

凌空手势识别综述

于汉超¹, 杨晓东^{1,2}, 张迎伟^{1,2}, 钟习^{1,2}, 陈益强^{1,2}

1. 中国科学院计算技术研究所; 移动计算与新型终端北京市重点实验室, 北京 100190
2. 中国科学院大学, 北京 100049

摘要 普适计算技术和可穿戴设备的快速发展对自然的手势识别技术提出了新的挑战: 应使用户尽可能摆脱对环境和输入设备的束缚, 与环境进行自然而有效的手势交互。凌空手势(mid-air gestures)识别是应对新挑战的一类有效的手势识别技术。本文首先对凌空手势的定义进行描述, 然后分析和总结现有的基于计算机视觉、超声波和电磁信号的三大类凌空手势识别技术, 并指出凌空手势识别技术的应用领域、开放性问题及未来发展方向。

关键词 人机交互; 凌空手势识别; 普适环境

普适环境是一个信息空间与物理空间融合的、充满计算与通信能力的环境。在该环境中, 计算以人为中心, 人机交互类似于人与人之间的自然交流方式, 用于计算的设备无处不在, 融入到人们生活的环境中, 并能够便捷地为人们提供所需要的服务; 人们可以随时随地透明地获得数字化的服务, 而使用计算设备的人却感知不到计算机的存在。麻省理工学院著名科学家 Michael Dertouzos 指出“未来的技术应该适应人的需求, 而不是与之相反”, 手势识别技术也应该能够主动去适应用户需求和环境, 而不应该要求用户和环境做出太多的变化。普适环境下的手势识别技术应主动感知与适应用户及其上下文环境, 使用户尽可能摆脱对输入设备、环境和规则的束缚, 与环境进行自然而有效的手势交互。

凌空手势(mid-air gestures)是一类非接触式的、能够使用户以徒手方式进行操作的空中手势, 其本质是一种不给用

户手势交互带来任何不便的自然人机交互方式。

凌空手势通过用户自然地利用手指、手腕和手臂动作表达其交互意图, 主要包括指、挥手、握拳、手掌转动等^[1], 具有更广阔的交互空间、更高的灵活度和更好的交互体验等特点。近年来, 在《钢铁侠》《少数派报告》等科幻电影中随处可见凌空手势交互的场景, 为电影添加了许多未来感和科技感(图1)。然而这种“未来”的交互手段已经悄然进入了人们的日常生活中, 例如微软公司(Microsoft Inc.)2009年发行的带有体感设备的Xbox, 突破传统的鼠标-键盘组合, 用户仅需要挥动手臂即可凌空操控游戏。凌空手势作为一种最贴近人-人交互的自然交互方式, 使得凌空手势识别技术成为人机交互(human computer interaction, HCI)研究中的热点问题之一。Kinect^[2]、Leap Motion^[3]、Soli^[4]等新型交互传感设备的出现, 为凌空手势识别注入了新的活力, 使其从传统的手掌、手



图1 《钢铁侠》、《少数派报告》中凌空手势操控场景
Fig. 1 Mid-air gestures in 《Iron Man》and 《Minority Report》

收稿日期: 2017-06-02; 修回日期: 2017-08-01

基金项目: 国家自然科学基金项目(61502456, 61572471); 国家国际科技合作专项(2014DFG12750)

作者简介: 于汉超, 助理研究员, 研究方向为人机交互、手势识别, 电子邮箱: yuhanchao@ict.ac.cn

引用格式: 于汉超, 杨晓东, 张迎伟, 等. 凌空手势识别综述[J]. 科技导报, 2017, 35(16): 64-73; doi: 10.3981/j.issn.1000-7857.2017.16.009

臂等大幅度动作的识别向手指精细动作的识别方向发展,为普通环境下的人机交互提供了更加自然和谐的解决方案。

凌空手势识别是普通环境下一种重要的新型交互技术,是手势识别未来的发展趋势。本文对凌空手势识别技术进行综述,分析和总结现有凌空手势识别方法、凌空手势的应用领域,并指出凌空手势识别存在的问题及未来发展方向。

1 凌空手势识别技术

凌空手势识别技术主要包括手势数据感知、手势分割、有效特征提取和手势识别模型构建等关键模块,依据凌空手势数据感知设备的不同,目前主要分为基于计算机视觉、超声波和电磁信号的三大类凌空手势识别技术。

1.1 基于计算机视觉的凌空手势识别

随着各种彩色摄像头和 Kinect 等深度摄像头的发展,基于计算机视觉的主动感知方式的凌空手势识别近年来越来越受到国内外研究人员的关注,并取得了大量的研究成果。

1.1.1 基于彩色摄像头的凌空手势识别

为了实现基于彩色摄像头的凌空手势识别,首先需要将人手从图像中分割出来,因此,基于肤色信息人手分割的手势识别技术受到了研究人员的青睐。Lars 等^[9]选择肤色特征,利用粒子滤波检测手掌及手指,之后用模板匹配的方法进行

手势识别。Bjom^[6]采用肤色与运动特征跟踪人手,使用最近邻分类器进行手势识别。Kaufmann 等^[7]通过肤色信息进行人手分割,提出了利用智能进化算法进行手势识别的方法。Weng 等^[8]通过贝叶斯肤色模型进行人手分割,采用多特征融合的方法进行手势识别。Flasinski 和 Myslinski^[9]构建了一个高斯肤色模型,并在此基础上提出了一种手势图解析分类法用于手势识别。任海兵、徐光祐和林学闯^[10]融合肤色、运动和边缘等多种信息提取能够反映人手结构特性的特征线条,通过抽取平移不变的平面模型参数进行手势识别。朱继玉等^[11]通过对视频流中空域和时域上的背景、运动、肤色等执行模糊运算,分割出人手,使用图像金字塔提取手势特征进行手势识别。杨波等^[12]利用高斯亮度模型分割肤色区域,通过“搜索窗口”筛选肤色区域实现手势定位,提出了一种基于手势空间分布特征的手势识别算法。Beh 等^[13]提出一种基于手部轨迹角度的凌空手势建模方法,利用隐马尔科夫模型实现了对动态凌空手势的精准识别。Song 等^[14]提出一种多阶段的随机森林算法,利用一个彩色摄像头同时识别手部形状并估计手部的深度位置,进而实现了在智能手机、智能手表和智能眼镜等终端设备上的凌空手势识别,如图 2 所示。目前来看,基于肤色信息人手分割的凌空手势识别方法较多,但此类方法易受到肤色的限制和环境光照的影响。



