

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)02-0349-25

论文引用格式: Wang Y W and Zhao X. 2026. Input-output modalities in augmented reality: a two-dimensional survey of human-computer interaction. Journal of Image and Graphics, 31(2):0349-0373(王怡雯, 赵玺. 2026. 输入输出双视角下的增强现实人机交互技术综述. 中国图象图形学报, 31(2):0349-0373)[DOI:10.11834/jig.250197]

输入输出双视角下的增强现实人机交互技术综述

王怡雯, 赵玺*

西安交通大学计算机科学与技术学院, 西安 710049

摘要: 增强现实(augmented reality, AR)技术通过三维空间感知与多模态交互机制,突破了传统二维交互的局限,其应用已渗透至医疗、教育、工业设计等领域。然而,现有研究多聚焦单一输入模态或特定应用场景,缺乏对多模态输入与跨感官反馈的系统性整合。本文构建了输入—输出双维度的AR交互技术全景框架:输入模态涵盖语音与姿态识别(包括眼动、头部及肢体动作),输出模态依据人类五感分类解析。该框架不仅为研究者提供技术链路图谱,还可通过系统性的技术缺口分析引导创新方向。研究发现,当前AR交互面临多模态融合复杂度高、隐私安全隐患、输入输出失衡及跨学科协同不足等核心挑战,需通过材料科学、认知神经科学、计算机伦理及计算机科学等领域的深度交叉,推动AR从界面增强向认知协同的范式跃迁。

关键词: 增强现实(AR); 人机交互; 多模态; 输入模态; 反馈

Input-output modalities in augmented reality: a two-dimensional survey of human-computer interaction

Wang Yiwen, Zhao Xi*

School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: Augmented reality (AR) technology has transcended the limitations of traditional 2D interactions relying on 3D spatial perception and multimodal interaction mechanisms. By seamlessly integrating digital content with the physical environment, AR systems offer users immersive experiences. In AR environments, users can navigate physical spaces while interacting with virtual objects through intuitive and natural behaviors. Such a demand has positioned AR technology at the forefront of human-computer interaction (HCI) research, attracting considerable academic and industrial attention across multiple fields. The rapid increase in consumer-grade AR devices—including Microsoft HoloLens, Meta Quest Pro, and Apple Vision Pro—has accelerated the technology's adoption across various scenarios. This widespread deployment requires robust interaction frameworks that support AR's unique operational demands. Different interaction modalities present diverse characteristics. For instance, head/face-based interaction is suitable for scenarios in which users' hands are occupied, though it may lead to limited precisions. By contrast, hand-based manipulation offers greater flexibility and is good for tasks requiring precise control. Each interaction modality relies on specific hardware configurations. Applications employing bare-hand interaction must consider camera placement to ensure complete hand tracking for accurate inter-

收稿日期: 2025-05-08; 修回日期: 2025-06-24; 预印本日期: 2025-07-01

* 通信作者: 赵玺 xi.zhao@mail.xjtu.edu.cn

基金项目: 国家自然科学基金项目(U23A20312); 国家重点研发计划资助(2022YFB3303200); 陕西省重点研发计划资助(2024SF-GJHX-27)

Supported by: National Natural Science Foundation of China (U23A20312); National Key R&D Program of China (2022YFB3303200); Key R&D Program of Shaanxi Province, China (2024SF-GJHX-27)

action. Meanwhile, applications utilizing gaze interaction require careful consideration of eye-tracking apparatus installation. To this end, selecting appropriate interaction methods on the basis of task requirements while designing corresponding hardware support constitutes one of the key challenges hindering the widespread adoption of AR technologies. Unlike traditional computing environments, AR interactions occur in dynamic 3D spaces and engage multiple human sensory channels simultaneously. Consequently, effective AR systems must integrate diverse input modalities (e. g. , hand gestures, voice commands, and eye tracking) with appropriate output feedback systems (e. g. , haptic responses, spatialized audio, and olfactory cues). However, existing research landscapes reveal significant gaps in addressing these complex requirements. Current studies predominantly focus on isolated input modalities——such as gesture recognition algorithms or eye-tracking precision——or narrow application-specific implementations (e. g. , museum guides or classroom applications). While these investigations provide valuable domain-specific insights, they fail to establish comprehensive frameworks that integrate multimodal information. More critically, the majority of AR HCI literature emphasizes input mechanisms while neglecting nonvisual feedback channels. This limitation results in a fragmented understanding of cross-sensory interaction paradigms, leaving critical questions regarding multimodal synchronization, sensory bandwidth constraints, and perceptual congruence. We propose a dual-dimensional analytical framework based on input and output modalities for AR interaction technologies. Through a comprehensive review of existing AR interaction paradigms——including their underlying principles, recent developments, core applications, and functional characteristics——we construct an integrated conceptual map to support future research. The input modalities are classified into speech recognition and gesture-based inputs, with the latter further subdivided into eye-, head motion-, and body movement-based interaction technologies on the basis of anatomical engagement. Output modalities are classified according to in accordance with the human sensory channels into vision, hearing, touch, smell, and taste. Our analysis reveals four fundamental challenges in contemporary AR interaction systems: 1) multimodal fusion complexity: real-time sensor fusion imposes significant computational demands. For instance, integrating high-frequency gesture tracking with voice recognition often exceeds the capabilities of current edge computing systems, resulting in perceptible visual latency. 2) Privacy-security concerns: AR devices equipped with cameras and microphones raise concerns about surveillance and data exposure. Meanwhile, neural interface technologies introduce new challenges surrounding the protection of neural data and cognitive privacy. 3) Input-output asymmetry: disproportionate development focus on input systems has created feedback channels incapable of matching the richness of user actions (e. g. , advanced gesture recognition paired with rudimentary vibration feedback). 4) Interdisciplinary fragmentation: gaps between materials science (for working on flexible electronics), neuroscience (for studying multisensory perception), and computer engineering (for optimizing rendering pipelines) hinder comprehensive system-level optimization and innovation. Addressing these challenges requires deep integration across traditionally siloed disciplines. Materials science must advance the development of flexible, biocompatible sensors; cognitive neuroscience should quantify thresholds for effective multisensory integration; computer science needs to create adaptive algorithms for latency compensation; and computer ethics must establish robust frameworks for governing neural data. Only through such synergistic efforts can AR evolve beyond its current role as an interface enhancement tool into a transformative platform for human-machine cognitive collaboration. In summary, the proposed framework offers dual value to AR researchers and developers, particularly newcomers. On the one hand, it organizes fragmented technologies into a coherent structure, accelerating learning and helping readers navigate complex technical domains while identifying key research questions. On the other hand, through critical analysis of technical bottlenecks——such as cross-modal latency and limited sensory bandwidth——it highlights underexplored areas and inspires innovative solutions, avoiding redundant efforts and low-level repetition of existing work.

Key words: augmented reality (AR); human-computer interaction; multi-modality; input modality; response

0 引言

增强现实(augmented reality, AR)通过三维建

模、虚实融合与智能交互技术,将虚拟信息叠加于真实环境,为用户提供多感官沉浸式体验。通过同步定位与地图构建(simultaneous localization and mapping, SLAM)、三维追踪及实时渲染等核心技术,增

强现实系统支持用户在动态三维空间中与虚实对象交互。与传统二维交互依赖固定屏幕和单一输入(如触控、键盘)不同,AR交互发生在动态的三维空间中,需要整合多样化的输入(如手势、眼动)与输出模式(如触觉、空间音频),这对交互设计提出更高层次的挑战。

随着商品级AR设备(如Microsoft hololens、meta quest)的普及,AR技术逐渐渗透到医疗(Anjali等,2022)、教育(Al-Ansi等,2023)、工业设计(Cheng等,2021a)等多个行业,其独特的交互模式和多样化的应用场景吸引了大量研究者的关注。然而,现有关于AR人机交互技术的综述多聚焦单一的输入模式(如手势识别(Gavgiotaki等,2023)、眼动追踪(Jin等,2024))或特定领域应用(如博物馆导览(Zhou等,2022));同时,这些综述多从用户输入的角度出发,忽视了除视觉之外的反馈模式,如触觉和嗅觉等其他感官模式(陶建华等,2022),缺乏对多种模式输入与跨感官反馈的全面性、系统性认识。

本文从输入—输出双维度构建AR交互技术的全景分析框架,对现有的增强现实交互技术和应用进行了梳理,介绍了各种模式的基本原理、当前的研究进展、主要的应用领域以及功能特点,为多模态融合和跨感官交互研究提供了复合型知识图谱。框架图如图1所示。输入模式方面,划分为基于语音识别的输入和基于姿态识别的输入。在基于姿态识别的输入模式中,根据主要关注的交互部位,进一步细分为基于眼部输入、基于头部运动输入以及基于肢体运动输入的增强现实交互技术。输出模式则依据人类五感(视觉、听觉、触觉、嗅觉和味觉)分类解析。这一全景视角可为增强现实领域的研究者与开发者(尤其是新进入者)提供双重支持:1)通过系统化的技术框架缩短领域学习周期,帮助读者跨越碎片化技术现状与复杂应用场景的认知壁垒,快速锚定研究问题的核心;2)基于对现有技术链路的梳理,既避免低水平重复工作,又通过分析技术瓶颈激发创新方向。

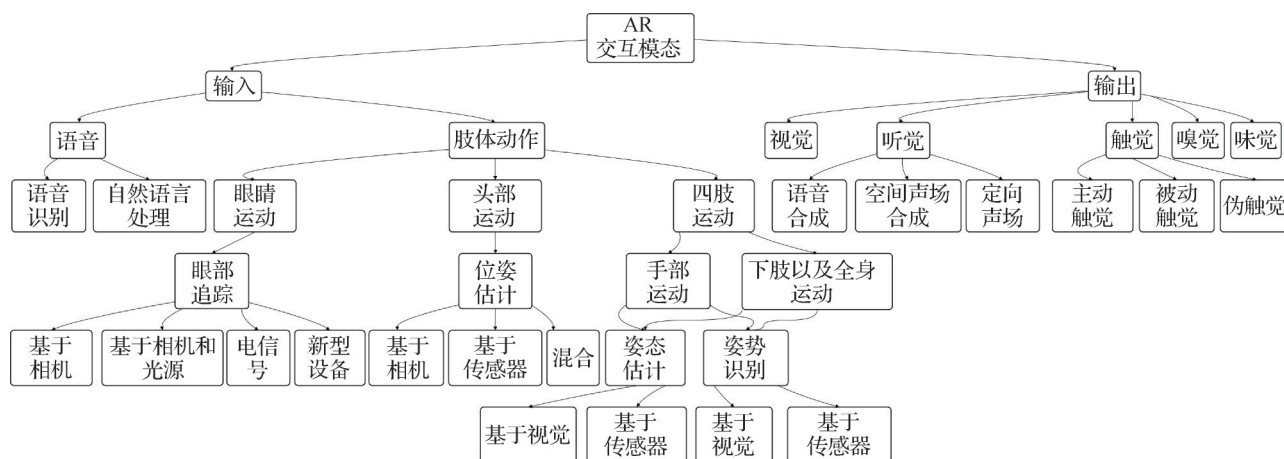


图1 输入—输出双维度的增强现实交互技术全景分析框架

Fig. 1 A dual-dimensional analytical framework based on input and output modalities for AR interaction technologies

1 增强现实应用中的输入模式

在现实世界中,人们常常通过声音或肢体动作与物理环境进行互动。例如,可以通过语音指令控制智能设备,或通过手势、视线等肢体动作表达意图。这些交互方式不仅直观自然,还蕴含了丰富的信息,例如用户的位置、身份,甚至是周围环境的特征。随着技术的发展,如何通过传感器和算法捕捉并识别这些声音和肢体动作信号,使机器能够理解

用户的意图并做出自然的响应,成为人机交互领域的研究热点。在增强现实(AR)领域,丰富用户的交互方式对于提升沉浸感和交互效率至关重要。然而,不同的交互方式各有特点,例如基于头面部的交互方式适合双手被占用的场景,但可能存在精度问题;而手部操作则更加灵活,适合需要精确控制的场景。同时,交互方式也依赖特定的设备,如使用裸手交互作为主要交互方式的应用需要考虑相机配置和安装位置来完整地捕捉手部信息以保证交互的准确性;以视线交互为主要交互方式的设备则需要考虑

眼动装置的架设等。因此,如何根据任务需求选择合适的交互方式,并设计相应的硬件支持,成为增强现实技术广泛应用的关键挑战之一。本节详细介绍基于姿态识别及基于语音识别的各种交互技术,包含其基本的技术原理、研究现状、应用领域以及特点分析,以期增强现实交互设计提供理论支持和实践指导。

1.1 姿态识别

肢体语言是人类重要的交互形式之一,包括眼动、头面部运动、手势或其他身体部位的运动等多种形式。通过单个部位的动作或多部位结合的动作,人们可以传达感觉、思想和情感(Noroozi等,2021)。20世纪以来,姿态识别技术不断发展,逐渐在增强现实系统中得到应用,从虚拟键盘(Lee和Lee,2018)到基于面部、身体动作识别技术的先进安全系统(Bhalla等,2021),姿态识别技术正在改变人们与数字世界的交互方式。在增强现实的应用场景下,根据主导交互的身体部位不同,衍生出多种基于姿态识别的系统。不同部位的数据表示方式各不相同,相应的识别算法和交互设计也存在差异,如表1所示。本节根据交互方式的主导身体部位以及数据表示的特点,分别介绍眼部、头部和肢体输入的姿态识别技术,包括原理、技术进展和应用案例。

1.1.1 眼部姿态识别

人类获取信息的主要途径依赖于视觉感知,而视线的方向和关注点能够直观反映用户的兴趣和意图。作为获取眼睛运动信息的基础技术,眼动追踪技术得到广泛的研究与应用(Jin等,2024)。

1)眼动数据表示。眼动跟踪是指自动检测瞳孔中心位置或识别三维视线方向及注视点的过程。这一技术的核心在于捕捉和分析眼动数据,而眼动数据通常由高频采样的注视点坐标序列(2D或3D)构成,具有高频低维的特征。基于时间和空间上的规律性,眼动数据可以划分为注视、扫视与平滑追踪等多种事件模式。这些模式反映了用户的注意力分布和视觉行为:注视通常用于目标选择和细节观察,扫视用于快速切换注意力,而平滑追踪则用于跟踪移动目标。

