软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.006217]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

动态手势理解与交互综述*

张 维[1,2], 林泽一[1,2], 程 坚[1,2], 柯铭雨[1,2], 邓小明[1], 王宏安[1]

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

[2](中国科学院大学 计算机学院, 北京 101408)

通讯作者: 邓小明,王宏安, E-mail: xiaoming@iscas.ac.cn, hongan@iscas.ac.cn

摘 要: 近年来,手势作为一种输入通道已在人机交互、虚拟现实等领域得到了广泛的应用,引起了研究者的关注.特别是随着先进人机交互技术的出现以及计算机技术(特别是深度学习、GPU 并行计算等)的飞速发展,手势理解和交互方法取得了突破性的成果,引发了研究的热潮.本文综述了动态手势理解与交互的研究进展与典型应用.首先阐述手势交互的核心概念,分析了动态手势识别与检测进展,而后阐述了动态手势交互在人机交互中的代表性应用,并总结了手势交互现状,分析下一步发展趋势.

关键词: 手势交互;动态手势理解;人机交互

中图法分类号: TP311

中文引用格式: 张维,林泽一,程坚,柯铭雨,邓小明,王宏安. 动态手势理解与交互综述.软件学报,2021. http://www.jos.org.cn/1000-9825/6217.htm

英文引用格式: Zhang W, Lin ZY, Cheng J, Ke MY, Deng XM, Wang HA. Survey of dynamic hand gesture understanding and interaction. Ruan Jian Xue Bao/Journal of Software, 2021 (in Chinese). http://www.jos.org.cn/1000-9825/6217.htm

Survey of Dynamic Hand Gesture Understanding and Interaction

ZHANG Wei[1,2], LIN Ze-Yi[1,2], CHENG Jian[1,2], KE Ming-Yu[1,2], DENG Xiao-Ming[1] , WANG Hong-An[1]

[1](Beijing Key Laboratory of Human-Computer Interactions, Institute of Software Chinese Academy of Sciences, Beijing 100190, China)

[2](School of Computer Science, University of Chinese Academy of Sciences, Beijing 101408, China)

Abstract: In recent years, hand gesture has been widely used in human-computer interaction, virtual reality and other fields as an input channel. Especially with the emergence of advanced technology of human- computer interaction and the rapid development of computer technology (especially deep learning, GPU and parallel computation technology), gesture understanding and interaction methods have made breakthroughs. This paper reviews the research progress of dynamic gesture understanding and and typical interaction applications. Firstly, the core concepts of gesture interactions are elaborated; secondly, the progress of dynamic gesture recognition and detection is introduced; thirdly, the representative applications of dynamic gesture interaction are elaborated; finally, the future development trend of gesture interaction is discussed.

Key words: hand gesture interaction; dynamic hand gesture understanding; human-computer interaction

1 引言

手势交互技术是通过捕获人手的肢体动作,并将其转化为相应的命令来对设备进行操作的技术,手势已成为继键盘、鼠标和触屏之后新型和主流的人机交互通道[1,2].手势交互系统主要分为几部分:人、手势输入设备、手势分析和识别设备、被操作的设备或界面.ACM SIGCHI 年会的相关论坛多次把自然人机交互限定为触摸和手势的交互方式.在 Gartner[3]发布的 2017 年度人机交互技术优先级矩阵中,增强现实(Augmented Reality)、混

* 基金项目: 国家重点研发计划(2018YFC0809300)
Foundation item: National Key Research and Development Project (2018YFC0809300)
收稿时间: 2020-02-21; 修改时间: 2020-05-10; 采用时间: 2020-07-10; jos 在线出版时间: 2021-01-15

合现实(Mixed Reality)、手势控制设备(Gesture Control Devices)、对话用户界面(Conversational User Interfaces)被认为是未来 5 到 10 年的主流应用新兴技术.

手势交互可广泛应用于虚拟现实[4]、汽车用户界面[5,6]、人与机器人交互[7]、生物医学等领域[2,8],下面列举几种代表性的手势交互应用:

(1)虚拟现实和增强现实.手势可以作为虚拟现实和增强现实系统的输入,使得在虚拟空间通过直观自然手势进行设计和操作成为可能,典型应用场景如虚拟装配[9,10]、虚拟设计等.在执行虚拟装配时,手势可对产品的零部件进行直接装配、定义零件间的装配关系、验证装配设计并校验操作的合规性.在虚拟设计中,可以用于相关产品或设施的人机工效评价、给出这些产品或设施设计的定量参考.

(2)汽车用户界面.通过手势输入可以操控车载信息系统,辅助驾驶员完成驾驶过程中的多项任务,如音响控制、温度调节、车身周围环境 3D 显示等,让驾驶员更加专心于驾驶,提升驾驶安全性.

(3)人与机器人交互.通过对手势的识别和理解,机器人可对人的手势进行模仿或给出适当反馈.

(4)生物医学,特别是人手运动定量分析.比如在人手运动功能康复评定中,可定量描述手功能障碍患者康复治疗效果,促进康复治疗方案.

手势交互也有着重要的研究价值.在手势交互技术方面,学术界较早研究了相应的交互基础理论与概念模型,例如:手部运动信息的感知和处理模式;通过用户观察和实验分析,将手势交互行为分解成为基本动作,研究这些基本动作的合理参数及范围;通过对稳定参数的收集和统计,确定参数与任务间的关系用于指导自然人机交互的相关设计.而学术界和工业界都研发了手势交互关键技术,例如手势的采集、识别、合成与理解技术,在三维空间数据的采集方面取得了较好进展,然而进一步提高其可靠性、精度、准确度、稳定性仍然存在诸多技术问题.为了能识别模糊的以及细微的手势,需要开展更高精度传感器的研究和开发.

手势交互仍面临一些研究问题需要解决:

(1)金手指问题[4,11].金手指(midas touch) 问题是指手势识别系统不能有效判别在人的连续运动中, 哪些动作是有意图的交互,哪些是下意识的动作, 或者不能明确判别一个手势的发起或结束,这对使用手势交互的系统造成极大的困扰,容易造成连续交互的中断.

(2)动态手势交互识别存在延时问题.目前的动态手势识别方法必须在一个相对完整的运动结束之后才能识别出手势的类别, 造成一定的延时.

(3)对于界面的哪一部分功能更适合用手势操控仍没有达成共识.

(4)使用手势交互需要根据应用场景的特点来选取手势,有些手势需要用户额外学习和记忆.当手势过多或者各类应用使用不同的手势,将增加用户的记忆负担.

(5)人手姿态和动作获取技术仍是限制精细手势交互界面的重要技术,精细操作虚拟场景中的对象目前还存在技术上的难度,这需要进一步提升交互手势动作捕获和识别的精度.

在手势交互研究领域,目前一些著名的研究机构如 CMU、Stanford、MIT、Microsoft、Apple、Google、Facebook、Intel、法国 INRIA、德国马普学会等都在进行许多有益的尝试,在手势交互设备、交互技术、以及应用方面已经有了不少的研究成果.国内,中科院、清华大学、浙江大学、北京航空航天大学、北京师范大学、北京理工大学等研究机构的研究人员也高度关注该领域的研究,在交互手势获取、动态手势理解及在虚拟现实和增强现实中的人机交互等领域产生了非常有价值的成果[7,12,13].

综上所述,手势交互及手势理解既有理论研究意义又有重要应用前景,已引起了国内外研究者的广泛关注.本文拟针对手势交互中的核心问题——动态手势的类型、交互应用中的手势选择、动态手势识别与检测方法,综述这些方面的代表性研究进展和思路,并概述动态手势交互应用,讨论动态手势与交互领域的未来发展方向,希望给从事手势交互研究和应用的人士提供一些参考.

2 手势交互中的核心概念

2.1 手势的概念

手势通常指的是人在使用手臂时,所体现出的具体动作与体位,分为动态手势和静态手势.它是人类最早使用的、至今仍被广泛运用的一种交际工具.手势可被赋予各种特定的含义,具有表现力强和灵活性度高的特点,手势既是人类表述情感及意图的自然手段,也是人类对外界加以影响的重要方式之一.

2.2 手势分类

2.2.1 基于运动特点的分类

根据手势的运动特点,手势可以分为静态手势(posture,也称手形)或动态手势.静态手势是主要依靠手部外形与轮廓来传达信息的方法,是一类特殊的动态手势.这些类型的手势也可以被称为姿势.在一个动态手势中,手的形状、位置都根据时间变化, 它包含了更加丰富且准确的信息与内容, 是一种人们常用的表达与交流方式.动态手势通常可以分为准备,开始,执行,结束,收回等五个阶段.