(a) 智能手机中缩放地图 (b) 手机游戏中进行射击 (c) 凌空控制电脑音量 (d) 凌空控制智能手表进行照片分享 (e) 凌空控制智能眼镜进行联系人选择

图2 利用设备自带的彩色摄像头实现的凌空手势识别应用
Fig. 2 Color camera based mid-air gesture applications

1.1.2 基于深度摄像头的凌空手势识别

基于光学和光波传感器进行深度信息获取和骨骼跟踪,进而实现基于手势识别的体感人机互动,是近年来手势识别领域的研究热点。微软的基于结构光编码及人体骨架实时跟踪技术的 Kinect 体感装置是这方面的代表性成果,其他还包括 PrimeSensor、Xtion PRO、SoftKinetic、Bumblebee 等,如图 3 所示,其特点是能够以相对较好的实时性获取深度图像序列,实现 3D 数据捕捉、人体识别和骨骼跟踪^[15],并可进一步支持复杂手势识别应用的研究。

基于深度摄像头的新型视觉方法主要是利用深度传感器获取的深度信息克服复杂环境与杂乱背景对手势分割的影响,同时能够很好地适应不同的光照条件^[16]。Ren 等^[17]提出了一个基于深度摄像头 Kinect 的对人手种类和形状变化不敏

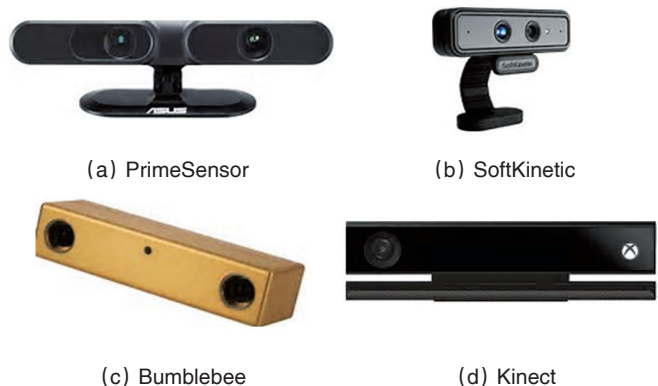
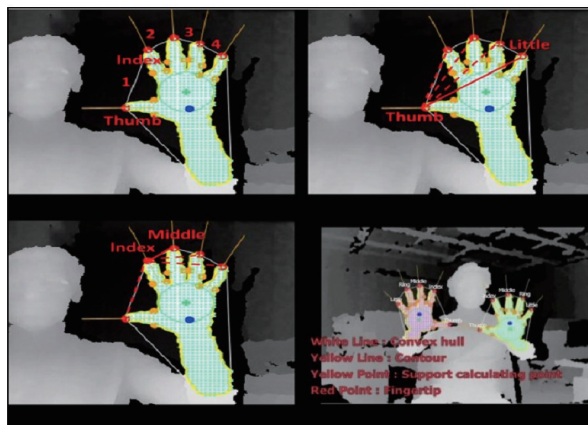


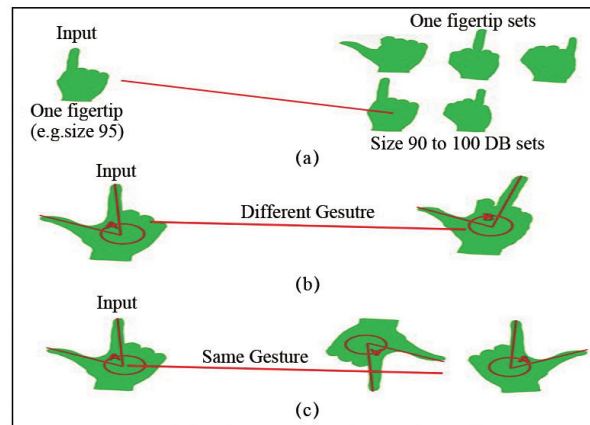
图3 深度摄像头
Fig. 3 Depth camera

感的手势识别系统,该系统同时采用深度信息和颜色信息从杂乱的环境中检测出人手,并利用形状距离度量方法对数字0~9和4种数学运算符号用14种手势进行识别。Wang等^[18]基于Kinect深度摄像头,将人手形状及相应的纹理和深度信息组合形成“超像素”,提出了一种基于超像素的手势距离度量方法,提高了手势识别的鲁棒性。Elmezain等^[19]使用Bumblebee立体摄像头,通过肤色信息提取人脸和人手,利用深度信息区分出人手,进而实现了基于隐马尔可夫模型(hidden Markov model, HMM)的孤立手势和连续手势识别。Sanchez等^[20]使用深度摄像头提出了一种新的人体手势识别方法,该方法包括光照不变的肤色分割,并支持机器人和人之间通过动作进行通信。以上方法虽然利用了深度信息,在一定程度上提高了人手分割效果,但仍然没有完全摆脱肤色的限制。

Molina等^[21]利用基于time-of-flight的深度摄像头,将在距离深度摄像头最近点20 cm以内的像素点作为人手进行分割,实现了一个用于控制窗口应用程序的手势识别系统,此方法对人手的分割虽然摆脱了肤色限制,但是当人手不位于人体最前面时将无法实现对人手的准确分割。文献[22]利用Kinect采集的深度数据进行手指检测,然后将手指检测的结果用于手势识别,如图4所示,但该方法在实际操作中,用户需要将手掌张开并正对深度摄像头才能较好地对指尖进行检测。因此,以上方法均没有完全摆脱人手姿态的限制。文献[23]为了解决上述问题,提出了一种手掌姿态自适应的单指尖动态识别方法,并将该方法应用到了数字0~9和字母a~z等36种凌空手势的识别^[24]和脑卒中疾病的检测预警中^[25-26],取得了较好的效果。



(a) 手指检测



(b) 基于手指检测的手势识别

图4 基于深度摄像头的手指检测和手势识别方法

Fig. 4 Finger detection and gesture recognition methods based on depth camera

1.2 基于超声波的凌空手势识别

基于计算机视觉的凌空手势识别易受可视角的范围限制和环境光照的影响,而超声波固有的特性能够很好地解决这些问题,因此,国内外学者广泛开展了基于超声波的凌空手势识别技术的研究。超声波是指频率超过20 kHz的声音信号(正常成年人的听力范围是20 Hz~20 kHz),利用超声波的多普勒效应(Doppler effect)和飞行时间(time of flight, TOF)可以表征手部的运动情况,从而实现对凌空手势的准确识别。值得关注的是,笔记本电脑、智能手机等常见终端内置的麦克风和扬声器即可发射和接收较低频率的超声波信号,为基于超声波的凌空手势识别技术研究推广提供了普适化的硬件基础。下面将介绍基于超声波的凌空手势识别技术的原理。

