

基于肌肉感知的手势交互模型^{*}

范俊君^{1,2}, 田丰^{1,2}, 黄进¹, 刘杰¹, 王宏安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 计算机与控制学院,北京 100049)

通讯作者: 田丰, E-mail: tianfeng@iscas.ac.cn

摘要: 在人机交互技术由以计算机演化为以人为中心的背景下,通过感知肌肉活动的手势识别方法,因其可穿戴性、隐式交互性和可靠性的特点在近几年得到了人机交互研究领域的高度关注.但目前的相关研究缺乏统一的语义分析模型和系统模型支持研究和开发.为此,分析讨论了交互手势的分类并归纳总结出适合肌肉感知方法的输入原语,提出基于肌肉感知的手势交互语义分析模型和分层处理的系统结构模型,旨在提高该类型交互应用的研究和开发工作效率.最后分析了办公室环境下的操作手势交互应用场景,给出了该语义分析模型和分层系统结构模型的应用实例.

关键词: 肌肉感知;手势交互;语义分析模型;系统结构模型;表面肌电信号

中文引用格式: 范俊君,田丰,黄进,刘杰,王宏安,戴国忠.基于肌肉感知的手势交互模型.软件学报,2018,29(Suppl.(2)):62-74. <http://www.jos.org.cn/1000-9825/18018.htm>

英文引用格式: Fan JJ, Tian F, Huang J, Liu J, Wang HA, Dai GZ. Gesture interaction model based on muscle sensing. Ruan Jian Xue Bao/Journal of Software, 2018,29(Suppl.(2)):62-74 (in Chinese). <http://www.jos.org.cn/1000-9825/18018.htm>

Gesture Interaction Model Based on Muscle Sensing

FAN Jun-Jun^{1,2}, TIAN Feng^{1,2}, HUANG Jin¹, LIU Jie¹, WANG Hong-An¹, DAI Guo-Zhong¹

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

²(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Gesture interaction based on muscle sensing has received great attention due to its wearability, implicit interaction, and reliability under the background of human-computer interaction technique shifting from computer-centered to human-centered. However, current researches lack of a unified semantic model and system model. This paper discussed the classification of interactive gestures, summarized the input primitives suitable for physiological computing technology, and proposed muscle sensing based gesture interaction semantic model and hierarchical processing a system model. Finally, this paper implemented a prototype of object operating gesture recognition under scenario of office environment.

Key words: muscle sensing; gesture interaction; semantic analysis model; system structure model; surface electromyography signal

人机交互(human-computer interaction,简称 HCI)是研究、设计、实现人与计算机之间通过人机界面进行信息交流的技术,其最终目标是实现人与机器自然地交流,从而为操作者提供更为直观、舒适的交互体验.由于计算机已经大量融入人类的社会生活中,人机交互活动日益成为了我们日常生活中不可或缺的组成部分.传统的人机交互方式,比如键盘、鼠标、遥控器、触摸屏等,都需要人去适应机器.特别在虚拟现实、增强现实等新型交互场景下,这些方式已经不能满足用户对自然人机交互的需求,突显出了非常大的局限性.而新一代的交互方

* 基金项目: 国家重点研发计划(2016YFB10011402); 中国科学院前沿科学重点研究计划(QYZDY-SSW-JSC041)

Foundation item: National Key Research and Development Plan (2016YFB10011402); Key Research Program of Frontier Sciences, CAS(QYZDY-SSW-JSC041)

收稿时间: 2017-06-01; 采用时间: 2018-11-16

式则侧重于各种新的、更加开放的输入渠道,比如语音、身体动作、手势、表情、生理信号等.交互界面正逐渐从人围绕着机器发展成为由机器围绕着人的以人为中心,人机交互朝着更加人性化,更加符合人的自然技能的方向发展.

生理计算(physiological computing)是一种人体生理与计算机的直接接口,它的目标是对来自人类的生物信号进行解析,并转换为计算机的实时输入来提高和改善交互体验,其中最具代表性的生理信号是脑电、肌肉电、心电、眼电、呼吸模式、皮肤电反应和血压等^[1].由于其可穿戴性、隐式交互性、移动性和可靠性的特点,使得基于生理计算的人机交互技术具有巨大的潜力,可以被大多数交互式应用和生理信息驱动式应用所支持和使用,将作为未来人机交互领域的一个重要方向.基于肌肉感知的手势交互技术是通过特定的设备获取手势动作时的肌肉生理数据,通过生理计算方法实现手势动作识别,是实现自然人机交互的一种重要方式^[2-11].在现有基于肌肉感知的手势交互技术中,通过表面肌电(surface electromyography)捕捉手部动作时肌肉活动产生的肌肉电信号进行手势识别是目前在人机交互研究中最受研究者关注并且应用最多的方法^[2].微软雷德蒙德研究院的Tan等人首先展示了使用表面肌电作为人机交互接口的可行性.通过使用成环状分布在小臂上的多通道肌电传感器获取肌电信号,有效地识别了5个手指的点击、抬起等动作,并使用该方法开发了便携可穿戴的肌电-机接口系统原型,成功开发了肌电吉他、随身听控制器等应用^[3,4].华盛顿大学Huang等人则把肌电传感器配置成环状分布佩戴在用户手腕上,探索了触屏手机环境下的拇指动作精细识别,包括有反馈和无反馈条件下的点击、滑动及九宫格解锁操作^[5].为了能够捕获更好更全面的手部肌电信号,德国卡尔斯鲁厄理工学院的Amma等人使用了8×24点阵的高密度肌电传感器获取整个手臂的肌电信号,实验证明这种方案可以获得非常高的手势识别率^[6].根据表面肌电手势识别原理,Thalmic Labs公司开发了肌电腕带MYO,并且在腕带中加入运动传感器.MYO腕带自带的应用支持演示文档控制、简单手势等操作,并且对外开放应用开发接口,是基于肌肉感知的手势交互技术从研究走向商业应用的第1步.同样,也有研究者利用新颖的传感器技术来实现手势交互应用.麻省理工学院Dementyev等人使用压力传感器设计了一条压力腕带感知手臂肌肉群的收缩状况,并利用这条腕带可以识别握拳、张手、掐指等多种手动动作^[7].卡内基梅隆大学的Zhang等人则利用肌肉收缩与放松时的电阻抗差异,使用高分辨电阻断层成像技术对手腕截面进行扫描的方式检测手部做动作时手臂肌肉电阻变化,开发了一套能实时识别手势动作的系统^[8].

东南大学的宋爱国团队在肌肉活动信号获取设备的研究和开发方面做了深入的研究^[9].中国科学技术大学的陈香教授在基于表面肌电的中国聋哑人手语识别、虚拟现实手势交互应用等方面也获得了不菲的成果^[10].中国科学院软件研究所田丰团队提取手势动作时的肌电信号用于判断手势的起始结束点,解决手势识别应用中手势与手势之间难以分割的难题^[11].