2)眼动追踪技术。根据核心设备的不同,眼动追踪技术主要分为以下4类:仅依赖相机的眼动追踪、依赖相机和光源的眼动追踪、依赖电信号的眼动追踪以及基于新型设备的眼动跟踪。

(1)仅依赖相机的眼动追踪是常见且成本较低的一种方案。原理是通过摄像头捕捉用户眼部的图像,利用计算机视觉算法(如特征点检测、模板匹配等)提取瞳孔、虹膜眼角的特征,进而估计视线方向。最常用的是瞳孔中心法,该方法通过摄像头获取眼睛的图像,然后从眼睛图像提取的瞳孔轮廓来计算瞳孔中心,进而推断每帧照片中眼睛的位移量,结合眼部的球体模型即可估计注视点和视线方向。早期的算法主要依赖边缘检测、阈值分割、椭圆拟合等传统手工特征检测并提取瞳孔区域,如ExCuSe(ellipse extraction for robust pupil detection)算法(Fuhl等,2015)等。近年来,基于神经网络的瞳孔检测技术得到了发展,例如Fuhl等人(2016)提出的双卷积神经网络PupilNet以及Lee等人(2020)提出的基于互信息学习的瞳孔追踪网络。得益于大规模眼动数据集的构建,跳过瞳孔检测直接从眼部的图像中预测视线方向(gaze estimation)的方法也逐渐受到关注,例如Zhang等人(2019)提出的直接从单眼图像中估计三维视线方向的MPIIGaze以及Cheng和Lu(2023)提出的基于双眼图像的视线回归网络DVGaze(dual-view gaze)。仅依赖相机的方法无须额外硬件,响应迅速,适用于消费级设备。但对光照敏感,当环境光照变化或头部剧烈运动时,相机拍摄到的图像质量会下降,从而导致追踪的精度下降。

(2)依赖相机和光源的眼动追踪。为了提高精度和稳定性,许多眼动追踪系统采用红外相机和红外光源的组合,用瞳孔角膜反射法估计视线。当红外光源照射用户眼部时,会产生明显的瞳孔反射(称为“普尔钦斑”),红外相机则捕捉这些反射点。算法先利用反射点与光源位置计算得到角膜曲率中心;然后利用图像处理技术计算获得瞳孔中心;最终通过角膜曲率中心与瞳孔中心的连线求得眼球光轴,并利用光轴和视轴之间的夹角计算得到真实的视线方向(Sigut和Sidha,2011)。在反射点检测问题上,已有方法常采用基于深度学习的策略,如Wu等人(2019)提出的3层卷积神经网络EyeNet以及Maquiling等人(2023)提出的瞳孔及角膜反射点检测网络V-ir-Net等。该方法对光照变化较为鲁棒,且能够实现高精度追踪。但仍存在若干局限:其一,头部自由运动会降低定位精度,需通过限制用户头部活动或加装固定装置(如AR/VR(augmented reality/virtual reality)头盔);其二,因需配置高性能红外

表 1 不同输入模态对比

Table 1 Comparison between different input modalities

输入模态	依赖设备	主要技术	算法/工具	数据集	交互设计	应用领域	
语音	麦克风	语音识别	Wav2Vec *(Baevski 等, 2020)	LibriSpeech ASR corpus*	显式为主 (设计明确指令集、身份验证及情绪分析)	制造、娱乐等	
		自然语言理解	NLTK、Spacy 等	ChineseNlpCorpus *			
眼部	RGB 相机	视线检测	PupilNet(Fuhl 等, 2016)	ETH-XGaze*	显隐结合 (物体的选择和操控; 触发系统的自适应行为; 反映用户生理、心理状态)	驾驶、制造、娱乐等	
			DVGaze*(Cheng 和 Lu, 2023)	EVE*			
	红外相机和光源		EyeNet(Wu 等, 2019)	MagicEyes			
			V-ir-Net*(Maquiling 等, 2023)				
	电信号		EOG-Assist*(Narayan 等, 2024)	BlinkDataset*			
	事件相机		Retina*(Bonazzi 等, 2024)	INI-30			
	无镜头相机		FlatTrack(Jain 等, 2025)	-			
头部	相机	相机位姿估计	ORB-SLAM*(Campos 等, 2021)	TUM*	显隐结合 (物体的指向、选择和操控; 触发系统的自适应行为; 身份验证等)	娱乐、机械等	
			DSO*(Engel 等, 2018)				
	惯性传感器		预积分(Forster 等, 2017)	-			
	磁力计		Kataria 等人(2020)	-			
肢体动作	混合		Vins-mono*(Qin 等, 2018)	EuRoC MAV Dataset*			
手部	相机	关节点检测	MediaPipe Hands (Zhang 等, 2020a)		Shrec17*、Egogesture*	显式为主 (虚拟以及真实物体的控制、编辑)	制造、医疗、康复、教育等
		手势识别					
	多传感器集成手套	手势识别	基于 IMU 数据(Kim 等, 2019)				
			基于 Flex 数据(Panda 等, 2021)				
	神经腕带	手势识别	sEMG NN(Kaifosh 等, 2024)	emg2pose*			
下肢	相机	步态识别	SMPLGait(Zheng 等, 2022)	Gait3D*	显式为主 (物体控制、身份验证)	身份验证、康复训练等	
全身	相机	关节点检测	OpenPose*(Cao 等, 2021)	Human3.6M*、UCF101*	显式为主	体感游戏等娱乐领域	
		动作识别	SlowFast*(Feichtenhofer 等, 2019)				
			DD-Net*(Yang 等, 2019)				

注: “*”表示开源; “-”表示无数据。

光源、摄像头及图像处理系统, 导致设备整体能耗与制造成本较高(Jin 等, 2024)。

(3) 依赖电信号的眼动跟踪。除了视觉信息, 还可以利用电信号来跟踪眼部运动信息(Xiong 等, 2024), 例如 Sun 等人(2023)提出的基于脑电图(electroencephalogram, EEG)的眼动跟踪系统 EEG-VET (EEG-based virtual eye tracker), Narayan 等人(2024)提出的基于眼电图(electrooculogram, EOG)

的眼动与眨眼追踪系统等。电信号的收集常依赖贴在皮肤表面的电极, 图 2 展示了采集 EEG 信号的一种装置。这种方法不受光照条件影响, 且能够实现高频率的精确采样, 但需要接触皮肤, 且对舒适度有影响, 因此在消费级的增强现实设备中较少使用。

(4) 基于新型设备的眼动跟踪。事件相机和无镜头相机(Boominathan 等, 2020)是近年来新兴的设备, 在眼动追踪问题上得到了应用, 如 Bonazzi 等人

(2024)提出的基于事件数据流的神经形态眼动追踪框架以及 Jain 等人(2025)提出的首个基于无镜头相机的紧凑型眼动追踪系统 FlatTrack 等。这两种设备具有更高的时间分辨率和更低的功耗,适用于移动设备和 AR/VR 头显。相较于传统固定帧率 RGB 相机,事件相机通过异步捕捉亮度变化生成稀疏事件流,其微秒级时间分辨率与低功耗特性可显著提升 XR(extended reality)眼动追踪性能。无镜头相机则突破光学镜头限制,采用相位/振幅掩膜结合计算成像技术,在实现超薄结构(适用于空间受限设备)的同时,因数据以非直观编码形式记录,天然具备隐私保护功能。

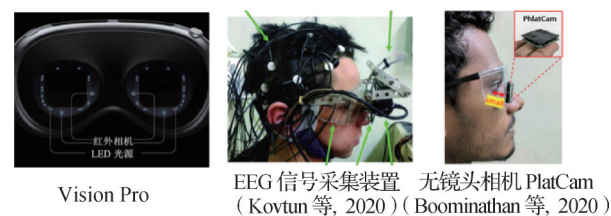


图2 用于眼动追踪的设备
Fig. 2 Eye-tracking devices

3)眼动追踪技术的应用及挑战。主流的基于眼动数据的交互设计有显式交互和隐式交互两种。显式交互关注诸如点选和方向控制等明确指令(如 Dwell-time 激活);隐式交互聚焦于交互系统的自适应和优化(如动态用户接口(user interface, UI)和注视驱动的渲染优先级调整)。

视线交互技术广泛应用于各类对象的选择与控制任务。在静态对象选择方面,研究者通过瞳孔追踪识别用户视线的聚焦与停留,实现对虚拟键盘字母或 UI 文本的精准选择(Mardanbegi 和 Pfeiffer, 2019)。对于动态目标,VRpursuits(Khamis 等, 2018)利用皮尔逊相关系数匹配眼球运动与目标轨迹,从而完成运动物体的选择。此外,视线还可用于虚拟对象的基础编辑操作(如旋转、缩放等),甚至控制物理设备的运动,例如通过眼动调整轮椅转向(Maule 等, 2016; Zhang 和 Hansen, 2019)或改变相机朝向(Caspi 等, 2018)。

视线也用于触发系统的自适应行为。在空间交互优化方面,系统可依据视线方向动态调整虚拟信息的位置与朝向(如驾驶、制造场景),确保关键内容始终位于用户视野范围内(Scholtes 等, 2019;

Ivaschenko 等, 2018)。在注意力感知方面,通过实时检测注视区域,系统可自适应调节界面亮度、滚动速度、消息提示及高亮显示,显著减少手动操作(Yan 等, 2023)。此外,视线还能辅助视图管理——基于瞳孔状态智能控制虚拟信息的展示时机与详细程度,从而优化用户认知负载(Lindlbauer 等, 2019)。在渲染效率提升领域,注视点渲染技术利用人眼中央凹区高分辨特性,优先渲染注视区域的高清图像(如 4D 光场算法(Sun 等, 2017)),而周边区域采用低分辨率处理(如景深滤波系统(Weier 等, 2018)),在保障视觉体验的同时显著降低计算开销。

视线交互还能通过监测用户状态提升交互体验。在社交场景中,系统可利用视线数据优化真人或虚拟角色之间的目光接触,从而增强多人协作、远程会议等场景的临场感与沉浸感(Jing 等, 2021)。同时,眼动特征可作为有效的生理心理指标:通过分析注视模式,系统能实时评估飞行员在高压环境下的压力水平与注意力状态(Stavrev 和 Ginchev, 2023);在教育领域,视线分布还能用于预测学习者的知识掌握程度与理解深度(Rau 等, 2024)。

视线交互凭借快速响应、高效操作以及自然直观等优势,成为人机交互领域的重要技术。它支持非接触式操作,有效降低用户负担,并能与语音、手势等模态结合,实现多模态协同交互。此外,该技术尤其适用于复杂环境或需要解放双手的场景。然而,眼动追踪技术的实际应用仍面临诸多挑战。其一,数据采集与处理难题:眼部运动微小且快速,易受头部移动、光照变化和设备抖动等噪声干扰;且用户的个体差异性大(如眼球结构、运动习惯、生理状态),要求算法具备更强的适应性和鲁棒性;同时,眼睛运动模式多样(如注视、扫视、平滑追踪),增加了算法设计与优化的复杂度。其二,交互设计困境:用户无意识的短暂注视可能触发错误操作,影响交互精准度,即“米达斯误触”;同时由于 AR/VR 环境中,固定焦平面显示易导致用户视觉调节与双眼集合角度不匹配,引发视觉疲劳和不适感(Xia 等, 2019),即调节—集合冲突。其三,成本与普及限制:依赖高精度传感器和计算单元,导致设备成本高昂,且存在能耗问题。

1. 1. 2 头部姿态识别

头部运动是传达用户注意力和交互意图的关键线索。头部追踪则是增强现实系统中获取头部运动

数据的关键技术,通过摄像机或内置传感器来检测用户的头部运动,为虚拟对象注册和交互提供了支持(Nwobodo等,2023)。

1)头部运动数据表示。头部的6自由度运动信息包括3个旋转角度(roll, pitch, yaw)和3个平移坐标(X, Y, Z)。相较于眼动数据,头部运动数据中频中维,数据流更平滑,但也易受抖动干扰。

2)头部跟踪技术。根据数据来源,主流的头部追踪技术可以分为3类:基于视觉的跟踪、基于传感器的跟踪和混合跟踪。

(1)基于视觉的跟踪。该方法通过分析相机捕获的图像来解算相机在全局坐标系下的绝对位姿,即相机位姿估计。相机姿态估计常用特征点匹配、直接法以及基于深度学习的方法。

特征点匹配。该方法首先检测图像中的特征点并找到两幅图像之间的特征匹配关系,然后使用运动恢复结构(structure from motion, SFM)方法来解算相机的姿态,最后使用重投影误差等来优化相机位姿(高翔等,2024),采用ORB(oriented FAST(features from accelerated segment test) and rotated BRIEF(binary robust independent elementary features))特征

的ORB-SLAM系列是代表工作(Campos等,2021),其他常用的特征子还包括SIFT(scale invariant feature Transform)、SURF(speed up robust feature)等。

直接法。该方法通过直接比较两幅图像之间的像素值差异来估计相机的姿态,并用光度误差进行位姿优化。直接法在光照变化、遮挡等情况下仍然有效,但由于需要处理大量的像素点,计算复杂度较高。基于稀疏直接法的视觉里程计DSO(direct sparse odometry)是其中的代表性工作(Engel等,2018)。

基于深度学习的方法。这类方法通过端到端神经网络直接回归相机位姿(Shavit和Ferens,2019),例如PoseNet系列工作(Kendall和Cipolla,2017)。在视觉头部跟踪问题中,基于相机位姿估计的inside-out(内向外)的定位模式是研究和行业发展的主流趋势,本文也主要关注这一模式,不对基于物体跟踪的outside-in(外向内)跟踪策略做过多介绍。

(2)基于传感器的跟踪。该方法利用固定在用户头部或头戴式设备上的传感器(如惯性测量单元(inertial measurement unit, IMU)、磁力计)来估计头部位姿。IMU是最常用的测量物体角速度以及加速

度的装置,结合线性运动假设和预积分计算等方法估计姿态(Forster等,2017)。IMU可提供高频率的位姿估计,但其数据存在噪声和漂移问题。部分IMU中还集成了磁力计,通过磁场来追踪物体位置和方向。传感器可以放置在用户的头部,磁场则由环境中的一组线圈产生。当用户移动头部时,传感器检测到磁场的变化,算法解析磁场变化来估计运动,例如Kataria等人(2020)提出的基于6自由度磁运动追踪系统。基于磁力计方法可以在任何方向上追踪运动,且不需要用户面对特定的方向。然而,磁追踪传感器在抖动方面表现不佳,精度随距离增加而下降,并且对电磁噪声敏感。

