

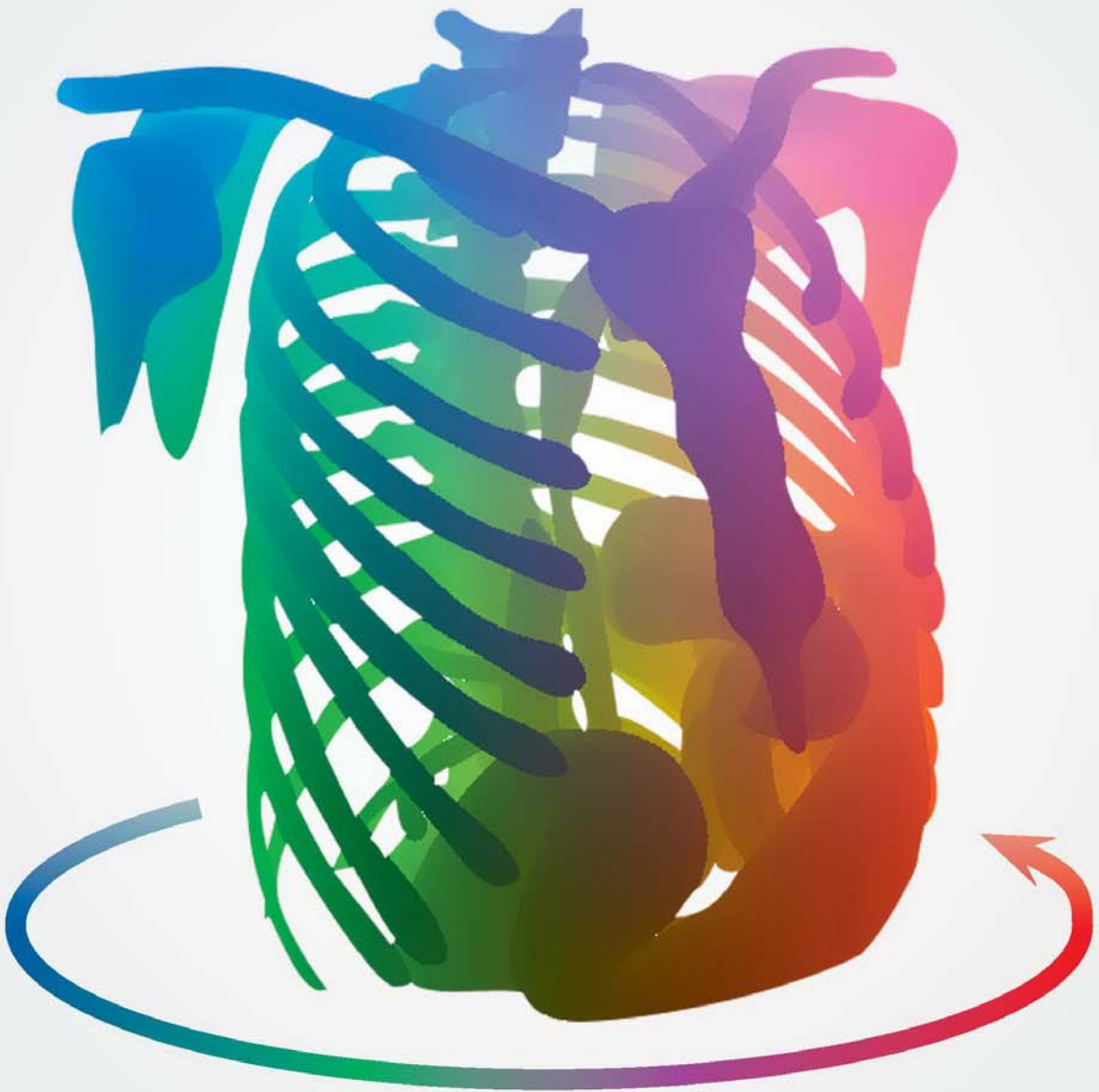


主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
02
VOL.20

ISSN1006-8961
CN11-3758/TB



等值面光线跟踪 P193



第20卷第2期（总第226期）

2015年2月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@radi.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 50.00元



多尺度活动网格在云场景仿真中的应用(第0202页)



自然环境视频中基于显著鲁棒轨迹的行为识别(第0245页)



虚拟手抓持力觉生成算法真实性的评价(第0280页)

图像处理和编码

引入神经网络中间神经元的快速小波图像压缩

张海涛, 张永霖 0159

拟正态分布扩散的图像平滑

周先春, 汪美玲, 周林锋 0169

可选特征的快速分形图像编码

袁宗文, 鲁业频, 杨汉生 0177

图像分析和识别

融合多尺度码本的全局编码图像分类

董振宇, 赵杰煜, 祝军 0183

计算机图形学

结合K-D树和Shell的快速动态等值面光线跟踪法

罗月童, 石放放, 张伟, 朱会国 0193

多尺度活动网格在云场景仿真中的应用

范晓磊, 张立民, 张建廷, 徐涛 0202

第十届和谐人机环境联合会议专栏

低资源语言的无监督语音关键词检测技术综述

杨鹏, 谢磊, 张艳宁 0211

快速离散Curvelet变换域的图像融合

杨勇, 董松, 黄淑英 0219

高效视频编码中变换跳过模式的快速选择

王宁, 张永飞, 樊锐 0229

主动学习的多标签图像在线分类

徐美香, 孙福明, 李豪杰 0237

自然环境视频中基于显著鲁棒轨迹的行为识别

易云, 王瀚漓 0245

特征变换和数据集分块的行人比对

沈洋, 林巍晓, 殷惠清 0254

移动手持环境下触摸式目标选择技术对比

辛义忠, 李岩, 袁伟强, 郑谦 0264

虚拟环境的用户意图捕获

程成, 赵东坡, 卢保安 0271

虚拟手抓持力觉生成算法真实性的评价

杨文珍, 许艳, 颜传武, 吴新丽, 张昊, 孟闯 0280

交互式乐器演奏的六自由度力觉渲染方法

童号, 史有姣, 王党校, 张玉茹 0288

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-scale moving grids and its application in clouds scene simulation(P0202)