多普勒效应由奥地利物理学家及数学家克里斯琴·约翰·多普勒(Christian Johann Doppler)于1842年提出,是指物体辐射的波长随波源和观测者的相对运动而产生变化。生活中常见此类现象,例如当一辆汽车迎面驶来时,会听到声

音越来越高;而当车离去时,声音会越来越低(声音的高低与频率有关)。如图5所示,在运动的波源前面,波被压缩,波长变得较短,频率变得较高,发生蓝移;在运动的波源后面会产生相反的效应,波长变得较长,频率变得较低,发生红移。根据波红(蓝)移的程度,可以计算出波源循着观测方向运动的

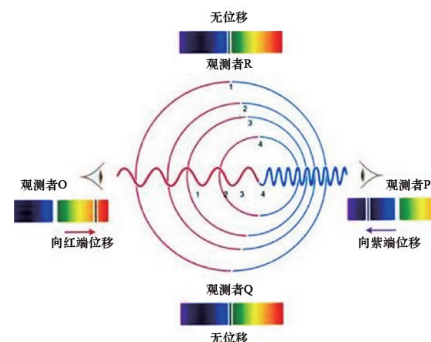


图5 多普勒效应

Fig. 5 Doppler effect

速度,如下所示:

$$f' = f \cdot \left(\frac{v \pm v_0}{v \mp v_0} \right) \quad (1)$$

式中, f 为信号基频,即发射的原始频率; v 为声波在空气中传播的速度; v_0 为物体相对声源运动的速度; f' 为多普勒效应之后接收到的频率。

当扬声器持续发射超声波信号时,用户进行手势运动将反射超声波;对超声波接收设备麦克风而言,人的手部将被视为声源,手部运动产生的多普勒效应将反映在接收到的超声波信号中。通过对接收的超声波信号进行时频分析,即可判断手势移动的方向和速度。如图6所示,当手势相对于麦克风运动时,将引起信号频率的含量发生变化。

除了利用多普勒效应,超声波测距技术同样被应用于凌空手势识别中。超声波测距技术主要利用信号在两个异步收发装置之间往返的飞行时间测量节点间的距离,如图7所示。由超声波发射器发射脉冲信号,通过探测发射脉冲和接收脉冲之间的相位差感知手势的轨迹,从而实现手势识别。

早在2009年就出现了利用超声波进行凌空手势识别的方法^[7],但由于需要特定的传感器搭建识别系统,难以进行实用和推广。如图8(a)所示,该装置由一个超声波发射器和3个超声波接收器组成,通过感知多个超声波接收装置的频率变化,实现凌空手势识别。随着个人计算机(PC)、智能手机、智能可穿戴设备等移动终端的硬件升级以及计算能力的提

升,面向移动终端的超声波凌空手势交互技术蓬勃发展。

在PC平台上,微软研究院 Gupta 等^[28]设计的 Soundwave 系统,如图8(b)所示,利用设备内置的传感器发射和采集 18 kHz 声波信号,通过信号的频域特征判断手势的方向和速度等属性,实现了对向前、向后、剪刀手等凌空手势的有效识别;Pittman 等^[29]在 Soundwave 系统的基础上提出了 Multiwave 系统,使用双声道扬声器分别发射不同频率的超声波,将麦克风采集到的频率偏移转换到欧式空间,从而生成手势的运动路径、表征相应的动态手势,实现了二维空间的凌空手势识别。

在智能手机 Android 平台上,Qifan 等^[30]设计的 Dolphin 系统,如图8(c)所示,利用内置的扬声器和麦克风发射和接收连续的 21 kHz 超声波信号,提取多普勒效应相关的频域特征,采用机器学习模型实现了多达 17 种凌空手势的识别;Wang 等^[31]利用智能手机的扬声器发射脉冲式超声波,通过比较原始信号与接收信号的相位,获得细粒度的手势运动方向和距离,实现了二维空间的凌空手势识别,如图8(d)所示;Nandakumar 等^[32]根据声纳原理发射超声波信号,同时利用麦克风采集手指反射的信号,利用无线通信中的正交频分复用 (orthogonal frequency division multiplexing, OFDM) 技术识别手势,在毫米级别的误差上实现了二维手势的跟踪,即使智能设备放在口袋中也能实现精准的手势跟踪。