(3)混合跟踪。该类方法结合视觉和传感器数据,利用多模态融合技术提高跟踪的准确性和鲁棒性(章国锋等,2024),如基于单目相机和IMU的视觉惯导里程计Vins-mono(Qin等,2018)。在AR场景下,头部追踪的算法最常用的就是基于视觉和IMU融合的相机追踪方案,包括Apple, Meta在内的头部厂商目前均在其设备端部署摄像头和IMU,融合视觉的精确环境感知和IMU对高速运动的快速响应,以减少延迟和漂移问题。

3)头部跟踪技术的应用及挑战。在AR交互设计层面,与眼动数据类似,头部数据也常用于驱动两类典型应用。显式交互关注指向选择虚拟对象等明确指令(如选择指令、运动控制等);隐式交互关注系统的自适应调整 and 用户体验优化(如根据头部位置优化渲染视场角)。

头部运动数据可用于实现选择和控制等基本功能。针对静态物体,研究者利用头部运动实现了光标移动和点选功能(Sancheti等,2018)。对于动态目标选择,SmoothMove系统通过匹配头部轨迹与目标运动轨迹实现精准选择(Esteves等,2017)。基于特定头部姿态(如点头、摇头、绕圈)的交互设计也得到关注,如GlassGesture(Yi等,2016)和HeadGesture(Yan等,2018)。此外,实时头部动作还可用于远程操控机器人运动(Shchekoldin等,2017)。

头部运动反映了用户的注意力分布,也用于优化系统性能和用户体验。例如通过预测观众的头部位姿提前渲染体素视频的相应视角,从而降低了计算负担(Gül等,2020);根据用户双手打字时的头部运动数据对输入的单词或字符进行推测(Slocum等,2023),从而提高响应速度等。部分研究还根据

头部运动来估计用户的压力状态 (Ferrarotti 等, 2024)、情绪状态 (Samanta 和 Guha, 2017), 以及作为生物特征来识别用户的身份 (Bhalla 等, 2021)。

头部姿态追踪使得用户可通过无意识的本体运动实现视场操控, 提供了一种低认知负荷的自然交互模式。然而, 当前头部追踪技术仍面临多重技术瓶颈。其一, 跟踪的稳定性问题。用户的头部快速旋转时容易引发特征描述子丢失或 IMU 漂移, 导致跟踪丢失, 使得虚实内容的融合不稳定。其二, 跟踪的实时性问题。AR 设备计算资源有限, 头部跟踪的实时性难以满足, 从而产生时延, 带来用户体验下降甚至安全问题 (如 VST (video see-through) 头显中延迟导致用户碰到障碍物)。其三, 交互意图识别的问题。与视线选择类似, 基于头部运动的选择等操作存在米达斯误触问题, 设计高效的机制来区分用户的意向性指令与无意识动作仍是一个研究难点; 同时, 还需要考虑持续头部姿势表达带来的疲劳问题。

1. 1. 3 肢体姿态识别

肢体动作 (如手势和姿态) 相比眼动和头部运动, 能够更直观地表达复杂的用户意图。在增强现实领域, 基于肢体运动的姿态估计与识别技术已成为重要的输入方式, 主要包括手部姿态交互、下肢姿态交互和全身姿态交互 3 类。手势作为仅次于面部表情的肢体语言载体, 凭借其高自由度的解剖结构和丰富的语义表达能力, 成为最具发展潜力的自然交互方式。考虑到下肢及全身姿态估计在数据表示、自由度及技术实现方面与手部姿态具有相似性, 本研究将重点分析基于手部姿态识别的增强现实交互技术, 包括其核心原理、研究进展和应用。同时, 简要探讨下肢和全身姿态识别的研究工作及应用。

1) 肢体运动数据表示。肢体交互作为最复杂的姿态交互形态, 其数据特性呈现低频高维的特征。目前学界普遍采用基于骨骼拓扑的运动链模型来建模人体运动, 如图 3 所示。该模型以树状图结构表达关节连接关系, 每个点代表特定关节并参数化其自由度范围。

2) 肢体姿态估计技术。姿态估计的核心任务在于通过回归方法重建肢体在空间中的完整运动学结构, 即身体关节点的三维坐标。当前的肢体姿态估计方法依据传感器差异可分为基于视觉的方法和基于可穿戴设备的方法两大类。

(1) 基于视觉的方法。通常使用 RGB 摄像头、

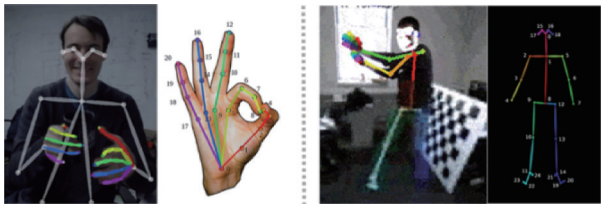


图 3 手部和身体关节点表示 (Cao 等, 2021)

Fig. 3 Keypoints of hand and body (Cao 等, 2021)

RGB-D 摄像头或红外摄像头捕获用户的手部和身体运动, 然后通过计算机视觉相关的算法对关键点进行检测和跟踪。根据算法思路可进一步细分为基于生成的方法和基于判别的方法。

基于生成 (generative-based) 的方法首先需要基于先验知识构建显式的手部或人体模型。目前常用的手部参数模型是 MANO (hand model with articulated and non-rigid deformations) (Romero 等, 2017), 常用的人体参数化模型是 SMPL (skinned multi-person linear model) (Loper 等, 2015), 模型通常会定义姿态参数 θ 以及形状参数 β 。然后通过梯度下降、粒子等算法估计最优参数, 从而得到手部或身体姿态。

基于判别 (discriminative-based) 的方法通过直接从大量的标注数据中学习从图像特征到关节位置的映射关系。早期的判别式的姿态估计主要是基于机器学习的算法, 如支持向量机、随机森林等 (Tang 等, 2013)。近年来, 深度学习方法通过训练深度卷积神经网络, 实现了端到端的关节点预测网络, 如 Google 的开源框架 MediaPipe 系列 (Zhang 等, 2020a) 以及 Openpose (Cao 等, 2021) 等。二者均可从单幅图像中估计手部及全身的关节点位姿态。也有部分工作从视频序列中实时进行人体的姿态估计 (Cao 和 Yan, 2024; Liu 等, 2019)。

基于生成的方法需要为每一帧重新计算模型参数, 速度较慢, 因此实时性能通常较差。此外, 模型每一帧的参数通常基于前一帧的参数进行初始化, 存在误差累积的问题。而判别类方法的模型参数虽然可离线训练并直接用于预测, 但需要大量标注数据来训练, 且在训练和测试的场景差异较大时, 姿态估计的质量也会受到影响。因此, 一些研究人员尝试将基于模型和数据驱动的方法相结合, 如 Baek 等人 (2019) 提出的基于神经渲染的三维手势估计框架。HoloLens 和 Quest 的手势追踪方案也是基于生

成和判别的混合方案。

(2) 基于可穿戴传感器的方法。这类方法通常使用弯曲传感器、IMU等传感器捕获用户的手部运动,然后通过机器学习或深度学习的方法对手部的关键点进行检测和跟踪。常用的传感器包括以下几种:基于弯曲的传感器、基于拉伸的传感器、基于惯性单元的传感器以及基于磁力的传感器(Rashid和Hasan, 2019)。这些穿戴式传感器主要以手套的形式集成并应用,如图4所示。更多基于可穿戴传感器的手部姿态估计算法可以参考Chen等人(2020)、Rashid和Hasan(2019)的工作。



图4 用于手势交互的穿戴式设备

Fig. 4 Wearable devices used for gesture-based interaction

3) 肢体动作识别技术。与肢体动作交互相关的另一个问题是如何将姿态转化为增强现实系统可理解的交互操作,即姿态识别问题,对应于计算机视觉中的手势识别与动作识别。该问题需要将肢体的外观或运动映射到预定义词汇集。需要注意的是,姿态识别的输入不一定是结构化的姿态数据,也可以是其他形式的数据,如图像、视频等。依据数据的来源,姿态识别的方法也可以分为两大类:基于视觉的方法和基于传感器的方法。

(1) 基于视觉的方法。目前主流的基于视觉的姿态识别方法是基于深度学习的方法(Rautaray和Agrawal, 2015)。这类方法通常利用大量标注数据集训练深度神经网络,直接从图像或视频中预测动作的类别,如Meta提出的估计人体姿态的快慢双通道网络SlowFast(Feichtenhofer等, 2019)。部分工作直接以3D关键点数据作为输入,通过训练深度神经网络实现手势识别,如Yang等人(2019)提出的轻量级骨架识别模型DD-Net(double feature double motion network)等。

(2) 基于传感器的方法。除了视觉方案外,也有研究者直接从传感器获取的数据中设计并提取特

征,实现了手势识别。例如Panda等人(2021)提出的基于手指弯曲编码的手势识别方案以及Kim等人(2019)提出的从IMU数据中提取时序信息的动态手势识别网络等。近年来,基于皮肤表面肌电信号(electromyogram, EMG)的手势识别技术也得到了应用(Jaramillo-Yáñez等, 2020),如Meta最新推出的神经腕带Orion(如图4所示)。通过捕捉皮肤表面的电脉冲,sEMG神经网络解码模型被转换成眼镜的输入指令,例如导航、手势检测(手指捏合和滑动)以及手写转录等(Kaifosh等, 2024)。

4) 肢体姿态识别技术的应用。手势识别在增强现实系统中的交互设计以显式方案为主。具体而言,手势可以操作虚拟物体,实现基本的选择、移动和缩放等基础的编辑工作(Fadzli和Ismail, 2019),也用于直接操控机械臂等真实物体(Murhij和Serebrenny, 2021),从而应用到工业装配、手术培训、运动康复、课堂教学和场景导航等各个领域(Gavgiotaki等, 2023)。手势还可以用来实现虚拟键盘输入,为增强现实下的文本输入提供更加自然的体验(Nooruddin等, 2020)。

相较于手部姿态应用的多样性,下肢姿态识别多用于运动康复领域。例如通过识别患者的下肢运动来进行运动纠正(Phan等, 2022)。也有研究者创新性地设计出腿部动作控制的3D鼠标的3DRudder(Minakata等, 2019)。最近一些工作发现下肢动作在用户身份验证上的应用,例如用于识别视频序列中的步态信息的双分支框架SMPLGait(skinned multi-person linear gait)(Zheng等, 2022)。相比人脸识别、指纹识别及虹膜识别,步态识别应用成本更低、识别距离更远、特征不易伪装以及环境适应更强(罗亚波等, 2025)。

全身的姿态输入则常用于娱乐领域,用于用户的体感输入。如Moll等人(2019)开发的一款节奏游戏HeadbangZ。该游戏可识别包含17个关键点类型的人体姿态,并通过分析连续视频帧中检测到的姿态来识别用户动作,特别是与音乐节拍同步的动作,从而给予奖励。

1.2 语音识别

声音作为一种重要的人机交互方式,在不同情境下传递丰富的信息,为用户提供了一种不占用双手、不遮挡视线的交互方式。在增强现实应用中,使用语音作为系统输入所依赖的主要技术就是语音识

别和自然语言理解。

1.2.1 语音识别技术

语音识别技术的核心是将人类语音信号转化为对应的文本(Aouam等,2018)。其流程通常从信号预处理开始:首先通过降噪算法去除环境噪声,随后将语音信号切割为短时帧(每帧约20~40 ms),并加窗处理以减少频谱泄漏。接下来,通过梅尔频率倒谱系数特征提取手段捕捉语音的频域特征,这些特征能够模拟人耳听觉特性,直接作为声学模型的输入。声学模型再将这些特征映射为音素或字词。传统声学模型多为隐马尔可夫模型(hidden Markov model, HMM)与高斯混合模型(Gaussian mixture model, GMM)的声学建模,但近年来深度学习方法逐渐成为主流,如DNN-HMM(deep neural network-HMM)(Xia等,2020)和Transformer(Dong等,2018)模型,它们能更有效地处理语音的时序特性和长距离依赖。与此同时,语言模型帮助系统理解词语之间的合理组合,提升识别准确率。传统 N -gram模型基于统计概率预测词序列,而现代方法采用神经网络(如RNN(recurrent neural network)或Transformer)捕捉长距离上下文关系(Zhou等,2024)。在解码阶段,声学模型和语言模型的输出会被结合,通过动态搜索算法(如维特比算法)找出最可能的文本序列。最后,进行拼写校正(如修正同音词错误)、标点符号恢复等后处理,确保输出文本符合语言习惯。近年来,端到端语音识别技术显著简化了传统流程。例如,SAN-CTC(self-attentional network for connectionist temporal classification)(Salazar等,2019)、RNN-T(Zhang等,2020b)以及Transformer架构的Wav2Vec 2.0(Baevski等,2020)等。

1.2.2 自然语言处理与理解

仅实现语音到文本的转化还不足以实现用户与增强现实系统之间的自然交互,还需理解语音指令背后的用户意图。这一过程需要结合自然语言处理(natural language processing, NLP)与自然语言理解(natural language understanding, NLU)技术。自然语言处理包括语音识别、文本分词和句法分析等基础任务。例如,NLP可将用户语音转化为文本,并对文本进行初步的结构化处理(如识别关键词、划分句子成分)。自然语言理解则是NLP的核心子领域,专注于从文本中提取深层语义信息,包括意图识别、实体抽取(如对象名称、位置参数)和上下文推理等。

例如,当用户说“把左边的桌子调成红色”,NLU需解析“左边”的空间指向、“桌子”的虚拟对象标识以及“调成红色”的操作意图。传统NLU依赖规则引擎或统计模型,但在增强现实场景中,用户指令常包含复杂的空间描述和多轮交互需求(如“旋转刚才选中的模型”)。随着大语言模型(large language models, LLMs)(如GPT-4(generative pre-trained Transformer)、DeepSeek)的发展,AR系统能够更好地理解模糊指令、支持上下文连贯的对话,甚至结合视觉信息实现多模态语义理解(Xu等,2025)。

