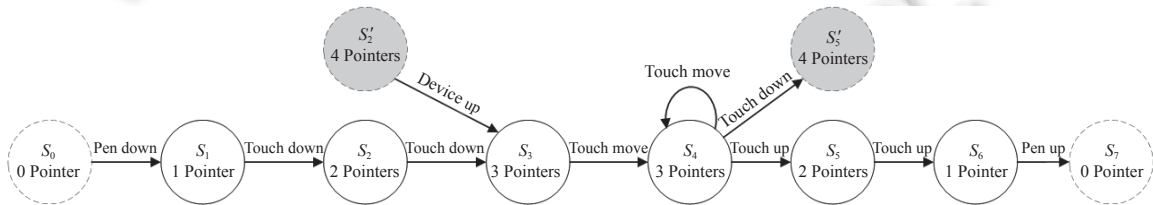


图 7 HoldPinch 混合交互原语状态迁移图

图 5 和图 7 用 3 个交互动作来描述 HoldTap 和 HoldPinch 交互原语的状态变迁, 但这两个迁移图同样可以用来描述 HoldHold、HoldRotate 交互原语. 为了充分识别并区分这些交互原语, 还需进一步挖掘交互动作的特征, 如将 device move 细化为有向的横、竖、撇、捺、折、圆弧运动. 很多图形识别和手势识别的研究工作都可以进一步区分具体的交互原语, 由于不是本文的研究重点, 不作深入探讨.

4 指-笔混合输入意图提取方法

4.1 基于 POMDP 的增量式交互意图提取方法

对于具有多种交互意图的交互原语, 需要获取更多的信息才能推断出用户最终的交互意图, 执行期望的交互

任务. 当前输入的具有多义性的交互原语, 可以从上下文中获取到更多的语义信息. 用户在本次交互之前输入的交互原语序列或者历史交互记录蕴含着许多有价值的信息, 对这些潜在信息的挖掘和利用是解决指-笔混合输入意图不确定性的关键, 而能够对部分可观测的时序数据建模的 POMDP 是解决这类问题的一种有效方法.

在用户利用指、笔持续输入的过程中, 我们得到的信息是不完整的, 利用部分信息提取到的用户意图具有很高的不精确性, 但随着交互原语的持续输入系统可以获取到的有价值信息越来越多, 交互意图逐步由不确定性向确定性转化. 因此, 利用 POMDP 进行意图推断的方法具有增量特征, 是一种增量式交互意图提取方法.

在人工智能领域, POMDP 描述了 Agent 在不确定性环境中的运行方式, 一般可以用一个 8 元组来定义, 即 $\langle S, A, T, R, \Omega, O, \gamma, b_0 \rangle$ [28]. 其中, S 表示状态空间, 是一组机器状态 s 的集合; A 表示动作空间, 是机器可能执行的动作 a 的集合; T 表示系统的状态转移函数集合, $T(s, a, s') = P(s'|a, s)$, 表示在当前状态 s 下执行动作 a 转移到状态 s' 的概率; R 表示回报函数的集合, $R(s, a)$ 表示在状态 s 下执行动作 a 所获得的即时收益; Ω 表示观测空间, 是所有观测 o 的集合; $O: S \times A \rightarrow \Pi(\Omega)$ 表示观测函数的集合, 观测函数用 $O(s', a, o) = P(o|s', a)$ 表示, 为执行动作 a 并且转移到状态 s' 的情况下得到观测 o 的概率; $\gamma \in [0, 1]$ 是折扣系数; b_0 表示初始信念状态.

在指-笔混合输入的系统中, 我们用 s_t 表示系统 t 时刻的不可完全观测状态, 它隐含了 3 个方面的信息: 用户的最终意图、当前输入的交互原语、历史交互记录等; 用 a_t 表示 t 时刻的机器动作, 它是用户的交互行为在系统中的映射, 表现为交互原语识别后所执行的交互任务; 用 o_t 表示 t 时刻所能观测到的状态特征, 如系统中可视化构件的选中状态、选中的数目、位置信息的变化等.