在智能可穿戴设备中,加州大学伯克利及戴维斯分校的

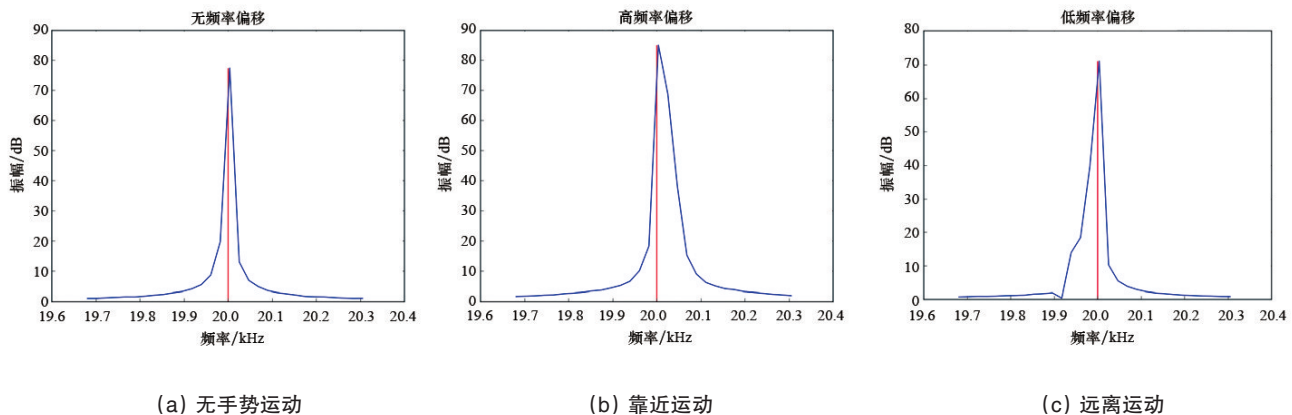


图6 手势的相对运动

Fig. 6 Relative movement of hand gestures

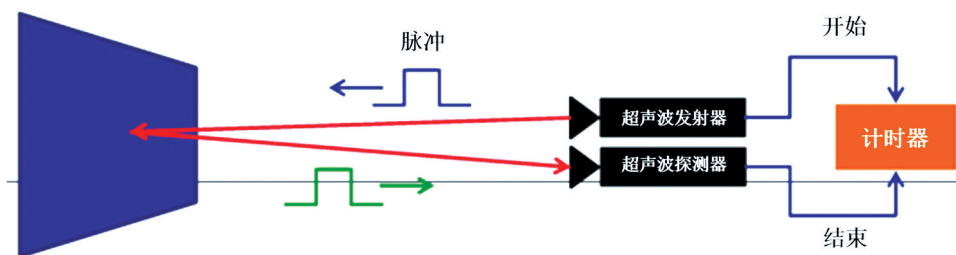


图7 超声波测距技术

Fig. 7 Ultrasound based distance measuring technology

研究人员研制出一种面向可穿戴设备的超声波微型芯片,能够支持三维空间中的凌空手势识别,极大扩展了可穿戴设备的交互空间,如图8(e)所示;中国科学院计算技术研究所的杨晓东等^[33]针对可穿戴设备的计算资源和功耗有限的问题,提出了基于Goertzel算法和隐马尔可夫模型的超声波凌空手势识别方法,不仅保证了手势识别的准确率,而且有效降低了手势识别过程中的资源消耗和功耗。

除了利用移动终端内置的传感器,挪威Elliptic Labs公

司^[34]利用设置在终端上方的两对40 kHz超声波发射器和接收器,根据用户手部操作对超声波产生的影响来确定手部的位置,实现了围绕设备180°的全方位三维凌空手势交互,如图8(f)所示。

总的来说,超声波手势识别技术能够在一定程度上克服环境光照的影响并扩大了手势操控的范围,但由于其鲁棒性较弱、信噪比需求较高等问题还有待进一步研究和解决,所以将其应用于实际生活仍有一段距离。

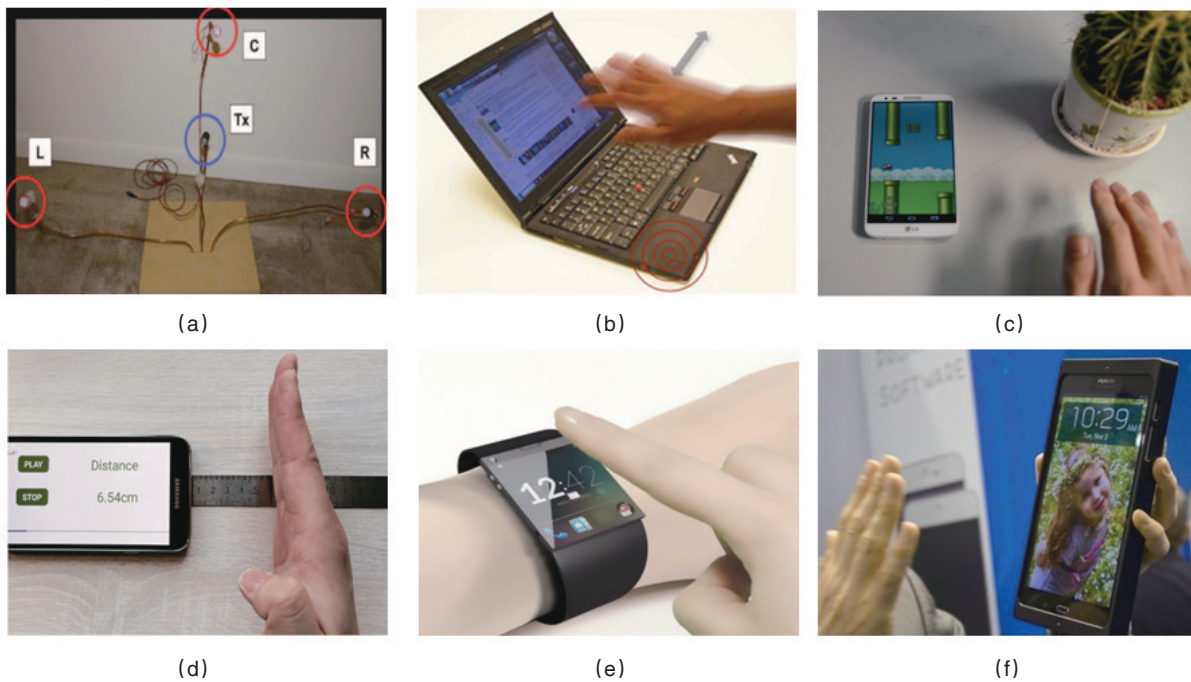


图8 基于超声波的凌空手势识别应用
Fig. 8 Ultrasound based mid-air gesture applications

1.3 基于电磁信号的凌空手势识别

早在20世纪30年代,雷达电磁波(radio frequency, RF)因具有不受光线、噪声、大气环境的影响,能够穿越大部分材料等优势,已经被应用于构建快响应、高精度的物体移动监测和追踪系统,例如,空中交通管制、潜水艇雷达、地理监测等。其原理如图9所示,利用原始RF信号与接收的反射信号之间的时间间隔、相位、频率差异及能量衰减等信息,可以推

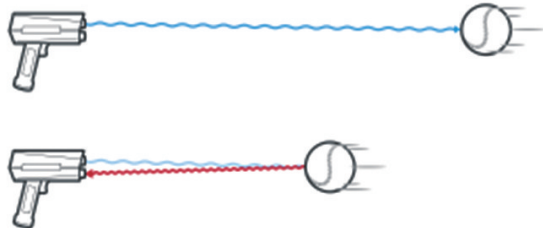


图9 电磁波基本原理
Fig. 9 Transmission of electromagnetic wave

测出目标物体的距离、速度、形状、大小、表面光滑程度和材料等特性。如果目标物体是人的手部,则接收到的RF反射信号是手部多个动态散射中心(如手掌和多个手指)反射信号的叠加;在特定的手势下,信号将呈现特定的模式,通过对叠加接收信号的分析 and 转换,进而实现对手势的识别。

在此方面最具有代表性的工作是谷歌Touchstone团队的Solli项目。由于移动可穿戴设备的微型化、计算资源有限、操控距离近(小于5 mm)等要求,现有的RF技术无法实现基于可穿戴设备的凌空手势识别,该团队针对以上问题,设计并实现了从底层硬件到顶层应用的一整套端到端的解决方案。该项目于2014年启动,历经两年时间发展,至2016年成功将芯片面积缩小了70%、功耗从1.2 W下降到0.054 W(下降了22倍)、算法效率提升了256倍,并将其应用于智能手表等微型的移动可穿戴设备中。