Action recognition from unconstrained videos via salient and robust trajectory (P0245)



Validation the Authenticity of Virtual fingers Grasping Haptic Algorithm(P0280)

Image Processing and Coding

- Fast wavelet image compression introducing neural network of relay neuron
Zhang Haitao, Zhang Yonglin 0159
- Image smoothing algorithm based on matching normal distribution diffusion
Zhou Xianchun, Wang Meiling, Zhou Linfeng 0169
- Optional features fast fractal image coding algorithm
Yuan Zongwen, Lu Yepin, Yang Hansheng 0177

Image Analysis and Recognition

- Image classification based on global coding combined with multi-scale codebook
Dong Zhenyu, Zhao Jieyu, Zhu Jun 0183

Computer Graphics

- Fast isosurface ray tracing method by combining K-D tree and Shell
Luo Yuetong, Shi Fangfang, Zhang Wei, Zhu Huiguo 0193
- Multi-scale moving grids and its application in clouds scene simulation
Fan Xiaolei, Zhang Limin, Zhang Jianting, Xu Tao 0202

Column of HHME'2014

- Survey on unsupervised spoken term detection for low-resource languages
Yang Peng, Xie Lei, Zhang Yanning 0211
- Image fusion based on fast discrete Curvelet transform
Yang Yong, Tong Song, Huang Shuying 0219
- Fast transform skip mode decision for high efficiency video coding
Wang Ning, Zhang Yongfei, Fan Rui 0229
- Online multi-label image classification with active learning
Xu Meixiang, Sun Fuming, Li haojie 0237
- Action recognition from unconstrained videos via salient and robust trajectory
Yi Yun, Wang Hanli 0245
- Pedestrian re-identification based on feature transformation and dataset classification
Shen Yang, Lin Weiyao, Yin Huiqing 0254
- Comparison of target selection techniques in mobile handheld environment
Xin Yizhong, Li Yan, Yuan Weiqiang, Zheng Qian 0264
- Intent understanding for virtual environments
Cheng Cheng, Zhao Dongpo, Lu Baoan 0271
- Validation the authenticity of virtual fingers grasping haptic algorithm
Yang Wenzhen, Xu Yan, Yan Chuanwu, Wu Xinli, Zhang Hao, Meng Chuang 0280
- Interactive simulation of music instrument playing by a six degree-of-freedom haptic rendering approach
Tong Hao, Shi Youjiao, Wang Dangxiao, Zhang Yuru 0288

中图法分类号: TP399 文献标识码: A 文章编号: 1006-8961(2015)02-0271-09

论文引用格式: Cheng C, Zhao D P, Lu B A. Intent understanding for virtual environments[J]. Journal of Image and Graphics, 2015, 20(2): 0271-0279.

[程成, 赵东坡, 卢保安. 虚拟环境的用户意图捕获[J]. 中国图象图形学报, 2015, 20(2): 0271-0279.][DOI:10. 11834/jig. 20150214]

虚拟环境的用户意图捕获

程成, 赵东坡, 卢保安

北京理工大学计算机学院智能信息技术北京市重点实验室, 北京 100081

摘要: **目的** 虚拟制造环境中需要复杂精确的 3D 人机交互。目前的虚拟环境(VE)的主要问题是人在交互过程中的认知和操作负荷太重,交互效率亟需提高。解决此问题的重要途径是提高机器的认知能力。**方法** 本文研究了用户意图的分析和抽取,并建立多通道用户意图理解的算法,以此来提高交互效率。**结果** 结合虚拟装配应用给出了典型意图的实验结果并给予分析。通过实验对多通道意图的可用性和可靠性,以及基于意图系统的实时性进行了评估。实验是虚拟装配空间中用户拾取对象意图的实验。当 3 维鼠标和对象距离为 5 000 mm 时,传统系统中操作平均耗时 5.344 7 s,而基于意图的系统中平均耗时 2.326 6 s。基于意图的系统极大地降低了操作的时间和复杂度。**结论** 采用意图驱动的系统情景转换能在虚拟环境工作中有效地降低人的认知负荷,并能很好地帮助系统开发者进行混成系统的建模和分析,降低开发的复杂度。实践结果表明用户意图理解的多通道模型和算法能极大地提高交互式系统的交互自然性和交互效率。该方法不仅适用于本文所用的虚拟装配系统,对于所有的虚拟环境应用场景都有同样的有效性。

关键词: 意图;人机交互;虚拟环境;眼动跟踪;多通道

Intent understanding for virtual environments

Cheng Cheng, Zhao Dongpo, Lu Baoan

Department of Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China

Abstract: **Objective** Virtual manufacturing environments require complex and accurate three-dimensional (3D) human-computer interaction (HCI). The main problem of current virtual environments (VEs) is the heavy user burden associated with the cognitive and motor operation aspects, as well as the improvement in HCI efficiency. This problem is solved by promoting the cognitive capability of the machine. This study investigates how user intents are analyzed and abstracted, as well as constructs multimodal intent understanding algorithms. **Method** Intent-based VE construction is practiced in a virtual assembly system. Experiments on typical intents are conducted and analyzed. A comprehensive evaluation of the usability and reliability of multimodal intent understanding is presented, and the intent-based VE system is demonstrated to be a real-time system. The experiment focuses on the intent of object picking in VE. **Result** When the distance between the 3D cursor and object is 5 000 mm, the operation costs 5.344 7 s on average in traditional systems, whereas it costs 2.326 6 s on average in intent-based systems. The intent-based system significantly reduces operation time and manipulation complexity. **Conclusion** Intent-driven scenario transition can significantly enhance the naturalness and efficiency of HCI, as well as effectively reduce the complexity of human-centered VE system analysis and development. Application of intent understanding

收稿日期:2014-07-23;修回日期:2014-11-21

基金项目:国家自然科学基金项目(61370135)

第一作者简介:程成(1966—),男,副教授,2003 年于中国科学院研究生院获计算机应用技术工学博士学位,主要研究方向为人机交互、虚拟环境。E-mail: guoguocheng@vip.sina.com

Supported by: National Natural Science Foundation of China (61370135)

demonstrates that multimodal intent models and algorithms can efficiently promote the naturalness and efficiency of HCI.