1.2.3 语音识别技术的应用及挑战

依托语音识别和自然语言理解技术,AR应用的设计人员可以根据应用场景设计指令系统以便用户发出语音指令。例如Aouam等人(2018)开发一个用于汽车零部件装配/拆卸的AR系统,该系统支持简单的关键词命令(如“显示发动机”)和复杂的句子命令(如“将前轮移动到左侧”)。Ballard(2018)提出一个用于航空航天维护的AR系统,该系统允许航空机械师能够在进行组装任务时通过语音指令检查天窗的扭矩值。除却明确指令集之外,部分研究工作专注于根据语音指令生成相应的虚拟对象或场景,如Weng等人(2024)提出的MR(mixed reality)环境中的3D内容创作工具Dream Mesh。

声音也可以反映用户的一些生理或心理特征,用于身份识别和情绪、压力分析等领域。Shang和Wu(2019)设计一种语音欺骗防御系统,该系统结合了用户体内部传播和体外空气传播的声音特征,确保语音输入来自佩戴头显的真实用户。Arushi等人(2021)提出一个用于实时分析演讲者的声音压力检测模型,该模型通过提取和识别显著的声音特征,识别用户压力状态的变化。

语音输入使用户能够以更直观、自然的方式与增强现实设备交互,无须依赖摄像头等其他复杂传感器。这种输入模态允许用户在不占用双手的情况下与系统交互,提高工作效率和安全性。同时,相比复杂的按键组合或手势操作,语音命令学习成本低,适用于老年人及新手用户。结合自然语言处理,语音交互模态可以提供个性化的响应。语音识别系统可以在几秒钟内识别并执行用户的命令,减少了交互所需的时间。特别是在需要快速决策的场景,用户可以通过一次语音命令执行多个操作,而无须逐个点击或输入指令。

然而,专业领域(如医疗、法律)的语音识别需通过领域微调和动态语言模型实现个性化适配。语音转文字的交互依赖于语音识别系统的准确性,而准确性对噪声和语言表达方式较为敏感。如,在嘈杂环境中(如公共场所、工厂车间或交通繁忙的街道),背景噪音会影响系统的准确性,导致误识别或无法识别用户的命令。语音识别系统有时难以区分同音词(如“花”和“画”)或处理模糊的语义表达,尤其是在自然语言处理能力较弱的情况下,可能会误解用户的意图。持续监听用户的语音命令会消耗大量电量,尤其是在移动设备上。为了节省电池,某些设备可能会采用“唤醒词”机制(如“Hey Siri”或“Alexa”),但这也会增加误触发的风险。

1.3 对比与总结

为了进一步讨论不同输入模态的特性,本文对提及的输入模态从依赖设备、主要技术、各技术的代表算法、数据集以及交互设计、应用领域等多个角度进行了总结,如表1所示,以便相关从业人员参考。

2 增强现实系统的输出模态

当人们用动作或语言与周围环境互动时,环境也会通过各种刺激来传递信号给予反馈。比如走路时,眼睛看到的東西会变化,告诉我们位置在移动;碰到障碍物时,皮肤能感觉到它的温度和材质。这些反馈让人们感知到自己的行动产生了什么效果。在增强现实的交互设计中,系统反馈同样不可或缺。从图形界面跳动的光标到错误提示音,反馈机制始终在弥合用户意图与系统状态的差距,帮助人们理解系统是否正常工作,或哪里需要调整(Bermejo和Hui,2021)。然而,当前多模态交互领域持续深入研究输入技术(Kim等,2021),对多感官输出的系统性探索却未获同等关注。目前视觉反馈仍是增强现实系统的主流输出,如在眼镜上显示3D模型或导航箭头等。尽管视觉反馈能给用户带来丰富的信息,但单一感官的局限性日益凸显。如,缺乏触觉导致虚拟操作失实,无环境声则会削弱空间临境感,同时嗅觉与味觉的缺席更使餐饮零售等增强现实体验流于表面。本节从视觉、听觉、触觉、嗅觉和味觉五大感官维度,梳理AR反馈技术的进展、应用与挑战。

2.1 视觉

视觉是AR系统最成熟的反馈通道,为用户提

供了丰富的信息和精准灵活的反馈,主要依赖可视化技术及适应性用户接口设计来实现。

2.1.1 可视化技术

增强现实中的可视化技术通常是指将虚拟信息按照一定的配准关系与真实环境融合并呈现给用户的过程(Zollmann等,2021)。这一过程包含3个阶段。1)数据表示。真实环境的结构化表达以及虚拟信息的生成,真实场景通常需要借助三维重建算法表示为点云、网格等结构化数据。2)滤波与映射。对虚拟信息进行筛选和调整,如基于用户位置计算目标重要性或采用渐隐函数动态调节可视化优先级。3)融合。采用直接叠加、透明度混合渲染或遮挡区域修复绘制等多种方式实现虚拟内容与真实环境的视觉整合。

2.1.2 适应性用户接口

在整个可视化的过程中,与交互设计最相关的部分是虚实融合中配准信息的生成与调整,尤其是虚拟信息与环境的动态适应,即如何根据用户状态以及环境信息来调整这些视觉增强信息的基本状态,如大小、位姿(Cheng等,2021b)、详细程度(Lindlbauer等,2019)及渲染质量(Jabbireddy等,2022)等。该问题在增强现实中称为适应性用户接口(adaptive user interface),得到了研究者的广泛关注。Belo等人(2022)系统地阐述了适应性UI设计中的基本组件和框架,并提供了很多动态调整的策略,如UI不应距离用户太远,UI应保持在用户视野范围内等。部分研究者关注到虚拟信息与真实物体之间的语义关联(Cheng等,2021b;Qian等,2022)。Lindlbauer等人(2019)基于认知负荷模型,动态简化非核心UI的细节层次,在任务复杂度峰值时将当前任务相关性较低的UI调整成缩略模式,避免认识过载降低用户效率。

2.1.3 视觉反馈技术的应用及挑战

视觉反馈在增强现实中得到广泛应用,提升了用户在工业制造、医疗、教育以及游戏等领域的体验和效率。在工业制造领域,AR技术借助真实的视觉反馈加速了设计、制造和检修等一系列任务。例如Shi等人(2017)开发的AR材质设计界面,通过实时环境光照渲染将设计材质叠加到物理设备上,增强了设计师对材质质感评估的沉浸感。Seelinger等人(2023)开发的直接在实物产品上可视化产品缺陷的质检系统,减轻了工人执行任务的负担。在教育培

训领域,AR 视觉反馈革新了技能训练模式,如人工智能(artificial intelligence, AI)驱动的 AR 手术教学系统通过强化学习生成 3D 指导轨迹,提高培训效率(Long 等, 2022)。视觉反馈还被用于优化运动表现与康复效果。Guinet 等人(2022)设计的 AR 步行系统通过多种视觉反馈特征(颜色变化、反馈速度等)提示患者移动时的速度以进行步态调整。

视觉反馈为用户提供了丰富的信息,给增强现实带来了无限的可能性。然而常用的 AR 设备(如头显等)通常视场角受限,佩戴时显眼且价格昂贵,用户接受度有限。同时,高精度的视觉渲染耗费计算资源,遮挡处理(如虚拟物体被真实物体遮挡时的自然渲染)和光照一致性(如虚拟光影与真实环境匹配)仍是技术难点。未来研究可聚焦于解决更大视场角和更紧凑设备设计之间的矛盾(例如 Mojo Vision 的电子隐形眼镜(Wiemer, 2021))以及 AI 驱动的实时渲染优化任务。

2.2 听觉

在增强现实(AR)交互系统中,听觉反馈通过模拟真实世界的声音传播规律,为虚拟信息提供空间化的感知维度,成为多模态交互的重要组成部分。随着可穿戴音频耳机的普及,声学增强现实逐渐得到了广泛的研究,为创造基于声音的系统反馈提供了解决方案。其技术实现主要依托语音合成与空间声场合成两大技术体系。

2.2.1 语音合成技术

该技术旨在通过给定文本合成清晰自然的语音,是语音、语言和机器学习领域的热门研究课题,并在有声读物服务、汽车导航服务等领域具有广泛的应用。传统 TTS(text-to-speech)技术大多采用两阶段流程:首先通过文本预处理(包括分词、词性标注和音素识别)生成梅尔频谱图等中间声学特征,随后由 WaveNet 等声码器转换为波形音频(Shen 等, 2018)。尽管该方案能产生高保真语音,但存在模块间误差累积和实时性受限的缺陷。近年来,以 Tacotron(Wang 等, 2017)为代表的端到端模型可直接将文本映射为音频波形,利用 Transformer 架构实现并行化生成,在提升效率的同时缓解了误差传播问题,逐渐成为主流(Tan 等, 2021)。

2.2.2 空间声场合成技术

该技术旨在以特定方式处理声音,使得通过耳机或耳塞传递的声音在频谱和双耳特性上模拟真实

世界中传入耳朵的声音特征。最近研究提出采用头传函数(head related transfer functions, HRTF)实现双耳空间化(Jayaram 等, 2023; Engel 等, 2022),其本质是建立声源空间坐标到双耳频谱特征的映射函数。为实现更高层级的沉浸感,有研究表明应将房间声学建模整合到空间声音合成中(Yang 等, 2022),建模时不仅需要重建环境几何结构,还需要估算室内表面材料的声吸收系数(Shih 和 Rowe, 2019),从而计算环境的脉冲响应模拟环境音色。部分研究关注定向声场,利用超声波干涉或声波聚焦技术,在特定位置形成可听声束,避免噪音干扰,同时保护了隐私(Berzborn 和 Vorländer, 2021)。

2.2.3 听觉反馈的应用及挑战

自然真实的音频信息是维持用户在增强现实(AR)/混合现实(MR)系统中高度沉浸感的关键要素,在游戏化交互、文化遗产导览、多模态感知等场景中展现出广泛的应用潜力。在文化遗产与导览领域,研究者通过融合定位技术与声音反馈提升用户体验。例如, Kaghat 等人(2020)开发的 SARIM(sound augmented reality interface for visiting museums)系统利用用户实时位置提供个性化的博物馆艺术品讲解,而 Sikora 等人(2018)基于智能手机的遗迹导览系统通过全球定位系统(global positioning system, GPS)和方向传感器重现历史声景,帮助游客沉浸式感知考古遗址的文化背景, Dam 等人(2024)用音频来增强用户在欣赏画作时的感受。在游戏与多模态交互领域,通过动态声效设计增强沉浸感。Sonic Interaction 团队开发的游戏“Audio Legend”以声音模拟敌人移动轨迹,玩家通过手持设备的声学交互完成攻击或防御动作(Rovithis 等, 2019); Li(2024)提出的 NAI(natural interaction-based audio interface)系统则利用手部接近虚拟物体时的音效反馈辅助空间感知,创造直观的互动体验。Guarese 等人(2024)开发的导航系统则通过手部移动实时调整音效参数,帮助用户逐步接近目标物体,为非视觉交互提供了新思路。声音信号也被用来辅助完成定位。例如, Yang 等人(2020)创新性地联合耳机与手机的惯性传感器,通过用户对 3D 音频的反馈动态校准定位误差。

声音反馈技术在视觉信息过载场景(如工业维修、外科手术),避免干扰用户对真实环境的观察,有效提升用户的关注度。同时,通过个性化的声学建

模和空间感知,提升了用户的情感共鸣与空间沉浸感。尽管优势显著,但当前技术仍面临关键挑战,如个性化 HRTF 校准困难、动态场景中多声源定位的稳定性差以及公共场合下音频外放造成的隐私泄漏和其他用户的干扰等。未来发展中,多源定位技术(超声波/IMU/视觉融合)、低生物特征依赖 HRTF 普适化等方向有望成为 AR 声音反馈的研究热点。

2.3 触觉

触觉反馈通过模拟力、振动或纹理,赋予虚拟对象真实存在感。例如,用户可以用裸手移动虚拟物体,或使用控制器在移动过程中“感受”虚拟物体的纹理、大小和重量等。其技术路径可分为主动触觉、被动触觉与伪触觉三大类。

2.3.1 主动触觉

主动触觉以机械装置为核心,通过振动、超声波、电刺激等多模态手段实现高自由度交互。图5展示了部分提供主动触觉反馈的设备。

在振动反馈领域,研究者开发了多种交互设备,如 Gabardi 等人(2016)设计的 Haptic Thimble 通过串联伺服电机调整指尖接触板方向,结合语音线圈振动复现表面纹理差异,适用于虚拟雕刻等精细操作。针对特殊需求场景, Schaack 等人(2019)开发的 Haptic Collar 通过颈部振动触觉编码(如单侧振动指示转向,双侧交替振动代表直行),实现无需视觉辅助的导航引导。超声波与电刺激技术突破了接触式交互的限制。UltraHaptics 的 256 阵元声波阵列聚焦声压波形成局部触觉“热点”,支持凌空操控虚拟界面元素(Carter 等,2013)。在电刺激领域, Sparkle 系统(Spelmezan 等,2016)利用高压电弧刺激指尖,结合红外接近感应实现悬停触觉反馈。热触觉方面,可穿戴的 TherModule 系统通过模块化温控单元在手腕传递火焰热浪或冰雪寒冷体验,显著增强影视娱乐的场景沉浸感(Maeda 和 Kurahashi,2019)。柔性材料和活性材料的突破为触觉设备微型化提供了新路径,这些技术以其紧凑、轻量或弹性形态,推动可穿戴触觉向自然交互演进(Yang 等,2021a),如基于形状记忆合金的皮肤触觉传感装置(Yu 等,2019),通过磁场控制流体黏度复现地面质地变化的磁流变触觉鞋(Yang 等,2021b)等。

2.3.2 被动触觉

被动触觉通过物理道具的空间映射提供触觉支持,核心在于利用真实物体的几何特性与用户动作



图5 新型的触觉增强设备

Fig. 5 Advanced haptic augmented devices

的匹配为虚拟交互赋予自然触感。

研究表明,物理表面(如桌面、墙面、地面)的存在可显著提升交互效率与沉浸感(Cheng 等,2022)。这一原理被扩展至复杂交互场景。Meier 等人(2021)开发的 TapID 系统通过腕带惯性传感器检测手指敲击桌面振动,结合头显追踪的手部姿态数据,在物理表面实现虚拟按钮的精准触控。Nagao 等人(2018)提出通过用户脚下的凸起状的木板提供被动触觉刺激,再结合通过头戴式显示器提供的楼梯和鞋子的视觉刺激,引发视觉—触觉互动,从而改变用户对地板形状的感知,模拟上下楼的触觉反馈。Cheng 等人(2015)设计的 TurkDeck 创新性地引入人类操作员实时调整道具布局,在有限物理空间内动态映射无限虚拟场景。他们提出的 SHP(sparse haptic proxy)系统采用半球形的触觉道具,使用一组简化的几何基元来表示多种不同的虚拟环境。通过“触觉重定向”技术,当用户接近虚拟目标时,系统自动引导至匹配部分(Cheng 等,2017)。