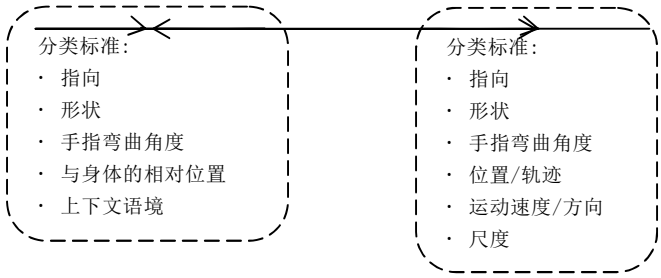

Fig. 1 Classification by gesture dynamics

图 1 按手势运动特点的分类

2.2.2 基于交互目的的分类

根据手势的交互目的,手势又可以分为交际性手势(communicative)或操纵性手势(manipulative).操纵性手势用手的运动来表示路径或者位置信息,而交际性手势往往跟演讲有关.交际性手势包含标志性手势、隐喻手势、调制符号手势、连贯手势,“Butterworth”手势和“Adaptors”手势.其中,标志性手势代表与演讲语义内容密切相关的意义,用来呈现在言语中唤起的对象的一些图形表征.例如,一个人在讨论一个从山上滚下来的物体时,会用手做一个滚的动作;隐喻手势是代表抽象内容的标志性手势,比如一个人会在决定的时候做出表示决定的切割手势;调制符号手势主要是对语音的补充,但也可以补充其他的沟通手段;连贯手势表示那些在语义上相关,但是在时间上不连续的手势,比如当一个演讲者被打断后,可以使用重复的手势表示相同的演讲内容以保证演讲的连续性;“Butterworth”手势是指在说话或者演讲中不合时宜的“错误”手势,比如一个人在演讲中回忆单词时摸头的手势;“Adaptors”指的是演讲中人们做出的无意义手势.

手势中有一类特殊的手势叫做指示手势,它用来表示预定动作的方向或者操纵的方向.指示手势根据上下文语境的不同,既可以分为交际性手势,又可以分为操纵性手势.在其表示预定动作方向时,可以将其看作交际性手势中的标志性手势,而在其表示操纵的方向时,往往将其看作操纵性手势.

2.2.3 基于指令的分类

根据指导手势表现的指令级别,手势可以分为规定的手势或自由形式的手势.规定的手势是那些定义了“手势字典”的手势.在使用之前,应用程序的用户必须了解这些手势,预定义手势可以触发对应的预定义的行动.指定的手势过多可能会增加用户的认知负荷:使用规定手势的应用要求用户学习和使用手势,但是这些手势可能是用户不愿意选择使用的.自由形式的手势通常不会触发特定的统一预定义动作.在交互上下文中,它们通常被复制到接口所用于的系统中,并且通常用于形成样条或曲面,或在虚拟空间中移动对象.这意味着它们不传

达手势的象征意义或隐喻,自由形式的手势的应用范围比较局限.

在本文中,手势主要指动态手势,包含单手或双手交互动作,对于包含手指动作的交互动作也属于手势.

2.3 手势获取涉及的技术

手势获取技术主要分为两类: 基于视觉的技术[11],主要靠摄像机进行手势识别与跟踪; 基于可穿戴设备的技术,要求对用户佩戴的手套、戒指、手镯、带加速计的腕带等设备输出的数据进行识别.这两类方法的主要区别在于:基于视觉的技术对用户是无干扰或者少干扰的,用户不需要佩戴任何传感器,可以用裸手进行交互,更符合用户的交互习惯,而基于可穿戴设备的技术则需要在手上安装其他传感器,对用户有一定的干扰.

2.3.1 基于视觉的技术

基于视觉的手势获取采用的设备通常可分为三类:彩色相机,红外/深度相机和运动捕捉设备.基于视觉的技术相比于基于可穿戴设备的技术,具有对用户基本无干扰和用户不容易疲劳等优点,但是这种技术通常容易受到光线和肤色等因素的影响.Leapmotion[14]和 Nimble VR 通过红外相机或深度相机可获取人手三维姿态,并可实时传输给虚拟现实头盔显示器如 Oculus Rift 等,促进在虚拟空间通过自然手势进行设计和操作. 基于视觉的交互手势姿态获取是目前动态手势识别的基础性技术之一,这类问题主要有两个思路: 基于模型优化的方法与数据驱动的方法.

基于模型优化的方法主要通过对手部各个关节弯曲角度等参数的直接捕捉与描绘,来探测并还原手势.模型的设计从人手的解剖结构入手,通过模拟骨骼的结构来设计物理模型,模型的结构与真实的人手越相似,手势的还原度就越高.通常所使用的人手模型分为两类,第一类模型使用圆柱体来模拟手指指节,而每个关节则以圆柱体的连接点代表[15].这种模型结构简单,参数直观,对于边界判断与碰撞检测均有不错的表现,但只限于对手势本身进行获取,如果想要精确地掌握手部皮肤、肌肉在动作当中的形态变化,就需要使用第二类模型,网格(mesh)结构模型[16],通过取点连线,对人手的表面进行直接描绘.相比于圆柱体模型,网格模型的还原度更高,但参数复杂.对于网格模型的优化将两种模型的设计思路融合在一起,既参考人手,尤其是骨骼的解剖结构来降低参数的个数,又设立单独的参数刻画表面的形态.具体来说,此种思路设计了个体差异参数向量与动作角度向量,前者主要负责刻画手的各个部分本身,后者主要负责确定这些部分之间的角度关系.通过这种方法,就可以既减少参数的个数,又能保证模型的细节尽量逼真.

交互手势姿态获取的另一个思路则是基于数据驱动.传统方法包括最近邻搜索、随机决策森林、具有隐含变量的回归森林等算法.最近邻搜索是将待识别的手势与数据库中的手势进行比对,将其识别为最接近的手势.一旦待识别的手势是一个数据库中所不包含的新的手势,则无法返回正确的结果.随机决策森林则用于对图像的像素进行分类,决定其所位于哪个关节点,再搭配均值漂移算法,进一步确定关节点.具有隐含变量的随机森林则是在回归森林中,凭借关节之间的隐含关系搜索关节点的坐标.传统方法需要大规模的标注数据集来进行训练,因此常常会使用合成数据集,由于合成数据集与真实数据的偏差,使得训练结果会受到严重影响,尤其是在真实数据上的泛化性能.目前,深度学习方法逐渐取代传统方法用于手势姿态获取.

基于深度学习的手势姿态获取基础模型主要包含卷积神经网络(CNN)和循环神经网络(RNN)等.基于 CNN 的深度学习模型的输入数据通常有深度图与彩色图像两种形式.对于深度图的输入,CNN 首先提取其特征,并提取关节点的热力图,并最终回归人手关节的三维位置[17].一些基于深度学习的方法,比如 Deep-prior,就可以从单个深度图像中提取手势,在这个方法的基础上,又可以增加强大的先验层、旋转、平移、缩放、在线增强等[18],也就是所谓的 Deep-prior++,进一步增强算法的性能.此外,迭代反馈[19]也是一种优化的方式.这种网络模型分为若干子网络.手势估计子网络从深度图中初步预测关节的位置,形成一个初步的手势,进而形成一个合成深度图,最后,由优化子网络通过合成深度图与初始深度图的比对,对检测结果进行修正.投影图[20]是另一种优化的方法.将深度图投影形成其三视图,在投影图上提取特征、绘制热力图再重新融合,学习最终的关节位置,使得探测的结果更加精确.

还可以通过将深度图像转化成点云[21],并进一步转化成体素[22]、点云网络来进行优化.这种方法体量小,运算便捷快速,精确度也很高; 抑或是采取数据扩充的思路,加入模拟数据[23]、多视角数据[24]、无监督数据[25]、

引入无对应深度图的纯关节点坐标[26],利用扩充之后的数据对网络进行训练,也能提升其预测关节点的精度;结合传统方法与深度学习方法的工作也取得了一定的成效,CNN 与均值漂移算法结合[27],CNN 与 PSO 运动约束优化算法结合[28],CNN 与关节点约束算法结合[29],都提高了关节点预测的精确度.

对于使用彩色图像进行输入的情况,基础的思路是通过 CNN 预测关节点的二维热力图,并试图还原相机角度、正则化关节坐标等参数,进而最终获得三维坐标[30].在此基础上,可以通过基于无标注训练样本的自助迭代训练[31]、运动学骨架拟合[32]、利用 VAE 网络模型进行跨模态训练[33]等方式,来扩充训练数据、细化坐标精度,提高手势识别的精确程度.