This system construction method can be used in any VE system.

Key words: intent; human-computer interaction; virtual environment; eye movement tracking; multimodal

0 引言

3 维人机交互是虚拟环境的一个重要特性。虚拟制造环境尤其需要复杂而精确的 3 维直接操作这种主要人机交互形式。直接操作是指用户利用数据手套、3 维鼠标等交互设备通过 3D 光标或虚拟手等用户隐喻在虚拟空间中操纵虚拟对象的拟实交互方式。与此同时,包括眼动在内的各交互通道的加入,使得虚拟环境参与者更高效更自然地与虚拟对象进行交互充满可能。当前的虚拟环境应用存在的一个最为重要的问题是,虽然可以建立多种交互通道进行交互,但是由于复杂的虚拟场景和大量的虚拟对象,用户在交互过程中利用比较单一的传统交互通道如 3 维鼠标、虚拟手套、键盘等形式,造成用户的认知和操作负荷过重,从而系统处理混成模糊的直接操作数据和多通道融合数据的能力严重不足。

以人为中心的软件构造是解决上述问题的根本途径^[1]。若要让虚拟环境人机交互更加自然高效和谐,就需要平衡计算机和人的认知能力,让计算机具有更强的感知和认知能力^[2]。其中一个最为关键的方面就是让计算机从多通道的混成输入中获取用户的交互意图^[3-5];另一个重要的方面是需要研究新的软件模型,突破先前的状态机模型的局限,实现一种意图驱动的交互式软件构造^[6-7]。目前有关意图的研究还比较少,本文试图首先给出此类研究的若干线索和简要的回顾。

在多通道研究方面,已有基于数据手套的手势和基于图像处理的裸手运动的手势分析的文献^[8-9]。在虚拟环境中虚拟手能极大地增强参与者的沉浸感和交互的真实性。虚拟手被用来表达各种手势以传达用户的交互意图,另外,直接用虚拟手去抓握诸如机械零件等虚拟对象,实现装配等精确操作。从 2003 年以来人们对眼动的研究兴趣逐渐高涨,出现了众多的在眼动相关方面的研究工作,尤其是将眼动,眼睛注视和扫视,作为人机交互手段的研究更令学者关注^[10-11]。但是在眼手协调和多通道融合方面的研究还没有见到。

在用户模型研究方面,经验看来它是未来以用户为中心的软件构造的基础设施和知识载体。Frank^[12]对经验知识的获取进行了较为深入的研究。他将经验所存在的上下文合理地分为外部上下文和内心上下文两个重要组成部分。对领域知识的形式化方法也是十分重要的,文献^[13]研究了领域知识的提取策略并在规划领域进行了尝试。文献^[14]研究了背景知识和复杂事件处理的融合,从而提高基于知识的事件处理系统的性能和可扩展性。

在人机交互的机器推理研究方面,Cowley^[15]认为对于意图推理研究,需要研究人基本的能力。他提出了一个对具体意识的心理学测试工具和方法,借此测试诸如在人类推理当中的意图作用。作者认为感觉能力较之知觉能力是更为重要的。就当前在机器智能和机器人方面的实践来看,机器算法可以感知外部环境并有相应的思考,不能仅仅是具有对场景中事件的反应,而是需要构建用户的认知模型,具有一种意图推理的能力。文献^[16]给出了一种结合情景演算和概率理论的逻辑框架去识别智能体意图的方法。此方法可以对有限数量行为情景中智能体的意图给出推断。

文献^[17]研究了 Agent 的意图模型和意图推理问题。Agent 的意图可以看成是对人的意图的形式描述。Agent 的意图模型和意图推理可以看成是对人的认知过程的一个建模。这是在意图推理方面的一个有益尝试,但在推理算法实时性方面还没有明确的实验结果。文献^[18]建立了意图的语义表示,研究了意图的合理性必要条件,提出了一些简明有效的约束条件来保证意图模型的可达性和非平凡性等现实要求。通过意图模态算子对经典的正规模态算子可能世界语义加以改善,是意图形式化的必要工作。

无疑一阶逻辑和模态逻辑是人机交互中的意图表达和推理方面的重要的理论基础和研究工具。唐稚松独创地构造了一种可执行时序逻辑 XYZ^[19],将逻辑的静态描述方面和动态执行机制有机地结合在一个系统中,在人机交互和动画等方面都展现了强大的应用能力。有关一阶逻辑的推理技术最近的

研究还可以参见文献[20],有关一阶模态逻辑的推理可以参见文献[21]。

1 虚拟环境中的意图表达

以人为中心的设计,包含非常多人的因素,需要建立在认知心理学理论的基础上。另外,由于虚拟环境有多种沉浸感水平,在不同水平的虚拟环境中用户所采用的交互设备是不一样的,因而对所表达内容的表达方式是不同的,也不同于现实世界中的人人交互模式,在以人为中心的研究中,又必须结合具体情况,研究方法上遵循从特殊到普遍的技术路线。以人为中心的设计和开发是解决此问题的根本途径。也即对人的意图、感知和交互行为模式等方面的观察和建模。

定义1 意图 意图是虚拟环境用户在交互活动中希望虚拟环境(VE)短时达到某一状态的心理活动,它决定于用户的当前心智状态和当前的环境情景,并通过多交互通道时序的混成事件来表达。