虽然基于生理计算的手势交互技术已经有了大量的研究和应用成果,但是研究人员在研究和开发应用时,往往根据自身的设定和交互环境进行研究,缺少具有针对性的手势交互语义分析模型和系统结构模型.传统的交互模型如MVC(model-view-controller)模型^[12]、PAC(presentation-abstraction-control)模型^[13]等由于过于宽泛而缺乏对实际应用的指导意义,无法很好地指导通过感知肌肉活动的手势交互应用在实际交互场景的设计和系统应用的开发.Pavlovic研究了基于机器视觉的手势识别和交互技术并提出了相应的手势交互系统模型,但由于图像数据与生理信号数据在数据获取设备、数据获取方法以及数据表达形式都存在上都存在巨大的差异,该模型也无法很好地应用到生理信号处理中来^[14].文献[15]提出了针对生理计算的交互工具箱PhysiolComp Toolkit,通过组件的方式根据需要集成生理计算系统需要的模块提高生理计算研究和应用的开发效率,但是缺少对手势交互场景的语义分析和系统模型的研究.

因此,本文在对交互手势的分类进行充分研究和合理扩展的基础上,分析了肌肉活动信号的交互手势输入原语,并提出了基于肌肉感知的手势交互语义分析模型和系统结构模型,为该类型手势交互应用的交互设计和系统实现提供相关的理论指导和支持.此外,本文调研了办公室环境下的常用操作手势,根据手势交互语义分析模型和系统结构模型实现了办公室日常操作手势识别系统的分析、设计和简单原型实现,并对系统进行了可用性和可靠性评估.经实践证明,该模型能够满足基于肌肉感知的手势交互应用的设计需求,对此类应用的开发和

评估具有一定的理论支持和指导作用.

1 交互手势分类与输入原语

手势是一种非语言的交流形式,通常被定义为手或者手与臂结合所产生的各种姿势和动作,以表达想法、情绪或者强调^[16].手势交流是人的本能,在学会语言和文字之前,已经能用肢体语言与人交流,是人们在日常生活中广泛适用的自然交流方式.随着计算机技术的不断发展,使得基于手势的人机交互技术成为可能.与已有成熟的交互方式相比,手势交互的技术更复杂,但同时有很多优势:学习成本较低,识别别人的自然手势,不像传统交互那样记住双击和鼠标左右点击的区别;可以脱离实体接触,实现远距离控制;交互动作更加丰富和自然,不同的操作有不同的手势,不像传统交互局限于点击和滑动等几种常用的操作;对用户正常活动影响较少,可以随时继续手势操作.因此,我们可以认为手势交互是继鼠标、键盘和触屏之后新的人机交互方式.对手势的交互予以进行分类和界定,是手势交互研究的重要内容.

在现有的手势分类中,最适合应用于人机交互领域的是 Pavlovic 的分类方法,根据手势的功能逐层把手势分成不同的类别^[14].随着手势交互研究和应用的深化和扩展,我们对该分类方法做了进一步的扩展和修改.根据手与目标关系的不同,把操作手势细分为触摸操作手势和抓握操作手势,如图 1 所示.

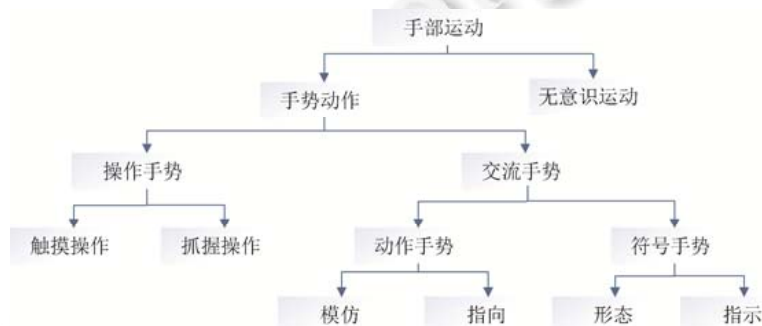


Fig.1 Gesture classification based on hand function

图 1 基于功能的手势分类

手部的运动可以被分为传递用户意图的手势动作和不传递用户意图的无意识手部运动两种情况.无意识手部运动是指不传达任何交流信息的动作,如走路时双手的摆动,手势准备前的手部动作等.在基于手势的人机交互应用中,往往需要事先约定手势,目的在于区分对人有意义的手势动作和没有意义的无意识手势.手势动作本身可以分成交流手势和物品操作手势.交流手势的目的是传递有用信息,根据其表现形态可以分为动作手势和符号手势.动作手势指动作本身就已经包含交流内容的手势,包括模仿(模仿某个动作)和指向(比如指向某个方位).符号手势是有着语言表达功能的手势,分别可以表示形态和表示指示.表示指示含义的手势可以通过手指的运动轨迹来传递表达意图(比如食指在空中画一个圆圈表示车轮),而表示形态类别的手势则通过模拟目标的形态来表达意图(比如伸出一个手指表示数字 1).这种类型的手势由于可以表示静态手势,易于定义和识别,也是人机交互中最常用的手势.交流手势在使用过程中通常是空手的,而操作手势则需要与物品接触.根据用户使用意图和功能的不同,操作手势又可以被分为触摸操作手势和抓握操作手势.触摸操作手势在操作物品时通常单指或多指对目标物品进行触摸、点击操作.例如在键盘上打字操作、按动开关等.而抓握操作手势往往以特定的抓握手势抓拿物品,通过操控该物品对目标进行操作.这时候用户抓拿在手中的物品可以看做是用户身体的延伸.例如用于写字的笔、用于吃饭的勺子.在不同的手势交互场景中需要定义不同的手势集,手势的分类往往影响着手势集的定义,它决定了手势特征空间和手势持续时间,这对手势的定义和识别起着至关重要的作用.

目前应用于手势识别的生理信号数据获取设备包括惯性传感器、肌电传感器、压力腕带等.用户的同一个输入动作在不同的交互设备上,系统的解析可能会是截然不同的.例如惯性传感器的输入数据经过手势交互

系统会映射为位置和方向,而肌电传感器的数据则映射为手势动作.而在不同的应用上下文中,用户做相同的一个动作,也可能表达的是完全不同的意思.比如用户伸出一个手指,在形态手势输入环境下,用户想表达的意思是输入单词“1”.而在指向手势输入环境下,用户想表达的则可能是指向手指的方位.

