

全息显控界面中多通道人机交互技术研究

史小龙¹, 李晓玲², 高虹霓¹, 韩枫²

(1. 空军工程大学, 西安 710051; 2. 西安交通大学, 西安 710049)

摘要: **目的** 解决传统座舱显控方式信息量大且操控复杂容易造成交互效率低的问题。**方法** 提出了一种在全息显控界面下融合手势及语音的多通道人机交互系统设计方法, 根据三维数字地图全息展示及交互特性, 进行了任务划分和整体设计。通过构建全息界面多通道交互系统实验台, 完成了基于 Kinect 的手势及语音模块和交互接口的设计, 实现了三维数字地图全息展示中多通道人机交互的准确控制。**结论** 工效实验表明, 该交互方法比传统操控方法的效率提高了约 17.5%。

关键词: 多通道交互; 数字地图; 全息界面; 人机工效

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2016)04-0120-05

Multimodal Interaction Technology in the Holographic Display and Control Interfaces

SHI Xiao-long¹, LI Xiao-ling², GAO Hong-ni², HAN Feng¹

(1. Air Force Engineering University, Xi'an 710051, China; 2. Xi'an Jiaotong University, Xi'an 710049, China)

ABSTRACT: It aims to solve the problems that traditional cockpit display-control system usually have too much complex information, which leads to the operators have to take a long time to be familiar with the interface and often cause the low efficiency. A new method of multimodal interaction system which integrated gesture and speech is proposed under the holographic display control interface. First, according to characteristics of 3D digital map holographic display and interaction, the system framework is designed and the task is allocation. Second, an experiment platform is established which with holographic interface and multimodal interaction, a multimodal interface is designed which can be controlled by gesture and speech based on Kinect. At last, the 3D digital map can be controlled precisely using multimodal interaction with holographic display. The experiment shows that the ergonomics of multimodal interaction increases by about 17.5% than traditional control ways.

KEY WORDS: multimodal interaction; digital map; holographic interface; ergonomics

近年来,全息显示技术在军事、医疗、工业、商业、娱乐等领域发展迅速,已经在大尺寸、真彩色、高分辨率、动态实时性等关键技术上取得了一些突破性成果。如英国 QinetiQ 公司的高分辨率全息立体显示^[1],德国 SeeReal 技术公司研发的三维全息摄影技术的立体显示器^[2],美国 MIT 空间成像研究组研制的全息立

体视频显示系统^[3]。将多通道人机交互技术与全息显示相结合也逐渐成为计算机通信领域的热点,如国外 MIT 的媒体实验室 SRI^[4],CMU 的交互系统实验室^[5],斯坦福的人工智能研究中心^[6],它们都在对多通道人机交互技术进行研究并取得了一定成果。国内的中科院^[7-8]、清华大学^[9-10]、北京大学^[11-12]、浙江大学^[13-14]等在

收稿日期:2015-10-15

基金项目:国家自然科学基金(61472314)

作者简介:史小龙(1989—),男,河南人,空军工程大学硕士生,主攻人机工程。

通讯作者:李晓玲(1971—),女,陕西人,博士,西安交通大学副教授,主要研究方向为工业设计和人机工程。

语言识别和动作识别方面也取得了显著成效。人机交互在语音交互、触控技术、姿态识别方面进展迅速,也在汽车领域得到应用,如特斯拉的 Model S 显控界面、AUDI 的 MMI 系统和 BMW 的 iDrive 人机交互系统。最近微软发布的 Microsoft HoloLens,更是完美地将多通道交互技术嵌入全息显示中,预示着人机交互 3.0 时代的到来。

1 关键技术

1.1 全息三维建模技术

全息投影技术利用干涉和衍射原理记录并再现物体发出的光的“全部信息”,提供人眼全部的深度要素,是真裸眼 3D 显示技术^[15]。三维数字地图具有信息量大、图形渲染要求高的特点,这对其开发平台有较高要求。Unity3D 是由 Unity Technologies 公司推出的多平台游戏开发引擎,可用于创建 3D 游戏、虚拟环境可视化、实时动画等多种类型^[16]。Unity3D 不但三维模型格式多,而且构建的三维场景便于渲染、利于跨平台兼容等,因此适合作为构建三维地图的场景平台。

信息化作战对未来战场的显示与指控系统提出了新的要求,三维数字地图全息展示能为精确打击的战术行动提供高度精确的局部地形、建筑物等实时信息,增强军队的战场态势感知能力。