通过指-笔混合输入系统的建模, 可将交互原语意图提取问题转化为 POMDP 决策问题. POMDP 具有出色的建模能力, 但其精确求解所带来的高计算复杂度使其难以应用; 近些年, 其近似求解方法取得了巨大进展并被广泛使用, 如 PBVI [29]、FBVI、Perseus、POMCP、DESPOT 等求解算法. 引入信念状态将 POMDP 转化为 Belief MDP 是使用最为广泛的一种求解方案, 很多求解算法都是基于该思想. 引入信念状态后, 不可完全观测的系统状态可以通过一个完整的动作、观测序列来逐步推断系统状态的变化, 这一前后相继的时序序列被称为历史, 可用 $h_t = \{a_1, o_1, a_2, o_2, \dots, a_t, o_t\}$ 表示, t 时刻的信念状态可表示为 $b_t = P(s_t|a_t, o_t, a_{t-1}, o_{t-1}, \dots, a_0, o_0, s_0)$. 此时, POMDP 问题求解的核心是寻求一个 (近似) 最优策略, 即将内部信念状态映射到动作的最优函数, 通常用 π 、 $\pi(b)$ 或 $\pi(b_0, h_t)$ 来表示, 并且有 $\pi(b) \in A$.

使用 POMDP 进行用户意图提取的最终目的, 是为了通过对交互意图的建模使系统选择最优的机器动作, 表现为实现折扣回报和的期望最大, 可用公式 (1) 来表示, 其中 π^* 为最优策略.

$$\pi^* = \arg \max E \left[\sum_{t=0}^{\infty} \gamma^t \sum_{s \in S} b_t(s) R(s, \pi(b_t)) \right] \quad (1)$$

4.2 意图提取示例

考虑到人的记忆容量和人的交互动作的复杂性, 有些指-笔混合交互原语会有多种交互意图, 接下来我们以指-笔并行交互原语 HoldLeft 为例, 展示意图提取方法的具体示例.

状态空间 $S = \{s_0, s_1, s_2\}$, 用来表征当前动作对后续交互原语 HoldLeft 多意图提取的影响, 空间中的状态可以表示为 $s_t = \langle Position, SelectionCount, Intention \rangle$, 蕴含了位置、可视化构件选择数和意图等特征. 在集合中, s_0 用于表示左对齐, s_1 表示撤销, s_2 表示返回上一界面 (或退出当前界面).

动作空间 $A = \{a_0, a_1, a_2, a_3\}$, 是机器动作的集合, 当交互原语具有唯一性时, 机器动作可以用交互原语来表示. 其中, a_0 表示点击动作, 用于选取构件或点击屏幕 (Tap 交互原语); a_1 表示移动动作, 可以用于改变构件的位置 (Pan 交互原语); a_2 表示批量选取动作, 可以同时选中多个构件 (Drag 交互原语); a_3 表示增量选取动作, 用于增选或减选构件 (HoldTap 交互原语).

观察空间 $\Omega = \{o_0, o_1, o_2, o_3, o_4, o_5\}$. 其中, o_0 表示观察到各个构件都处于常态, 即没有构件被选中并且没有构件的位置刚刚发生改变; o_1 表示观察到有一个构件被选中; o_2 表示观察到多个构件被选中; o_3 表示观察到多个构件以某构件的左边为参照进行了移动; o_4 表示观察到系统返回到上一次操作的状态, 主要表现为选中项目数和位置

信息; o_5 表示观察到退出了当前界面.

状态转移概率 $T(s,a,s')$: 机器动作直接影响系统状态的变化, 反映了当前动作对用户最终意图的期望; 构造的完整的状态转移矩阵如表 2 所示.

观测转移概率 $O(s',a,o)$: 观测转移概率反映了执行动作 a 后转移到状态 s' 后得到相关观测的概率, 完整的观测转移概率矩阵如表 3 所示.

表 2 状态转移概率矩阵

初始状态	动作	最终状态		
		s_0	s_1	s_2
s_0	a_0	0.2	0.4	0.4
s_1		0.2	0.4	0.4
s_2		0.4	0.1	0.5
s_0	a_1	0.4	0.3	0.3
s_1		0.3	0.4	0.3
s_2		0.2	0.4	0.4
s_0	a_2	0.85	0.1	0.05
s_1		0.85	0.1	0.05
s_2		0.9	0.05	0.05
s_0	a_3	0.8	0.15	0.05
s_1		0.45	0.45	0.1
s_2		0.8	0.15	0.05

表 3 观测转移概率矩阵

初始状态	动作	最终状态					
		o_0	o_1	o_2	o_3	o_4	o_5
s_0	a_0	0.1	0.3	0.2	0.4	0	0
s_1		0.2	0.2	0	0.2	0.4	0
s_2		0.2	0.3	0	0	0	0.5
s_0	a_1	0.4	0	0	0.4	0.2	0
s_1		0.1	0.1	0.1	0.2	0.4	0.1
s_2		0.1	0.1	0.05	0.3	0.05	0.4
s_0	a_2	0	0.1	0.3	0.4	0.1	0.1
s_1		0.1	0.1	0.3	0.1	0.3	0.1
s_2		0.2	0.1	0.2	0.1	0.1	0.3
s_0	a_3	0	0.3	0.3	0.3	0.05	0.05
s_1		0	0.2	0.2	0.2	0.4	0
s_2		0.1	0.1	0.4	0	0	0.4

回报 $R(s,a)$: 如在当前状态 s 下采取动作 a 后得到了期望结果, 则回报值+10; 否则, 根据差异程度给予 1、-10 和-1 的回报值, 具体的回报函数值参见表 4.