感知是虚拟对象在活动期内的各种实时性计算及其反馈表象。虚拟对象通过对场景中对象间时空关联的判定捕获场景中特定的动态特性或与其他对象间的复杂关系。以虚拟装配应用为例,在交互中需要及时判断零件对象之间的相对位置关系和空间方向一致性关系等。另一个需要建立的概念是:交互模式是对用户在特定的虚拟环境中完成特定任务过程中操作行为的时空抽象。主要特征是多种通道信息的混成和交互行为的时间序列。交互模式反映的是用户操作的共性。特有的交互模式本质上是产生于特有的与领域相关的交互需求。捕获交互模式的目的是获取其中蕴含的交互语义信息。通过模式可以表达如并行操作、紧随操作等特定交互语义内容。用户通过特定的交互手段在虚拟环境中完成一项任务时都会包含一个或多个交互模式。如在第4节实验中,用眼动注视点来选取场景中的物体,然后结合3维鼠标对选定物体进行变换、装配等操作。眼动通道和传统交互通道相结合的多通道融合交互系统,能够让用户以更简单自然的方式交互,提高交互效率。

在虚拟装配大流程中包含各个活动情景,每个活动中又包含一个任务序列,事实上意图在用户发生思维跳跃时系统将脱离原有任务序列,转入一个

新的活动情景。一个工作流程可能是个重复性的过程,它可以通过意图和交互模式加以形式化。关于虚拟装配的一个一般化的状态变迁如图1所示。椭圆节点表示一个意图,圆角长方形节点表示交互模式,六角形节点表示虚拟环境的感知及其反馈。图1给出了一个活动情景的结构和情景变迁,循环表示在像虚拟装配这样的交互过程中重复性任务的存在,分支描述用户思维的跳跃既意图驱动的变迁。对于装配流程可以如图1所示将其形式化为每个活动包含4个意图驱动的交互任务序列。

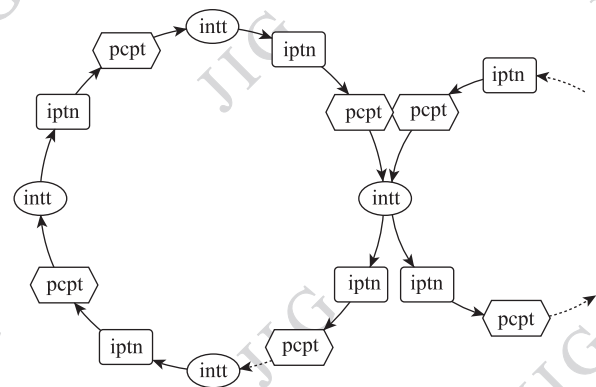


图1 意图驱动状态转换图

Fig. 1 Intent-driven state transition diagram

虽然意图非常多,本文给出最紧要和最典型的交互,说明意图和意图序列的表达,描述,捕获和实时推理。对于虚拟装配这样的复杂系统,涉及的用户有很多类型,有机械设计人员、装配工艺人员、接受装配培训的生产和维修人员,也有虚拟环境系统的设计开发人员以及研究制造集成的科研人员等。通过对众多机械设计人员和研究员以及有相关方面经历的虚拟环境研究人员的调查,所获得虚拟装配系统中用户意图及分类如表1所示。

表1中第1列是在虚拟装配应用中的所有典型用户意图。意图的含义基本可以从名字中反映出来。例如“装配意图”是描述用户在虚拟装配流程执行中试图进行零件的装配,表现在对零件的抓取后移动并接近装配体,它将由接近感知所体现并产生零件与装配体之间的聚合关系;“特征匹配”意图是描述用户在虚拟装配特定环境中试图利用装配特征进行精确定位,表现在对零件接近感知后的平移操作搜索特征对的匹配,它将由特征匹配感知所体现并产生零件间的约束依赖关系;其他的一些意图语义比较直接,此处不再赘述。

表 1 意图分类
Table 1 Intent category

用户意图	类				
	领域专门	VE 专门	过程转换	活动转换	状态转换
装配	√		√		
特征匹配		√			√
对齐		√			√
面贴合		√			√
并发装配	√		√		
特殊装配方式	√	√			√
拆卸	√		√		
屏蔽约束		√			√
碰撞避免		√			√
拆卸部分装配件	√			√	
重装部分装配件	√			√	
运动仿真	√		√		
系统仿真	√		√		
再设计	√		√		
加工仿真	√		√		
有限元分析	√		√		

意图的有效性主要体现在两个方面,一是意图的可用性,二是意图对用户交互的支持程度。通过进行意图的实验可以比较基于事件的操作和基于意图操作之间的不同效果。

2 意图理解

意图和操作模式需要通过算法实时捕获,以便虚拟环境制导状态转换并推理特定对象的特定行为。意图将引发系统在流程中活动间的转换,也可能产生单个任务节点内的状态跳转。如何得到意图判断和推理的规则去捕获意图是需要解决的问题。这个问题不能从软件构造理论和工程学中寻找答案,而是需要借助认知心理学的研究方法和工具,从实验中分析人的思维和行为,抓住交互中典型的通道时序特征进行模型检测。基于实验获得判定准则,进而整理成为时序通道融合的规则。

2.1 用户意图的多通道表达

根据行为理论,人的认知和行为是关联的。在

认知活动中也配合有外在的行为表现,而外部行为也反映了人的内在的思维活动。用户的意图是通过其在虚拟环境中的交互行为所表达出来的。由于虚拟环境有不同的存在形态,所以在研究意图表达时必须结合具体情况具体分析。本文主要结合桌面虚拟环境系统中的情形,基于本文所构造的虚拟装配原型系统 Interaction3D 分析意图的表达。

Interaction3D 的虚拟空间中有一双虚拟手或 3D 光标与 3 维交互设备的力矩相连,虚拟手随着用户双手对 3 维鼠标的操作在空间中进行自由的 3 维运动,形成各种手势,感知虚拟对象的位姿并精确抓握和捏拿等。眼动通道对双摄像机获得的眼部图像进行立体匹配,识别人眼图像特征并获得瞳孔和反光点的位置,通过线性预测算法实时计算将瞳孔和反光点位置信息转换为人眼对屏幕的注视点,由此获得所注视虚拟环境场景中的虚拟对象或对象上的具体特征。图 2 为双手和眼动通道在虚拟装配中的意图表达示例。各个独立通道在任何一个特定时刻会处于一个特定状态。多个通道在同一时刻点上状态