2.3.2 基于可穿戴设备的技术

基于可穿戴设备的手势获取设备通常包括手套、基于肌电图(EMG)的设备、加速度计、标记点(Marker)、基于射频识别(RFID)的设备、陀螺仪和加速度计等,其中数据手套和肌电图设备是最常用的两种可穿戴手势获取设备.

数据手套可以提供用户手势的姿态跟踪信息,如手指是怎样弯曲的、两个手指是否重叠、遮挡等,它有两种基本类型:弯曲感觉手套和压力手套.弯曲感觉数据手套是被动的输入设备,用来检测用户的手形和特定的手势.它的一个主要优点是它能提供大量的自由度信息,使得它不但可以识别各手势和手形,而且可以给 3D 应用提供用户的手姿态.不过,这类手套穿戴不舒服,需要根据用户手的物理参数的不同,进行参数校准.压力手套系统是一种用于判别用户是否有两个或多个指尖发生触碰的输入设备.这类手套的每个指尖有一个导电材料,这样用户可以通过两个手指的捏、压,产生一个电路连接.这类设备通常用来在物体选取、模式转换和其他一些技术中,用于判断用户的抓取和捏压手势.压力手套非常轻,能够降低用户的疲劳度,也可以应用于双手交互. 代表性的手势交互设备有 Soli[34], MYO 腕带[35].Soli 通过微型雷达获取的空中手势信号,识别为一系列交互手势,它更擅长处理动态手势.MYO 通过臂带上传感器获取的肌肉生物电变化,判断佩戴者的意图.此外,使用固定在手套上的光学标记点[36],与多台摄像机配合,可以完成高精度的手势采集.

基于可穿戴设备的技术具有健壮性好的优点,但是可穿戴设备往往对使用者有所限制,并且使用者容易疲劳等缺点,容易影响用户的交互体验.

2.4 手势在应用中的功能

本节关注手势在具体应用当中承担的功能.一个手势到底表达了什么意思,属于哪个类别,并不是一成不变的,手势本身的功能往往会因为开发者的事先设计、用户的使用体验以及技术水平的制约发生一定的改变,进而间接影响手势的语境分类[8].在每个特定的应用场景,研究人员往往会选用不尽相同的手势,通常没有一种压倒性被选择的手势,成为这种应用场景的专属手势.尽管如此,随着研究的日益深入与广泛,我们可以发现一些手势类型选择方式还是更受到研究人员的青睐.如对于基于手势的界面,由于深度摄像机与动作捕获设备的普及,使用基于视觉的技术已成为主流;而从手势的交互目的分类上看,视觉识别多用于操纵、指示、交互型手势;3D 建模往往使用自由型手势,通过 3D 相机和动作传感器进行手势理解;可穿戴式手势交互设备可以更准确地捕捉手势复杂组合以及信号型手势.

2.4.1 手势的语境分类

手势按照语境分类,可以先归为主要的四大类[8]: 指示型、互动型、操纵型、信号型.考虑到这四大类手势有一些十分常见的组合,因此将其进一步归为 11 类,分别为单纯指示、指示+互动型、指示+信号型、指示+操纵+互动型、指示+操纵+信号+自由型、单纯操纵、操纵+信号型、单纯自由型、自由+信号型、单纯信号型、信号+节奏型,其中节奏型手势主要用于音乐相关的应用.

在被调研的文献中,指示型与信号型的手势最为常见.信号型手势是指将特定手势作为信号,触发一系列事先设定好的复杂抽象功能的手势,其所表达的意义与这些手势通常的意义往往大不一样,这一点与动作直观、意思直接的操纵型手势显著不同.互动型手势是指与界面上的虚拟元素,如按钮等,进行互动的手势,与虚拟元素互动时,并不会对元素本身进行改变与编辑,这一点与用来直接编辑元素,如翻译、调整大小、旋转等的操纵型手势不同.

2.4.2 指示型手势及其组合

指示型手势用于点选,其与不同的手势组合承担不同的功能.与互动型手势组合,可以做到对元素的先点选、再编辑；与信号型手势组合,可用于元素的点选与操作；与互动型、操纵型手势组合,可以对元素完成点选、切换、互动的一系列动作,这种手势组合的典型例子就是操纵游戏人物；有时候,指示型手势与互动型、操纵型、信号型、自由型手势联用,可以用于 3D 建模.

2.4.3 自由型手势及其组合

通常用于 3D 建模与非接触控制,在这种情境下,自由型手势用来移动虚拟物体或是控制机械手.而对于一些意义较为复杂间接的功能,则往往与信号型手势进行联用.

2.4.4 操纵型手势及其组合

操纵型手势用以对对象进行直接编辑,比如翻译、旋转、调整大小等.

2.4.5 信号型手势及其组合

信号型手势用来触发预设的通常较为抽象的功能,其核心在于需要预先规定手势与所触发的功能的联系.特别的,有些手势本身带有一些意义,则需要辨析其语义归属.如果这些手势用来执行其通常的含义,则是操纵型,如果是用来作为一个信号,触发预设的一系列操作的,则是信号型.

2.5 手势识别引擎

手势识别引擎是指将手势识别的核心技术整理成软件开发包的形式,供各种应用程序调用手势识别核心技术,开发手势交互相关的各种应用软件.手势识别引擎核心技术模块主要是由以下几个部分组成:(1)手势定义,此定义用于匹配输入设备的行为,以便进行手势分析；(2)手势分析,运用手势分割和手势识别等技术实时分析从输入设备获取的数据；(3)后期处理,对手势分析返回的结果进行后期处理,使其更容易被其它模块使用；(4)结果反馈,将处理后的结果映射到已经定义好的手势并将其实时返回给引擎的调用者.

3 动态手势识别与手势检测进展

由于深度摄像机与动作捕获设备的普及,使用基于视觉的手势交互技术已成为主流,本节主要介绍基于视觉的动态手势识别和检测代表性工作和进展. 动态手势识别的流程通常可以分为手势分割(或检测模块)模块和手势识别模块(如图 2 所示).手势分割模块基于传感器原始数据进行候选手势提取,包含手势的起始和终结时刻检测；手势识别模块主要对分割好的候选手势片段进行分类,最后给出每个候选手势片段的动作类别,通过后处理给出动态手势的定位和类别信息.

Fig. 2 Flow chart of dynamic hand gesture recognition

图 2 动态手势识别流程图

3.1 基于不同模态的手势识别与手势检测

3.1.1 基于 RGB 图像与视频的手势识别与手势检测算法

我们首先分别介绍基于 RGB 图像与视频的手势识别、手势检测的代表性思路,然后按照根据是否使用深度学习分别介绍基于传统方法和基于深度学习的代表性工作.传统机器学习方法主要有动态时序规整(DTW),隐马尔科夫模型(HMM),条件随机场(CRF)和随机森林(RF)方法.基于深度学习的方法主要有基于 LSTM 的方法和基于 CNN 的方法.

对于手势识别问题,无论是基于传统机器学习的方法,还是基于深度学习的手势识别方法,一般都需要先提

取出视频中手部的位置,也称为人手检测.传统人手检测有基于手部肤色的方法和基于手部运动信息的方法.基于手部肤色的方法利用手部肤色与背景颜色信息的差异来进行手部的分割,但是这种方法对背景光照,颜色信息比较敏感.基于手部运动信息的方法利用手部相对于背景的运动信息来进行手势分割,这种方法需要背景信息大致不变,鲁棒性较差.随着近年来深度学习的发展如 Faster RCNN 和 SSD 等物体检测算法越来越多的被应用到手势分割上,这种方法具有精度高,鲁棒性好等优点.

手势识别方法主要流程分为两个阶段:1) 利用传统特征提取方法或深度学习特征提取方法提取手势的特征; 2) 将提取的特征输入到分类器中进行手势分类.

手势检测方法主要分为两种方式:基于候选动作片段(action proposal)的手势检测方法和基于样本类间差异的方式.基于候选动作片段的手势检测方式可以分为以下四个步骤:1) 利用时序滑动窗口等方式提取动作片段; 2) 利用传统特征提取方法或深度学习特征提取方法提取每个片段的特征; 3) 将每个片段的特征输入分类器中进行动作片段与背景片段分类; 4) 将每个动作片段的特征输入分类器中进行动作分类和起始帧的微调; 5)利用 NMS 等方法剔除重复片段.