2.3.3 伪触觉

伪触觉技术通过视觉线索与用户动作的耦合,在不依赖外部装置的条件下诱导触觉幻觉,其核心机制源于人类多感官整合中视觉信号的主导性(Ujitoko 和 Ban,2021)。

研究表明,通过调整虚拟对象的视觉形变、位移比例或运动轨迹,可显著改变用户对重量、刚度、粗糙度等触觉属性的感知。例如, Samad 等人(2019)通过控制手部渲染位移与实际移动的比例(C/D),给用户不同的感知重量, C/D 比越小,物体感觉越重。类似地,减少虚拟弹簧视觉形变可使感知刚度提升(Jang 和 Lee,2014)。

这种视觉主导效应广泛应用于触觉属性模拟。Punpongsanon 等人(2015)提出的空间增强现实系统 SoftAR, 可视化了一个二维的物体表面变形, 通过视觉上的变化来影响用户对物体柔软度的感知。Kawabe(2020)提出用手部的动作幅度来渲染伪触觉硬度。当手部移动时, 控制的布料对象会相应地水平变形。通过调整变形—距离比和泊松比, 有效模拟了布料的硬度。Ban 等人(2014)不仅控制了虚拟物体的表面变形, 还同时调整了抓取虚拟物体的手指姿态, 以增强伪触觉效果。在复合触觉增强方面, Strandholt 等人(2020)设计的“工具重定向”技术动态地偏移物理锤与虚拟锤的位置映射, 使敲击动作的物理冲击同步率提升, 任务完成时间缩短。

2.3.4 触觉反馈的应用及挑战

触觉反馈是一个多学科领域, 通过刺激人类的触觉和运动通道, 其在一定程度上弥补了视觉和听觉通道不足, 给 AR 生态系统引入了新的活力。触觉反馈可用于增强传统交互任务, 如键按钮或打字等任务的沉浸感(Meier 等, 2021)。在日常生活中, 触觉可以在现实世界中为视障人士导航, 通过触觉刺激给用户方向上的提示(Schaack 等, 2019)。在低成本医疗培训中, 伪触觉技术用来模拟软组织表面的硬度信息和肿瘤硬度, 以及相应的凹凸来帮助定位皮肤下的动脉或器官, 例如腹腔镜手术及其培训(Fakhoury 等, 2017; Li 等, 2016)。在远程机器人抓取任务中, 通过叠加视觉压力光晕可以增强操作者对抓取力量的感知, 从而减少抓取力并降低操作者的认知负荷(Haruna 等, 2021)。康复领域, 在上肢训练中结合视觉运动辅助, 通过应用微小的偏移来实现患者对身体的控制感, 改善中风患者关节活动度(Li 等, 2019; Li 等, 2022); 娱乐场景中, 触觉可以用来提升观众的场景沉浸感(Maeda 和 Kurahashi, 2019), 而驾驶模拟器的方向盘抖动反馈使“路面颠簸”感觉更加真实(Wang 等, 2021), 以及虚拟环境中上下楼梯的真实感体验(Nagao 等, 2018)。

然而, 触觉系统仍面临显著挑战: 精密机械结构与控制模块导致开发成本高昂, 且设备体积与重量较大(如外骨骼装置), 限制了移动场景适用性(Yin 等, 2021); 设备的机械运动和电流刺激引起用户不适甚至损伤。尽管被动触觉和伪触觉技术降低了硬件依赖和用户不适, 其效果仍受场景不匹配(物理道具与虚拟物体之间的不匹配)以及个体感官差异(视

觉—触觉整合阈值波动)等的影响, 限制了触觉作为主流交互输出模态的推广标准化。未来研究需探索多模态增强策略, 例如结合微振动或局部温度控制等主动触觉、柔性电子皮肤和热触觉融合等; 同时需制定安全边界, 以避免反馈力度过大导致用户损伤。

2.4 嗅觉

近年来, 增强现实(AR)和虚拟现实(VR)将视觉、听觉和触觉体验数字化的努力越来越成功。作为人类感知世界的重要化学感官的嗅觉, 对人类生理和心理具有双重影响, 与传统视觉、听觉和触觉同样重要。不同于视觉或听觉的物理信号传递, 嗅觉的独特性在于直接连接大脑的情感与记忆中枢。气味通过分子扩散激活鼻腔神经元, 能够触发快速的情感反应和生动的记忆回溯。研究表明, 人类可识别的气味种类超过 1 万亿种, 远超视觉和听觉的感知极限(Bushdid 等, 2014)。

2.4.1 嗅觉合成技术

研究者通过多学科交叉提出多种解决方案(章国锋 等, 2024)。例如, Yang 等人(2023)利用仿生纤维膜和静电场加速蒸发技术, 实现 3 s 内快速响应、无线可控的气味生成系统。柔性电子技术的突破进一步推动了设备微型化, 如 Liu 等人(2023)研制的 mm 级柔性气味发生器阵列, 功耗低的同时保持了 ms 级响应速度。这些创新不仅缩小了设备体积, 还显著提升了气味生成的精度与多样性。在交互设计方面, Patnaik 等人(2019)开发的 viScent 系统将 6 通道超声波雾化器与可视化界面结合, 通过气味—视觉融合增强数据感知; Brooks 等人(2020)利用特定化学物质直接激活鼻腔三叉神经, 创造出虚拟温度感知体验, 拓展了嗅觉反馈的感官维度。一些新型的嗅觉增强设备如图 6 所示。

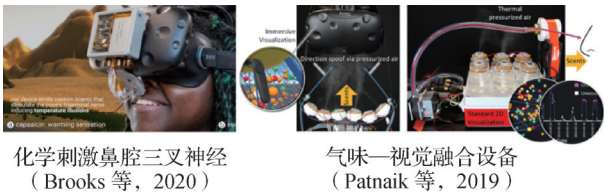


图6 新型的嗅觉增强设备

Fig. 6 Advanced augmented olfaction devices

2.4.2 嗅觉反馈的应用及挑战

嗅觉反馈技术正在教育、医疗和工业等领域展现潜力。在教育场景中, AR 系统可通过实时气味提

示帮助植物学学习者辨识不同花卉特征(Liu等, 2023);消防训练中模拟的烟味显著提升了演练的真实性(Liu等, 2022)。医疗应用方面,定制化气味方案被用于嗅觉障碍康复训练,并通过触发情感记忆辅助心理治疗(Lopes等, 2024)。工业领域, Wang等人(2018)提出的嗅觉增强维护系统,可通过识别机械故障产生的异常气味(如润滑油变质)实现早期预警。Liu等人(2022)开发的触觉—嗅觉感应阵列被用于救援任务中,能在无视觉环境下通过物体硬度与气味特征实现较高的识别准确率。

尽管取得显著进展,当前技术仍受限于硬件体积、气味残留和算法模型不足,很大程度上仍处于研究阶段(Kerruish, 2019)。主流系统仅支持6~10种基础气味,复杂气味的动态合成仍依赖实验室环境。未来突破或将聚焦于两个层面:在硬件层面,可穿戴设备的柔性化与微型化趋势明显,皮肤界面集成技术有望实现更自然的嗅觉交互;在软件层面,神经科学与人工智能的结合至关重要——通过解析嗅觉感知的神经编码机制,构建从分子结构到主观体验的映射模型。

2.5 味觉

味觉反馈机制十分复杂。人类味觉的形成是一个多感官整合的过程,主要通过舌头上的味蕾完成。味蕾中的特殊细胞通过表面的味毛接触食物中的化学物质,咸味和酸味直接由离子(如钠、氢离子)进入细胞触发电信号,甜、苦、鲜味则通过细胞表面的蛋白质受体传递信号。这些信号经神经传递至大脑,结合嗅觉(如食物香气)、触觉(如口感)和视觉(如颜色)信息,最终形成综合的“味道”体验(Kandel等, 2000)。

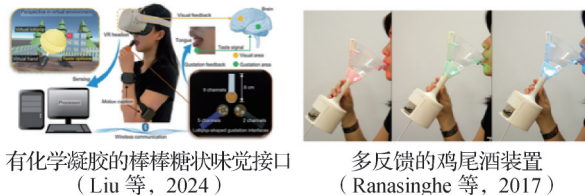
2.5.1 味觉合成技术

当前数字味觉的实现主要依赖三大技术路线:化学物质释放、电刺激调控和温度模拟,研究者通过跨学科创新不断拓展技术边界。一些新型的味觉增强设备如图7所示。

在化学释放领域, Liu等人(2024)开发的棒棒糖状味觉接口具有代表性。其内部嵌入装载化学物质的琼脂糖凝胶袋,通过电压控制离子电泳速率,调节化学物质释放浓度,模拟真实舔舐糖果的味觉渐变过程。电刺激技术则更注重非侵入式干预,例如 Miyashita等人(2023)设计的带电筷子通过电流驱动食物中钠离子定向迁移,提升了低盐食物的咸味感



电刺激增强盐味感知的智能碗和筷子 (Miyashita等, 2023) 温度调节模拟甜味 (Ranasinghe和Do, 2016a)



有化学凝胶的棒棒糖状味觉接口 (Liu等, 2024) 多反馈的鸡尾酒装置 (Ranasinghe等, 2017)

图7 新型的味觉增强设备

Fig. 7 Advanced augmented gustation devices

知强度。Ranasinghe和Do(2016b)设计了“数字棒棒糖”系统,通过调节施加于舌头的电流强度、频率和极性,成功诱发出酸、咸、苦、甜4种基本味觉。他们还设计了一种名为Vocktail(虚拟鸡尾酒)的互动饮品器具(Ranasinghe等, 2017),利用视觉(投射在饮料上的RGB灯光)、味觉(舌尖上的电刺激)和嗅觉刺激(由微型气泵释放)相结合,共同创造出一种虚拟的风味感受。温度调控路径同样取得突破, Ranasinghe和Do(2016a)利用帕尔贴元件阵列对舌尖快速加热或冷却,发现特定温变模式可模拟甜味感受,为无化学干预的味觉生成提供了新思路。近年来,柔性电子技术进一步推动设备革新,如 Mukashev等人(2023)开发的TactTongue平台通过高分辨率电刺激模拟口腔液体流动的触觉—味觉复合体验。

2.5.2 味觉反馈的应用以及挑战

味觉反馈技术在医疗健康、食品工业和教育培训领域展现出独特价值。例如,增强现实系统可帮助味觉障碍患者重建饮食体验;低盐饮食者借助电刺激餐具既能减少钠摄入,又可维持味觉满足感(Miyashita等, 2023)。但其商业化仍面临多重壁垒。然而,该技术仍面临多重制约:首先,人类对味觉神经编码机制的认知尚未完善,现有系统仅能模拟基础味型,复杂风味(如烤肉焦香)的数字化仍依赖实验室环境;其次,个体差异要求设备具备个性化校准能力;此外,化学残留、多味觉通道串扰等问题也限制着连续体验的可靠性。同时,味觉模拟设备成本高昂,且微型化与安全性难以兼得,带电设备的口腔长期接触可能引发黏膜刺激或金属离子渗透风险。技术实用化需在硬件革新与算法建模两端同步推

进。硬件层面,纳米级分子控制技术可能突破化学释放的精度限制,例如通过微流控芯片实现百万级风味分子的动态混合;柔性电子与生物相容材料的结合,则有望开发出可植入式味觉调控贴片。软件层面,需建立从刺激参数到主观感知的量化模型,这要求神经科学与人工智能深度协作:通过脑机接口

捕捉味觉皮层的神经活动特征,结合深度学习解析多模态感官信号的耦合规律。

2.6 对比与总结

本文对提及的各种输出模态进行对比与总结,分析不同模态的主要技术以及应用领域,如表2所示。

表2 不同输出模态之间的对比
Table 2 Comparison between different output modalities

输出模态	主要技术	应用领域
视觉	可视化、自适应用户接口	工业制造、医疗、教育、游戏等
听觉	语音合成、空间声场合成	旅游导览、游戏
触觉	主动触觉、被动触觉、伪触觉	打字辅助、视障导航、医疗康复
嗅觉	嗅觉合成	教育、医疗、工业
味觉	味觉合成	医疗、食品、教育培训

3 挑战与未来研究方向

尽管现有多种模态的增强现实交互技术取得了一定的发展与进步,但该领域仍面临着一些值得深入研究的挑战。

1)多模态融合是增强现实交互领域的大势所趋,但各种模态数据的耦合关系复杂。多模态融合需解决跨模态数据的时空异构问题。以眼动追踪为例,其数据有效性依赖头部运动补偿(如通过SLAM算法重建观察者位姿),并需消除躯干、下肢动作的干扰噪声。在AR/VR自由探索场景中,动态多变量(如头部偏转速率、身体位移)导致数据关联性建模难度指数级上升。同时,不同模态存在固有特性冲突,例如手势输入为离散事件(如抓取动作),而步态轨迹为连续信号。现有算法难以平衡模态间的优先级(如语音中断手势操作的阈值设定),更缺乏冲突消解的统一框架。

2)增强现实技术给交互带来了无限的可能,但也引入了新的隐私与安全问题。AR设备的全景感知能力(如环境扫描、生物特征识别)使其成为高隐蔽性数据采集工具,攻击者可利用移动AR技术劫持设备传感器,规模化窃取用户的敏感信息(如虹膜特征、空间行为模式)。同时,AR界面设计存在“感官操控”伦理风险。例如,通过空间音频定向诱导用户偏离预定路径或利用高频率视觉闪烁引发运动病