人的手部由手掌、手指等多个运动单元组成,因此Solli接

收手部反射的雷达信号是多个离散的、动态散射中心反射信号的叠加。电磁雷达信号的反射参数不仅依赖于信号频率,并且随着手部纹理的变化而改变,通过对信号反射参数的描述,能够实现对手部的形态估计和运动跟踪,如图10所示,其中 r 和 ρ 为电磁雷达信号的反射参数。

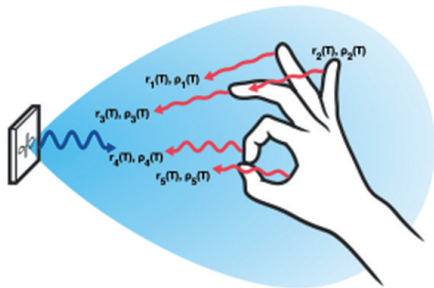


图10 Soli接收的信号是多个散射中心反射信号的叠加
Fig. 10 Entire hand is illuminated by Soli and a superposition of reflections from multiple dynamic scattering centers is measured

与在空间维度上进行目标识别的经典雷达技术不同,Soli提出了基于高时间分辨率的感知机制,通过超高帧速测量人手的反射信号,检测微小且复杂的手部动作和手势,进而提取与之相对应的精细信号变化。为了分析手部反射的信号,Soli以一定时间间隔重复发射雷达信号,发射的雷达信号频率从1 kHz逐渐增加至10 kHz;在每一个信号发射周期内,按照相应的采样频率接收反射信号。与雷达信号重复发射的频率相比,人的手部运动相对较慢,可以认为反射信号的特性在单个雷达重复时间间隔内近似为恒定。通过对单个雷达重复间隔内信号的处理,提取瞬时的手部姿势、方向等特征;综合多个连续间隔内提取的瞬时特征,利用其变化情况获得手势的动态运动,例如手部移动的速度等,如图11所示。

与传统的基于图像的凌空手势识别方法不同,Soli不基于特征推断手部的骨骼结构或者手指的空间方向,而是将提取的特征直接应用于机器学习模型中,与更传统的统计特征一起表征手势,并进行相应的识别。这样的设计能够有效地减少计算负荷,即使在智能手表或物联网中一些低功率设备

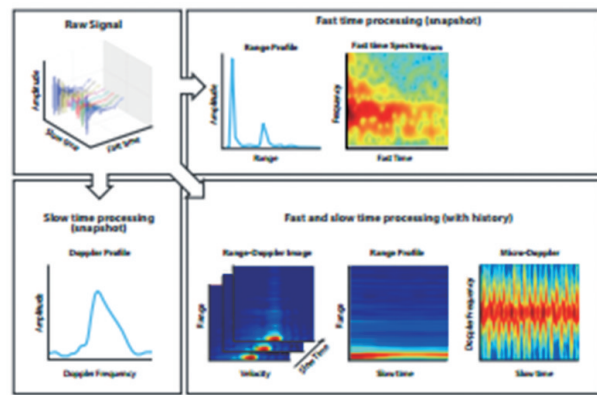


图11 Soli进行信号处理的两个时间粒度
Fig. 11 Combination of slow time and fast time processing

上也能够实现基于Soli的凌空手势识别。另外,为了保证凌空手势识别模型在计算速度、运算空间和泛化能力等方面的性能,Soli采用随机森林(random forest)作为分类模型;并利用贝叶斯过滤器(Bayesian filter),基于最近预测结果加权平均的时序先验和基于应用的情境先验,对识别的结果进行修正,进一步提升手势识别的准确率和鲁棒性。

Soli识别凌空手势的关键是通过提取信号特征对手部运动的动作、范围和速度等属性进行表征,因此Soli擅长识别动态手势(action gesture)而非静态手势(sign gesture)。为此,Soli提出了“虚拟工具(virtual tools)”的概念,通过手指运动模拟按钮、滑块等工具的使用过程,如图12(a)“虚拟按钮(virtual button)”、图12(b)“虚拟滑块(virtual slider)”所示。另外,对一些传统的动态凌空手势,如图12(c)、(d)所示,Soli同样能够表现出良好的性能。

总的来说,Soli为基于电磁信号的凌空手势识别这一全新的人机交互方式提供了从底层硬件到顶层应用的、端到端的完整解决方案,对未来该领域的工作具有先驱指导意义,为微型设备上凌空手势的识别提供了可能。然而,如何克服电磁信号本身所固有的衰减、多路、碰撞等问题,以及如何设计更加行之有效的机器学习模型等,均是未来研究者亟需解决的问题。

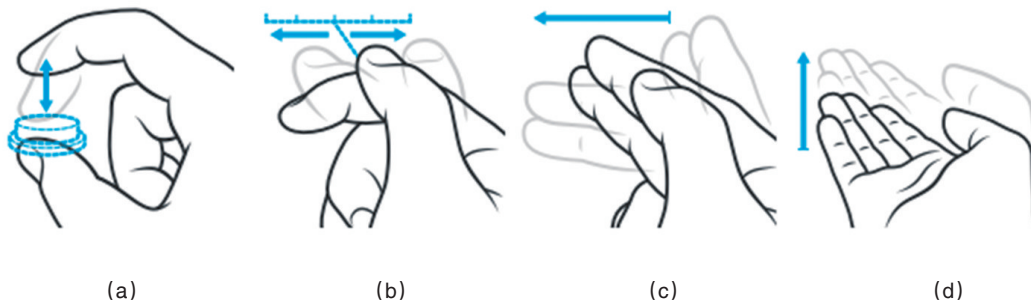


图12 动态凌空手势集合
Fig. 12 Dynamic mid-air gesture set

2 凌空手势识别技术的应用

凌空手势为用户提供了一种更加自然、方便的交互方式,能够解决娱乐、日常生活、医疗等领域中某些传统接触式交互方式无法满足的需求,下面将介绍几类凌空手势识别技术的具体应用。

2.1 凌空手势识别在日常生活中的应用

凌空手势识别在日常生活中的应用主要体现在对电视、平板电脑等常用设备的自然操控以及对汽车的安全辅助驾驶中。

三星电子、TCL集团股份有限公司、乐视致新电子科技(天津)有限公司等公司在其电视产品中加入了凌空手势控制功能(如三星智能电视 ES8000),通过电视自带的体感摄像头捕获用户挥手、握拳、移动、拖拽等常见的凌空手势,实现对电视的节目切换、音量调节等操控,如图 13 所示。阿尔卡特公司(Alcatel)设计了一款适应于厨房使用情境(烹饪时双手沾满油渍、面粉等,同时需要查阅菜谱,基于鼠标、键盘、触摸屏的传统操控方式已难以满足)的平板电脑 OneTouch Xess,利用深度摄像头实现凌空手势操控,如图 14 所示。