基于类间差异的手势检测方法主要分为以下几个阶段:1) 利用传统特征提取方法或深度学习特征提取方法提取手势的特征; 2) 利用动作片段与背景片段特征之间的差异性分离动作片段与背景片段,主要有基于先验知识的方式和基于 connectionist temporal classification (CTC)分类器的方式; 3) 利用分类器对每个片段的特征进行分类.

1)传统的手势识别和手势检测的方法

基于传统机器学习的方法在 2012 年之前使用的较多,传统机器学习方法主要有动态时序规整(DTW),隐马尔科夫模型(HMM),条件随机场(CRF)和随机森林(RF)方法.DTW 为一个模板匹配算法,这种方法实现简单,不需要训练,但是需要高精度模板来进行匹配.HMM 和 CRF 方法都是基于概率模型的算法,这两种方法都能很好的提取动态时序信息.RF 算法作为一种常用机器学习算法,主要使用集成树状分类器.

隐马尔可夫模型(HMM)是一种广泛应用于手势识别的模型,手势识别模型被假设为参数未知的马尔可夫过程,利用具有转移和发射概率的隐状态网络表示可观察符号序列的统计行为,可用于利用可观测数据识别隐藏参数后的模式识别.基于 HMM 的动态手势识别方法主要利用输入图像的时空特征.Chen 等人[37]利用傅里叶描述符和基于光流的运动分析分别表征空间和时间特征.该算法通过对手部的实时跟踪,从复杂背景中提取手部形状.基于 HMM 的识别器识别给定模式的最佳似然手势模型.手势与参考模式之间的变化会降低手势与模型之间的可能性.对于直接三维连续手势识别,可以方便地利用速度、轨迹等低层次的运动特征来检测定位的突变[38].Elmezain 等人[39]提出了一种利用 HMM 实时识别连续数字手势的系统,首先从深度域的时空轨迹生成方位动态特征,然后将其量化为码字.连续手势的分割是在零码字检测的基础上进行的,零码字实际检测到了手势的静态速度和端点.

动态时间归整(DTW)是一种动态规划应用,在手势识别和检测中得到了广泛的应用.DTW 通过计算待匹配信号之间的时间转换来进行输入手势时间对齐和归一化.在动态手势识别方法中,DTW 被广泛应用于在时间域中寻找匹配的手势片段.为了提高识别精度,Keskin 等人[40]提出了一种基于 DTW 的图像模型三维数字识别预聚集技术.为了利用每个手部的讨论性,Arici 等人[41]提出了一种加权 DTW 方法,该方法通过优化判别比对关节进行加权,以改进三维手臂手势识别.在开始-结束手势识别的背景下,Reyes 等人[42]提出了一种在 DTW 框架中使用特征权重的开始-结束手势识别方法.在传统的开始-结束 DTW 算法中,为了提高成本-距离计算的效率,提出了特征加权方法.

此外,还有一些基于随机森林和支持向量机的手势识别方法.Dong 等人[43]提出了一种基于手部深度图像的方法,这种方法首先利用 RF 来将深度图像的每个像素分为 11 类,然后利用分层模型搜索方法计算出每个像素的关节的方向,最后将关节的方法作为特征输入到 RF 中进行手势识别.Song 等人[44]将 RGB 视频每一帧中的信息转化成一个标准的人体姿态模型,然后利用 HOG 提取模型中的特征并输入到 SVM 中进行分类.

传统机器学习方法对训练数据和计算力要求都不太高,但是精度通常没有基于深度学习的方法高.随着近

年来数据的不断增长和计算技术的提高,使用传统机器学习方法的研究已经明显减少,但在计算资源受限的场景下,传统的手势识别和检测方法仍可起到重要的作用.

2)基于深度学习的手势识别和手势检测方法

基于深度学习的手势识别方法主要分为基于 LSTM 的方法和基于 CNN 的方法.

由于 LSTM 能很好地对手势时序信息进行建模和识别,基于 LSTM 的手势识别逐渐成为的主流方法.然而由于 RGB 图像的分辨率较大,直接将 RGB 图像作为特征输入到 LSTM 网络并不可行,一般利用 CNN 网络逐帧或分片段提取图像或者视频的特征,然后将视频对应的特征序列输入 LSTM.Molchanov 等人[45]提出一个端到端的手势识别架构,首先利用 C3D[46]来提取每个视频片段的特征,然后将这些特征输入到 RNN 中提取时序特征,最后将每个 RNN 输出的特征经过 softmax 层变换后输入到 Connectionist temporal classification(CTC)层中从而获取每个手势片段的起始位置、结束位置和手势类型.Camgoz 等人[47]也提出了一个类似的架构,区别是他们利用 CNN 来提取每一帧图片的信息并且利用的是双流 LSTM 架构.Cui 等人[48]也使用了类似的架构并且使用了分阶段优化方法以获取更加准确的分类与检测结果.Cao 等人[49]提出了一个第一人称视角下的手势识别数据库 EgoGesture 并且提出了一个 Recurrent 3D Convolutional Neural Networks 架构.该架构首先将视频分为一些视频片段,然后将每个视频片段使用 3DCNN 提取特征,最后将特征输入到一个 Spatiotemporal Transformer(STT)模块中,STT 模块的目的是为了将不同帧中的手部变换到同一个视角下,从而缓解第一人称视角下镜头视角变换的影响,最后利用 LSTM 提取每个视频片段之间的时序信息用来分类.

基于 CNN 的方法采用是双流输入的卷积神经网络结构,将 RGB 图像和光流图像分为作为 CNN 网络的两流输入,在特征层融合两个通道的特征作为特征输入到分类器中进行手势分类.Narayana 等人[50]改进了这一方法,将原始 RGB 图像输入变成原始图像加上左右手图像的三个输入,使模型能够更加注重手部信息从而提高手势识别精度.Bambach 等人[51]使用朴素贝叶斯方法提取动态手势片段的候选区域(action proposal),然后利用 CNN 对每个手势片段提取特征并分类,最后利用每个片段的分类得分确定最后得到的片段,他们还在自己提出的 EgoHands 数据库中验证了这个方法.Rogez 等人[52]提出了一个两阶段的抓握识别系统:第一阶段基于深度和 RGB 信息进行目标检测和分割, 第二阶段使用网络提取特征然后用 SVM 进行抓握动作分类.Joshi 等人[53]在 CVPR2017 中提出分层贝叶斯网络进行动态手势识别.Hu 等人[54]提出了一个 3D 分离卷积神经网络进行动态手势识别,3D 分离卷积把一个 3D 卷积变成一个 3D Depth-wise 卷积和一个 3D Point-wise 卷积,从而降低模型的复杂性.

3.1.2 基于手部姿态的手势识别与手势检测算法

基于手部姿态的手势识别方法是利用手部的关键点的信息来进行手势识别,相比于基于 RGB 图像与视频的手势识别方法,手部姿态不受背景信息的影响,能够更好的关注到手部的位置与运动信息, 是一种具有较大发展潜力的方法.基于手部姿态的手势识别主要分为三个步骤:首先,利用手部姿态检测方法获取手部的姿态信息(请参考 2.3 节);然后,利用传统特征提取方法或深度学习特征提取方法提取手部姿态的特征;最后,将提取的特征输入到分类器中进行手势分类.基于手部姿态的手势识别方法也分为基于传统机器学习方法和基于深度学习的方法.

基于传统机器学习的方法通常利用 Fisher Vector(FV)[55]或者直方图的方法构造出手部姿态的特征,然后利用 GMM 或者 CRF 等方法提取出时序特征,最后输入分类器中进行手势分类.Smedt 等人[56]使用三个向量来表示手部的运动方向信息,旋转信息和手部的形状信息,并且使用时间金字塔(Temporal Pyramid,TP)方法来聚合不同时间尺度上的手部信息,并利用 FV 和 GMMs 方法来编码这些特征,最后输入到 SVM 进行训练和分类.Zhao 等人[57]提出了一种基于骨架的动态手势识别方法.该方法提取了四种手部形状特征和一种手部方向特征,用时间金字塔(TP)表示手部形状 Fisher Vector 和手方向特征,得到最终的特征向量,并将其输入线性 SVM 分类器进行识别.Boulahia 等人[58]从指尖、掌心和手腕的三维坐标中提取出 HIF3D 特征表示手部形状的高阶信息,并加上时间金字塔来捕获时间信息,最后使用 SVM 对得到的特征进行分类.相比于深度学习方法,传统机器学习方法需要事先提取手动构造特征,这种特征往往没有深度学习自动提取的特征好,从而最后分类的效果也不如深

度学习的方法好.