表 4 回报函数值矩阵

状态	动作			
	a_0	a_1	a_2	a_3
s_0	-10	-1	10	1
s_1	1	1	-10	-1
s_2	10	1	-10	-1

R 平台已经实现了 grid、enum、twopass、witness、incprune、sarsop 等多种 POMDP 求解算法, 该示例直接使用 pomdp 包中的相应函数进行求解. 将上述的状态空间 S 、动作空间 A 、观测空间 Ω 、状态转移概率矩阵 T 、观测转移概率矩阵 O 、回报函数 R 、折扣因子 γ 、初始置信状态概率 $b_0 = P(s_0) = (1/3, 1/3, 1/3)$ 等数据以指定的格式代入 R 平台 pomdp 包中的 POMDP 函数, 构建问题模型, 然后利用 solve_POMDP 函数进行模型求解.

采用默认的 grid (一种 PBVI 实现算法) 求解方法和 7 位精度进行计算, 结果表明我们构造的模型能够收敛, 总的期望回报为 18.949661, 信念迭代次数为 10017, 并得到了 α 向量矩阵和最优策略. 利用内置的函数 plot_policy_graph, 得到了如图 8 所示的策略图. 图中共有 7 个节点, 每个节点代表一个置信状态, 代理从图中标记有“initial belief”节点 6 开始, 初始时刻 3 种置信具有同等概率, 基于回报函数矩阵等数据得到最优策略是执行动作 a_3 ; 节点之间的圆弧代表观测, 根据当前观测可以转到置信状态 5 或状态 7; 继续执行类似操作, 直到再次回到状态 6, 问题被重置. 策略图不仅能够描述施加动作后观测特征之间的转移, 还能描述信念状态之间的转换; 图中没有机器动作 a_1 , 这与模型中状态转移矩阵、观测转移矩阵和回报函数矩阵的具体数据有关, 它是当前数据的一种最优解.

从该原语交互意图提取示例可以看出, 本文提出的基于 POMDP 的增量式意图求解方法可以解决多义性指-

笔混合输入交互原语的交互意图提取问题. 虽然该示例相对于真实场景过于简单, 选择的求解算法也不一定是最优方案, 但计算结果的收敛特征及其他结果数据都佐证了该方法的可用性.

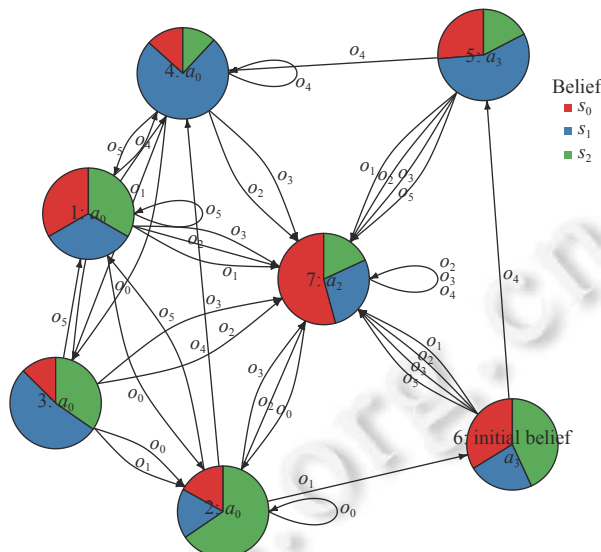


图 8 策略图

5 用户评估实验

为了评估本文提出模型的可行性和先进性, 我们将它与传统指笔交互方法进行了对比, 其中传统指笔交互是指支持指笔串行的触控交互. 本文通过两种交互方法在 4 个具体任务上的完成时间和点击功能区 (组件) 次数来评价两种方法的工作效率.

5.1 被试及实验环境

我们招募了 20 名被试, 12 名男性, 8 名女性, 年龄为 23–39 岁 (平均年龄为 28.5 岁), 惯用手均为右手, 均有平板和手写笔的使用经验.

我们的实验评估是在华为 WGR-W19 平板电脑上进行的, 该设备配套 HarmonyOS 2.0 操作系统, Huawei Kirin 9000E 处理器, 8 GB 运行内存, 2560×1600 分辨率, 12.6 英寸触摸屏. 同时该设备配备了第二代 M-Pencil 手写笔, 程序基于安卓开发.