在汽车安全驾驶方面,传统的人机交互技术需要用户通过按键或触摸屏幕的方式与手机和车机进行交互,使得驾驶员不得不从前方道路转移视线且至少占用单手进行触摸输入,从而导致危险的“盲目”驾驶,目前因驾驶时进行触摸式手机输入导致的交通事故率已经上升到第4位;相比之下,凌空手势交互只需驾驶员自然地做出手势,无需转移注意力就能获得手机或车机的反馈,如图 15 所示。此方面的代表性工作是中国科学院计算技术研究所和深圳市车音网科技有限公司联合承担的科技部国际科技合作专项项目,将基于超声

波的凌空手势交互与语音识别技术相结合,实现了基于自然人机交互技术的安全驾驶辅助系统,使驾驶时所有操控符合“手眼同向”的安全驾驶原则,极大地保障了行车安全。

2.2 凌空手势识别在医疗领域中的应用

医疗领域中,凌空手势识别在方便医生诊疗和辅助患者康复方面发挥了重要作用。

医生在手术过程中,由于无菌的严格要求,外科医生在术中进行病人资料查阅时难以通过鼠标、键盘、触摸屏等直接操控电脑,而凌空手势识别技术能够帮助医生无需接触即可进行操控,能够极大提高手术效率。例如,协助外科医生浏览 MRT 图像的 Gestix 系统等^[35-37](图 16),都是该方面比较成熟的产品,已应用于实际的医疗环境中。

在辅助患者康复方面,主要是针对身体运动功能损伤的病人,帮助他们进行康复训练,或者在康复期间辅助进行日常活动。此方面具有代表性的是 Kuno 等^[38]设计的一款支持凌空手势操控的轮椅(图 17),用户只需要通过手势就能控制轮椅的活动,方便了病人的日常生活;为缩短截肢病人义肢训练的时间,Phelan 等^[39]开发了一种肌电义肢手臂训练原型系统,融合了虚拟现实技术,可使病人在虚拟场景中开展义肢训练,提高病人康复的主动性。

2.3 凌空手势识别在智能机器人领域中的应用

随着机器人技术的发展,智能机器人已经逐渐成为了人们日常生活中的重要组成部分,而凌空手势不仅提供指令类信息,而且能够提供含有空间信息的操控命令,在与智能机器人的交互中具有显著的必要性^[40]。例如,常见的凌空手势“指”,不仅可以为机器人提供移动位置的指令,而且能够明确指出机器人位置移动的方向,如图 18 所示。目前,比较成



图 13 智能电视
Fig. 13 Smart TV



图 14 OneTouch Xess
Fig. 14 OneTouch Xess

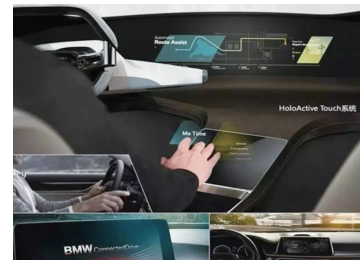


图 15 安全驾驶
Fig. 15 Safe Driving

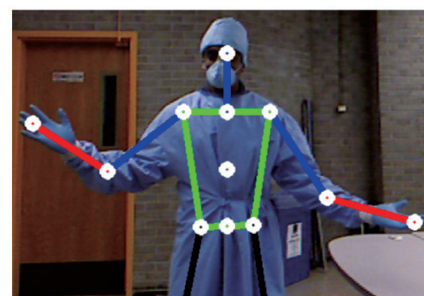


图 16 医疗领域凌空操控系统
Fig. 16 Mid-air gesture in medica applications

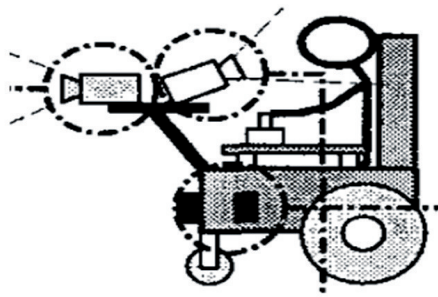


图 17 凌空手势操控轮椅

Fig. 17 Mid-air gesture controlled wheelchair

熟的产品主要有:深圳优必选科技有限公司(UBTECH)推出的仿“人脑”智能机器人Alpha,支持用户通过凌空手势操控Alpha完成股市、天气、航班等信息查询和拨打电话等功能,如图19所示;ZeroUI公司推出了面向儿童的智能机器人Ziro,

通过“挥手”手势便可指挥其进行旋转,如图20所示。

2.4 凌空手势识别在娱乐领域的应用

凌空手势识别技术的进步加速了体感游戏的发展。体感游戏是通过肢体动作变化和肢体感受进行操控的一类新型电子游戏,不仅能运动健身、有益于身心健康,而且能够让玩家“全身”投入到游戏当中,享受体感互动的体验。Leap Motion、Microsoft Kinect 等体感设备的出现开启了体感游戏的新篇章(图21)。带有体感设备Kinect的Xbox 360游戏机是微软于2009年推出的一款体感游戏机,截至2012年9月,Xbox 360全球累积销售量达到7000万部,《生化奇兵》《荒野大镖客:救赎》《水果忍者》等都是经典的Xbox 360游戏;《Boom Ball》《Dropchord》《Block54》是依靠Leap Motion进行操控的体感游戏,在为用户带来娱乐体验的同时,也展现了Leap Motion独特的凌空手势操控体验。



图 18 人与智能机器人进行凌空手势交互
Fig. 18 Mid-air gesture interaction between human and smart robot

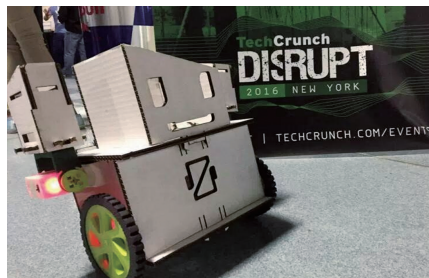


图 19 仿“人脑”智能机器人阿尔法
Fig. 19 Intelligent robot Alpha



图 20 面向儿童的智能机器人Ziro
Fig. 20 Intelligent robot Ziro



(a) 基于Kinect的体感游戏



(b) 基于Leap Motion 的体感游戏

图 21 体感游戏界面

Fig. 21 Motion sensing game interface

3 开放性问题

3.1 凌空手势交互范式设计

凌空手势交互作为最接近“人-人交互”的方式之一,“自然性”是其优于其他交互方式的主要特点。“自然性”在此表现为交互方式符合自然人的交互行为习惯,即凌空手势的交互语义集合是自然人手交互语义集合的子集,若以庞大而复杂的自然人手交互语义集合作为凌空手势交互设计的参考,容易造成凌空手势的自然性和识别准确性之间的冲突。