为了避免设备特定物理特性和输入环境的差异,本文参照文献[17]中输入原语(Input Primitive,简称 IP)的概念,并结合交互手势分类和生理数据获取设备抽象出一个完备的原语集合,其中包括指示手势、形态手势、模仿手势、指向手势、触摸操作手势和抓握操作手势这 6 类输入原语,见表 1.每个输入原语是来自不同通道的独立的、最小的、不可分割的原子操作,可以实例化为一个四元组(用户动作,数据表示,使用通道,时间标签),它体现了某一时刻来自某个输入通道的用户输入动作与一种内部数据表示联系在一起,在一定的应用上下文中有着特定的交互意义.输入原语是通道无关的,相同的输入原语可以从不同的物理设备通道获得,比如肌电传感器设备和截面电阻腕带设备做预定义的手势动作时都可以产生形态手势类输入原语.同时输入原语也是应用无关的,同一个输入原语在不同的应用上下文中完全可以被解释为完全不同的操作.

Table 1 Classification of input primitive of gesture interaction based on muscle sensing

表 1 基于肌肉感知的手势交互输入原语分类

输入原语类别	设备通道	动作	数据表达	说明
指示手势	惯性传感器	手指笔划	Point $i i=1,2,\dots,n$	手指或者手部的空间轨迹序列
形态手势		预定义的手势	Posture	命令或者单词
模仿手势	肌电传感器、	手势动作	Posture $i i=1,2,\dots,n$	系统预定义的变化手序列
指向手势	截面电阻腕带、 压力腕带、	指点	Vector	向量代表指向的方向,可以是 2 维或 3 维方向
触摸操作手势	推/拉/按/点/划	Posture	用户操作对象时使用的手势
抓握操作手势		抓/握/捏/夹	Posture	用户抓拿物品的手势动作

2 手势交互语义分析模型

基于肌肉感知的手势交互技术使用的输入通道可能是单通道,也可能是多通道单种输入设备,甚至可能是多通道多种输入设备.比如在 Saponas 等人的肌肉-机接口研究^[3,4]中,使用了多通道的肌电采集设备作为输入设备.而在聋哑人手语识别应用^[9]中,由于需要同时获取手势动作信息和位移信息,因此同时采用了惯性传感器和肌电采集设备作为输入设备.为了摆脱生理数据获取设备的特定物理特性和操作方式上的差异,将多种生理信号获取设备的信息整合统一处理,利用交互原语和交互任务串联设备数据和应用,同时考虑交互环境和上下文信息以进行更加精确的计算和分析,达到更好的手势交互效果,本文提出了基于肌肉感知的手势交互语义分析模型,如图 2 所示.从语义分析模型可以看到,交互设备管理构件与具体的交互设备直接相关,通过手势输入设备采集用户手势动作数据.而语法分析构件、语义分析构件和语义分析构件是交互语义分析的核心构件,逐步进行抽象整合提取出交互任务,将交互任务发送给应用程序构件处理.最后应用程序构件通过交互设备管理构件中的输出设备将处理结果反馈给用户,形成交互闭环.

把手势交互语义分析模型分割成负责各个级别功能的构件部分,并定义各个构件部分的功能及其关系,可以让研究和开发人员减少重复工作,理清交互应用各个部分之间的关系,从而专注于交互逻辑的分析和设计,提高工作效率.

(1) 交互设备管理构件

交互设备管理构件负责各种输入/输出设备的管理,实现用户动作到模型内部形式再到用户感知的转换.输入设备包括各种生理数据输入设备,如肌肉电传感器、手臂横截面电阻腕带、惯性传感器等.输出设备包括显示器、扬声器、震动马达等.

当用户做出手势后,交互设备管理构件通过不同的交互设备提取最原始的数据流,如惯性传感器的加速度数据、肌电传感器的肌电信号数据、手腕横截面电阻数据等.当应用程序处理完成操作命令,需要将处理结果反馈给用户时,又通过交互设备管理构件驱动输出设备,使用户感知处理结果.由于具体的数据获取方法或设备驱动方法与交互设备密切相关,通常需要设备厂商提供的相应的设备驱动程序和开发接口,因此对每一个交互

设备需要对应一个构件.如肌电传感器构件实现的是对肌电传感器的控制,完成动作到肌电信号的转换;惯性传感器实现惯性传感器的控制,完成动作到加速度数据的转换;显示器构件实现显示器的驱动,完成用户反馈命令到用户视觉信息的转换.对每个物理设备建立独立的构件是为了便于设备的增删和变更,以及保证输入/输出通道的完整性.

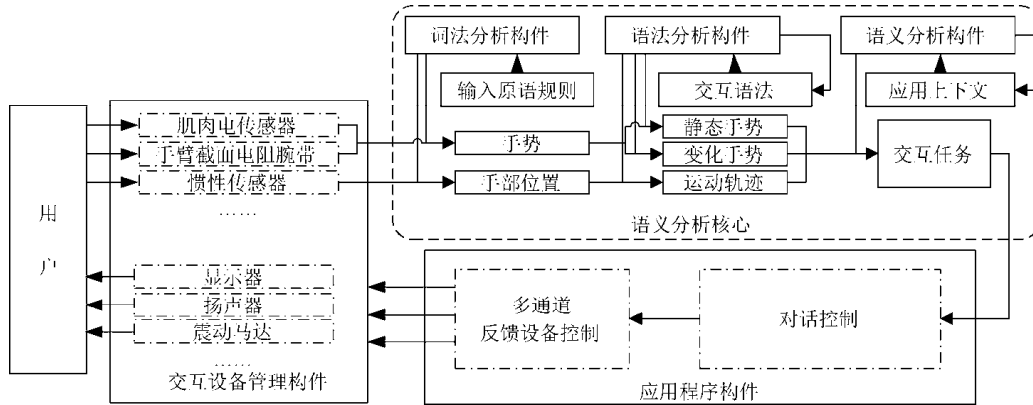


Fig.2 Gesture interaction semantic model

图2 手势交互语义分析模型

(2) 词法分析构件

词法分析构件负责对交互设备管理构件输入的原始数据流进行处理,并根据输入原语规则对来自不同交互设备的数据进行词法层级的整合和映射,使得输入设备的数据信息具有各自的词法解释.词法分析构件完成用户到计算机的词法输入转换,把计算机输入设备的原始数据转换成在一定的应用上下文有着特定的交互意义的输入原语.输入原语规则决定了把设备、数据、动作和时间实例化成输入原语的方法,实现从设备到输入原语的翻译工作.例如,用户使用指点手势指点方向时,词法分析构件对肌电传感器的输入实例化为输入原语表达(Point, EMG, Vector, time).

(3) 语法分析构件

语法分析构件负责来自任务的多个输入原语的交互信息进行整合,根据交互语法推断用户的手势操作交互意图.交互语法规则是指用户和计算机进行交互的操作规则,定义了指定应用场景下允许用户进行操作的手势集合以及进行相关手势所需要的条件.比如做手势动作的持续时间、速度、动作幅度等.手势持续时间过短、速度过快或者幅度过小等因素容易造成手势不能被系统正确识别.因此,交互语法规则的定义直接影响到最终手势判断的准确性和效率.