的组合构成一个特定的状态向量。这里的时间点不是物理时间,而是逻辑时间点,是在虚拟环境系统中捕获的一些特殊事件的时刻。一组特定的连续时间点上的通道状态向量短序列包含用户的交互意图信息,连续的长序列构成一个特定的交互模式。根据笔者项目组的长期实践经验和观察,意图在连续的5个时间点内都能得到很好的表达,在本文后面部分将讨论不超过5个状态向量的意图分析。

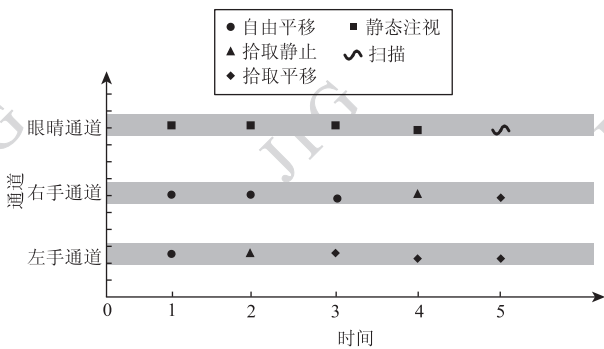


图2 多通道交互的时序描述

Fig. 2 Temporal illustration of multimodal interaction

定义2 特征向量 特征向量是描述虚拟环境某一时刻多维状态的向量。表达为 $\langle t, Sce, Obj_1, Obj_2, Obj_3, Tsk, Stt, Mod_1, Mod_2, Mod_3 \rangle$ 。其中 t 是当前时间, Sce 表示当前场景, Obj_i 表示交互中所涉及的虚拟对象, Tsk 说明用户当前的交互任务, Stt 表示系统当前状态, Mod_i 表示通道的状态。

通道有一个通道状态集。对于眼动通道,其状态主要分为:注视静物(GazeStat),注视动物(GazeDyn),跟随注视(Flw),扫视(Scan),左右手通道的状态包括:空手静止(FreStat),空手平移(FreTran),空手旋转(FreRot),指点手势(PntGst),状态切换(SwtTr),约束屏蔽(DisCn),抓取静止(GpSt),抓取平移(GpTran),抓取旋转(GpRot)。在每个时间点上,各通道都有自己特有的状态属性,在这些通道之间存在着重要的协作信息,比如手眼的协调操作,双手的协调操作等。在装配场景中双手协同装配操作。双手各拾取一个零件,移动到特定的装配空间并进行装配操作,用户的特征匹配意图包含的特征向量组如下所示:

$$\begin{aligned} &\langle t_1, Stk, _, _, o_1, Ass, Apr, FreTran, FreTran, Scan \rangle \\ &\langle t_2, Stk, o_1, _, o_1, Ass, Apr, PntGst, FreStat, GazeSt \rangle \\ &\langle t_3, Stk, o_1, o_2, o_2, Ass, Apr, GrspSt, PntGst, GazeSt \rangle \end{aligned}$$

$$\langle t_4, Asm, o_1, o_2, _, Ass, Apr, GrspSt, GrspSt, Flw \rangle$$

$$\langle t_5, Asm, o_1, o_2, _, Ass, Apr, GpTran, GpTran, Scan \rangle$$

(1)

一个困难的问题是何时捕获特征向量组或构造特征向量。从Norman循环中可以看到,所谓的意图表达,实际上一定是发生在虚拟对象感知表象事件之后。当虚拟对象感知表象出现且通道的状态改变被捕获到,就满足特征向量创建的条件,系统就可以创建一个新的特征向量序列。

2.2 基于时序逻辑的意图分析

结合应用领域知识和需求,更重要的是结合人的认知模式和习惯去构造软件。在每个交互关键环节中,首先根据心理学分析,抽取人在现实世界中完成操作任务过程中的认知要素,结合虚拟环境特殊场景和特定的交互设备,虚拟环境中交互任务完成的认知过程建模。进一步针对执行的典型交互任务分析它的具体特征向量短序列,建立若干推理规则,从交互前端特征向量短序列中分析出用户意图。建立一个观察点的概念,所谓观察点是在特征向量组中某些特征量之间的一种时空关系,是交互中的一个时序特征。

用线性时序逻辑工具对特征向量组进行描述和分析。时序逻辑的算子主要包括 O 、 $\square$ 、 $\diamond$ 、 U 、 W 、 $\odot$ 、 $\square$ 、 $\diamond$ 。其含义参见文献[19]。约定参考时刻总是所获得的有意义的特征向量组中中间一帧的时刻。通过实验,经过整理统计和分析抽象,最后给出各个观察点的时序逻辑表达。以下一些表达式都是来自Interaction3D一些观察点的时序表达实例。

一个观察点刻画了交互意图表达中的某个交互要素,任意一个意图是蕴含在多个这样的特性中的。意图捕获算法的具体描述不再列出。主要内容是捕获特征向量和对观察点集合的验证计算。

为配合此算法,针对眼动的灵活性问题采取了一些策略。首先排除用户发呆的情形,在此情形中,用户的手通道将是处于静止状态,因而虽然眼动可能是存在的,但算法不产生特征向量;另外一种需要考虑的情形是眼动的无意识扫描。有实验证明人眼在工作中实际上是在不断运动的,很多运动属于无意识的眼动。算法根据当前交互任务的内在需要,分析眼动的类型。眼动判定函数为

$$F(v, x, y) = (w_1 + w_2) \cdot ECC(v, x, y) = (w_1 + w_2) \cdot Vel(v) \cdot CT(Scope) \quad (2)$$