因此,如何设计更加自然且易识别的交互范式,一直是凌空手势交互的重要问题之一,在此方面,交互界面范式设计、凌空手势集合设计等方面值得关注。

3.2 用户个性化问题

精准的凌空手势识别是良好交互体验的前提和基础,然而凌空手势很大程度上依赖于用户习惯,易导致识别模型面对新用户时的“漂移”现象,影响识别效果。凌空手势的主体是人的手部,手作为人类身体最灵活的部分之一,具有很高

的自由度,并且凌空手势在空中操作,具有较大的波动性。因此即使定义了凌空手势规范,交互过程中手移动的速度、幅度和角度等同样会导致执行同一种手势时“轨迹”的差异,甚至具有不可再现性。除了主观因素造成的手势差异,一些客观因素也会影响手势交互的性能,例如在基于表面肌电的解决方案中,肌肉、毛发、汗液等因素会对表面肌电信号产生较大的影响。

此外,由于在手势识别模型的训练阶段将所有用户的凌空手势数据进行采集、标定和训练是不现实的,训练得到的只是能够适用于部分用户的离线模型,如何动态调整模型以适应特定的用户习惯,是一个值得探索的新方向。在此意义上,用户间的迁移学习、用户无关的特征提取等方法值得关注。

3.3 从单模态到多模态

传感技术的发展推动着凌空手势交互技术的革新,为之提供了新的解决方案。但与此同时,任何单个模态的感知手段都存在固有的缺陷和不足,限制了以此为基础的凌空手势交互的自然性和用户体验,例如,基于计算机视觉的手势识别方法在精细手势的识别上存在不足,因此,融合多模态传感器的凌空手势识别方法成为一个值得关注的研究方向。从传感的角度看,不同尺度的手部运动采用不同模态的传感器进行针对性感知,能够极大丰富手势集合;同时,不同模态数据间的相互协同,也有助于提高手势识别的精准度。尽管目前已经有关于此方面的相关工作,如将摄像头与加速度计结合起来,但尚处于起步阶段。

4 结论与展望

随着普适计算技术的发展,计算资源将以各种形态融入到生活的每个角落,使人们的日常生活充满着人机交互。凌空手势交互将更加通用自然,通过自适应交互对象,用户可以在不同的应用界面和场景之间实现无缝切换和自由交互;同时,凌空手势交互将更加精准细微,通过与力学触觉反馈等技术的结合,实现高精度凌空交互,逐渐取代日常的输入设备,例如键盘、鼠标、遥控器甚至汽车方向盘等,用户依靠双手即可进行自然交互;另外,凌空手势交互将更加丰富多样,随着虚拟现实技术、增强现实技术以及3D立体成像技术的发展,凌空手势识别技术将与之深度融合,实现3D图形的设计、虚拟场景下的自然互动等,从而更加激发人们的主动性和创造力。

总之,凌空手势作为最自然的人机交互方式之一,将在未来自然人机交互中发挥不可替代的作用。

参考文献(References)

[1] Cabreira A T, Hwang F. An analysis of mid-air gestures used across three platforms[C]//Proceedings of the 2015 British HCI Conference. New York: ACM, 2015: 257-258.
[2] Zhang Z. Microsoft kinect sensor and its effect[J]. IEEE Multimedia, 2012, 19(2): 4-10.

[3] Khademi M, Mousavi Hondori H, McKenzie A, et al. Free-hand interaction with leap motion controller for stroke rehabilitation[C]//Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems. New York: ACM, 2014: 1663-1668.
[4] Lien J, Gillian N, Karagozler M E, et al. Soli: Ubiquitous gesture sensing with millimeter wave radar[J]. ACM Transactions on Graphics (TOG), 2016, 35(4): 142.
[5] Bretzner L, Laptev I, Lindeberg T. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering [C]//Fifth IEEE International Conference on Automatic Face and Gesture Recognition. Washington D C: IEEE, 2002: 423-428.
[6] Stenger B. Template-based hand pose recognition using multiple cues [C]//ACCV 2006: Computer Vision. Berlin Heidelberg: Springer, 2006: 551-560.
[7] Kaufmann B, Louchet J, Lutton E. Hand posture recognition using real-time artificial evolution[C]//Applications of Evolutionary Computation. Berlin Heidelberg: Springer, 2010: 251-260.
[8] Weng C, Li Y, Zhang M, et al. Robust hand posture recognition integrating multi-cue hand tracking[C]//Edutainment 2010: Entertainment for Education. Digital Techniques and Systems. Berlin Heidelberg: Springer, 2010: 497-508.
[9] Flasiński M, My-liński S. On the use of graph parsing for recognition of isolated hand postures of Polish Sign Language[J]. Pattern Recognition, 2010, 43(6): 2249-2264.
[10] 任海兵, 徐光祐, 林学闯. 基于特征线条的手势识别[J]. 软件学报, 2002, 13(5): 987-993.
Ren Haibin, Xu Guangyou, Lin Xueyin. Hand gesture recognition based on characteristic curves[J]. Journal of Software, 2002, 13(5): 987-993.
[11] 朱继玉, 王西颖, 王威信, 等. 基于结构分析的手势识别[J]. 计算机学报, 2006, 29(12): 2130-2137.
Zhu Jiyu, Wang Xiyang, Wang Weixin, et al. Hand gesture recognition based on structure analysis[J]. Chinese Journal of Computers, 2006, 29(12): 2130-2137.
[12] 杨波, 宋晓娜, 冯志全, 等. 复杂背景下基于空间分布特征的手势识别算法[J]. 计算机辅助设计与图形学学报, 2010, 22(10): 1841-1848.
Yang Bo, Song Xiaona, Feng Zhiquan, et al. Gesture recognition in complex background based on distribution features of hand[J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(10): 1841-1848.
[13] Beh J, Han D, Ko H. Rule-based trajectory segmentation for modeling hand motion trajectory[J]. Pattern Recognition, 2014, 47(4): 1586-1601.
[14] Song J, Sörös G, Pece F, et al. Real-time hand gesture recognition on unmodified wearable devices[C/OL]. [2017-03-31]. <http://www.vs.inf.ethz.ch/publ/papers/soeroesg-cvpr2015-MobileGestures.pdf>.
[15] Kar A. Skeletal tracking using microsoft kinect[R/OL]. [2017-03-31]. http://people.eecs.berkeley.edu/~akar/IITK_website/cs397/Skeletal%20Tracking%20Using%20Microsoft%20Kinect.pdf.
[16] Doliotis P, Stefan A, McMurrough C, et al. Comparing gesture recognition accuracy using color and depth information[C]//Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments. New York: ACM, 2011: 20.
[17] Ren Z, Meng J, Yuan J, et al. Robust hand gesture recognition with kinect sensor[C]//Proceedings of the 19th ACM International Conference on Multimedia. New York: ACM, 2011: 759-760.
[18] Wang C, Liu Z, Chan S C. Superpixel-based hand gesture recognition with kinect depth camera[J]. IEEE Transactions on Multimedia, 2015, 17(1): 29-39.
[19] Elmezain M, Al-Hamadi A, Appenrodt J, et al. A hidden markov model-based isolated and meaningful hand gesture recognition[J]. Interna-