1.2 Kinect 多通道人机交互技术

微软的 Kinect 是基于 NUI 骨骼跟踪和 Microsoft Speech API 语音识别开发而成的一种体感交互设备。NUI 骨骼跟踪技术是通过所采集的深度信息并利用机器学习分类出大量的人体轮廓图形,然后在已建立的分类库中基于贝叶斯理论确定当前像素点所对应的人体关节。骨骼数据采用基于事件的方式通过 SkeletonStream 传输,在初始化并打开骨骼跟踪后,可从 SkeletonStream 获得骨骼数据, SkeletonStream 产生的每一帧数据都是一个骨骼对象集合^[17]。集合中的每一个元素都包含描述骨骼位置和骨骼关节的数据,利用 Microsoft Speech API 可以做到弱语音命令识别, Kinect 中音频前置处理主要是通过 KinectAudioSource 来完成的^[18]。KinectAudioSource 处理后的音频数据利用 Microsoft.Speech 类库进行语音识别, Microsoft.Speech 类库的关键数据对象是 SpeechRecognitionEngine,它负责预处理经 Kinect 优化后的音频数据流,在分析语法表达式 Grammar 和置信

度阈值后,解释匹配出最合适的语音命令。

2 多通道交互系统的整体设计

2.1 系统整体框架

系统整体框架见图 1,它是基于三维数字地图的全息界面多通道交互系统整体框架图,主要包括任务划分、全息显示以及多通道交互 3 个部分。根据任务特性将系统任务分为连续任务和离散任务,连续任务主要指鹰眼漫游,离散任务指地标搜索、导航。操控者根据任务划分选择是借助语音还是手势输出指令来调取数字地图相关功能;全息显示模块将这些功能的可视化界面信息反馈给操控者;多通道交互主要包括手势和语音的交互。鹰眼漫游主要包括地图场景的缩放、平移和旋转,如缩放和平移是通过左手臂的静态手势实现,旋转是通过右手臂的静态手势实现,而路径导航和地标搜索则可以直接通过口述起始地点或地标名称来实现。

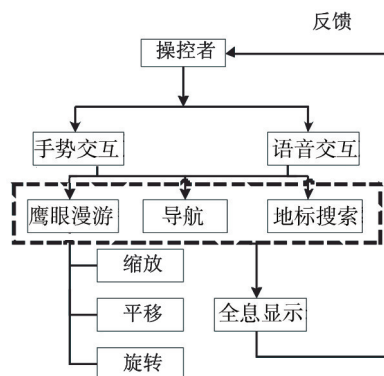


图1 系统整体框架

Fig.1 The system framework

开发工具包括地图场景创建工具 Unity 3D、手势识别开发包 Kinect for windows SDK 和语音识别开发包 Microsoft Speech SDK 等,整体系统使用 C# 语言编写。

2.2 全息展示多通道交互实验台的构建

该实验台由全息投影系统、多通道交互系统、控制系统 3 个部分组成,全息界面多通道交互实验台见图 2。

1) 全息投影系统主要由高亮度、高分辨率的工程投影仪和全息膜组成,用于裸眼三维显示各种座舱显控界面及视景。

2) 多通道交互系统主要指 Kinect 以及场景渲染

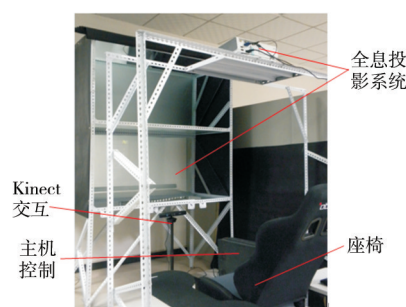


图2 全息界面多通道交互实验台

Fig.2 Experimental platform of holographic interface and multimodal interaction

软件等,包括实现手势和语音交互的软硬件,通过编写C#代码,调用经过Kinect二次封装的.Net类库来实现手势、语音对数字地图的鹰眼漫游、导航、目标搜索等交互功能。

3) 主机控制系统指通过中央控制主机,完成整个实验台软、硬件的整合和实验数据的分析、处理。

3 系统实现

构建基于Unity3D的地图场景模型后,参考MSDN中Kinect的相关说明,编写装载数据的结构体或类。然后调用Kinect.dll中的C++非托管代码初始化Kinect并运行。最后将读取的数据装入构造好的容器,以供高层函数或Unity3D调用。这样就可以在Unity3D中基于Kinect来构建手势和语音的交互逻辑了。