基于深度学习的识别方法通常将人手姿态信息输入到 RNN 或者 CNN 网络中直接进行分类.Devineau 等人[59]提出了一个新的卷积神经网络基于骨架信息进行动态手势分类,该网络采用并行卷积处理手部骨架序列.在该方法中,将手势序列按维度分成 66 个向量,每个向量各输入一个具有三个分支的 CNN,然后将每个 CNN 的输出连接起来使用全连接层进行分类.Chen 等人[60]提出了一种基于骨架信息的手势识别运动特征增强递归神经网络.提取手指运动特征描述手指运动,利用全局运动特征表示手部骨骼的全局运动,然后将这些运动特征与骨架序列一起输入一个双向递归神经网络(RNN),可以增强 RNN 的运动特征,提高分类性能.近年来由于注意力机制(attention mechanism)在计算机视觉和自然语言处理领域的兴起,也有研究者将其用到手势识别领域.Hou 等人[61]提出了 Spatial-Temporal Attention Res-TCN 网络,该网络在训练主要卷积网络的同时训练了一个权重卷积网络,对卷积网络中每一步的输出的特征给出一个权重,从而实现了注意力机制.除了直接将姿态信息输入网络,也有些方法先从姿态信息中人工提取特征,再输入网络进行手势识别.Avola 等人[62]提取手部关节点间的角度作为特征,然后输入 DLSTM 进行手势识别.

3.2 常用动态手势数据集

为了对动态手势识别算法进行可靠的测试和比较,研究者们已经建立了一些手势识别数据集[63].本节将回顾公开可用的手势数据集.表 1 列出了手部姿势、手势数据库以及下载链接.表 2 详细描述了这些数据集,可用的手势类别数、实验对象和样本的数量,包含的数据类型和动作类别.

Table 1 Publically available dynamic hand gesture datasets with sources, description in Table 2

表 1 公开可用的动态手势数据集及其来源,说明见表 2.

编号	名称,年	地址
1	ChaLearn gesture data[64], 2011	http://gesture.chalearn.org/data
2	ChaLearn multi-modal gesture data[65],2013	http://sunai.uoc.edu/chalearn/
3	NATOPS aircraft handling signals database[44],2011	http://groups.csail.mit.edu/mug/natops/
4	Sebastien Marcel hand posture and gesture datasets[66-69], 2001	https://www.idiap.ch/resource/gestures/
5	NVIDIA Dynamic hand gesture dataset[59], 2016	https://research.nvidia.com/publication/online-detection-and-classification-dynamic-hand-gestures-recurrent-3D-convolutional
6	GUN-71[52], 2015	http://www.gregrogez.net/research/egovision4health/gun-71/
7	EgoHands[51], 2015	http://vision.soic.indiana.edu/projects/egohands/
8	DHG-14/28[70], 2016	http://www-rech.telecom-lille.fr/DHGdataset/
9	SHREC 2017[71], 2017	http://www-rech.telecom-lille.fr/shrec2017-hand/
10	LMDHG[58], 2017	https://www-intuidoc.irisa.fr/english-leap-motion-dynamic-hand-gesture-lmdhg-database/
11	EgoGesture Dataset[72,49], 2018	http://www.nlpr.ia.ac.cn/iva/yfzhang/datasets/egogesture.html
12	Daily hand-object actions dataset[73], 2018	https://guiggh.github.io/publications/first-person-hands/
13	The yale human grasping dataset[74], 2015	http://grasp.xief.net/

Table 2 Descriptions of publically available dynamic hand gesture datasets, following the order of **Table 1**

表 2 对公开可用的动态手势数据库的描述(与表 1 中的顺序相同)

No.	规模	设备	提供的数据类型	特点
1	ChaLearn Gesture Challenge, 50,000 samples	Kinect	RGB+Depth	从九个不同的领域选择的手势,手势类型多,包含肢体动作
2	20 classes, 27 subjects, 13,858 samples	Kinect	RGB+Depth+Audio	意大利手语手势,包含肢体动作
3	24 classes, 20 subjects, 9600 samples	Bumblebee 2 stereo camera	RGB+Depth +mask+pose	飞机和航空母舰甲板通信手势,包含肢体动作

4	4 classes, 57 samples	unknown	RGB	人机交互手势
5	25 classes, 20 subjects, 1532 samples	DS325 and DUO 3D camera	RGB+Depth +stereo-IR	汽车辅助驾驶所使用的交互手势
6	71 classes, 8 subjects, 12,000 frames	Intel's Senz3D	RGB+Depth	第一人称拍摄的日常抓握动作,估计出了受力点和力的方向,但并不准确
7	4 classes, 4 subjects, 130,000 frames	Google Glass	RGB+mask	第一人称拍摄两人进行四种游戏时执行的动作,人工标注了双手的像素级掩码(mask)
8	14/28 classes, 20 subjects, 2800 samples	Intel Real Sense Depth camera	Depth+pose	交互手势,包含粗粒度与细粒度
9	14/28 classes, 28 subjects, 2800 samples	Intel Real Sense Depth camera	Depth+pose	交互手势,包含为粗粒度与细粒度
10	13 classes, 21 subjects, 608 samples	Leap Motion	pose	交互手势,包含双手手势
11	83 classes, 50 subjects, 24,161 samples	Intel RealSense SR300	RGB+Depth	交互手势,包含双手手势
12	45 classes, 6 subjects, 1175 samples	Intel RealSense SR300	RGB+Depth +pose+object pose	第一人称拍摄的日常动作,包含了交互物体的 3D 模型和姿态
13	33 classes, 4 subjects, 18,210 samples	RageCams	RGB+pose	第一人称抓握动作

4 动态手势交互应用

手势交互界面将推动虚拟现实、移动终端等研究和应用,促进这些产业的发展.具体而言,手势交互技术具有许多典型应用(如表 3 所示).

4.1 3D 建模

手势交互用于 3D 建模技术,可以在计算机生成的环境中自然地创建、操作和修改 3D 模型.3D 建模技术主要有以下应用方向:三维建筑城市规划、电缆应用设计、计算机辅助设计[75]、计算机辅助操作、虚拟陶器[34]等.此类技术对于未来的先进制造尤其实用,比如,手势交互结合其它虚拟现实设备,可用来进行各种虚拟设计(汽车设计和飞机设计[14]等), 相比于传统的 CAD 技术,由于系统的投入性和交互性,它能更好的满足设计要求.

4.2 数据输入和身份验证

数据输入与身份验证技术主要包含下述几类应用方向:电子身份证明、计算机输入、手写识别、手写输入[86]等.在这些交互应用中,手势被用来输入信息到计算机系统中,通过使用专用的指定手势,由界面设计师或用户定义[81,86].

Table 3 Representative gesture interaction interfaces, systems and applications in recent years

表 3 近年来有代表性的手势交互界面系统及应用

名称	关键交互技术	交互工具	特点
朱英杰等[9]	3D 建模技术、虚拟显示技术、机器学习	光学运动捕捉系统、数据手套、头盔显示器	虚拟装配系统,可实现人与虚拟环境间的复杂交互,难以实现精确操作
BodyAvatar[75]	3D 建模技术	Kinect	3D建模工具,允许人直观地使用身体表达 3D 形状
Kevin P. Pfeil 等[76]	3D 交互技术	Kinect	无人机控制系统,包含多种类型的手势
Vinayak 等[77]	3D 建模技术	Kinect	虚拟陶器应用程序
HoloDesk[78]	全息技术	光学透视显示器、Kinect	全息交互系统,无需使用任何穿戴设备

Soli[34]	电磁回波探测、机器学习	毫米波雷达	使用毫米波捕捉手部信息,鲁棒性好,精确到亚毫米级,能耗低
Mime[79]	计算机视觉技术	Mime	使用自制设备 Mime 的手势交互系统,对光线不敏感,精度高
Barehanded Music[80]	机器学习	Depthsense325 传感器	虚拟钢琴系统,只能用于平面交互,鲁棒性不高
SketchingWithHands[81]	3D 建模技术	Leap Motion、手写笔	三维手绘系统,能方便画出 3D 手持物品的草图
ATK[82]	机器学习	Leap Motion	空中打字应用
Jian Cui 等[83]	3D 建模技术	Leap Motion	使用 Leap Motion 构建的虚拟建模系统,用户无需学习即可使用
任镆等[12]	3D 建模技术	Leap Motion	古建筑三维场景快速搭建系统,直观、便捷、高效
Hui Liang 等[84]	3D 建模技术	Leap Motion	基于手势的木偶故事系统
GazeTap[85]	眼球跟踪技术、触控板技术	MeVisLab	减轻医生双手负荷
GBIS[34]	手势交互技术	摄像机、数据手套	实时跟踪,精度高、可靠性好
黄琦等[86]	触控板技术、几何识别技术、三维草图建模技术	触控板	简单高效设计复杂三维草图
Shen 等[87]	手势交互技术	Leap Motion	稳定的双手虚拟现实交互

4.3 操作/导航

操作与导航技术主要有以下几个应用方向:与显示/投影设备交互、增强现实(VR)/虚拟现实(AR)交互[78]、应用程序导航/选择、机器人交互[76]等.无论手势是用于导航二维屏幕或应用程序,还是与 AR、VR 或 3D 空间交互,这些界面的一个共同特点是,几乎所有使用的手势都是预先规定和预定义的; 它们要么是由界面设计师定义的,要么是用户最初可以为某些操作建议首选的手势.自由形式的手势主要用于交互空间的导航任务,如移动鼠标光标,或移动一个被拾取的对象.此外,大多数界面是多模态的,这意味着手势的应用范围有限,因此在适当的情况下应该使用其他模式.