(4) 语义分析构件

语义分析构件根据手势变化,结合应用上下文信息,经过一系列逻辑判断来推断用户的手势操作意图,即生成交互任务.语义分析构件的主要目的就是避免在出现二义性手势时的错误判断造成用户意图的误判.比如用户做的手势是伸出两个手指,在形态手势类别里伸出两个手指可能表示输入单词“Yes”,也可能表示输入数字“2”,甚至还可能是其他意思.这时候语义分析构件就要根据应用上下文推断用户的真正意图.另外,语义分析构件分析得到的语义结果也可能对应用上下文产生影响,比如上下文环境的变化,语境切换等.

(5) 应用程序构件

应用程序构件包含对话控制和多通道反馈设备控制两个模块.对话控制模块接收用户提交的任务,并将处理结果通过多通道反馈控制设备将反馈及时反馈给用户.比如在手势问答应用中用户使用手势回答计算机的问题,计算机在接收到手势输入后通过手环上的震动马达反馈给用户通知用户手势输入完成,通过显示器通知用户回答是否正确.从多个通道予以用户处理过程和结果的反馈,使用户获得更好的交互体验.

3 分层处理的系统结构模型

基于肌肉感知的手势交互技术在近几年得到了大量研究学者的关注并在空中手势识别、义肢控制等方面取得了不菲的研究成果和应用.从人机交互的角度来看,该方法主要研究问题可概括为以下几个方面.

- 如何选择多通道多模态的生理信号类型和部署相应的传感器以获得识别率和穿戴体验的折衷?
- 如何从大量的生理数据中抽取有用信息用于手势交互?
- 如何选择或者构建高效可靠的手势识别算法?
- 怎么样合理地设计手势交互应用场景,并将我们的手势交互技术融入其中?
- 如何合理地衡量其可用性和有效性?

针对以上问题,我们提出了基于肌肉感知的手势交互技术的分层处理模型,主要包括硬件层、方法层、平台层和应用层 4 个处理层级.每个处理层级负责解决特定的问题和处理任务,如图 3 所示.

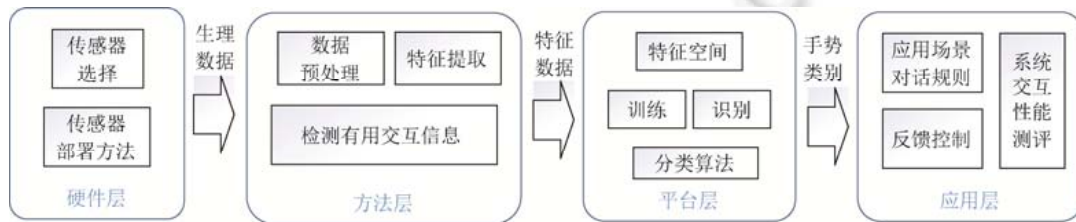


Fig. 3 Layered structure of gesture interaction system model

图 3 手势交互系统模型分层结构

硬件层包括生理数据传感器和传感器穿戴部署方法两部分.基于肌肉感知的手势交互方法可能用到一种或者多种生理数据,选择什么样的生理数据以及如何部署获取这些生理上数据的传感器是构建手势交互系统方法中基础的因素.目前获取手势生理数据的传感器包括弯曲传感器、惯性传感器、肌电传感器等,常见的传感器部署方式有手套、指关节环带、腕带、臂环等方式.传感器的选择和配置方案既要以最大程度获取能反映手势的生理信号,同时也要考虑到应用场景的限制和用户的体验问题.因此,在实际应用中的传感器选择方案和穿戴部署方法往往是功能性和用户体验的折衷.

方法层主要包含生理数据处理的方法,通过从硬件层获取的生理信号特征差异可以体现手部不同的动作.因此,在生理数据采集的基础上,方法层的任务是对连续生理数据进行预处理、融合分析,检测有用的交互信息并提取出一组能表征信号本质特征的数据对手势信号进行描述,为不同手势的识别提供数据准备.方法层包含数据预处理、有用交互信息提取、特征提取 3 部分.

其中数据预处理模块负责对原始生理数据进行初步的处理,如消除工频干扰、滤波去噪声等.检测有用交互信息模块负责手势分割工作,确定手势的开始时间点和结束时间点并从连续的生理信号数据中分割出代表手势的数据段,为特征提取模块提供可用数据.特征提取模块主要是从时域、频域、时频域和非线性动力学等方面进行分析,提取能够反映信号特征和物理意义,并且分离度大、鲁棒性好、运算复杂度低的模式特征.

平台层主要包含手势特征空间和手势分类算法,负责接收上一层生成的特征数据,对特征数据进行特征训练和识别,输出手势识别结果.对于手势识别系统而言,特征提取是基础,有效地进行识别是关键,选择了好的分类器可以为基于生理数据的手势交互系统提供更可靠的手势识别.随着模式识别理论不断发展,越来越多的分类器被应用到生理数据的模式识别中.目前,常用于生理计算的分类方法主要有:隐马尔可夫模型、bayes 决策、极大似然分类器、神经网络、支持向量机等.平台层具有输入通道无关性和应用无关性.输入数据的类型和处理方法不会对平台层造成影响,因此平台层不需要关注交互系统的输入设备选取.同样,平台层也不需要关注具体手势交互应用场景的设计.应用场景发生变化时,平台层不需要发生改变.

应用层根据应用场景对手势识别进行映射,推断用户在当前应用场景下使用该手势的操作意图.应用场景手势映射模块包含指定应用场景下手势集的定义、手势任务映射规则等.在不同的应用场景下,同一个手势可

能代表不同的用户意图.比如伸出一个手指,用户有可能想表示数字 1,也有可能表示指向某个方向.应用层同时也还包括对系统的交互性能测评,比如手势识别率、操作复杂度等.

4 应用实例研究

通过识别用户的在特定应用场景下的操作手势,可以推断用户在该场景下的交互行为及其交互的物品并记录用户活动^[18].为了验证本文提出的手势交互语义分析模型及系统模型的可靠性和有用性,本文提出了利用手势交互技术实现办公室环境下日常工作活动的跟踪和记录.

首先对办公室环境下的日常操作手势进行了调研,共记录了北京市一个开放办公室的 15 个员工(13 名男员工,2 名女员工)的日常办公行为,其中每个员工的工位都放置了台式办公电脑和一些日常办公用品,如笔、纸、水杯等.在两天的正常办公时间(早上 9 点至下午 5 点)里,每隔 1 小时观察 1 次,记录每个员工的当时行为.最后对观察结果进行统计,并选取了其中 6 种频率最高的操作手势行为,包括操作鼠标、打字、操作手机、阅读、喝水和写字,并且总结了每种操作手势的具体行为方式.