基于本文模型实现的用户界面如后文图 9 所示, 传统指笔交互用户界面与支持指笔并行交互用户界面基本一致, 区别仅在于工作区指笔交互原语支持力度不同. 用户界面由功能区、工作区、状态栏和任务控制区这 4 部分构成. (1) 功能区基于工具栏和菜单实现了用户办公环境下的一些常用功能, 包括重置、图形样式 (矩形、椭圆、菱形)、选择模式 (单选、多选)、复制、粘贴、删除、文本颜色、填充颜色、对齐分布、设置等功能, 并且文本颜色、填充颜色、对齐分布都存在二级菜单. (2) 工作区用于图形的绘制和编辑. (3) 状态栏主要呈现工作区图形个数、位置、大小, 并可以实现位置、大小的再编辑. (4) 任务控制区位于界面右下角, 用于 4 个任务的切换、执行和校验. 在该应用程序中, 将侧边指式交互原语与笔式交互原语融合生成了指-笔并行交互原语.

5.2 实验流程

整个实验将持续 20 min, 具体的实验过程如下.

(1) 被试需要在主试的指导下熟悉用户界面和两种交互方法, 特别是并行混合交互原语, 主要包括等宽等高图形绘制、快速复制、多选与减选、锚定对象切换、6 种对齐和 2 种均匀分布.

- (2) 被试需要严格按照实验要求, 逐个完成具体的实验任务.
- (3) 被试完成任务后, 需要填写一份用户体验反馈表. 该表由用户基本信息、标准系统可用性量表 (system usability scale, SUS)^[30]和反馈建议这 3 部分构成.

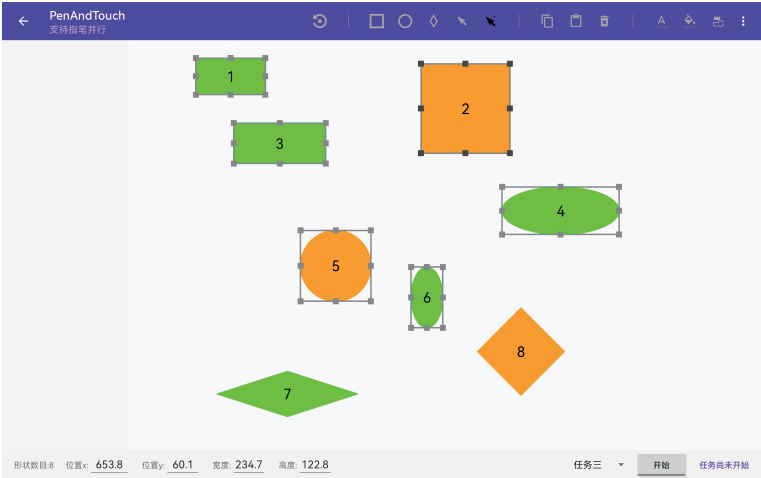


图 9 本文指笔混合交互界面

5.3 实验任务

为了进一步验证本文模型的可行性和先进性, 我们在 APP 上分别实现了支持传统指笔操作和指笔并行操作的两个用户界面, 用户可以使用不同的交互方式完成相同的功能. 根据日常办公需要, 我们设计了 4 个实验任务. 这些实验任务针对 6 种常用图形对象, 涵盖了创建、复制、多选、锚定对象切换、对齐、均匀分布等常用操作.

- (1) 创建图形: 新建 6 个图形 (矩形、椭圆、菱形、正方形、圆形和正菱形各 1 个).
- (2) 复制图形: 红色图形复制 1 个, 绿色图形复制 2 个, 蓝色图形复制 3 个.
- (3) 切换基准图形并左对齐: 在已选中的一组图形上切换锚定对象并实现左对齐 (选择所有绿色图形, 并以图形 1 为基准实现左对齐).
- (4) 多个图形对齐与均匀分布: 3 种不同颜色图形以指定图形为基准实现各自的对齐与分布操作 (红色图形以图形 1 为基准右对齐, 灰色图形以图形 3 为基准底端对齐, 蓝色图形以图形 2 为基准垂直居中对齐并平均分布).