3.1 场景模型的构建

3.1.1 地图模型的导入

将地图模型文件作为asset导入Unity3D,在asset菜单中找到城市的prefab文件,拖入hierarchy菜单即可。导入Unity3D后的地图场景见图3。

3.1.2 导航界面的绘制

针对Unity特有的导航系统,绘制导航地图界面,见图4,绘制步骤如下。

1) 获取地图。调整工作视角为俯视,在Hierarchy面板中新建plane物体,将其拉伸至刚好能够覆盖场景平面的大小,按照plane大小来截取屏幕图片,获得导航地图的图片。

2) 绘制导航地图。在OnGUI()函数中使用GUILDrawTexture(),以小地图的高宽为矩形大小,并将小地图当作材质填充画笔,将小地图放置于屏幕右下角。

3) 将场景坐标和导航地图坐标对应。读取小地图的高宽像素值,读取plane的高宽像素值,利用两者的



图3 地图场景

Fig.3 Map scene



图4 导航地图界面

Fig.4 Interfaces of navigation map

比例关系,将操作者场景中的位置映射到导航地图中。

4) 读取规划路径点。利用脚本系统的变量传递特性,读取路径规划脚本中的Path.corners信息。

5) 构建连线函数DrawLine。利用旋转画布函数GUIUtility.RotateAroundPivot()和GUILDrawTexture()来代替连接两点的函数。

6) 利用DrawLine函数,参照坐标映射方法,连接获得的坐标点,并将它显示在GUI中。

3.2 手势交互的实现

这里把手势交互区定义为手腕和肩部的深度数据相差0.23 m的位置区域。为了缓解手悬空造成的交互疲劳,采用静态手势代替动态手势执行相关任务,即将手势设置成为单值变量,利用程序循环代替手势移动,从而改变任务驱动变量。实现手势交互实现地图漫游功能的主要步骤如下。

1) 构建容器。新建Unity脚本文件,参照MSDN中Kinect文档,编写NUI_SKELETON_DATA, NUI_SKELETON_FRAME等重要的数据结构体容器和枚举类型。

2) 获取原始函数。利用C#标签特性,从Kinect10.dll中引用NuiInitialize(), NuiSkeletonGetNextFrame()等关键函数,并利用这些函数向构建的容器输出骨骼数据备用。骨骼数据如下:

```
int hr = OriginalMethods.NuiSkeletonGetNextFrame(100, ref skeletonFrame)
```

3) 应用数据。向Unity场景中固定的GameObject(如摄像机)添加一个脚本,循环读取骨骼数据,便于将数据应用于该场景。

4) 交互逻辑判断。首先判断右手手腕与右肩深度距离,大于0.23 m则判断右手手腕与右手肘的X向和Y向距离,任一方向大于0.08 m就使摄像机向该向旋转。同理使左手控制摄像机的位移效果。

3.3 语音交互的实现

语言具有较大的复杂性和机器学习的难度^[19]。为

了获得较好的语音交互体验,将本系统中置信度设置为70%,操作者站立点距离 Kinect 传感器 1.2 m~1.5 m,选择较简单且能明确区分的单词作为语音命令。实现语音交互的主要步骤如下。

1) 实现语音识别基础。编译一个 dll 文件,在该文件中利用 Speech API 中的 API for Speech Recognition 的关键接口,如 ISpRecognizer 和 ISpRecoContext 等,建立 InitSpeech Recognizer(), LoadSpeech Grammar() 等封装函数以便在 Unity 脚本中使用。

2) 在 Unity 中使用 Kinect 语音数据。参照手势交互实现中骨骼数据的获取,在 Unity 中编写容器,获取 Kinect.dll 中的语音函数,初始化识别引擎。

3) 创建语音识别语法文件。利用 xml 文件建立识别目标的短语或者单词,然后由 LoadSpeechGrammar() 加载进入初始化后的识别引擎。

4) 设置导航系统。利用 Unity 中的 Navigation 模块,创建导航物体 agent。

5) 语音导航。使 agent 的 destination 属性或者 Warp() 函数的参数跟随语音识别结果作相应变化,即可实现语音地标搜索功能。

在创建语音识别语法文件中,"start from" 命令设定如下:

```
<item>
    <tag>Here</tag>
    <one-of>
        <item> here </item>
    </one-of>
</item>
```