4.1 交互语义分析

本应用定义了用户的 6 种交互行为,包括操作鼠标、打字、操作手机、阅读、喝水和写字.其操作对象分别是鼠标、键盘、手机、A4 纸、水杯和笔.操作动作包括点击(键盘、鼠标、手机)、抓拿(杯子、鼠标、手机)、捏(笔).特别地,由于鼠标和手机两个交互对象既需要被抓拿在手里,又需要点击进行操作,因此其交互动作包含了抓拿和点击两类动作.由于抓拿鼠标和手机的手势是相对稳定的,而点击动作幅度较小,因此其交互动作以抓拿为主.按照我们对操作手势分类的定义,操作鼠标、操作手机、阅读、喝水和写字这种的操作手势可以归类为抓握操作手势,而打字可以归类为触摸操作手势.从语义分析模型的角度来看,这 6 种交互行为分别对应语义分析模型最终需要提取的 6 种交互任务,而输入原语包括抓握操作手势和触摸操作手势两类.

通过以上交互行为分析并且对照表 1 的输入原语分析结果,我们发现办公室环境下的操作手势交互动作可以通过肌电传感器、截面电阻腕带、压力腕带等手势输入设备进行输入数据采集.考虑到交互手势行为的复杂性对手势输入设备的要求,同时穿戴输入设备对交互体验带来的影响尽可能小,我们采取了成熟度和可靠性较高的肌电采集设备作为手势交互输入设备.因此,在语义分析模型词法规则下,输入原语可以实例化为手持操作和触摸操作手势两大类,见表 2.

Table 2 Input primitive of manipulating gesture under office environment

表 2 办公室环境下操作手势输入原语

行为名称	动作描述	输入原语描述	手势类别
打字	双手在键盘上打字	<click, posture, EMG, timestamp>	触摸操作手势
操作鼠标	使用惯用手握住鼠标并移动、点击	<grasp, posture, EMG, timestamp>	
操作手机	拿着手机,在屏幕上滑动和点击	<graspPoint, posture, EMG, timestamp>	
阅读	拿着纸质文件阅读	<nip, posture, EMG, timestamp>	抓握操作手势
喝水	拿起水杯喝水然后放回	<hold, posture, EMG, timestamp>	
写字	使用笔在平放桌面的本子上写字	<Write, posture, EMG, timestamp>	

从设备获取到原始肌电信号之后,通过词法分析构件得到输入原语(包含动作类型、原始信号所代表当前时刻的操作手势、代表肌电通道的标记、当前时间),其中操作手势由手势库预定义;语法分析构件对一定时间段内的输入原语进行组合分析,根据交互语法规则定义当前操作手势代表的行为.比如保持抓拿类似球形物品的手势超过 3s 则认为用户是在操作类球形物品,用户在 5s 的时间内持续使用手指进行点击操作则认为用户在使用点击手势操作物品.语法分析构件能够获取用户当前的操作手势类型,但是由于缺乏上下文无法进行具体操作任务的推断;而语义分析构件就是结合操作手势类型和应用上下文推断用户具体的操作任务.在办公室环境下本文定义了 6 种交互任务,用户操作类球形物品则表示用户正在操作的是鼠标,点击操作手势则表示用户正在操作的是键盘;语义分析完成之后将交互任务送达应用程序控制构件,由应用程序记录当前用户执行的操作任务并在特定的时间进行统计分析,通过显示器反馈给用户.

4.2 交互系统分析与设计

完成交互语义分析和定义之后,下一步工作是交互系统的设计与实现.按照前面章节定义的分层处理的系统结构模型方法,本文分别从生理数据设备采集与布置方法、生理数据处理方法、手势识别算法以及应用设计方法4个层面入手,设计一套基于肌肉感知的手势交互系统原型.

(1) 生理数据设备与配置方法

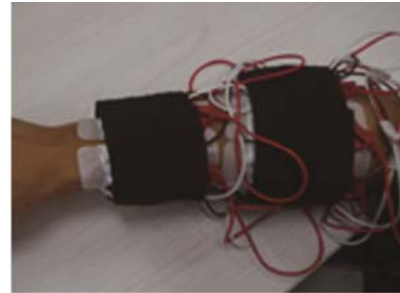
在这套手势交互系统里,本文采用 BIOPAC 公司的 MP150^[19]作为表面肌电采集系统.MP150 是一个商用的生理信号采集分析系统,能采集脑电、心电、肌电、眼电等多种人体生理信号,其外观如图 4(a)所示.肌肉的采样频率设置为 1 000Hz,并过滤掉 10-500Hz 以外的频率信号以及 50Hz 的工频干扰.MP150 通过 USB 与 PC 相连接,原始数据将被传送到 PC 上并在 PC 上进行分析处理.



(a) EMG 数据获取设备- BIOPAC 的 MP150



(b) 氯化银电极片及其部署方式



(c) 连接了 MP150 设备的双臂环部署电极穿戴原型

Fig.4

图 4

在现有的表面肌电采集传感器穿戴方法中,常用的是采用腕带的部署方式^[5]或者臂环的部署方式^[3,4].为了能够采集更全面的肌电信号的同时获得较好的穿戴体验,我们设计了一种双臂环的肌电极穿戴方式—7 对电极均匀分布成环状称为上臂环戴在小臂的肌肉肚上,5 对电极均匀布置成环状称为下臂环戴在手腕上.电极片布置方式和穿戴效果如图 4(b)、图 4(c)所示.

(2) 生理数据分析处理方法

为了能够从连续的肌电信号中提取能表征手势的特征,我们需要把连续信号分割成方便计算机处理的时间不相关数据段.在分割数据段的分割处理中,Saponas 等人分别采用了 64ms 和 125ms 的时间窗长度^[3,4].而我们的前期研究中,我们发现过短的时间窗长度难以捕获完整的手势动作,特别在手势发生变化时识别错误率会明显提高,而当用 500ms 时间窗把肌电信号数据分割成数据段时能取得最好的手势识别效果.

因此,我们对每一个手势动作提取一个 500ms 时间窗的数据段,并从数据段中提取各个通道的均方根^[20]、频谱能量^[21]以及过零率^[22]共 234 个特征.其中均方根类特征提取 12 肌电通道的均方根值及通道之间的均方根比率,共 78 个特征;频谱能量类特征包含 12 肌电通道里每个通道的 12 等比频段的频域能量积,共 144 个特征;过零率则包含 12 个肌电通道肌电信号的过零次数共 12 个特征.均方根特征和频谱能量特征在 Saponas 的肌肉

机接口研究^[3,4]中被证明了其可以用于表征手势动作,而过零率也常被用于信号分析中.在本文的预实验中这三类特征组合在支持向量机等简单分类器中表现出了稳定可靠的识别效果.