5.4 实验结果

本文交互方式与传统交互方式的任务完成时间存在显著差异, 相较于传统交互方式, 本文交互方式任务完成时间更短. 如图 10 所示, 传统交互方式完成任务 1~4 的平均时间分别为 58.21 s (SD=10.306)、30.04 s (SD=8.021)、20.79 s (SD=5.973)、44.83 s (SD=15.781), 而本文所提出的交互方式的用时分别为 24.79 s (SD=8.490)、20.20 s (SD=3.810)、19.81 s (SD=5.350)、44.19 s (SD=9.982). 由于数据服从正态分布, 我们使用配对 t 检验方法对两种交互方式在 4 个任务上的完成时间进行了显著性分析. 结果表明, 本文交互方式与传统交互方式在任务 1 和任务 2 上存在显著性差异 (任务 1: $t=14.846, p<0.05$; 任务 2: $t=5.171, p<0.05$), 但是在任务 3 和 4 上未发现显著性差异 (任务 3: $t=0.468, p=0.323$; 任务 4: $t=0.148, p=0.442$). 结合数据, 我们发现本文并行交互原语相较于传统操作耗时更短, 耗时差异的大小与并行交互原语所替代的传统操作的复杂程度有关. 例如, 在任务 1 中进行等宽等高图形绘制时, 传统交互方式需要借助属性面板修改宽高值来实现, 而本文方法可以利用并行交互原语 PanHold 直接达成目标, 所以耗时更短; 但在完成任务 4 中的对齐、平均分布等操作时, 由于传统交互方式的操作实现并不复杂, 不能很好地体现出本文交互方式的优越性, 这也从另一个方面证实了本文交互方式在相对复杂的操作程序中优势更为明显.

我们还对两种交互方式点击功能区(组件)次数进行了分析,两者之间也存在着显著差异,本文交互方式相交更优.如图11所示,传统交互方式完成任务1~4过程中点击功能区次数均值分别为3.05(SD=0.223)、10.05(SD=1.317)、1.00(SD=0.324)、5.40(SD=0.754),而本文交互方式的点击次数均值分别为3.45(SD=0.826)、0.00(SD=0.000)、0.15(SD=0.366)、0.55(SD=1.395).我们同样使用配对 t 检验方法对两种交互方式在4个任务上的点击功能区次数进行显著性分析,结果表明,两种交互方式在这4个任务上都存在显著性差异(任务1: $t=-2.027$, $p<0.05$; 任务2: $t=34.130$, $p<0.05$; 任务3: $t=7.768$, $p<0.05$; 任务4: $t=24.718$, $p<0.05$).这进一步表明了本文提出的交互方式在点击功能区次数上存在明显优势,从而使得用户使用本模型进行界面设计时可以大大节省屏幕空间.例如,在任务2中,本文交互方式的点击功能区次数可以为0,即用户可以完全依赖并行交互原语实现相应操作,而无需点击功能区,大大节省了屏幕空间.同时,我们注意到任务1中 t 值符号为负,这与任务设计和应用程序的实现有关.

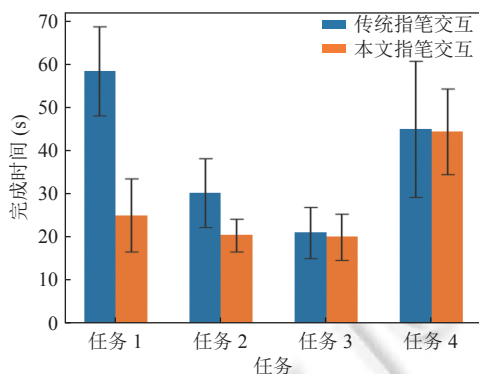


图10 两种交互方式完成任务时间对比

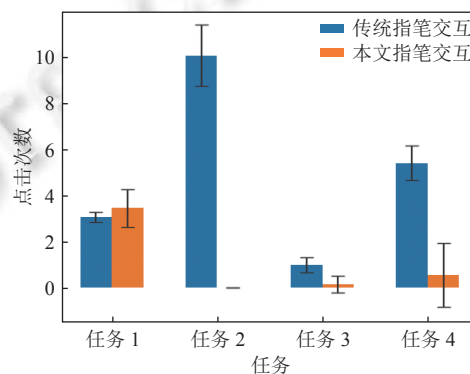


图11 两种交互方式点击功能区次数对比

5.5 用户反馈

在被试完成实验后,我们还通过标准的SUS量表收集了他们对本文模型应用的主观评价.

SUS量表统计结果显示,本模型应用的SUS平均分为85.1(SD=10.590),对应百分等级为96.3%,表示本模型优于96.3%的同类产品,再次肯定了本文模型的可用性和先进性.同时,我们利用量表中Q4和Q10两个子问题组成的子量表求得了系统的易用性平均分为88.1(SD=10.319);利用Q1、Q6、Q9组成的子量表得到了系统满意度平均分为86.3(SD=16.058).SUS子量表的结果表明,我们在用户实验中使用的并行交互原语很容易记忆,利用本文模型构建的应用获得了较高的满意度评价.