导致生理不适。现行AR设计规范尚未建立隐私安全保护的相关条例和感官刺激的安全阈值标准。

3)当前交互系统呈现输入模态多元化与输出通道单一化的结构性失衡,这凸显了触/嗅/味觉等跨通道反馈机制构建的迫切需求。当前AR系统呈现“强输入—弱反馈”特性,大量研究聚焦如何丰富用户的输入手段,而忽视了交互中的反馈。除视觉增强之外,触觉/嗅觉/味觉反馈仍停留于实验室原型阶段。尽管跨感官反馈在医疗(如手术刀触觉导航)、教育(如气味辅助历史场景还原)领域潜力显著,其工程化落地受限于硬件微型化与多模态编码标准缺失。同时,多感官反馈需遵循认知神经规律(如视听时延的可察觉阈值),但现有系统难以保障跨模态同步性,导致用户体验割裂。

4)增强现实技术的突破需要材料科学、认知神经科学与计算机学科的深度协同,这些学科的融合将重构AR系统的人机协同范式,推动从界面层向认知层的体系升级。作为融合环境感知、多模态交互与跨感官反馈的复杂系统,AR的终极目标是逼近甚至超越人类与真实世界的交互体验,而这需3类学科的闭环支撑:材料科学革新硬件载体,例如光场透明显示材料需匹配人眼焦距调节机制,柔性电子皮肤则需兼容触觉神经的灵敏度阈值;认知神经科学解码交互本质,通过脑电、眼动等生物信号识别用户意图,并建立多感官反馈的认知一致性模型;计算机科学构建智能实现:开发实时生物信号解码算法、

多模态融合引擎等。当前,跨学科壁垒仍阻碍技术闭环,如生物信号解码延迟、神经兼容材料缺位等问题。突破上述瓶颈,需建立跨学科联合实验范式(如脑机接口与触觉材料的共设计流程),最终使AR从“视觉增强工具”进化为“认知增强伙伴”。

4 结 语

增强现实技术通过虚实融合与多模态交互重构了人机协同范式,但其发展仍受限于多维技术瓶颈。本文提出的输入—输出双维度分析框架,系统梳理AR交互技术的模态特征与应用场景,填补了现有研究对跨感官反馈机制关注的缺失。研究发现,多模态数据的时空异构性导致融合困难,例如眼动追踪需动态补偿头部运动噪声,而离散手势与连续步态信号间的优先级冲突亟待统一建模框架。此外,AR设备的全景感知能力引发隐私泄露与感官操控风险,需建立数据安全规范与多模态刺激阈值标准。当前,触/嗅/味觉反馈的工程化滞后于输入技术创新,需突破硬件微型化与神经兼容性设计难题。未来研究应聚焦跨学科协同机制:材料科学革新神经兼容型交互载体(如柔性电子皮肤),认知神经科学解码多感官反馈的耐受边界,计算机科学构建实时生物信号解析引擎。唯有打破学科壁垒,才能实现AR从“工具”到“认知伙伴”的终极进化,为医疗、教育等领域提供更自然的沉浸式交互体验。

参考文献(References)

- Al-Ansi A M, Jaboo M, Garad A and Al-Ansi A. 2023. Analyzing augmented reality (AR) and virtual reality (VR) recent development in education. *Social Sciences and Humanities Open*, 8 (1): #100532 [DOI: 10.1016/j.ssaho.2023.100532]
- Anjali, Singh G and Sandhu J K. 2022. Virtual and augmented reality technology for the treatment of mental health disorders: an overview//*Proceedings of 2022 International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Kharagpur, India: IEEE: 1-5 [DOI: 10.1109/ICCCNT54827.2022.9984551]
- Aouam D, Benbelkacem S, Zenati N, Zakaria S and Meftah Z. 2018. Voice-based augmented reality interactive system for car's components assembly//*Proceedings of the 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. Tebessa, Algeria: IEEE: 1-5 [DOI: 10.1109/PAIS.2018.8598516]
- Arushi, Dillon R and Teoh A N. 2021. Real-time stress detection model and voice analysis: an integrated VR-based game for training public speaking skills//*Proceedings of 2021 Conference on Games (CoG)*. Copenhagen, Denmark: IEEE: 1-4 [DOI: 10.1109/CoG52621.2021.9618989]
- Baek S, Kim K I and Kim T K. 2019. Pushing the envelope for RGB-based dense 3D hand pose estimation via neural rendering//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA: IEEE: 1067-1076 [DOI: 10.1109/CVPR.2019.00116]
- Baevski A, Zhou Y, Mohamed A and Auli M. 2020. Wav2vec 2.0: a framework for self-supervised learning of speech representations//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates, Inc.: 12449-12460
- Ballard B. 2018. Giving voice to augmented reality [EB/OL]. [2025-06-27]. <https://www.aerospacemanufacturinganddesign.com/article/giving-voice-to-augmented-reality>
- Ban Y K, Narumi T, Tanikawa T and Hirose M. 2014. Controlling perceived stiffness of pinched objects using visual feedback of hand deformation//*2014 IEEE Haptics Symposium (HAPTICS)*. Houston, USA: IEEE: 557-562 [DOI: 10.1109/HAPTICS.2014.6775516]
- Belo J M E, Lystbæk M E, Feit A M, Pfeuffer K, Kán P, Oulasvirta A, et al. 2022. AUIT—the adaptive user interfaces toolkit for designing XR applications//*The 35th Annual ACM Symposium on User Interface Software and Technology (UIST'22)*. Association for Computing Machinery. New York, USA: [s. n.]: 1-16 [DOI: 10.1145/3526113.3545651]
- Bermejo C and Hui P. 2021. A survey on haptic technologies for mobile augmented reality. *ACM Computing Surveys (CSUR)*, 54(9): #184 [DOI: 10.1145/3465396]
- Berzborn M and Vorländer M. 2021. Directional sound field decay analysis in performance spaces. *Building Acoustics*, 28(3): 249-263 [DOI: 10.1177/1351010X20984622]
- Bhalla A, Sluganovic I, Krawiecka K and Martinovic I. 2021. MoveAR: continuous biometric authentication for augmented reality headsets//*Proceedings of the 7th ACM on Cyber-Physical System Security Workshop*. Hong Kong, China: Association for Computing Machinery: 41-52 [DOI: 10.1145/3457339.3457983]
- Bonazzi P, Bian S Z, Lippolis G, Li Y W, Sheik S and Magno M. 2024. Retina: low-power eye tracking with event camera and spiking hardware//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Seattle, USA: IEEE: 5684-5692 [DOI: 10.1109/CVPRW63382.2024.00577]
- Boominathan V, Adams J K, Robinson J T and Veeraraghavan A. 2020. PhlatCam: designed phase-mask based thin lensless camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

- 42(7): 1618-1629 [DOI: 10.1109/TPAMI.2020.2987489]
- Brooks J, Nagels S and Lopes P. 2020. Trigeminal-based temperature illusions//Proceedings of 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, USA: Association for Computing Machinery: 1-12 [DOI: 10.1145/3313831.3376806]
- Bushdid C, Magnasco M O, Vossall L B and Keller A. 2014. Humans can discriminate more than 1 trillion olfactory stimuli. *Science*, 343 (6177): 1370-1372 [DOI: 10.1126/science.1249168]
- Campos C, Elvira R, Rodríguez J J G, Montiel J M M and Tardós J D. 2021. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 37(6): 1874-1890 [DOI: 10.1109/TRO.2021.3075644]
- Cao X W and Yan W Q. 2024. Pose estimation for swimmers in video surveillance. *Multimedia Tools and Applications*, 83 (9): 26565-26580 [DOI: 10.1007/s11042-023-16618-w]
- Cao Z, Hidalgo G, Simon T, Wei S E and Sheikh Y. 2021. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1): 172-186 [DOI: 10.1109/TPAMI.2019.2929257]
- Carter T, Seah S A, Long B, Drinkwater B and Subramanian S. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces//2013 annual ACM Symposium on User Interface Software and Technology. St. Andrews Scotland, UK: Association for Computing Machinery: 505-514 [DOI: 10.1145/2501988.2502018]
- Caspi A, Roy A, Wuyyuru V, Rosendall P E, Harper J W, Katyal K D, et al. 2018. Eye movement control in the Argus II retinal-prosthesis enables reduced head movement and better localization precision. *Investigative Ophthalmology and Visual Science*, 59(2): 792-802 [DOI: 10.1167/iov.17-22377]
- Chen W Y, Yu C C, Tu C Y, Lyu Z, Tang J, Ou S Q, et al. 2020. A survey on hand pose estimation with wearable sensors and computer-vision-based methods. *Sensors*, 20(4): #1074 [DOI: 10.3390/s20041074]
- Cheng D W, Wang Q W, Liu Y, Chen H L, Ni D W, Wang X M, et al. 2021a. Design and manufacture AR head-mounted displays: a review and outlook. *Light: Advanced Manufacturing*, 2(3): 350-369 [DOI: 10.37188/lam.2021.024]
- Cheng L P, Ofek E, Holz C, Benko H and Wilson A D. 2017. Sparse haptic proxy: touch feedback in virtual environments using a general passive prop//Proceedings of 2017 CHI Conference on Human Factors in Computing Systems. Denver, USA: Association for Computing Machinery: 3718-3728 [DOI: 10.1145/3025453.3025753]
- Cheng L P, Roumen T, Rantzsch H, Köhler S, Schmidt P, Kovacs R, et al. 2015. TurkDeck: physical virtual reality based on people//2014 Annual ACM Symposium on User Interface Software and Technology. Charlotte, USA: Association for Computing Machinery: 417-426 [DOI: 10.1145/2807442.2807463]
- Cheng Y F, Luong T, Fender A R, Strelt P and Holz C. 2022. Comfortable user interfaces: surfaces reduce input error, time, and exertion for tabletop and mid-air user interfaces//2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Singapore, Singapore: IEEE: 150-159 [DOI: 10.1109/ISMAR55827.2022.00029]
- Cheng Y F, Yan Y K, Yi X, Shi Y C and Lindlbauer D. 2021b. SemanticAdapt: optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections//The 34th Annual ACM Symposium on User Interface Software and Technology. Virtual Event, USA: Association for Computing Machinery: 282-297 [DOI: 10.1145/3472749.3474750]
- Cheng Y H and Lu F. 2023. DVGaze: dual-view gaze estimation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 20575-20584 [DOI: 10.1109/ICCV51070.2023.01886]
- Dam A, Lee Y, Siddiqui A, Lages W S and Jeon M. 2024. Audio augmented reality using sonification to enhance visual art experiences: lessons learned. *International Journal of Human-Computer Studies*, 191: #103329 [DOI: 10.1016/j.ijhcs.2024.103329]
- Dong L H, Xu S and Xu B. 2018. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada: IEEE: 5884-5888 [DOI: 10.1109/ICASSP.2018.8462506]
- Engel I, Goodman D F M and Picinali L. 2022. Assessing hrtf preprocessing methods for ambisonics rendering through perceptual models. *Acta Acustica*, 6: #4 [DOI: 10.1051/aacus/2021055]
- Engel J, Koltun V and Cremers D. 2018. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3): 611-625 [DOI: 10.1109/TPAMI.2017.2658577]
- Esteves A, Verweij D, Suraiya L, Islam R, Lee Y and Oakley I. 2017. SmoothMoves: smooth pursuits head movements for augmented reality//Proceedings of 2017 Annual ACM Symposium on User Interface Software and Technology. Québec City, USA: Association for Computing Machinery: 167-178 [DOI: 10.1145/3126594.3126616]
- Fadzli F E and Ismail A W. 2019. VoxAR: 3D modelling editor using real hands gesture for augmented reality//Proceedings of the 7th IEEE Conference on Systems, Process and Control (ICSPC). Melaka, Malaysia: IEEE: 242-247 [DOI: 10.1109/ICSPC47137.2019.9067992]
- Fakhoury E, Culmer P and Henson B. 2017. The impact of visual cues on haptic compliance discrimination using a pseudo-haptic robotic system//Proceedings of 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO). Macau, China: IEEE: 1719-1725 [DOI: 10.1109/ROBIO.2017.8324666]
- Feichtenhofer C, Fan H Q, Malik J and He K M. 2019. SlowFast networks for video recognition//Proceedings of 2019 IEEE/CVF Inter-