(3) 手势识别算法

使用合适的模式分类方法对表征手势的特征进行分类,是手势交互系统设计非常重要的一步.在前期研究中,本文使用 Weka 工具箱^[23]中的几种常用机器学习算法进行比较,包括逻辑回归、K 近邻算法、支持向量机、决策树、朴素贝叶斯网络,以及随机森林.

(4) 应用对话设计

根据调研结果,我们定义了办公室环境下的 6 种操作手势交互行为,分别是操作鼠标、在键盘上打字、操作触屏手机、阅读文献、使用杯子喝水、使用签字笔写字.操作手势交互行为的操作要求与用户调研时观察的行为一直.手势识别算法识别出用户的操作之后,将交互任务发送到应用层,由应用层进行对话控制.在本文的办公室环境下操作行为记录应用中,应用层的工作是保存交互任务记录以及根据记录做对话逻辑控制并通知到用户,如发现用户在一段时间内没有喝水的操作时提醒用户多喝水,在用户指定的时间提醒用户进行某项特定的操作,在下班的时候对用户的操作行为作出统计分析结果等.

根据以上的分析设计,本文提出办公室环境下的操作手势交互系统框架如图 5 所示.在日常办公情景下,多通道肌电采集设备采集用户手臂 12 通道的肌电信号,然后用滑动时间窗将连续肌电信号数据分割成时间长度为 500ms 的数据段作为独立样本.

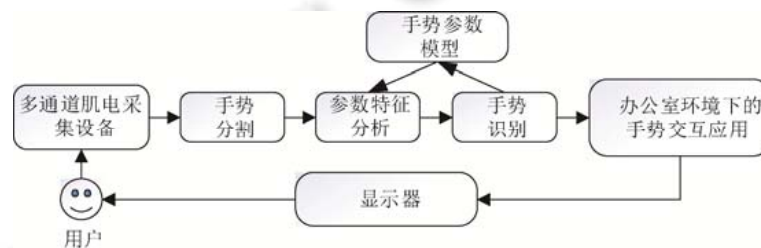


Fig.5 Dataflow of manipulating gesture interaction system under office enviroment

图 5 办公室环境下的操作手势交互系统数据流程

在训练阶段中,从每个训练样本中提取表征动作的特征用于随机森林算法的训练,生成手势参数模型.在测试识别阶段,测试样本特征被送入识别算法中与参数模型比较,最终推断出样板特征表征的操作手势类型.识别结果被送入交互应用中进行交互对话控制模块.最后,交互应用模块将对话输出到显示器,提示用户对话结果,完成交互闭环.

4.3 用户实验过程

共有 6 名志愿者(4 名男性和 2 名女性,平均年龄为 26 岁)参加了数据采集实验.所有志愿者均为右利手,无任何神经肌肉系统疾病史,日常工作环境为办公室.其中 3 人使用过 MYO 手环等肌电手势识别相关产品,另外 3 人没有接触过肌电任何相关产品或者实验.

实验开始前,志愿者坐在电脑办工作面前上,在实验人员的帮助下穿戴好肌电数据采集设备并告知实验的流程与风险.如图 6 所示,实验环境与调研中的办公室环境类似,桌子上放置着鼠标、键盘、触屏手机、A4 纸张、签字笔以及装有水的杯子.

开始实验时,志愿者可以随意选择图 6 中操作鼠标、使用键盘打字、操作触屏手机、阅读文献、使用杯子喝水和使用签字笔写字这 6 种操作中的任意一种,然后按照自己的习惯自主进行操作.操作开始时每次操作时间至少持续 3s.每种操作被要求执行总共 20 次,每两次操作之间志愿者可以自行选择停顿休息.实验过程被记录下来用于做数据处理的时候标记操作样本.

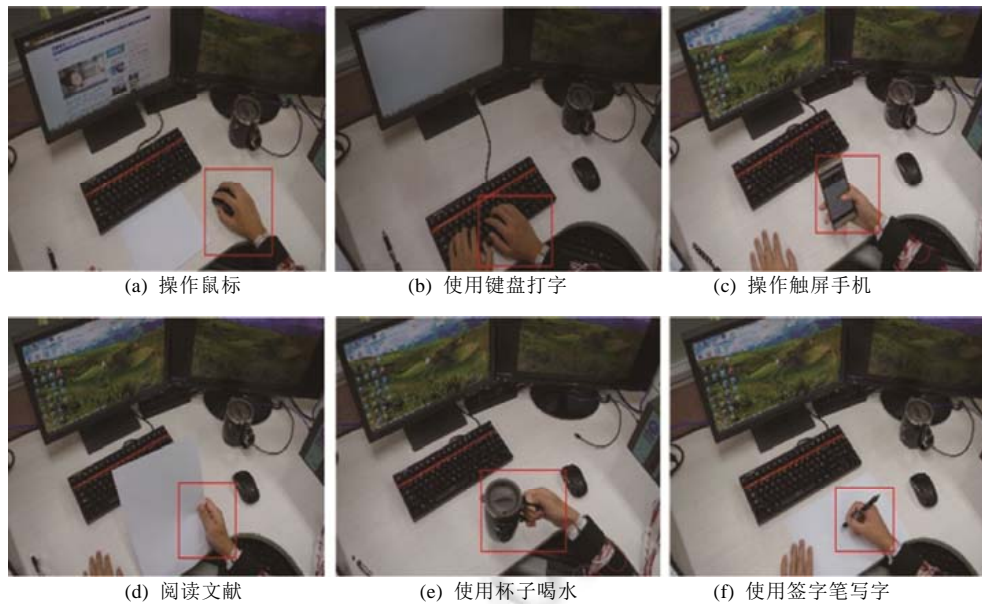


Fig.6 Manipulating gesture defined based on the field study in an office

图 6 基于办公室调研的操作手势

4.4 结果分析和讨论

通过用户实验,本实验得到了每个志愿者每种操作手势 20 次的肌电信号原始数据.根据实验过程的标记结果从每次操作的原始数据中提取 2 个长度为 500ms,同时间隔为 500ms 的数据段作为代表本次操作手势的样本数据,因此从实验原始数据中提取了 2560 个样本.从每个样本中提取出信号特征用于分类器训练和验证,以衡量本实验提出的基于组合特征和随机森林分类算法的手势识别模型的有效性和可用性.在验证方法中,本实验采用了十折交叉验证法,即 90% 的样本用于训练模型,剩余 10% 的样本用于测试.由于不同的用户之间肌电信号存在个体差异,为了减小个体差异带来的影响,本文采用了用户独立的验证方法,即每个用户的数据独自用于训练和测试.