语音导航中,可将导航目的地参数设定为“我的位置”:

```
if (result == "Here") {
    Vector3 here= new Vector3 (main_camera.transform.
position.x, 6.3f,main_camera.transform.position.z );
agent.destination = here;}
```

4 人机交互工效评价

4.1 实验方案

实验设计的连续任务和离散任务各一组,每组进行3次实验。被试分别通过传统键鼠交互和多通道交互对数字地图进行操作,记录任务完成时间,以评价全息显控界面的交互工效。被试由受训的和未训练的

地标进行搜索,受训者指的是用鼠标操控,已提前进行多次操作训练的被试;未训练者指的是用手势和语音操控,在实验前只进行简短的操作说明就进入实验的被试。传统鼠标交互是指通过滚动鼠标滚轮来缩放地图,敲击键盘的方向键来平移地图,单击鼠标左键来浏览三维地图场景,以及使用其配合键盘向地图搜索框输入信息来搜索地标;多通道交互是指通过握拳、前推等配合来平移、缩放、浏览地图,用语音命令来搜索地标和启动导航。全息界面多通道交互工效测试见图5。



图5 全息界面多通道交互工效测试
Fig.5 The ergonomics testing of holographic interface and multimodal interaction

4.2 数据分析

对比全息界面多通道交互方式与传统交互方式完成任务所用的时间,地标搜索任务实验和漫游导航任务实验的对比见图6-7。

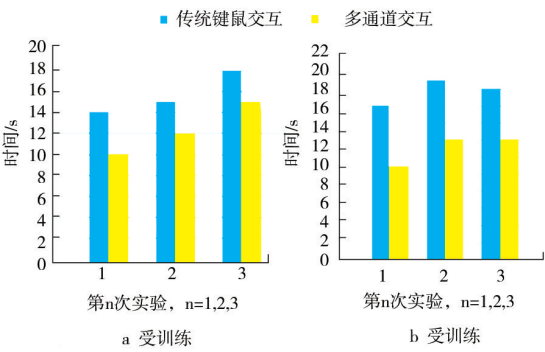


图6 地标搜索任务实验对比
Fig.6 Comparison of landmark searching

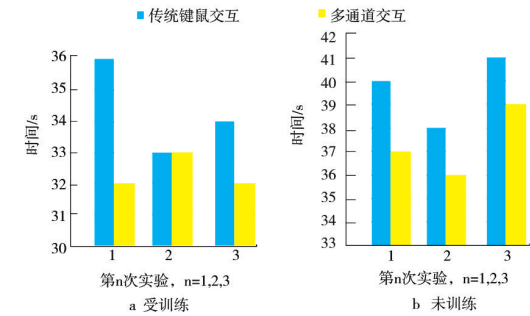


图7 漫游导航任务实验对比
Fig.7 Comparison of roaming and navigation

由地标搜索任务实验对比可知,受训练的被试用多通道交互比用传统的键鼠交互快21%,未训练的被试相应地快37%。虽然受训练的被试使用传统的键鼠交互方式的耗时比未训练的被试少很多,但在用多通道交互时他们所用的时间总体差距不大。

由漫游导航任务实验对比可知,在按指定导航路径鹰眼漫游任务中,任务相对较复杂,传统的鼠标交互和多通道交互所用的时间都相对较长。无论受训练的还是未训练的被试,多通道交互所用时间比传统的鼠标交互大约快6%,这主要是随着任务的复杂,交互动作不够准确,被试认知负荷的增加,以及语音和手势识别准确率的降低导致的。从整体上看,多通道交互的效率比传统的键鼠交互提高约17.5%。

5 结语

全息界面下的多通道交互方式和传统鼠标等操作方式相比,它具有操作直观、快捷,舒适度较高,便于上手等特点,能提高人机交互工效。鉴于本次实验结果,为了进一步增加多通道交互较传统交互的优势,未来在改善多通道交互的鲁棒性方面需要考虑引进更先进的手势、语音识别算法,并在实验台中布置2个甚至更多的Kinect来实现多通道交互。

参考文献:

- [1] DIELISSSEN C. Development of SODI, IPE and DIRSOL[C]//61st International Astronautical Congress 2010. Prague: Czech Republic, 2010.
- [2] ZSCHAU E, MISSBACH R, SCHWERDTNER A. Generation, Encoding and Presentation of Content on Holographic Displays in Real Time[C]//Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV. Orlando: SPIE, 2010.
- [3] JOLLY S, BARABAS J, SMALLEY D. Progress in Updatable Photorefractive Polymer-Based Holographic Displays via Direct Optical Writing of Computer-Generated Fringe Patterns[C]//SPIE Practical Holography XXVII: Materials and Applications. San Francisco: SPIE, 2013.
- [4] THOMAZ A. Teachable Robots: Understanding Human Teaching Behavior to Build More Effective Robot Learners[J]. Artificial Intelligence, 2008(2): 716—737.
- [5] KIM S, DEY A. AR Interfacing with Prototype 3D Applications Based on User-Centered Interactivity[J]. Computer-Aided Design, 2010(4): 373—386.
- [6] JOHNSTON M, BANGALORE S. Finitestate Multimodal Parsing and Understanding[C]//Proceedings of the 18th Conference on Computational Linguistics, 2000.
- [7] 路璐, 田丰, 戴国忠. 融合触、听、视觉的多通道认知和交互模型[J]. 计算机辅助设计与图形学学报, 2014, 26(4): 654—661.
- [8] LU Lu, TIAN Feng, DAI Guo-zhong. A Study of the Multimodal Cognition and Interaction Based on Touch, Audition and Vision[J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(4): 654—661.
- [8] 敖翔, 王绪刚, 戴国忠. 基于多通道融合连续手写识别纠错方法[J]. 软件学报, 2007, 18(9): 162—173.
- [9] AO Xiang, WANG Xu-gang, DAI Guo-zhong. Error Correction of Continuous Handwriting Recognition by Multimodal Fusion[J]. Journal of Software, 2007, 18(9): 162—173.
- [9] 王平, 陆继翔, 余隋怀. 云终端语音交互中改进型谱减法语音增强算法[J]. 计算机集成制造系统, 2013, 19(7): 21—25.
- [9] WANG Ping, LU Ji-xiang, YU Sui-huai. Improved Spectral Subtraction Speech Enhancement Algorithm in Cloud Terminal Speech Interaction[J]. Computer Integrated Manufacturing Systems, 2013, 19(7): 21—25.
- [10] 薛为民. 基于认知机制的情感虚拟人交互技术研究[J]. 北京联合大学学报, 2010, 24(4): 1—6.
- [10] XUE Wei-min. Research on Affective on Virtual Human Interaction Based on Cognitive Mechanics[J]. Journal of Beijing Union University, 2010, 24(4): 1—6.
- [11] 张勇, 刘轶, 刘宏. 结合人耳听觉感知的两级语音增强算法[J]. 信号处理, 2014, 30(4): 363—373.
- [11] ZHANG Yong, LIU Yi, LIU Hong. A Two-stage Speech Enhancement Algorithm Combined with Human Auditory Perception[J]. Signal Processing, 2014, 30(4): 363—373.
- [12] 卢可可, 刘冠阳, 张玉茹. 基于力学交互的高速率精准操作技能训练方法[J]. 计算机工程与应用, 2011, 47(11): 215—218.
- [12] LU Ke-ke, LIU Guan-yang, ZHANG Yu-ru. Precise High-speed Motor Skill Training Based on Haptic Interaction[J]. Computer Engineering and Applications, 2011, 47(11): 215—218.
- [13] PAN Z, LI Y, ZHANG M. A Real Time Multicue Hand Tracking Algorithm Based on Computer Vision[C]//Virtual Reality Conference. Boston: IEEE, 2010.
- [14] PAN Z, CHEOK A D, YANG H. Virtual Reality and Mixed Reality for Virtual Learning Environments[J]. Computers & Graphics, 2006, 30(1): 20—28.
- [15] ZHAO Y, WEI Y, CUI X. 3D Display Technology in Medical Imaging Field[C]//2013 IEEE International Conference on Medical Imaging Physics and Engineering. Shenyang: IEEE, 2013.
- [16] 潘丹. 基于unity3D的手术训练游戏的设计与实现[D]. 济南: 山东大学, 2014.
- [16] PAN Dan. The Design and Implementation of a Virtual Surgery

计中占有着举足轻重的地位,其能够直观地反映出平面作品本身所要传达的语义,能赋予画面鲜明的风格特色和强烈的视觉冲击力。为此,进一步深入探讨不同地域、文化、时代背景之下的色彩语义,并对色彩语义的影响因素以及色彩语义感应进行研究总结,以更好地把握好色彩语义在现代平面广告设计中的应用规律,从而为现代平面设计师提供更多和更有意义的创作方法。