手势交互也可以与 AR、VR 技术结合,在教育教学[9]中发挥重要的作用[85],例如在医学手术教学与规划过程,将病患的情况呈现在医生眼前,通过自然手势规划手术方案,有助于手术过程的效率; 在对人类有害的危险场合,通过在虚拟现实系统中人手遥控操作机器人或机器手完成相应的任务.

此外,目前机械手已在工业生产线、危险环境操作等领域中得到了广泛的应用,但与正常人手迥异的结构使得其应用存在一定的局限.如何将人手姿态自然地迁移到机械手上,也是值得研究者努力的方向.

4.4 非接触控制

非接触控制技术主要包含下述几类应用方向:控制音乐录制[80]、游戏控制、家电控制、车载交互系统控制、机器人操纵等.这些应用程序中使用的手势与用于与不同表示类型交互的手势类似,混合使用了一些预定义的手势来触发必须学习的预定义动作,以及用于在两个预定义手势之间导航的自由形式的手势.自由形式的手势通常更多地用于机器人或游戏控制.在手势提取过程中,指定的手势偶尔会考虑用户的偏好,通常用于家电控制手势的定义.

汽车用户界面也是非接触控制技术的重点应用领域[5,6].许多科技公司与汽车厂商合作研发了基于手势控制的车载信息系统,通过车载摄像头对特定手势进行识别,完成原本通过汽车仪表盘上各种旋钮和按钮所完成的功能,提高驾驶安全.

手势交互也可以用于辅助型应用技术.当老年用户与“生活辅助环境”中提供帮助的电子设备、计算机或机器人进行交互时,其交互方式可以通过指定的手势被简化.目前,在这些应用程序中尚没有明确的手势使用模式,而主要使用开发者预定义的各种手势.

4.5 手势交互手柄

重量更轻、功率更小、探测更精确的传感器使得基于手势的人机交互动作更加自由、内容更加丰富、交

互更加快捷方便.手势交互手柄作为技术、创意与设计的集大成者,不仅要考虑对手势的探测与分类是否快速准确,更要考虑用户在使用中是否舒适.新一代的手势交互设备不再拘泥于几个固定的手势,而是对整个手进行实时的跟踪与重建——用户的手是什么姿态,系统里呈现的“手”就是什么姿态.接下来将介绍两种商业化手柄,来说明这一手势交互的新应用场景.

Valve Knuckles 手柄,内置了电容压力传感器、加速度传感器与光学测距传感器,实现对于用户手指位置与姿态、动作力度的精确捕捉.具体来说,每个手柄上的 87 个传感器可以实时跟踪手的位置,手指的位置,手的运动和手指对手柄的压力,它可以捕捉到手的位置和每个手指的动作.这款手柄不再需要人手主动抓握,而是用绑带固定在手掌上,用户只需要自然地做出动作,从而在一定程度上实现了“无感交互”.该款手柄通过 SteamVR2.0 进行开发,将手的动作抽象为六种,分别是布尔,代表动作是否发生;单值,代表动作的幅度;二维向量,代表二维平面上的运动;三维向量,代表三维空间内的运动;姿态,代表手(手柄)的位置和旋转;骨架,代表每只手的关节信息,结合内置的手部模型,可以直接还原手的姿态.

Oculus-Quest 手柄,主要配合 VR 眼镜使用.它的手柄需要用户主动握持,通过 VR 眼镜上的远红外相机捕捉手柄的位置,从而实现用户与机器的交互.同时,其 VR 眼镜上的相机也可以探测到用户的手并识别常见手势,比如捏取、滑动、指向、旋转等,通过开发者在 Unity 平台上的设置,来决定手势探测的开始条件、结束条件以及每种手势的交互场景、对应操作.

在 VR,AR 等领域,类似设备的出现可明显降低用户进行手势交互的学习成本.

4.6 医疗康复

康复评定在是康复医学领域发挥着很重要的作用,三维动态手势运动获取和分析可为人手肢体障碍人士的康复评定提供科学的数据基础和辅助分析工具.例如,对于脑卒中等疾病引起的手部肢体运动功能障碍的康复治疗主要是通过康复训练进行重建,以改善患者手部的运动功能[7].传统的康复治疗缺乏手部肢体训练定量参数和康复效果关系的客观数据,难以通过对训练参数进行优化以获得理想的康复治疗方案.为此,研究三维动态手势运动获取和分析具有重要的研究和应用价值.

5 总结与讨论

手势作为一种输入通道已在人机交互、虚拟现实等领域得到了广泛的应用,引起了研究者的关注.特别是随着先进人机交互技术的出现以及计算机技术(特别是深度学习、GPU 并行计算等)的飞速发展,手势理解和交互方法取得了突破性的成果.手势交互现阶段主要的问题,从宏观上说,在于对于界面的哪一部分功能适合用手势操控仍没有达成共识;而从微观上说,在于每种应用对应什么样的手势,仍然莫衷一是.

1)还没有形成手势交互界面中统一手势选择标准和框架.对于调查用户偏好的研究,有建立框架与标准的趋势,主要问题:样本过少、涉及应用的范围过窄、没有细致考量对于被试背景的影响.

2)目的相似的解决方案手势不一定相似.被调研的工作在引述相关工作时,只考察了其应用开发的目标,而鲜有提及这些工作与自身的工作使用的手势是否相近及选择的依据. 若想要构建一个通用性强的框架与标准,则应当考虑每个应用场景下的各种工作有何规律,而在这一场景下被使用的手势有是否有规律.在这种规律的基础上,制定出的统一框架才更易被人接受.

3)语境分类相关的问题.语音与按钮都可以辅助手势表达意思,但是这些元素的使用并不仅仅是简单的“相辅相成”的关系.在实际工作中,何种元素被置于何种地位,往往依赖于技术发展的水平.对于同时使用手势与语音的场景,尽管两者都可能起到了传递信息的作用,但是所传递的信息仍然不尽相同.对于这种场景,如果能够分别看待语音与手势所起到的作用,不仅可以增进对手势在交互中的作用的了解,对手势其他方面的研究也大有裨益.

4)对技术的依赖.每种界面使用何种技术进行手势的捕捉与处理,还没有一套统一的标准,在这种情况下,尤其是大样本数的调查缺乏,使得各个工作的结论说服力不强.如果能够真正探明在每种应用场景之下,人群对手势选择的具有怎样的本能反应,会极具指导意义.

随着传感器技术和人工智能技术的发展,手势交互界面将会越来越直观和自然,在虚拟现实和增强现实等中的应用会更加普及.

References:

[1] Wachs JP, Kölsch M, Stern H, Edan Y, Vision-based hand-gesture applications. Communications of the Acm, 2011. 54(2): 60-71.

[2] Xia SH, Gao L, Lai YK, Yuan MZ, Chai JX. A survey on human performance capture and animation. Journal of Computer Science and Technology, 2017, 32(3):536-554.

[3] Gartner 2017, Gartner's Top 10 Strategic Technology Trends for 2017, https://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/

[4] Zhang FJ, Dai GZ, Peng XL. A survey on human-computer interaction in virtual reality. Scientia Sinica Informationis, 2016, 46(12): 1711-1736 (in Chinese with English abstract).

[5] Huang J, Han DQ, Chen YN, Tian F, Wang HA, Dai GZ. A survey on human-computer interaction in mixed reality. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(6): 869-880 (in Chinese with English abstract).