$$Vel(v) = \text{sig}[(v - \delta_{v1})(v - \delta_{v2})] \quad (3)$$

$$CT(Scope) = \text{sig}(Scope - d) + 2 \quad (4)$$

参数 w_1, w_2 代表左右手通道的活动标识, 如果左手通道有大于阈值的运动发生, 则置 w_1 为 0.5, 否则为 0, w_2 代表右手类似描述。当两只手都处于无操作状态时, 眼动将被视为无意识运动。函数 $ECC(v, x, y)$ 根据眼的注视点位置参数和位移速度判定眼动的各种情形。其中参数 v 是眼动速度, x, y 为眼注视在屏幕上的位置坐标。函数 $Vel()$ 检测眼睛视线在屏幕上视点的运动速度。如式(3)所示,

$$F(t, v, x, y) = \begin{cases} 0 & w_1 = 0, w_2 = 0 \text{ or } (x, y) \text{ is out of space} \\ 1 & w_1 = 0.5, w_2 = 0.5 \wedge v < \delta_{v1} \wedge Scope \leq d \\ -3 & w_1 = 0.5, w_2 = 0.5 \wedge \delta_{v1} \leq v < \delta_{v2} \wedge Scope \leq d \\ 3 & w_1 = 0.5, w_2 = 0.5 \wedge v > \delta_{v2} \wedge Scope > d \end{cases} \quad (5)$$

首先在构造意图系统时进行意图的二义性分析, 排除所有的意图不确定性。这是通过对每个状态的意图集合中的至少一个不同意图构造取值的观察点来达到的。由于对每个流程中的每个交互任务, 其场景中的状态数量是有限的。任何一个意图特征向量短序列中的观察点数也是有限的, 其由观察点组合构成的经验规则因而数量有限, 实际情况是, 针对每个系统状态, 所要检验的意图是很少的, 对这些模型检测一遍, 实践证明实时性是能得到保证的。

3 意图理解实验及分析

实验目的是验证虚拟环境中基于意图交互的可行性和可靠性, 以及对基于意图的多通道虚拟系统的效率评估。本文系统中有十几个交互意图, 考虑到论文篇幅所限, 也为了研究和叙述的方便, 在这里对一个典型的意图, 也就是对象拾取意图, 进行实验和分析。关于对象拾取意图的形成在论文原理部分有详细说明, 在此通过基于用户双目注视跟踪的意图捕获过程, 来说明基于意图的多通道虚拟交互系统。此外和传统的基于 3 维鼠标的虚拟交互系统, 在准确性和操作时间的损耗上进行对比。

3.1 实验参与者

实验参与者是 20 名在校大学生, 其中 16 名男

函数 $Vel()$ 判定眼动扫描速度是否在两个阈值 δ_{v1}, δ_{v2} 的范围内, $\delta_{v1} < \delta_{v2}$, 当眼睛处于静物注视状态时, 其扫描速度会小于 δ_{v1} , 当用户在做跟随注视时其眼动速度会介于 δ_{v1}, δ_{v2} 两者之间, 而当眼睛处于扫视状态时其速度会大于 δ_{v2} 。

符号函数 $\text{sig}()$ 给出是否在阈值间的判定。函数 $CT()$ 是对眼注视虚拟对象范围的一个判断, 函数分析所观察到的眼睛在空间中运动范围的特性给出眼动的分类。参数 $Scope$ 是注视点间的位移量, d 为一个阈值。因而, 眼动判定函数可以很好地将无意识眼动、静物注视、跟随注视和扫视很好地区分出来, 即

生, 4 名女生。所有的参与者都是用右手完成实验操作。将 20 名参与者自然分成两个实验小组。每个小组每名参与者需要做两组实验, 在此之前参与者有 20 min 学习的时间。

3.2 实验装置

实验使用 Intel(R) Xeon(R) CPU 的个人台式机, 内存(RAM)是 4.0 GB, 32 位 Windows 7 操作系统。3 维鼠标是 3Dconnection 的 SpaceMouse Plus, 3 维鼠标提供虚拟场景中 6 自由度的 3D 交互, 结合若干个按键, 实现 3 维场景的漫游和仿真物体的控制。双目的屏幕注视点用桌面系统的眼动跟踪仪 EyeTribe 获取。EyeTribe 放置在显示屏下方, 通过 USB3.0 接口连接到计算机, 每次使用前需要一个简单的标定过程, 实时捕获用户双目在屏幕上的注视点。利用上述硬件, 在 Open Inventor5.0 的平台上搭建虚拟装配系统 Interaction3D, 该系统是虚拟装配的交互式仿真系统, 实现 3 维空间中零件模型的虚拟装配、装配仿真、意图捕获等交互功能。Interaction3D 系统是用 C++ 实现, 运行环境是 Visual C++ 6.0。

3.3 实验设计

实验目标是检验基于普通的 PC 和桌面交互设备虚拟装配环境中用户意图的捕获, 并比较有意意图系统和传统无意图系统的性能。在实验中用户通过直接操作和眼动两种自然方式表达交互意图, 系统

自动产生各种复杂交互语义,如图3所示。



图3 虚拟装配系统操控

Fig. 3 Interface for Interaction3D

在虚拟装配场景中,摆放着若干零件对象,一个代表3维鼠标控制的3维光标以及一个代表眼睛注视点的2维光标。在桌面上空间中设置两条蓝色标志线,相距5 000 mm或3 000 mm,初始时刻将待拾取零件摆放在远端水平线位置,将3维光标摆放在近端水平线位置,如图4所示。3维光标将跟随用户3维鼠标操作在空间中移动。注视点2维光标反映眼动视线来自眼动跟踪装置的用户。该实验中参与者主要的任务是在虚拟装配环境中完成零件拾取操作。

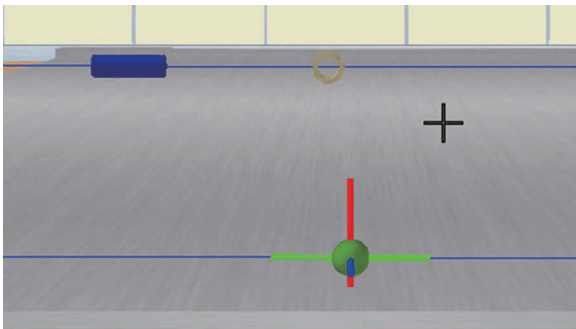


图4 虚拟环境中意图实验情景设计

Fig. 4 Intent experiment design for VE

实验分为两组。一组是参与者仅使用3维鼠标操作完成零件的拾取任务。在仅使用3维鼠标作为拾取手段的实验中,用户推动3维鼠标扭矩盖,移动3维光标到零件的某个特定距离范围内,系统感知到用户的拾取意图,点亮零件包围盒,如图5所示。当用户按下鼠标键确认后,则3维光标绑定零件,如图6所示。另一组是参与者同时使用眼动和3维鼠标完成拾取任务。用户用眼动注视光标指点选取零件对象,当注视坐标落入某个零件对象的临界区域