参考文献:

- [1] 陈蕾.色彩设计与应用[M].北京:中国纺织出版社,2013.
CHEN Lei.Color Design and Application[M].Beijing: China Textile Press,2013.
- [2] 金荣淑.设计中的色彩心理学[M].北京:人民邮电出版社,2013.
JIN Rong-shu.Design Color Psychology[M].Beijing: People Post Press,2013.
- [3] 吕莎莎.色彩语义在平面广告设计中的应用研究[D].青岛:青岛理工大学,2014.
LYU Sha-sha.Study on the Color Semantics in Graphic Design in the Application Mode[D].Qingdao:Qingdao Technological University,2014.
- [4] 甘忆辛.色彩语义在平面广告设计中的应用探析[J].艺术科技,2015(6):200.
GAN Yi-xin.The Color Semantic Application of Graphic Design[J].Art Science and Technology,2015(6):200.
- [5] 宫凯.平面设计中视觉审美元素的构建[J].建材与装饰,2013(32):17—18.
GONG Kai.Construction of the Visual Aesthetic Elements of Graphic Design[J].Construction Materials & Decoration,2013(32):17—18.
- [6] 王晓燕.平面设计中图形与色彩的运用分析[J].大舞台,2014(8):41—42.
WANG Xiao-yan.Application Analysis of Graphic Design Graphics and Color[J].Big Stage,2014(8):41—42.
- [7] 潘海生,杜晓雨.平面设计原理[M].北京:北京理工大学出版社,2013.
PAN Hai-sheng, DU Xiao-yu.Principles of Graphic Design [M].Beijing:Beijing Institute of Technology Press,2013.
- [8] 赵春雷.平面设计中视觉审美元素的构建[J].电脑知识与技术,2014(14):397—398.
ZHAO Chun-lei.Construction of the Visual Aesthetic Elements of Graphic Design[J].Computer Knowledge and Technology,2014(14):397—398.
- [9] 吴昊晨.平面设计中的色彩设计[J].黑龙江科技信息,2015(7):51.
WU Hao-chen.Color Design of Graphic Design[J].Heilongjiang Science and Technology Information,2015(7):51.
- [10] 董茜.浅谈平面设计中视觉审美元素的构建[J].魅力中国,2014(10):70.
DONG Qian.Construction of the Visual Aesthetic Elements of Graphic Design[J].Charming China,2014(10):70.
- [11] 张蕊.探析“红色”在现代平面设计艺术中的应用[D].济南:山东大学,2013.
ZHANG Rui.The Application of "Red" in the Graphic Design [D].Jinan:Shandong University,2013.
- [12] 佐佐木刚士.配色设计原理[M].北京:中国青年出版社,2013.
SASAKI T.Color Design Principle[M].Beijing: China Youth Press,2013.
- [13] 谢文婷.关于平面设计中视觉审美元素的构建[J].中国包装工业,2015(2):146—148.
XIE Wen-ting.Construction of the Visual Aesthetic Elements of Graphic Design[J].China Packaging Industry,2015(2):146—148.
- [14] 朱华,张歆.当代平面设计中隐性色彩信息的价值探究[J].包装工程,2013,34(2):13—16.
ZHU Hua, ZHANG Xin.An Exploration on the Value of Implicit Color Information in Contemporary Graphic[J].Packaging Engineering,2013,34(2):13—16.
- [15] Medical Imaging by Off-the-shelf Input Devices[J].Computers in Biology and Medicine,2010,40(3):350—358.
- [16] 谭浩,赵丹华,赵江洪.面向复杂交互情境的汽车人机界面设计研究[J].包装工程,2012,33(18):26—30.
TAN Hao, ZHAO Dan-hua, ZHAO Jiang-hong.Research on Automotive Human Machine Interface Design Based on Complex Interaction Context[J].Packaging Engineering,2012,33(18):26—30.
- [17] 陈燕军,汪地,杨浩.基于Kinect的手术辅助系统研究[J].计算机技术与发展,2014,24(9):81—88.
CHEN Yan-jun, WANG Di, YANG Hao.Research on Operation Auxiliary System Based on Kinect[J].Computer Technology and Development,2014,24(9):81—88.
- [18] GALLO L, MINUTOLO A.A User Interface for VR Ready 3D

(上接第124页)

Training Game Based on Unity 3D[D].Jinan: Shandong University,2014.