[6] Yu HC, Yang XD, Zhang YW, Zhong X, Chen YQ. A review on the recognition of mid-air gestures. Science & Technology Review, 2017, 35(16): 64-73 (in Chinese with English abstract).

[7] Guo XH, Wang J, Xu GH. The latest progress in the research of hand function rehabilitation robot. Chinese Journal of Rehabilitation Medicine, 2017, 32(2):235-240 (in Chinese).

[8] Vuletic T, Duffy A, Hay L, McTeague C, Campbell G, Grealy M. Systematic literature review of hand gestures used in human computer interaction interfaces. International Journal of Human-Computer Studies, 2019. 129:74-94.

[9] Zhu YJ, Li CP, Ma WL, Xia SH, Zhang TL, Wang ZQ. Interaction feature modeling of virtual object in inmmersive virtual assembly. Journal of Computer Research and Development, 2011, 48(7): 1298-1306 (in Chinese with English abstract).

[10] Xu YH, Li JR. Research and implementation of virtual hand interaction in virtual mechanical assembly. Machinery, Design & Manufacture, 2014, 5: 262-266 (in Chinese with English abstract).

[11] Wu HY, Zhang FJ, Liu YJ, Dai GZ. Research on key issues of vision-based gesture interfaces. Chinese Journal of Computers, 2019,32(10): 2030-2041 (in Chinese with English abstract).

[12] Ren P, Zhou MQ, Fan YC, Qian L, Shui WY. A rapid ancient architecture modeling method facing the gesture interaction. Transactions of Beijing Institude of Technology, 2018, 38(4): 412-416,436 (in Chinese with English abstract).

[13] Wang XH, Hua W, Bao HJ. Design and development of a gesture-based interaction system for multi-projector tiled display wall. Journal of Computer-Aided Design & Computer Graphics, 200719(3): 318-322, 328 (in Chinese with English abstract).

[14] Weichert F, Bachmann D, Rudak B, Fisseler D. Analysis of the accuracy and robustness of the leap motion controller. Sensors, 2013, 13(5):6380-6393.

[15] Oikonomidis I, Kyriazis N, Argyros AA. Efficient model-based 3D tracking of hand articulations using kinect. British Machine Vision Conference (BMVC), 2011,3(1).

[16] Romero J, Dimitrios Tzionas, Black JM. Embodied Hands: Modeling and capturing hands and bodies together. SIGGRAPH Asia 2017.

[17] Tompson J, Stein M, LeCun Y, Perlin K. Real-time continuous pose recovery of human hands using convolutional networks. ACM Trans. Graph, 2014. 33(5): 1-10.

[18] Oberweger M, Lepetit V. Improving fast and accurate 3D hand pose estimation. Proceedings of the IEEE International Conference on Computer Vision, 2017: 585-594

[19] Oberweger M, Wohlhart P, Lepetit V. Training a feed back loop for hand pose estimation. Proceedings of the IEEE International Conference on Computer Vision, 2015: 3316-3324

[20] Ge LH, Liang H, Yuan JS, Thalmann D. Robust 3D hand pose estimation in single depth images: from single-view cnn to multi-view cnns. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3593-3601

[21] Ge LH, Cai YJ, Weng GW, Yuan JS. Hand pointnet: 3D hand pose estimation using point sets. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8417-8426

[22] Moon GS, Chang YJ, Lee KM. V2v-posenet: voxel-to-voxel prediction network for accurate 3D hand and human pose estimation from a single depth map. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5080-5088

[23] Rad M, Oberweger M, Lepetit V. Feature mapping for learing fast and accurate 3D pose inference from synthetic images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4663-4672

[24] Poier G, Schinagl D, Bischof H. Learning pose specific representations by predicting different views. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 60-69

[25] Baek SR, Kim KI, Kim TK. Augmented skeleton space transfer for depth-based hand pose estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8330-8339

[26] Dibra E, Wolf T, Oztireli C, Gross M. How to refine 3D hand pose estimation from unlabelled depth data?. 2017 International Conference on 3D Vision(3DV). IEEE, 2017: 135-144

[27] Wan CD, Probst T, Van Gool L, Yao A. Dense 3D regression for hand pose estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5147-5156

[28] Ye Q, Yuan S, Kim T K. Spatial attention deep net with partial pso for hierarchical hyrid hand pose estimation. European conference on computer vision. Springer, 2016: 346-361.

[29] Malik J, Elhayek A, Nummari F, Varanasi K, Tamaddon K, Heloir A, Stricker D. DeepHPS: End-to-end estimation of 3D hand pose and shape by learning from synthetic depth. 2018 International Conference on 3D Vision (3DV), Verona, 2018: 110-119, doi: 10.1109/3DV.2018.00023.

[30] Zimmermann C, Brox T. Learning to estimate 3D hand pose from single RGB images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[31] Simon T, Joo H, Matthews I, Sheikh Y. Hand keypoint detection in single images using multiview bootstrapping. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1145-1153

[32] Mueller F, Bernard F, Sotnychenko O, Mehta D, Sridhar S, Casas D, Theobalt C. GANerated hands for real-time 3D hand tracking from monocular RGB. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 49-59

[33] Spurr A, Song J, Park S, Hilliges O. Cross-modal deel variational hand pose estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-98

[34] Lien J, Gillian N, Karagozler ME, Amyhood P, Schwesig C, Olson E, Raja H, Poupyrev I. Soli: Ubiquitous gesture sensing with millimeter wave radar. ACM Transactions on Graphics, 2016, 35(4):1-19.

[35] Nymoen K, Haugen MR, Jensenius AR. MuMYO - Evaluating and Exploring the MYO armband for musical interaction. International Conference on New Interfaces for Musical Expression. The School of Music and the Center for Computation and Technology (CCT), Louisiana State University, 2015.

[36] Han SC, Liu BB, Wang R, Ye YT, Twigg CD, Chen K. Online optical marker-based hand tracking with deep labels. ACM Transactions on Graphics, 34(7), 2018. Article No: 166.

[37] C.W. Ng, S. Ranganath, Real-time gesture recognition system and application. Image and Vision Computing. 20 (2002) 993–1007.

[38] Cheng H, Yang L, Liu ZC, A survey on 3D hand gesture recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2015. 26: 1-1.

[39] Elmezain M, Al-Hamadi A, Appenrodt J, Michaelis B. A hidden markov model-based continuous gesture recognition system for hand motion trajectory. 19th International Conference on Pattern Recognition. 2008: 1–4.

[40] Keskin C, Cemgil AT, Akarun L. DTW based clustering to improve hand gesture recognition. Proceedings of Human Behavior Understanding, 2011: 72–81.

[41] Arici T, Celebi S, Aydin AS, Temiz TT, Robust gesture recognition using feature pre-processing and weighted dynamic time warping. Multimedia Tools Application, 2014,72(3): 3045–3062.

[42] Reyes M, Dominguez G, Escalera S, Featureweighting in dynamic timewarping for gesture recognition in depth data. IEEE International Conference on Computer Vision Workshops, 2011: 1182–1188.

[43] Dong C, Leu MC, Yin Z. American sign language alphabet recognition using microsoft Kinect. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2015.

[44] Song Y, Demirdjian D, Davis R. Tracking body and hands for gesture recognition: natops aircraft handling signals database. In Proceedings of the 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011). Santa Barbara, CA, 2011.

[45] Molchanov P, Yang X, Gupta S, Kim KW, Tyree S, Kautz J. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[46] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M, Learning spatiotemporal features with 3d convolutional networks. IEEE International Conference on Computer Vision, 2015: 4489-4497.

[47] Camgoz N C, Hadfield S, Koller O, Bowden R. SubUNets: End-to-end hand shape and continuous sign language recognition. IEEE International Conference on Computer Vision (ICCV), 2017.

[48] Cui Rp, Liu H, Zhang CS, Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1610-1618.

[49] Cao CQ, Zhang YF, Wu Y, Lu HQ, Cheng J. Egocentric gesture recognition using recurrent 3d convolutional neural networks with spatiotemporal transformer modules. IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2017.

[50] Narayana P, Beveridge RJ, Draper BA. Gesture recognition: focus on the hands. IEEE Conference on Computer Vision and Pattern Recognition, 2018. 5235-5244.

[51] Bambach S, Lee S, Crandall DJ, Chen Y. Lending a hand: detecting hands and recognizing activities in complex egocentric interactions. Proceedings of the IEEE International Conference on Computer Vision. 2015: 1949-1957.

[52] Rogez G, Supancic JS, Ramanan D. Understanding everyday hands in action from RGB-D images. Proceedings of the IEEE International Conference on Computer Vision. 2015: 3889-3897.