内,并保持某一 Δt 时间间隔,则系统感知到该零件被注视,将该零件的包围盒点亮。此时若用户按下3维鼠标上的某一确认键后,系统获得一个拾取意图,3维鼠标自动绑定该零件,此时可以通过控制3维鼠标来实现对被拾取零件的移动、旋转、装配等操作。若零件被用户成功拾取,则认为用户意图被准确捕获。

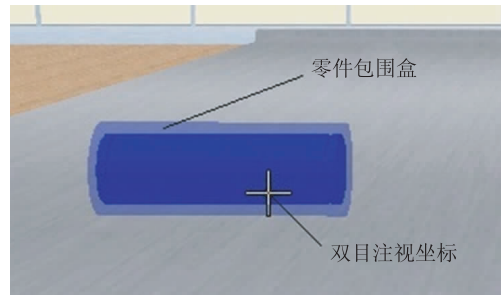


图5 双目注视选取零件对象

Fig. 5 Eye gaze picking a part object

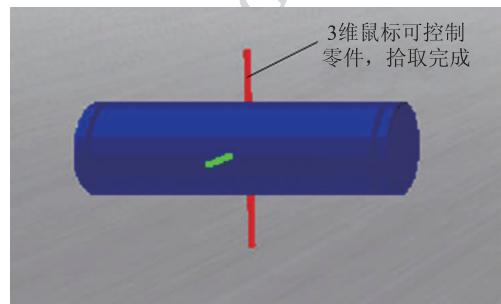


图6 零件对象被拾取

Fig. 6 The condition of an object is picked

实验前对参与者讲解清楚虚拟装配场景和任务目标,交互设备的使用方法和眼动捕捉对参与者的基本要求,给参与者同样的时间熟悉系统和操作。由实验设计者观察每一组中每一位参与者的实验过程,记录操作时间和成功次数。统计每组的平均操作和相关参数。

3.4 实验过程

在第1组实验中,指定拾取的零件和3维鼠标的距离设定为3 000 mm,第1组10名实验参与者用3Dconnexion SpaceMouse控制3维光标在虚拟场景中移动至被拾取的零件附近,若零件包围盒点亮,此时可按下3Dconnexion SpaceMouse上的确认键2,表示拾取完成。计时从按下1键开始,到按下确认键2结束。每个参与者完成10次实验,记录各项实验

数据。然后再将虚拟场景中 3 维鼠标距离目标零件的初始值设定为 5 000 mm,重复上述实验过程,记录各项实验数据。

在第 2 组实验中,10 名参与者需要在相同的实验场景中完成拾取操作,增加眼动注视选取目标零件,当 2 维眼动光标落在零件图像识别圈内零件包围盒点亮,此时参与者可按下 3Dconnexion Space-Mouse 上的确认键,表示拾取该对象,实验结束。每个参与者做 10 次拾取操作,记录实验结果数据。然后再将虚拟场景中 3 维鼠标距离目标零件的初始值设定为 5 000 mm,重复上述实验过程,记录各项实验数据。

在整个实验过程中,系统会自动保存计时值,实

验人员自始至终监督参与者的实验操作,记录眼动注视选取指定零件的准确性和错误率。统计两个小组的实验数据均值作为实验结果。

3.5 实验结果分析

单独使用手动交互设备的操作是当前典型传统系统的事件驱动工作方式,本文实现的包括眼动的多通道系统是意图驱动的新工作方式。通过一组性能量度对此两种系统进行实验比较。RT 量度是 Real-time 的缩写,ACCY 是 Accuracy 即操作精确度的缩写,Op-Cpx 是 Operation complexity 即操作复杂度的缩写,SM 是 Semantics 即语义的缩写,而 Op-time 是 Operation time 的缩写。实验结果如表 2 所示。

表 2 实验数据
Table 2 Experiment data

系统	RT	ACCY/%	Op-Cpx/个	SM/个	Op-time	
					距离/mm	平均耗时/s
传统系统	Yes	100	5	1	3 000	3.980
					5 000	5.345
意图系统	Yes	100	1.5	1	3 000	2.267
					5 000	2.327

两种系统的实时性都很好,眼动设备的实时性和稳定性也确保了在拾取意图中系统的性能。可以从表中前两项结果中看出。操作复杂度表现在用户在拾取操作中的深度认知和控制难度上。数据表示在操作中平均切换 3 维鼠标运动的方向次数。传统系统中此数据较大,基于意图的系统中此数得到极大降低。语义数据这里都是 1,因为该实验只有一个基本的拾取任务。

可以看出在 3 维鼠标距离被拾取零件距离设定为 3 000 mm 的情况下,基于 3 维鼠标的传统虚拟装配系统将 3 维鼠标移动到特定零件附近的平均耗时为 3.982 0 s,而在基于意图的多通道虚拟装配系统完成拾取操作平均耗时为 2.267 s;当距离为 5 000 mm 时,传统系统中操作平均耗时 5.344 7 s,而基于意图的系统中它的平均耗时为 2.326 6 s。此组数据说明眼动意图获取较之手动事件处理的拾取任务有明显的省时效果,且意图获取不再与距离远近相关。