图 7(a)显示了本文实现的操作手势识别系统实验中不同志愿者的操作手势的识别率,其中平均识别率达到了 81.4% (SD3.3),最高识别率为 86.64%.其中操作手势为操作鼠标(OM)、使用键盘打字(TP)、操作触屏手机(OS)、阅读文献(RP)、使用杯子喝水(DK)、使用签字笔写字(WR).实验结果说明,系统能够较好地识别用户的操作手势动作,并且具有较高的鲁棒性.为了判断对肌电产品的熟悉度是否对手势动作的识别率造成影响,我们把志愿者中曾经使用过肌电相关产品的志愿者(组 A,组员包括编号为 1、2、5 的志愿者)和首次接触肌电的志愿者(组 B,组员包括编号为 3、4、6 的志愿者).通过对比我们发现,组 A 的平均识别率为 82%,而组 B 的平均识别率为 80.7%,组 A 的识别率略微高与组 B 的识别率.这说明当用户使用我们的手势交互系统自然地做操作手势交互时,对系统设备的熟悉程度能带来识别率的轻微提升,但不是提升识别率的关键因素.

操作手势之间动作幅度及操作力度差别越大,被系统正确识别的概率越高.如图 7(b)所示,在 6 种操作手势中,操作鼠标的用户平均识别率为 88%,而使用键盘打字的用户平均识别率仅有 66%.部分操作手势在执行操作时手指拥有比较高的灵活度,执行过程的不同时间段反映在肌电信号上变化比较大,容易被系统错误识别为其他的操作手势.比如使用键盘打字操作容易被识别为写字、喝水等操作.另外,有些操作手势在动作上有较高的相似性,也容易出现识别混淆的情况,比如操作鼠标和操作手机两种操作.因此,在设计操作手势集的时候,要注意避免过于相似的手势和灵活度过大的手势出现在同一手势集里,以达到提高系统可用性和鲁棒性的目的.

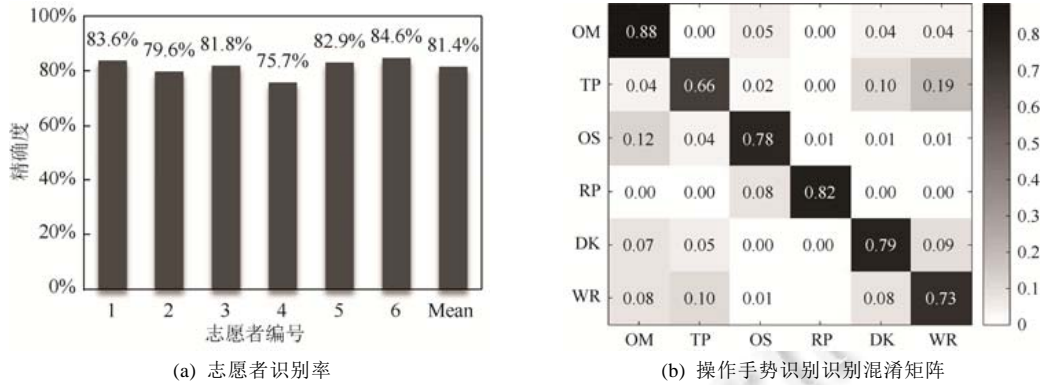


Fig. 7
图 7

(1) 传感器配置方式 vs. 识别率

在人机交互领域的研究和应用中,肌电传感器通常是配置成腕带的形式佩戴在手腕上或者是配置成臂环的形式佩戴在手臂上.在实例应用研究中,本文首次提出了一种新的肌电传感器配置方式——双臂环.这种方式既可以获取手臂更多位置的肌电信号,同时也便于穿戴.

通过实验数据集的实验结果发现,只在手腕配置传感器的情况下系统获得的手势识别率是 77.2%,只在手臂配置的情况下系统获得的手势识别率是 78.9%.而采用双臂环配置传感器的情况下,无需做参数调整或者特征优化,系统的手势识别率可以提升到 81.4%.这说明本文提出的传感器配置方式是有利于采集更加全面,更能代表手势的肌电信号,要优于传统的腕带或者单臂环配置方式.

(2) 训练样本 vs. 识别率

由于本文提出的手势交互系统是用户依赖的,这意味着每个针对每个新用户都需要先做训练采集数据才可以正常使用系统.为了研究训练样本与识别率的关系,分别使用从 3 个~40 个样本用于训练得到的识别率进行研究,如图 8(a)所示.从图中可以看到,当训练样本数从 3 个~15 个的区间里,识别率快速从 25% 提升到 80%,而之后再增加训练样本也难以带来识别率的提升.因此,本系统只需要少量的训练数据,即少量的训练时间,就可以达到比较稳定的识别率,不会因为大量的训练而对用户体验造成严重的影响.

(3) 分类算法 vs. 识别率

分类算法是手势识别系统的关键因素,选择了合适的分类算法可以提高手势识别率,带来手势交互系统使用体验的提升,因此选择适当的分类算法是手势交互系统应用研究重要的一个环节.本文在实验数据集的基础上实现并比较了几个较为常用的分类算法,包括逻辑回归、K 近邻算法、支持向量机、决策树、朴素贝叶斯网络,以及随机森林.图 8(b)显示了在办公室环境下的操作手势数据集在各个分类算法下的识别率比较结果.其中基于序列最小优化算法(sequential minimal optimization,简称 SMO)的支持向量机(support vector machine,简称 SVM)和随机森林算法(random forest)能取得较好的识别效果.尤其是随即森林算法达到了 81.4% 的手势识别率,因此在本文交互系统设计中采用了随机森林作为手势分类算法.

5 结论和展望

本文在总结和目前基于肌肉感知的手势交互相关工作的基础上,分析了交互手势的分类,归纳了该类型手势交互原语,并提出了基于肌肉感知的手势交互语义分析模型和系统结构模型,解决了以往研究成果中描述语言不统一,开发设计过程过于依赖应用场景的问题.从应用实例研究中可见,该模型能够从设计、开发到评价整个工作过程进行有效的指导.