- national Conference on Computer Vision (ICCV). Seoul, Korea (South); IEEE: 6202-6211 [DOI: 10.1109/ICCV.2019.00630]
- Ferrarotti A, Baldoni S, Carli M and Battisti F. 2024. Stress assessment for augmented reality applications based on head movement features. *IEEE Transactions on Visualization and Computer Graphics*, 30(10): 6970-6983 [DOI: 10.1109/TVCG.2024.3385637]
- Forster C, Carlone L, Dellaert F and Scaramuzza D. 2017. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1): 1-21 [DOI: 10.1109/TRO.2016.2597321]
- Fuhl W, Kübler T, Sippel K, Rosenstiel W and Kasneci E. 2015. ExCuSe: robust pupil detection in real-world scenarios//*Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns (CAIP)*. Valletta, Malta: Springer International Publishing: 39-51 [DOI: 10.1007/978-3-319-23192-1_4]
- Fuhl W, Santini T, Kasneci G and Kasneci E. 2016. PupilNet: convolutional neural networks for robust pupil detection [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/1601.04902.pdf>
- Gabardi M, Solazzi M, Leonardi D and Frisoli A. 2016. A new wearable fingertip haptic interface for the rendering of virtual shapes and surface features//*Proceedings of 2016 IEEE Haptics Symposium (HAPTICS)*. Philadelphia, USA: IEEE: 140-146 [DOI: 10.1109/HAPTICS.2016.7463168]
- Gao X, Li M H and Shen S H. 2024. Large-scale structure from motion: a survey. *Journal of Computer-Aided Design and Computer Graphics*, 36(7): 969-994 (高翔, 李梦晗, 申抒含. 2024. 大规模场景运动恢复结构研究综述. *计算机辅助设计与图形学学报*, 36(7): 969-994) [DOI: 10.3724/SP.J.1089.2024.2024-00132]
- Gavgiotaki D, Ntoa S, Margetis G, Apostolakis K C and Stephanidis C. 2023. Gesture-based interaction for AR systems: a short review//*Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments*. Corfu, Greece: Association for Computing Machinery: 284-292 [DOI: 10.1145/3594806.3594815]
- Glauser O, Wu S, Panozzo D, Hilliges O and Sorkine-Hornung O. 2019. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Transactions on Graphics (ToG)*, 38(4): 1-15
- Guarese R, Pretty E, Renata A, Polson D and Zambetta F. 2024. Exploring audio interfaces for vertical guidance in augmented reality via hand-based feedback. *IEEE Transactions on Visualization and Computer Graphics*, 30(5): 2818-2828 [DOI: 10.1109/TVCG.2024.3372040]
- Guinet A L, Bouyer G, Otmane S and Desailly E. 2022. Visual feedback in augmented reality to walk at predefined speed cross-sectional study including children with cerebral palsy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30: 2322-2331 [DOI: 10.1109/TNSRE.2022.3198243]
- Gül S, Bosse S, Podborski D, Schierl T and Hellge C. 2020. Kalman filter-based head motion prediction for cloud-based mixed reality//*Proceedings of the 28th ACM International Conference on Multimedia*. Seattle, USA: Association for Computing Machinery: 3632-3641 [DOI: 10.1145/3394171.3413699]
- Haruna M, Kawaguchi N, Ogino M and Koike-Akino T. 2021. Comparison of three feedback modalities for haptics sensation in remote machine manipulation. *IEEE Robotics and Automation Letters*, 6(3): 5040-5047 [DOI: 10.1109/LRA.2021.3070301]
- Ivaschenko A, Khorina A and Sitnikov P. 2018. Accented visualization by augmented reality for smart manufacturing applications//*Proceedings of 2018 IEEE Industrial Cyber-Physical Systems (ICPS)*. St. Petersburg, Russia: IEEE: 519-522 [DOI: 10.1109/ICPHYS.2018.8390759]
- Jabbireddy S, Sun X T, Meng X X and Varshney A. 2022. Foveated rendering: motivation, taxonomy, and research directions [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/2205.04529.pdf>
- Jain P, Nazar A M, Khan S S, Mitra K and Chakravarthula P. 2025. FlatTrack: eye-tracking with ultra-thin lensless cameras//*Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. Tucson, USA: IEEE: #50 [DOI: 10.1109/WACVW65960.2025.00050]
- Jang I and Lee D. 2014. On utilizing pseudo-haptics for cutaneous fingertip haptic device//*2014 IEEE Haptics Symposium (HAPTICS)*. Houston, USA: IEEE: 635-639 [DOI: 10.1109/HAPTICS.2014.6775529]
- Jaramillo-Yáñez A, Benalcázar M E and Mena-Maldonado E. 2020. Real-time hand gesture recognition using surface electromyography and machine learning: a systematic literature review. *Sensors*, 20(9): #2467 [DOI: 10.3390/s20092467]
- Jayaram V, Kemelmacher-Shlizerman I and Seitz S M. 2023. HRTF estimation in the wild//*Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. San Francisco, USA: Association for Computing Machinery: 1-9 [DOI: 10.1145/3586183.3606782]
- Jin X, Chai S Y, Tang J, Zhou X D and Wang K. 2024. Eye-tracking in AR/VR: a technological review and future directions. *IEEE Open Journal on Immersive Displays*, 1: 146-154 [DOI: 10.1109/OJID.2024.3450657]
- Jing A, May K, Lee G and Billingham M. 2021. Eye see what you see: exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Frontiers in Virtual Reality*, 2: #697367 [DOI: 10.3389/frvir.2021.697367]
- Kaghat F Z, Azough A, Fakhour M and Meknassi M. 2020. A new audio augmented reality interaction and adaptation model for museum visits. *Computers and Electrical Engineering*, 84: #106606 [DOI: 10.1016/j.compeleceng.2020.106606]
- Kaifosh P, Reardon T R and Ctrl-Labs at Reality Labs. 2024. A generic non-invasive neuromotor interface for human-computer interaction

- [J/OL]. [2025-11-17]. Nature.
<https://www.nature.com/articles/s41586-025-09255-w>
- Kandel E R, Schwartz J H, Jessell T M, Siegelbaum S, Hudspeth A J and Mack S. 2000. Principles of Neural Science. New York, USA: McGraw-Hill
- Kataria A, Ghosh S and Karar V. 2020. Data prediction of electromagnetic head tracking using self healing neural model for head-mounted display. Romanian Journal of Information Science and Technology, 23(4): 354-367
- Kawabe T. 2020. Mid-air action contributes to pseudo-haptic stiffness effects. IEEE Transactions on Haptics, 13(1): 18-24 [DOI: 10.1109/TOH.2019.2961883]
- Kendall A and Cipolla R. 2017. Geometric loss functions for camera pose regression with deep learning//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6555-6564 [DOI: 10.1109/cvpr.2017.694]
- Kerruish E. 2019. Arranging sensations: smell and taste in augmented and virtual reality. The Senses and Society, 14(1): 31-45 [DOI: 10.1080/17458927.2018.1556952]
- Khamis M, Oechsner C, Alt F and Bulling A. 2018. VRpursuits: interaction in virtual reality using smooth pursuit eye movements//Proceedings of 2018 International Conference on Advanced Visual Interfaces. Castiglione della Pescaia Grosseto, Italy: Association for Computing Machinery: #18 [DOI: 10.1145/3206505.3206522]
- Kim J C, Laine T H and Åhlund C. 2021. Multimodal interaction systems based on internet of things and augmented reality: a systematic literature review. Applied Sciences, 11(4): #1738 [DOI: 10.3390/app11041738]
- Kim M, Cho J, Lee S and Jung Y. 2019. Imu sensor-based hand gesture recognition for human-machine interfaces. Sensors, 19(18): #3827 [DOI: 10.3390/s19183827]
- Kovtun R, Radchenko S, Netreba A, Sudakov O, Natarov R, Dyka Z, et al. 2020. Exploiting EEG signals for Eye motion tracking//2020 IEEE East-West Design & Test Symposium (EWDTS). [s.l.]: IEEE: 1-5
- Lee K I, Jeon J H and Song B C. 2020. Deep learning-based pupil center detection for fast and accurate eye tracking system//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer International Publishing: 36-52 [DOI: 10.1007/978-3-030-58529-7_3]
- Lee T H and Lee H J. 2018. A new virtual keyboard with finger gesture recognition for AR/VR devices//Proceedings of the 20th International Conference on Human-Computer Interaction. Interaction Technologies. Las Vegas, USA: Springer International Publishing: 56-67 [DOI: 10.1007/978-3-319-91250-9_5]
- Li J Y. 2024. Beyond sight: enhancing augmented reality interactivity with audio-based and non-visual interfaces. Applied Sciences, 14(11): #4881 [DOI: 10.3390/app14114881]
- Li M, Guo W L, He B, Xu G H, Yuan H, Xie J, et al. 2019. Motion assistance and resistance using pseudo-haptic feedback for upper-limb rehabilitation//Proceedings of 2019 International Conference on Ubiquitous Robots (UR). Jeju, Korea (South): IEEE: 328-333 [DOI: 10.1109/URAI.2019.8768499]
- Li M, Sareh S, Xu G H, Ridzuan M B, Luo S, Xie J, et al. 2016. Evaluation of pseudo-haptic interactions with soft objects in virtual environments. PLoS One, 11(6): #e0157681 [DOI: 10.1371/journal.pone.0157681]
- Li Z P, Jiang Y, Zhu Y H, Chen R J, Wang R T, Wang Y T, et al. 2022. Modeling the noticeability of user-avatar movement inconsistency for sense of body ownership intervention. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2): #64 [DOI: 10.1145/3534590]
- Lin B S, Lee I J, Yang S Y, Lo Y C, Lee J and Chen J L. 2018. Design of an inertial-sensor-based data glove for hand function evaluation. Sensors, 18(5): #1545
- Lindlbauer D, Feit A M and Hilliges O. 2019. Context-aware online adaptation of mixed reality interfaces//The 32nd Annual ACM Symposium on User Interface Software and Technology. New Orleans, USA: Association for Computing Machinery: 147-160 [DOI: 10.1145/3332165.3347945]
- Liu M W, Zhang Y J, Wang J C, Qin N, Yang H, Sun K, et al. 2022. A star-nose-like tactile-olfactory bionic sensing array for robust object recognition in non-visual environments. Nature Communications, 13(1): #79 [DOI: 10.1038/s41467-021-27672-z]
- Liu S G, Li Y and Hua G G. 2019. Human pose estimation in video via structured space learning and halfway temporal evaluation. IEEE Transactions on Circuits and Systems for Video Technology, 29(7): 2029-2038 [DOI: 10.1109/TCSVT.2018.2858828]
- Liu Y M, Park W, Yiu C K, Huang X C, Jia S X, Chen Y, et al. 2024. Miniaturized, portable gustation interfaces for VR/AR/MR. Proceedings of the National Academy of Sciences of the United States of America, 121(49): #e2412116121 [DOI: 10.1073/pnas.2412116121]
- Liu Y M, Yiu C K, Zhao Z, Park W, Shi R, Huang X C, et al. 2023. Soft, miniaturized, wireless olfactory interface for virtual reality. Nature Communications, 14(1): #2297 [DOI: 10.1038/s41467-023-37678-4]
- Long Y H, Cao J F, Deguet A, Taylor R H and Dou Q. 2022. Integrating artificial intelligence and augmented reality in robotic surgery: an initial dVRK study using a surgical education scenario//2022 International Symposium on Medical Robotics (ISMR). Atlanta, USA: IEEE: 1-8 [DOI: 10.1109/ISMR48347.2022.9807505]
- Loper M, Mahmood N, Romero J, Pons-Moll G and Black M J. 2015. SMPL: a skinned multi-person linear model. ACM Transactions on Graphics, 34(6): #248 [DOI: 10.1145/2816795.2818013]

- Lopes M K S, De Jesus Jr B J, Rosanne O M, Pardini S, Appel L, Smith C, et al. 2024. Stop to smell the virtual roses: a mixed-methods pilot study on the impact of multisensory virtual reality nature experiences on feelings of relaxation. *Frontiers in Virtual Reality*, 5: #1451704 [DOI: 10.3389/frvir.2024.1451704]
- Luo Y B, Liang X Y, Zhang F and Li C R. 2025. Gait recognition combining deep residual and multi feature fusion. *Journal of Image and Graphics*, 30(5): 1466-1478 (罗亚波, 梁心语, 张峰, 李存荣. 2025. 结合深度残差与多特征融合的步态识别方法. *中国图象图形学报*, 30(5): 1466-1478) [DOI: 10.11834/jig.240152]
- Maeda T and Kurahashi T. 2019. TherModule: wearable and modular thermal feedback system based on a wireless platform//*Proceedings of the 10th Augmented Human International Conference*. Reims, France: Association for Computing Machinery: #14 [DOI: 10.1145/3311823.3311826]
- Maquiling V, Byrne S A, Nyström M, Kasneci E and Niehorster D C. 2023. V-ir-Net: a novel neural network for pupil and corneal reflection detection trained on simulated light distributions//*Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*. Athens, Greece: Association for Computing Machinery: #23 [DOI: 10.1145/3565066.3608690]
- Mardanbegi D and Pfeiffer T. 2019. EyeMRTK: a toolkit for developing eye gaze interactive applications in virtual and augmented reality//*2019 ACM Symposium on Eye Tracking Research and Applications*. Denver, USA: Association for Computing Machinery: #76 [DOI: 10.1145/3317956.3318155]
- Maule L, Fornaser A, Leuci M, Conci N, Da Lio M and De Cecco M. 2016. Development of innovative hmi strategies for eye controlled wheelchairs in virtual reality//*Proceedings of the 3rd International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Lecce, Italy: Springer International Publishing: 358-377 [DOI: 10.1007/978-3-319-40651-0_29]
- Meier M, Streli P, Fender A and Holz C. 2021. TapID: rapid touch interaction in virtual reality using wearable sensing//*Proceedings of 2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. Lisboa, Portugal: IEEE: 519-528 [DOI: 10.1109/VR50410.2021.00076]
- Minakata K, Hansen J P, Mackenzie I S, Bækgaard P and Rajanna V. 2019. Pointing by gaze, head, and foot in a head-mounted display//*2019 ACM Symposium on Eye Tracking Research and Applications*. Denver, USA: Association for Computing Machinery: #69 [DOI: 10.1145/3317956.3318150]
- Miyashita H, Kaji Y and Sato A. 2023. Electric salt: tableware design for enhancing taste of low-salt foods//*The 36th Annual ACM Symposium on User Interface Software and Technology*. San Francisco, USA: Association for Computing Machinery: #32 [DOI: 10.1145/3586182.3616626]
- Moll P, Leibetseder A, Kletz S, Lux M and Muenzer B. 2019. Alternative inputs for games and AR/VR applications: deep headbanging on the web//*Proceedings of the 10th ACM Multimedia Systems Conference*. New York, USA: Association for Computing Machinery: 320-323 [DOI: 10.1145/3304109.3323832]
- Mukashev D, Ranasinghe N and Nittala A S. 2023. TactTongue: prototyping electrotactile stimulations on the tongue//*Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. San Francisco, USA: Association for Computing Machinery: #105 [DOI: 10.1145/3586183.3606829]
- Murhij Y and Serebrenny V. 2021. Hand gestures recognition model for augmented reality robotic applications//*The 15th International Conference on Electromechanics and Robotics "Zavalishin's Readings"*. Ufa, Russia: Springer International Publishing: 187-196 [DOI: 10.1007/978-981-15-5580-0_15]
- Nagao R, Matsumoto K, Narumi T, Tanikawa T and Hirose M. 2018. Ascending and descending in virtual reality: simple and safe system using passive haptics. *IEEE Transactions on Visualization and Computer Graphics*, 24(4): 1584-1593 [DOI: 10.1109/TVCG.2018.2793038]
- Narayan B S, Janamanchi S, Harsha A, Nitheezkant R, Jonna P and Rao M. 2024. EOG-assist: real-time EOG-based eye tracking for assistive systems//*Proceedings of 2024 International Conference on Control, Automation, Robotics and Vision (ICARCV)*. Dubai, United Arab Emirates: IEEE: 795-800 [DOI: 10.1109/ICARCV63323.2024.10821594]
- Nooruddin N, Dembani R and Maitlo N. 2020. HGR: hand-gesture-recognition based text input method for AR/VR wearable devices//*Proceedings of 2020 IEEE International conference on Systems, Man, and Cybernetics (SMC)*. Toronto, Canada: IEEE: 744-751 [DOI: 10.1109/SMC42975.2020.9283348]
- Noroozi F, Corneanu C A, Kamińska D, Sapiński T, Escalera S and Anbarjafari G. 2021. Survey on emotional body gesture recognition. *IEEE Transactions on Affective Computing*, 12(2): 505-523 [DOI: 10.1109/TAFFC.2018.2874986]
- Nwobodo O J, Wereszczyński K and Cyran K. 2023. A review on tracking head movement in augmented reality systems. *Procedia Computer Science*, 225: 4344-4353 [DOI: 10.1016/j.procs.2023.10.431]
- Panda A K, Chakravarty R and Moulik S. 2021. Hand gesture recognition using flex sensor and machine learning algorithms//*Proceedings of 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. Langkawi Island, Malaysia: IEEE: 449-453 [DOI: 10.1109/IECBES48179.2021.9398789]
- Patnaik B, Batch A and Elmqvist N. 2019. Information olfaction: harnessing scent to convey data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 726-736 [DOI: 10.1109/TVCG.2018.2865237]
- Phan H L, Le T H, Lim J M, Hwang C H and Koo K I. 2022. Effectiveness of augmented reality in stroke rehabilitation: a meta-analysis.