实验结果表明,基于意图的多通道虚拟装配系统能极大地提高系统的操作效率。就单个任务实验看,眼动通道的应用已经有良好的可用性。

4 结 论

3D 交互中,用户意图理解是以人为中心的软件构造的重要体现。实验结果表明,捕获用户意图能极大地降低虚拟环境用户的认知和操作负荷。结合对人机交互要求极高的虚拟装配应用,实现一种基于用户意图的软件构造。在多通道建设方面,在意图捕获方法和交互模式抽象方面,以及在基于时序逻辑的意图推理方面进行了一些尝试,实践证明在虚拟制造环境中理解人的意图,采用意图驱动的系统情景转换至少在两个方面有重要意义。一是在虚拟环境工作方面能有效地降低人的认知负荷,二是在虚拟环境制造的软件和系统构造方面能很好地帮助开发者进行混成系统的建模和分析,降低开发的

复杂度。目前有关意图的捕获和基于意图的软件构造的研究论文还非常少,如何在性能上给出一套系统的科学的评估量度还是有待进一步研究的方面。另外,在虚拟制造中人的认知规律和认知能力的提升方面应该有更多的认知心理学和机械设计及计算技术方面的学科融合。此类系统研究的关键问题是用户模型和用户的操作模式的认知心理学实验研究的介入和加强。意图表达的准确性是需要深入研究和分析的,有哪些因素影响意图的准确性,如何改善意图的准确性都是下一步作者准备做的工作。

参考文献 (References)

- [1] Ford N. Cognitive styles and virtual environments[J]. Journal of the American Society for Information Science, 2000, 51(6): 543-547. [DOI: 10.1002/(SICI) 1097-4571 (2000) 51: 6 <543::AID-ASIS>3.0.CO;2-S]
- [2] Dey A K. Modeling and adding intelligibility to human activity [C] //International Symposium on ubiquitous virtual reality. Gwangju, South Korea: IEEE, 2008; 5-8. [DOI: 10.1109/ISUVR.2008.25]
- [3] Bianchi-Berthouze N, Lisetti C L. Modeling multimodal expression of user's affective subjective experience[J]. User Modeling and User-adapted Interaction, 2002, 12(1): 49-84. [DOI: 10.1023/A:1013365332180]
- [4] Keates S, Langdon P, Clarkson P J, et al. User models and user physical capability[J]. User Modeling and User-adapted Interaction, 2002, 12(2-3): 139-169. [DOI: 10.1023/A:1015047002796]
- [5] Dumas B, Beat S, Denis L. Fusion in multimodal interactive systems: an HMM-based algorithm for user-induced adaptation[C] // The 2012 ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Copenhagen, Denmark; 2012, 15-24. [DOI: 10.1145/2305484.2305490]
- [6] Janse B, Belpaeme T. A model for inferring the intention in imitation tasks[C] //The 15th IEEE International Symposium on Robot and Human Interactive Communication. Piscataway, NJ, USA: IEEE, 2006; 6 pp.
- [7] Varol H A, Sup F, Goldfarb M. Multiclass real-time intent recognition of a powered lower limb prosthesis[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(3): 542-551. [DOI: 10.1109/TBME.2009.2034734]
- [8] Wan H, Gao S, Peng Q. Virtual grasping for virtual assembly tasks[C] //Proceedings of the 3rd International Conference on Image and Graphics. Los Alamitos, CA, USA: IEEE Computer Soc., 2004: 448-451.
- [9] Yi B, Harris F C, Ling W, et al. Real-time natural hand gestures[J]. Computing in Science & Engineering, 2005, 7(3): 92-96. [DOI: 10.1109/MCSE.2005.58]
- [10] Zhu Z, Qiang J. Eye gaze tracking under natural head movements [C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA: IEEE, 2005, 918-923. [DOI: 10.1109/CVPR.2005.148]
- [11] Kumar M, Klingner J, Puranik R, et al. Improving the accuracy of gaze input for interaction[C] //Proceedings of the Eye Tracking Research & Application. New York, USA: ACM, 2008, 65-68. [DOI: 10.1145/1344471.1344488]
- [12] Nack F. Capturing experience-a matter of contextualising events [C] //Proceedings of the ACM SIGMM Workshop on Experiential Telepresence. New York: Association for Computing Machinery, 2003: 53-64. [DOI: 10.1145/982484.982492]
- [13] Wu X, Jiang Y, Ling Y. Strategy of extracting domain knowledge for STRIPS world [J]. Journal of Software, 2007, 18(3): 490-504. [DOI: 10.1360/jos180490]
- [14] Teymourian K, Rohde M, Paschke A. Fusion of background knowledge and streams of events[C] //Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems. New York: Association for Computing Machinery, 2012; 302-313. [DOI: 10.1145/2335484.2335517]
- [15] Cowley M. The relevance of intent to human-android strategic interaction and artificial consciousness[C] //The 15th IEEE International Symposium on Robot and Human Interactive Communication. Piscataway: IEEE, 2006, 6 pp.
- [16] Demolombe R, Otermin Fernandez A M. Recognition of the intention to perform a procedure: a method based on probabilities [C] // Proceedings of the International Conference on Autonomous Agents. New York: Association for Computing Machinery, 2005; 1247-1248.
- [17] Zhu Z, Chen X, Zhou Y. A new formal system of intention consequences lm5c[J]. Journal of Software, 2002, 13(7): 1271-1277.
- [18] Hu S, Shi C. An improved twin-subset semantic model for intention of agent[J]. Journal of Software, 2006, 17(3): 396-402. [DOI: 10.1360/jos170396]
- [19] Tang Z S. Temporal Logic Programming and Software Engineering [M]. Beijing: Science Press, 1999. [唐稚松. 时序逻辑程序设计 with 软件工程[M]. 北京: 科学出版社, 1999.]
- [20] Xu G, Zhang J. First-order logic reasoning support for the semantic web[J]. Journal of Software, 2008, 19(12): 3091-3099. [DOI: 10.3724/SP.J.1001.2008.03091]
- [21] Zhang X, Zhang Z, Sui Yu, et al. Fuzzy reasoning based on first-order modal logic[J]. Journal of Software, 2008, 19(12): 3170-3178. [DOI: 10.3724/SP.J.1001.2008.03170]