6 总结与讨论

本文结合人机界面和信息技术的发展趋势,分析了笔式交互、触摸交互和指笔混合输入技术的发展现状,并对各自的优越性和存在的问题进行了探讨.指笔组合交互正在成为一种新的交互手段,引起了学术界和工业界的广泛关注,但主要集中在串行的分工合作和特定任务的并行实现方式上,缺乏对一般情境下指笔并行交互问题的深入探讨.基于指-笔混合输入的时空连续性,本文提出了一个指-笔混合输入的交互模型.该模型以指-笔混合输入为切入点,通过定义指-笔并行混合输入交互原语来拓展触屏设备的交互能力,解决了因交互手势复杂而引起的不确定问题;通过构建基于POMDP的增量式意图提取方法,进一步解决了多语义交互原语意图不确定性问题,并结合交互原语和意图提取的示例,更加清晰地展示了模型的应用方式.最后,通过用户评估实验,进一步验证了指-笔混合输入交互模型的优越性和可行性.

虽然用户评估实验验证了本文模型的可行性和先进性,但本文的工作依然存在几点不足.(1)现有主流的移动操作系统如Android、HarmonyOS等并未支持指、笔区分功能,它们的事件机制与本文的交互原语并不完全一致,指笔并行功能的实现难度较高,因此,本文示例仅实现了部分并行交互原语.目前还没有完美支持指-笔混合交互

的系统^[7],从操作系统层面研究指笔混合交互问题将成为本文今后的一个主要研究方向。(2) POMDP 在构建模型时具有很优秀的表达力,其近似求解算法很多,不同的求解方法的适应场景也有所差异,但本文仅以 PBVI 算法为例进行了展示,并未就哪种求解算法更适合意图提取方法进行探讨。(3) 本文的实验任务虽然从任务绩效上都达到了预期目标,但应用本文方法完成任务 1 的点击功能区次数偏高,这是因为本文方法和传统方法理论上应该具有相同的点击次数,但被试对本文方法的熟练程度不如传统方法,容易出现失误而重复操作。同时实验中针对 6 种常用图形对象的 4 种常用任务取得好的效果,但在未来更大规模的实验中可进一步扩展操作对象和任务类型。

References:

- [1] Tian F. Research on Post-WIMP software interface [Ph.D. Thesis]. Beijing: University of Chinese Academy of Sciences, 2003 (in Chinese with English abstract).
- [2] Dai GZ, Tian F. Pen-based User Interface. 2nd ed., Hefei: China University of Science and Technology Press, 2014 (in Chinese).
- [3] Tong QS, Zhang ZQ, Huang J, Tian F, Liu J, Dai GZ. Scenario design tool based on hybrid input. *Ruan Jian Xue Bao/Journal of Software*, 2019, 30: 48–61 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19017.htm>
- [4] Brandl P, Forlines C, Wigdor D, Haller M, Shen C. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In: *Proc. of the 2008 Working Conf. on Advanced Visual Interfaces*. Napoli: ACM, 2008. 154–161. [doi: 10.1145/1385569.1385595]
- [5] Zhang Y, Pahud M, Holz C, Xia HJ, Laput G, McGuffin M, Tu X, Mittereder A, Su F, Buxton W, Hinckley K. Sensing posture-aware pen+touch interaction on tablets. In: *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems*. Glasgow: ACM, 2019. 55. [doi: 10.1145/3290605.3300285]
- [6] Guiard Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, 1987, 19(4): 486–517. [doi: 10.1080/00222895.1987.10735426]
- [7] Hinckley K, Yatani K, Pahud M, Coddington N, Rodenhouse J, Wilson A, Benko H, Buxton B. Pen + touch = New tools. In: *Proc. of the 23rd Annual ACM Symp. on User Interface Software and Technology*. New York: ACM, 2010. 27–36. [doi: 10.1145/1866029.1866036]
- [8] Pfeuffer K, Hinckley K, Pahud M, Buxton B. Thumb + pen interaction on tablets. In: *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*. Denver: ACM, 2017. 3254–3266. [doi: 10.1145/3025453.3025567]
- [9] Hinckley K, Yatani K, Pahud M, Coddington N, Rodenhouse J, Wilson A, Benko H, Buxton B. Manual deskterity: An exploration of simultaneous pen+touch direct input. In: *Proc. of the 2010 CHI Extended Abstracts on Human Factors in Computing Systems*. Atlanta: ACM, 2010. 2793–2802. [doi: 10.1145/1753846.1753865]
- [10] Romat H, Pietriga E, Henry-Riche N, Hinckley K, Appert C. SpaceInk: Making space for in-context annotations. In: *Proc. of the 32nd Annual ACM Symp. on User Interface Software and Technology*. New Orleans: ACM, 2019. 871–882. [doi: 10.1145/3332165.3347934]
- [11] Srinivasan A, Lee B, Henry Riche N, Drucker SM, Hinckley K. InChorus: Designing consistent multimodal interactions for data visualization on tablet devices. In: *Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems*. Honolulu: ACM, 2020. 1–13. [doi: 10.1145/3313831.3376782]
- [12] Pfeuffer K, Dinc A, Obernolte J, Rivu R, Abdrabou Y, Shelter F, Abdelrahman Y, Alt F. Bi-3D: Bi-manual pen-and-touch interaction for 3D manipulation on tablets. In: *Proc. of the 34th Annual ACM Symp. on User Interface Software and Technology*. ACM, 2021. 149–161. [doi: 10.1145/3472749.3474741]
- [13] Tian F, Mu S, Dai GZ, Wang HA. Research on a pen-based interaction paradigm in Post-WIMP environment. *Chinese Journal of Computers*, 2004, 27(7): 977–984 (in Chinese with English abstract). [doi: 10.3321/j.issn:0254-4164.2004.07.015]
- [14] Li Y, Guan ZW, Dai GZ. Modeling Post-WIMP user interfaces based on hybrid automata. *Ruan Jian Xue Bao/Journal of Software*, 2001, 12(5): 633–644. <http://www.jos.org.cn/1000-9825/20010501.htm> [doi: 10.13328/j.cnki.jos.2001.05.001]
- [15] Li Y, Lu H, Zhang HM. Optimistic programming of touch interaction. *ACM Trans. on Computer-human Interaction*, 2014, 21(4): 24. [doi: 10.1145/2631914]
- [16] Liu J, Huang J, Tian F, Hu WP, Dai GZ, Wang HA. Hybrid gesture interaction model in the continuous interaction space. *Ruan Jian Xue Bao/Journal of Software*, 2017, 28(8): 2080–2095 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5123.htm> [doi: 10.13328/j.cnki.jos.005123]
- [17] Romat H, Henry Riche N, Hinckley K, Lee B, Appert C, Pietriga E, Collins C. ActiveInk: (Th) inking with data. In: *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems*. Glasgow: ACM, 2019. 42. [doi: 10.1145/3290605.3300272]
- [18] Cami D, Matulic F, Calland RG, Vogel B, Vogel D. Unimanual pen+touch input using variations of precision grip postures. In: *Proc. of the 31st Annual ACM Symp. on User Interface Software and Technology*. Berlin: ACM, 2018. 825–837. [doi: 10.1145/3242587]