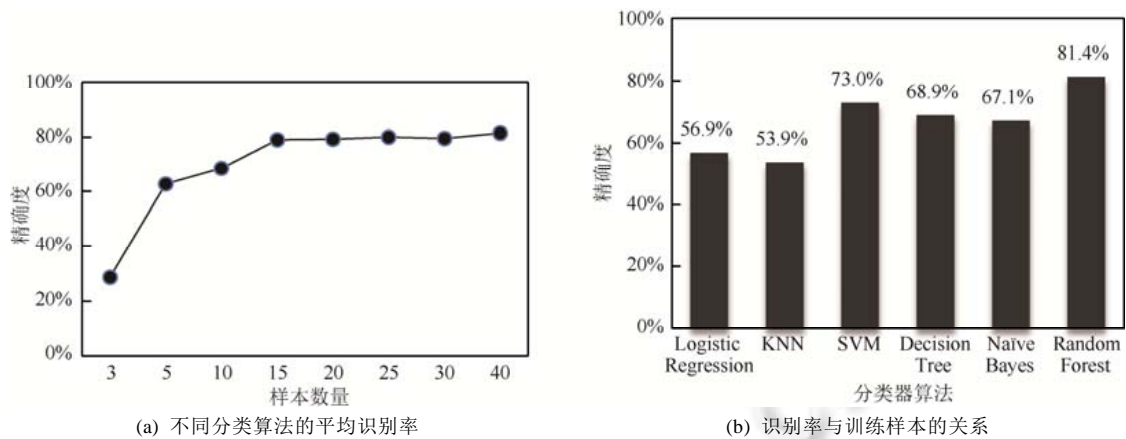


Fig.8

图 8

在目前而言,可用于获取手势交互的肌肉活动生理数据的方法相对较少,其中肌肉电信号由于能直接反应手部活动,因此最受研究者关注并且应用也最为广泛.另外,惯性传感器、压力传感器、人体组织电阻传感器等在近几年中也开始被用于手势识别应用,并取得不错的效果.因此我们有理由相信随着传感器技术的发展,会有越来越多的新技术可用于手势交互相关的生理信号获取和识别.

限于现有的设备,本文对于交互设备的归纳可能不完备,但是出现新的生理信号采集设备并不影响模型整体框架.本文所提的模型支持多通道输入/输出,但是模型并没有对多通道融合问题进行深入的分析讨论.多通道交互技术是未来发展的趋势,这也是未来可以继续开展的工作方向之一.

References:

- [1] Allanson J, Fairclough SH. A research agenda for physiological computing. *Interacting with Computers*. 2004,16(5):857–78.
- [2] Disselhorst-Klug C, Schmitz-Rode T, Rau G. Surface electromyography and muscle force: Limits in sEMG–force relationship and new approaches for applications. *Clinical Biomechanics*, 2009,24(3):225–235.
- [3] Saponas TS, Tan DS, Morris D, *et al.* Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In: *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2008)*. 2008, 515–524.
- [4] Saponas TS, Tan DS, Morris D, *et al.* Enabling always-available input with muscle-computer interfaces. In: *Proc. of the 22nd Annual ACM Symp. on User Interface Software and Technology*. ACM, 2009. 167–176.
- [5] Huang D, Zhang X, Saponas TS, *et al.* Leveraging dual-observable input for fine-grained thumb interaction using forearm EMG. In: *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2015)*. 2015. 523–528.
- [6] Amma C, Krings T, Böer J, *et al.* Advancing muscle-computer interfaces with high-density electromyography. In: *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2015)*. 2015. 929–938.
- [7] Dementyev A, Paradiso JA. WristFlex: Low-Power gesture input with wrist-worn pressure sensors. *ACM Symp. on User Interface Software and Technology*. ACM, 2014. 161–166.
- [8] Zhang Y, Xiao R, Harrison C. Advancing hand gesture recognition with high resolution electrical impedance tomography. *Symp. on User Interface Software and Technology*. ACM, 2016.
- [9] He LS, Ni HY, Song AG. A portable electromyograph (EMG) collection method and circuit realization. *Journal of Electronic Measurement and Instrument*, 2006,20(2):70–74 (in Chinese with English abstract).
- [10] Zhang X, Chen X, Li Y, *et al.* A framework for hand gesture recognition based on accelerometer and EMG sensors. *IEEE Trans. on Systems Man and Cybernetics—Part A Systems and Humans*, 2011,41(6):1064–1076.
- [11] Chen Y, Su X, Tian F, *et al.* Pactolus: A method for mid-air gesture segmentation within EMG. In: *Proc. of the CHI Conf. Extended Abstracts on Human Factors in Computing Systems*. ACM, 2016. 1760–1765.

- [12] Krasner GE, Pope ST. A cookbook for using the model-view controller user interface paradigm in Smalltalk. *Journal of Object-Oriented Programming*, 1988,1(3):26-49.
- [13] Coutaz J. Software architecture modeling for user interfaces. *Encyclopedia of Software Engineering*, 1993.
- [14] Pavlovic V, Sharma R, Huang TS. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 1997,19(7):677-695.
- [15] Chen YN, Lu L, Hu WP. PhysiolComp toolkit: Analysis and design for a physiological computing based human-computer interaction toolkit. *Chinese Journal of Computers*, 2015,38(12):2464-2476 (in Chinese with English abstract).
- [16] Goldinmeadow S. The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 1999,3(11):419-429.
- [17] Dong SH, Wang J, Dai GZ. *Human-Computer Interaction and Multimodal User Interface*. Beijing: Science Press, 1999 (in Chinese).
- [18] Paulson B, Cummings D, Hammond T. Object interaction detection using hand posture cues in an office setting. *Int'l Journal of Human-Computer Studies*, 2011,69(1-2):19-29.
- [19] <https://www.biopac.com/product-category/research/systems/mp150-starter-systems/>
- [20] Lee JC, Tan DS. Using a low-cost electroencephalograph for task classification in HCI research. In: *Proc. of the 19th Annual ACM Symp. on User Interface Software and Technology*. ACM, 2006. 81-90.
- [21] Aasman J, Mulder G, Mulder LJ. Operator effort and the measurement of heart-rate variability. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1987,29(2):161-170.
- [22] Inbar GF, Paiss O, Allin J, *et al.* Monitoring surface EMG spectral changes by the zero crossing rate. *Medical and Biological Engineering and Computing*, 1986,24(1):10-18.
- [23] Witten IH, Frank E, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd., Morgan Kaufmann Publishers, 2011.

附中文参考文献:

- [9] 何乐生,倪海燕,宋爱国.一种便携式肌电信号(EMG)提取方法及其电路实现. *电子测量与仪器学报*,2006,20(2):70-74.
- [15] 陈毅能,路璐,胡伟平. *PhysiolComp Toolkit: 一个生理计算交互工具箱的分析与设计*. *计算机学报*,2015,38(12):2464-2476.
- [17] 董士海,王坚,戴国忠. *人机交互和多通道用户界面*.北京:科学出版社,1999.



范俊君(1986-),男,广东清远人,硕士,CCF 学生会员,主要研究领域为人机交互,图像处理,生理计算.



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技术,虚拟现实.



黄进(1985-),男,助理研究员,CCF 学生会员,主要研究领域为人机交互技术,图形图像处理.



刘杰(1981-),男,博士,高级工程师,CCF 专业会员,主要研究领域为人机交互技术,图形图像处理.



王宏安(1963-),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为实时智能,用户界面.



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