- tional Journal of Electrical, Computer, and Systems Engineering, 2009, 3(3): 156-163.
- [20] Gonzalez-Sanchez T, Puig D. Real-time body gesture recognition using depth camera[J]. Electronics Letters, 2011, 47(12): 697-698.
- [21] Molina J, Escudero-Viñolo M, Signoriello A, et al. Real-time user independent hand gesture recognition from time-of-flight camera video using static and dynamic models[J]. Machine Vision and Applications, 2013, 24(1): 187-204.
- [22] Lee U, Tanaka J. Finger identification and hand gesture recognition techniques for natural user interface[C]//Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction. New York: ACM, 2013: 274-279.
- [23] 于汉超, 唐晓庆, 刘军发, 等. 手掌姿态自适应的指尖鲁棒跟踪方法[J]. 计算机辅助设计与图形学学报, 2013, 25(12): 1793-1800.
Yu Hanchao, Tang Xiaoqing, Liu Junfa, et al. Robust single fingertip tracking method based on palm posture self-adaption[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(12): 1793-1800.
- [24] Yu H C, Chen Y Q, Liu J F, et al. Adaptive and iterative online sequential ELM based multi-degree-of-freedom gesture recognition system[J]. IEEE Intelligent Systems, 2013, 28(6): 55-59.
- [25] Yu H C, Yang X D, Chen Y Q, et al. strDoctor: Indicate Stroke for Elderly through Body Sensing Game[C]//Proceedings of the 12th IEEE International Conference on Ubiquitous Intelligence and Computing. Beijing: IEEE, 2015: 360-363.
- [26] Chen Y Q, Yu H C, Miao C Y, et al. Using motor patterns for stroke detection[R/OL]. [2017-03-31]. http://www.ntulily.org/wp-content/uploads/journal/Using_motor_patterns_for_stroke_detection_accepted.pdf.
- [27] Kalgaonkar K, Raj B. One-handed gesture recognition using ultrasonic Doppler sonar[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D C: IEEE, 2009: 1889-1892.
- [28] Gupta S, Morris D, Patel S, et al. Soundwave: Using the doppler effect to sense gestures[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM, 2012: 1911-1914.
- [29] Pittman C, Wisniewski P, Brooks C, et al. Multiwave: Doppler Effect Based Gesture Recognition in Multiple Dimensions[C]//Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York: ACM, 2016: 1729-1736.
- [30] Yang Q F, Tang H, Zhao X B, et al. Dolphin: Ultrasonic-based gesture recognition on smartphone platform[C]//Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on. Washington D C: IEEE, 2014: 1461-1468.
- [31] Wang W, Liu A X, Sun K. Device-free gesture tracking using acoustic signals[C]//Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking. New York: ACM, 2016: 82-94.
- [32] Nandakumar R, Iyer V, Tan D, et al. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking[C]//Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2016: 1515-1525.
- [33] 杨晓东, 陈益强, 于汉超, 等. 面向可穿戴设备的超声波手势识别方法[J]. 计算机科学, 2015, 42(10): 20-24.
Yang Xiaodong, Chen Yiqiang, Yu Hanchao, et al. Ultrasonic waves based gesture recognition method for wearable equipment[J]. Computer Science, 2015, 42(10): 20-24.
- [34] Dahl T, Ealo J L, Pazos-Ospina J, et al. High-resolution ultrasonic gesture tracking systems for future portable devices[C]//Ultrasonics Symposium (IUS), 2012 IEEE International. Washington D C: IEEE, 2012: 150-153.
- [35] Wachs J P, Stern H I, Edan Y, et al. A hand gesture sterile tool for browsing MRI images in the OR[J]. Journal of the American Medical Informatics Association, 2008, 15(3): 321-323.
- [36] Ruppert G C S, Reis L O, Amorim P H J, et al. Touchless gesture user interface for interactive image visualization in urological surgery[J]. World Journal of Urology, 2012, 30(5): 687-691.
- [37] Jacob M G, Wachs J P. Context-based hand gesture recognition for the operating room[J]. Pattern Recognition Letters, 2014, 36: 196-203.
- [38] Phelan I, Arden M, Garcia C, et al. Exploring virtual reality and prosthetic training[C]//2015 IEEE Virtual Reality (VR). Washington D C: IEEE, 2015: 353-354.
- [39] Phelan I, Arden M, Garcia C, et al. Exploring virtual reality and prosthetic training[C]//2015 IEEE Virtual Reality (VR). Washington D C: IEEE, 2015: 353-354.
- [40] Lee S W. Automatic gesture recognition for intelligent human-robot interaction[C]//Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on. Washington D C: IEEE, 2006: 645-650.

A review on the recognition of mid-air gestures

YU Hanchao¹, YANG Xiaodong^{1,2}, ZHANG Yingwei^{1,2}, ZHONG Xi^{1,2}, CHEN Yiqiang^{1,2}

1. Institute of Computing Technology, Chinese Academy of Sciences; Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The rapid development of the ubiquitous computing and wearable devices witnesses a new challenge in the natural hand gesture recognition: to free the users from the constraints of the environment and the devices and help the users interact with the environment in a natural and effective way. And the mid-air gesture recognition is one of the effective methods, capable of dealing with the challenge. This paper describes the definition of the mid-air gesture at first, and then analyzes and summarizes the existing hand gesture recognition methods, based on the computer vision, the ultrasonic signal and the electromagnetic wave. At last, this paper discusses the applications of the mid-air gesture recognition, some open questions and the development in the future.

Keywords human computer interaction; mid-air gestures recognition; ubiquitous environment

(责任编辑 刘志远)