[53] Joshi A, Ghosh S, Betke M, Sclaroff S, Pfister H. Personalizing gesture recognition using hierarchical bayesian neural networks. IEEE Conference on Computer Vision and Pattern Recognition 2017. 455-464.

[54] Hu ZX, Hu YM, Liu J, Wu B, Han DM, Kurfess T. 3D Separable Convolutional Neural Network for Dynamic Hand Gesture Recognition. Neurocomputing, 2018. 318: 151-161.

[55] Sánchez J, Perronnin F, Mensink T, Verbeek J. Image classification with the fisher vector: theory and practice. International Journal of Computer Vision, 2013, 105(3):222-245.

[56] Smedt DQ, Wannous H, Vandeborre JP. Heterogeneous hand gesture recognition using 3D dynamic skeletal data. Computer Vision and Image Understanding, 2019.

[57] Zhao D, Liu Y, Li GC. Skeleton-based Dynamic Hand Gesture Recognition using 3D Depth Data. Electronic Imaging, 2018

[58] Boulahia SY, Anquetil E, Multon F, Kulpa R. Dynamic hand gesture recognition based on 3d pattern assembled trajectories. Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2017: 1-6.

[59] Devineau G, Moutarde F, Xi W, Yang J. Deep learning for hand gesture recognition on skeletal data, the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018):106-113.

[60] Chen XH, Wang GJ, Guo HK, Zhang CR, Wang H, Zhang L. Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition. IEEE International Conference on Image Processing, 2017.

[61] Hou JX, Wang GJ, Chen XH, Xue JH, Zhu R, Yang HZ. Spatial-Temporal attention Res-TCN for skeleton-based dynamic hand gesture recognition. European Conference on Computer Vision. Springer, Cham, 2018.

[62] Avola D, Bernardi M, Cinque L, Foresti LG, Massaroni C. Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Transactions on Multimedia, 2018, 21(1): 234-245.

[63] Pisharady PK, Saerbeck M, Recent methods and databases in vision-based hand gesture recognition: a review. Computer Vision & Image Understanding, 2015. 141(C): P. 152-165.

[64] Guyon I, Athitsos V, Jangyodsuk P, Escalante HJ. The ChaLearn gesture dataset (CGD 2011). Machine Vision and Applications, 2014, 25(8): 1929-1951.

[65] Escalera S, Gonzàlez J, Baró X, Reyes M, Lopes O, Guyon I, Athistos V, Escalante HJ, Multi-modal gesture recognition challenge 2013: dataset and results. Acm International Conference on Multimodal Interaction 2013.

[66] Triesch J, Christoph M. Robust classification of hand postures against complex backgrounds. Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996:170-175.

[67] Triesch J, Christoph M. A system for person-independent hand posture recognition against complex backgrounds. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(12): 1449-1453.

[68] Marcel S, Bernier O. Hand posture recognition in a body-face centered space. Proceedings of the Conference on Human Factors in Computer Systems (CHI), 1999.

[69] Marcel S, Bernier O, Viallet JE, Collobert D. Hand gesture recognition using input/ouput hidden markov models. Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (AFGR), 2000.

[70] Smedt QD, Wannous H, Vandeborre JP, Skeleton-based dynamic hand gesture recognition. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016.

[71] Smedt QD, Wannous H, Vandeborre JP, Guerry J, Bertrand LS, Filliat D, SHREC'17 Track: 3d hand gesture recognition using a depth and skeletal dataset. 10th Eurographics Workshop on 3D Object Retrieval, 2017.

[72] Zhang YF, Cao CQ, Cheng J, Lu HQ. EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition. IEEE Transactions on Multimedia (T-MM), 2018, 20(5): 1038-1050.

[73] Garcia-Hernando G, Yuan SX, Baek SR, Kim TK. First-person hand action benchmark with RGB-D videos and 3D hand pose annotations. IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[74] Bullock IM, Feix T, Dollar AM. The yale human grasping dataset: grasp, object, and task data in household and machine shop environments. The International Journal of Robotics Research, 2015, 34(3): 251-255.

[75] Zhang YP, Han T, Ren ZM, Umetani N, Tong X, Liu Y, Shiratori T, Cao X. BodyAvatar: creating freeform 3d avatars using first-person body gestures. Proceedings of the 26th annual ACM symposium on User interface software and technology. 2013: 387-396.

[76] Pfeil KP, Koh SL, LaViola JJ, Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. Proceedings of the 2013 international conference on Intelligent user interfaces, 2013. 257-266.

[77] Vinayak, K. Ramani K. Extracting hand grasp and motion for intent expression in mid-air shape deformation: a concrete and iterative exploration through a virtual pottery application. Computers & Graphics, 2016. 55: 143-156.

[78] Hilliges O, Kim D, Izadi S, Weiss M, Wilson DA. HoloDesk: Direct 3D interactions with a situated see-through display. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012: 2421-2430.

[79] Colaço A, Kirmani A, Yang HS. Mime: Compact, Low-Power 3D gesture sensing for interaction with head-mounted displays. Proceedings of the 26th annual ACM symposium on User interface software and technology, 2013: 227-236.

[80] Liang H, Wang J, Sun Q, Liu YJ, Yuan JS, Luo J, He Y. Barehanded music: real-time hand interaction for virtual piano. Proceedings of 20th Acm Siggraph Symposium on Interactive 3D Graphics and Games, 2016: 87-94.

[81] Kim, YK, Bae SH, and Acm, SketchingWithHands: 3D sketching handheld products with first-person hand posture. Proceedings of the 29th Annual Symposium on User Interface Software and Technology. 2016: 797-808.

[82] Yi X, Yu C, Zhang MR, Gao SD, Sun K, Shi YC. ATK: enabling ten-finger freehand typing in air based on 3d hand tracking data. Proceedings of the 28th Annual ACM Symposium on User Interface Software, 2015: 539-548.

[83] Cui J, Fellner DW, Kuijper A, Sourin A. Mid-air gestures for virtual modeling with leap motion. Springer International Publishing, 2016.

[84] Liang H, Chang J, Kazmi IK, Zhang JJ, Jiao PF. Hand gesture-based interactive puppetry system to assist storytelling for children. The Visual Computer. 2016, 33(4): 517-531

[85] Hatscher B, Luz M, Nacke LE, Elkmann N, Müller V, Hansen C. GazeTap: Towards hands-free interaction in the operating room. Proceedings of the 19th ACM International Conference on Multimodal Interaction. 2017: 243-251.

[86] Sun SQ, Zhang LS. Three-dimension sketch design oriented to product innovation. Computer Integrated Manufacturing Systems, 2007(02): 224-227, 274 (in Chinese with English abstract).

[87] Shen JC, Luo YL, Wu ZK, Tian Y, Deng QQ. CUDA-based real-time hand gesture interaction and visualization for CT volume dataset using leap motion. The Visual Computer., 2016. 32(3): 359-370.

附中文参考文献:

[4] 张凤军,戴国忠,彭晓兰.虚拟现实的人机交互综述.中国科学:信息科学,2016,46(12):1711-1736.

[5] 黄进,韩冬奇,陈毅能,田丰,王宏安,戴国忠.混合现实中的人机交互综述.计算机辅助设计与图形学学报,2016,28(06):869-880.
[6] 于汉超,杨晓东,张迎伟,钟习,陈益强,凌空手势识别综述.科技导报,2017,16:64-73
[7] 郭晓辉,王晶,徐光华,手部功能康复机器人研究最新进展.中国康复医学杂志,2017,2:235-240
[9] 朱英杰,李淳芃,马万里,夏时洪,张铁林,王兆其,沉浸式虚拟装配中物体交互特征建模方法研究.计算机研究与发展,2011,48(7): 1298-1306.
[10] 绪玉花,李静蓉,面向虚拟装配的虚拟手交互技术研究.机械设计与制造,2014,5
[11] 武汇岳,张凤军,刘玉进,戴国忠.基于视觉的手势界面关键技术研究.计算机学报,2009,32(10):2030-2041.
[12] 任镆,周明全,樊亚春,钱露,税午阳,面向手势交互的古建场景快速搭建方法.北京理工大学学报,2018,38(04):412-416,436.
[13] 王修晖,华炜,鲍虎军,面向多投影显示墙的手势交互系统设计与实现.计算机辅助设计与图形学学报,2007(03):318-322,328.
[86] 黄琦,孙守迁,张立珊,面向产品创新的 3 维草图设计技术研究.计算机集成制造系统,2007(02):224-227,274.