- 3242652]
- [19] Shi YC. Interpreting user input intention in natural human computer interaction. In: Proc. of the 26th Conf. on User Modeling, Adaptation and Personalization. Singapore: ACM, 2018. 277–278. [doi: [10.1145/3209219.3209267](https://doi.org/10.1145/3209219.3209267)]
 - [20] Yu K, Chen L, Chen B, Sun K, Zhu S. Cognitive technology in task-oriented dialogue systems: Concepts, advances and future. Chinese Journal of Computers, 2015, 38(12): 2333–2348 (in Chinese with English abstract). [doi: [10.11897/SP.J.1016.2015.02333](https://doi.org/10.11897/SP.J.1016.2015.02333)]
 - [21] Chu-Carroll J, Brown MK. Tracking initiative in collaborative dialogue interactions. In: Proc. of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conf. of the European Chapter of the Association for Computational Linguistics. Madrid: ACL, 1997. 262–270. [doi: [10.3115/976909.979651](https://doi.org/10.3115/976909.979651)]
 - [22] Levin E, Pieraccini R, Eckert W. Learning dialogue strategies within the Markov decision process framework. In: Proc. of the 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proc. Santa Barbara: IEEE, 1997. 72–79. [doi: [10.1109/ASRU.1997.658989](https://doi.org/10.1109/ASRU.1997.658989)]
 - [23] Yi X, Yu C, Shi YC. Bayesian method for intent prediction in pervasive computing environments. Scientia Sinica Informationis, 2018, 48(4): 419–432 (in Chinese with English abstract). [doi: [10.1360/N112017-00228](https://doi.org/10.1360/N112017-00228)]
 - [24] Bai HY, Cai SJ, Ye N, Hsu D, Lee WS. Intention-aware online POMDP planning for autonomous driving in a crowd. In: Proc. of the 2015 IEEE Int'l Conf. on Robotics and Automation. Seattle: IEEE, 2015. 454–460. [doi: [10.1109/ICRA.2015.7139219](https://doi.org/10.1109/ICRA.2015.7139219)]
 - [25] Nguyen THD, Hsu D, Lee WS, Leong TY, Kaelbling LP, Lozano-Perez T, Grant AH. CAPIR: Collaborative action planning with intention recognition. In: Proc. of the 7th AAAI Conf. on Artificial Intelligence and Interactive Digital Entertainment. Stanford: AAAI Press, 2011. 61–66.
 - [26] Wu T, Wang CJ, Xie JY. Research on cyber attack intention recognition based on partially observable Markov decision process. Journal of Nanjing University (Natural Sciences), 2010, 46(2): 122–130 (in Chinese with English abstract).
 - [27] Dong D, Feng ZQ, Tian JL. Smart beaker based on multimodal fusion and intentional understanding. In: Proc. of the 6th Int'l Conf. on Computing and Data Engineering. Sanya: ACM, 2020. 206–211. [doi: [10.1145/3379247.3379257](https://doi.org/10.1145/3379247.3379257)]
 - [28] Young S, Gašić M, Thomson B, Williams JD. POMDP-based statistical spoken dialog systems: A review. Proc. of the IEEE, 2013, 101(5): 1160–1179. [doi: [10.1109/JPROC.2012.2225812](https://doi.org/10.1109/JPROC.2012.2225812)]
 - [29] Pineau J, Gordon G, Thrun S. Point-based value iteration: An anytime algorithm for POMDPs. In: Proc. of the 18th Int'l Joint Conf. on Artificial intelligence. Acapulco: Morgan Kaufmann Publishers, Inc., 2003. 1025–1030.
 - [30] Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Usability Evaluation in Industry. London: CRC Press, 1996: 189–194. [doi: [10.1201/9781498710411](https://doi.org/10.1201/9781498710411)]

附中文参考文献:

- [1] 田丰. Post-WIMP软件界面研究 [博士学位论文]. 北京: 中国科学院研究生院, 2003.
- [2] 戴国忠, 田丰. 笔式用户界面. 第2版, 合肥: 中国科学技术大学出版社, 2014.
- [3] 全青山, 张宗琦, 黄进, 田丰, 刘杰, 戴国忠. 基于混合输入的场景设计工具. 软件学报, 2019, 30: 48–61. <http://www.jos.org.cn/1000-9825/19017.htm>
- [13] 田丰, 牟书, 戴国忠, 王宏安. Post-WIMP环境下笔式交互范式的研究. 计算机学报, 2004, 27(7): 977–984. [doi: [10.3321/j.issn:0254-4164.2004.07.015](https://doi.org/10.3321/j.issn:0254-4164.2004.07.015)]
- [16] 刘杰, 黄进, 田丰, 胡伟平, 戴国忠, 王宏安. 连续交互空间下的混合手势交互模型. 软件学报, 2017, 28(8): 2080–2095. <http://www.jos.org.cn/1000-9825/5123.htm> [doi: [10.13328/j.cnki.jos.005123](https://doi.org/10.13328/j.cnki.jos.005123)]
- [20] 俞凯, 陈露, 陈博, 孙锴, 朱苏. 任务型人机对话系统中的认知技术——概念、进展及其未来. 计算机学报, 2015, 38(12): 2333–2348. [doi: [10.11897/SP.J.1016.2015.02333](https://doi.org/10.11897/SP.J.1016.2015.02333)]
- [23] 易鑫, 喻纯, 史元春. 普适计算环境中用户意图推理的Bayes方法. 中国科学: 信息科学, 2018, 48(4): 419–432. [doi: [10.1360/N112017-00228](https://doi.org/10.1360/N112017-00228)]
- [26] 吴涛, 王崇骏, 谢俊元. 基于部分可观测马尔可夫决策过程的网络入侵意图识别研究. 南京大学学报(自然科学版), 2010, 46(2): 122–130.



仝青山(1983—), 男, 博士, 讲师, CCF 专业会员, 主要研究领域为人机交互, 软件工程.



黄进(1985—), 男, 博士, 副研究员, CCF 专业会员, 主要研究领域为人机交互, 图形图像处理.



康文惠(1993—), 男, 博士生, CCF 学生会员, 主要研究领域为人机交互, 机器学习.



田丰(1976—), 男, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为人机交互, 虚拟现实.



付强(1991—), 男, 博士生, 主要研究领域为人机交互, 计算机仿真.



戴国忠(1944—), 男, 研究员, 博士生导师, CCF 杰出会员, 主要研究领域为人机交互, 计算机图形学.