- Applied Sciences, 12(4): #1848 [DOI: 10.3390/app12041848]
- Punpongsanon P, Iwai D and Sato K. 2015. SoftAR: visually manipulating haptic softness perception in spatial augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 21 (11): 1279-1288 [DOI: 10.1109/TVCG.2015.2459792]
- Qian X, He F M, Hu X Y, Wang T Y, Ipsita A and Ramani K. 2022. ScalAR: authoring semantically adaptive augmented reality experiences in virtual reality//*Proceedings of 2022 CHI Conference on Human Factors in Computing Systems*. New Orleans, USA: Association for Computing Machinery: #65 [DOI: 10.1145/3491102.3517665]
- Qin T, Li P L and Shen S J. 2018. VINS-mono: a robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4): 1004-1020 [DOI: 10.1109/TRO.2018.2853729]
- Ranasinghe N and Do E Y L. 2016a. Virtual sweet: simulating sweet sensation using thermal stimulation on the tip of the tongue//*The 29th Annual ACM Symposium on User Interface Software and Technology*. Tokyo, Japan: Association for Computing Machinery: 127-128 [DOI: 10.1145/2984751.2985729]
- Ranasinghe N and Do E Y L. 2016b. Digital lollipop: studying electrical stimulation on the human tongue to simulate taste sensations. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(1): #5 [DOI: 10.1145/2996462]
- Ranasinghe N, Nguyen T N T, Liangkun Y, Lin L Y, Tolley D and Do E Y L. 2017. Vocktail: a virtual cocktail for pairing digital taste, smell, and color sensations//*Proceedings of the 25th ACM International Conference on Multimedia*. Mountain View, USA: Association for Computing Machinery: 1139-1147 [DOI: 10.1145/3123266.3123440]
- Rashid A and Hasan O. 2019. Wearable technologies for hand joints monitoring for rehabilitation: a survey. *Microelectronics Journal*, 88: 173-183 [DOI: 10.1016/j.mejo.2018.01.014]
- Rau T, Koch M, Pathmanathan N, Isenberg T, Weiskopf D, Sedlmair M, et al. 2024. Understanding collaborative learning of molecular structures in AR with eye tracking. *IEEE Computer Graphics and Applications*, 44(6): 39-51 [DOI: 10.1109/MCG.2024.3503903]
- Rautaray S S and Agrawal A. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1): 1-54 [DOI: 10.1007/s10462-012-9356-9]
- Romero J, Tzionas D and Black M J. 2017. Embodied hands: modeling and capturing hands and bodies together. *ACM Transactions on Graphics*, 36(6): #245 [DOI: 10.1145/3130800.3130883]
- Rovithis E, Moustakas N, Floros A and Vogklis K. 2019. Audio legends: investigating sonic interaction in an augmented reality audio game. *Multimodal Technologies and Interaction*, 3(4): #73 [DOI: 10.3390/mti3040073]
- Salazar J, Kirchhoff K and Huang Z H. 2019. Self-attention networks for connectionist temporal classification in speech recognition//*Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK: IEEE: 7115-7119 [DOI: 10.1109/ICASSP.2019.8682539]
- Samad M, Gatti E, Hermes A, Benko H and Parise C. 2019. Pseudo-haptic weight: changing the perceived weight of virtual objects by manipulating control-display ratio//*Proceedings of 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland, UK: Association for Computing Machinery: #320 [DOI: 10.1145/3290605.3300550]
- Samanta A and Guha T. 2017. On the role of head motion in affective expression//*Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, USA: IEEE: 2886-2890 [DOI: 10.1109/ICASSP.2017.7952684]
- Sancheti K, Krishnan K S, Suhaas A and Suresh P. 2018. Hands-free cursor control using intuitive head movements and cheek muscle twitches//*Proceedings of 2018 IEEE Region 10 Conference*. Jeju, Korea (South): IEEE: 356-361 [DOI: 10.1109/TENCON.2018.8650532]
- Schaack S, Chernyshov G, Ragozin K, Tag B, Peiris R and Kunze K. 2019. Haptic collar: vibrotactile feedback around the neck for guidance applications//*Proceedings of the 10th Augmented Human International Conference*. Reims, France: Association for Computing Machinery: #12 [DOI: 10.1145/3311823.3311840]
- Scholtes M, Seewald P and Eckstein L. 2019. Implementation and evaluation of a gaze-dependent in-vehicle driver warning system//*Proceedings of 2018 AHFE International Conference on Human Factors in Transportation on Advances in Human Aspects of Transportation*. Orlando, USA: Springer International Publishing: 895-905 [DOI: 10.1007/978-3-319-93885-1_84]
- Seelinger A, Cheng L and Netland T. Augmented reality for industrial quality inspection: an experiment assessing task performance and human factors. *Computers in Industry*, 2023, 151: #103985 [DOI: 10.1016/j.compind.2023.103985]
- Shang J C and Wu J. 2019. Enabling secure voice input on augmented reality headsets using internal body voice//*Proceedings of the 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. Boston, USA: IEEE: 1-9 [DOI: 10.1109/SAHCN.2019.8824980]
- Shavit Y and Ferens R. 2019. Introduction to camera pose estimation with deep learning [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/1907.05272.pdf>
- Shchekoldin A I, Shevyakov A D, Dema N U and Kolyubin S A. 2017. Adaptive head movements tracking algorithms for AR interface controlled telepresence robot//*Proceedings of the 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*. Miedzyzdroje, Poland: IEEE: 728-733 [DOI: 10.1109/MMAR.2017.8046918]
- Shen J, Pang R M, Weiss R J, Schuster M, Jaitly N, Yang Z H, et al.

2018. Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada: IEEE: 4779-4783 [DOI: 10.1109/ICASSP.2018.8461368]
- Shi W W, Wang Z Y, Sezgin T M, Dorsey J and Rushmeier H E. 2017. Material design in augmented reality with in-situ visual feedback//2017 Eurographics Symposium on Rendering: Experimental Ideas and Implementation. Helsinki, Finland: Eurographics Association: 93-103 [DOI: 10.2312/sre.20171198]
- Shih O and Rowe A. 2019. Can a phone hear the shape of a room?//Proceedings of the 18th International Conference on Information Processing in Sensor Networks. Montreal, Canada: Association for Computing Machinery: 277-288 [DOI: 10.1145/3302506.3310407]
- Sigut J and Sidha S A. 2011. Iris center corneal reflection method for gaze tracking using visible light. IEEE Transactions on Biomedical Engineering, 58(2): 411-419 [DOI: 10.1109/TBME.2010.2087330]
- Sikora M, Russo M, Derek J and Jurčević A. 2018. Soundscape of an archaeological site recreated with audio augmented reality. ACM Transactions on Multimedia Computing, Communications, and Applications, 14(3): #74 [DOI: 10.1145/3230652]
- Slocum C, Zhang Y C, Abu-Ghazaleh N and Chen J S. 2023. Going through the motions: AR/VR keylogging from user head motions//The 32nd USENIX Security Symposium. Anaheim, USA: USENIX Association: 159-174 [DOI: 10.5555/3620237.3620247]
- Spelmezan D, Sahoo D R and Subramanian S. 2016. Sparkle: towards haptic hover-feedback with electric arcs//The 29th Annual ACM Symposium on User Interface Software and Technology. Tokyo, Japan: Association for Computing Machinery: 55-57 [DOI: 10.1145/2984751.2985702]
- Stavrev S and Ginchev D. 2023. Prototype monitoring system detecting mental workload during pilot simulation training//Proceedings of 2023 International Conference on Military Technologies (ICMT). Brno, Czech Republic: IEEE: 1-6 [DOI: 10.1109/ICMT58149.2023.10171264]
- Strandholt P L, Dogaru O A, Nilsson N C, Nordahl R and Serafin S. 2020. Knock on wood: combining redirected touching and physical props for tool-based interaction in virtual reality//Proceedings of 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, USA: Association for Computing Machinery: 1-13 [DOI: 10.1145/3313831.3376303]
- Sun Q, Huang F C, Kim J, Wei L Y, Luebke D and Kaufman A. 2017. Perceptually-guided foveation for light field displays. ACM Transactions on Graphics, 36(6): 1-13 [DOI: 10.1145/3130800.3130807]
- Sun R, Cheng A S K, Chan C, Hsiao J, Privitera A J, Gao J L, et al. 2023. Tracking gaze position from EEG: exploring the possibility of an EEG-based virtual eye-tracker. Brain and Behavior, 13(10): #e3205 [DOI: 10.1002/brb3.3205]
- Tan X, Qin T, Soong F and Liu T Y. 2021. A survey on neural speech synthesis [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/2106.15561.pdf>
- Tang D H, Yu T H and Kim T K. 2013. Real-time articulated hand pose estimation using semi-supervised transductive regression forests//Proceedings of 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE: 3224-3231 [DOI: 10.1109/iccv.2013.400]
- Tao J H, Wu Y C, Yu C, Weng D D, Li G J, Han T, et al. 2022. A survey on multi-modal human-computer interaction. Journal of Image and Graphics, 27(6): 1956-1987 (陶建华, 巫英才, 喻纯, 翁冬冬, 李冠君, 韩腾, 等. 2022. 多模态人机交互综述. 中国图象图形学报, 27(6): 1956-1987) [DOI: 10.11834/jig.220151]
- Ujitoko Y and Ban Y. 2021. Survey of pseudo-haptics: haptic feedback design and application proposals. IEEE Transactions on Haptics, 14(4): 699-711 [DOI: 10.1109/TOH.2021.3077619]
- Wang J, Erkoyuncu J and Roy R. 2018. A conceptual design for smell based augmented reality: case study in maintenance diagnosis. Procedia CIRP, 78: 109-114 [DOI: 10.1016/j.procir.2018.09.067]
- Wang P Y, Xu C H, Wang P Y, Huang H Y, Chang Y W, Cheng J H, et al. 2021. Game illusionization: a workflow for applying optical illusions to video games//The 34th Annual ACM Symposium on User Interface Software and Technology. Virtual Event, USA: Association for Computing Machinery: 1326-1344 [DOI: 10.1145/3472749.3474824]
- Wang Y X, Skerry-Ryan R J, Stanton D, Wu Y H, Weiss R J, Jaitly N, et al. 2017. Tacotron: towards end-to-end speech synthesis [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/1703.10135.pdf>
- Weier M, Roth T, Hinkenjann A and Slusallek P. 2018. Foveated depth-of-field filtering in head-mounted displays. ACM Transactions on Applied Perception, 15(4): #26 [DOI: 10.1145/3238301]
- Weng S C C, Chiou Y M and Do E Y L. 2024. Dream mesh: a speech-to-3D model generative pipeline in mixed reality//Proceedings of 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR). Los Angeles, USA: IEEE: 345-349 [DOI: 10.1109/AIxVR59861.2024.00059]
- Wiemer M. 2021. Mojo vision: designing anytime, anywhere AR contact lenses with mojo lens//Proceedings of SPIE AVR21 Industry Talks II. Virtual Event: SPIE: #117640Z [DOI: 10.1117/12.2597476]
- Wu Z Y, Rajendran S, Van As T, Badrinarayanan V and Rabinovich A. 2019. EyeNet: a multi-task deep network for off-axis eye gaze estimation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South): IEEE: 3683-3687 [DOI: 10.1109/iccvw.2019.00455]
- Xia W Y, Qiu W and Feng X C. 2020. Research on speech accurate recognition technology based on deep learning DNN-HMM//The 11th International Symposium on Multispectral Image Processing and

- Pattern Recognition. Wuhan, China: Society of Photo-Optical Instrumentation Engineers: 312-315 [DOI: 10.1117/12.2539467]
- Xia X X, Guan Y Q, State A, Chakravarthula P, Rathinavel K, Cham T J, et al. 2019. Towards a switchable AR/VR near-eye display with accommodation-vergence and eyeglass prescription support. *IEEE Transactions on Visualization and Computer Graphics*, 25(11): 3114-3124 [DOI: 10.1109/TVCG.2019.2932238]
- Xiong D Z, Zhang D H, Chu Y Q, Zhao Y W and Zhao X G. 2024. Intuitive human-robot-environment interaction with EMG signals: a review. *IEEE/CAA Journal of Automatica Sinica*, 11(5): 1075-1091 [DOI: 10.1109/JAS.2024.124329]
- Xu F, Zhou T Y, Nguyen T, Bao H H, Lin C and Du J. 2025. Integrating augmented reality and LLM for enhanced cognitive support in critical audio communications. *International Journal of Human-Computer Studies*, 194: #103402 [DOI: 10.1016/j.ijhcs.2024.103402]
- Yan Y K, Yu C, Yi X and Shi Y C. 2018. HeadGesture: hands-free input approach leveraging head movements for HMD devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4): #198 [DOI: 10.1145/3287076]
- Yan Y X, Bai X L, He W P, Wang S X, Zhang X Y, Wang P, et al. 2023. A novel adaptive visualization method based on user intention in AR manual assembly. *The International Journal of Advanced Manufacturing Technology*, 129(9): 4705-4730 [DOI: 10.1007/s00170-023-12557-w]
- Yang F, Wu Y, Sakti S and Nakamura S. 2019. Make skeleton-based action recognition model smaller, faster and better//*Proceedings of the 1st ACM International Conference on Multimedia in Asia*. Beijing, China: Association for Computing Machinery: #31 [DOI: 10.1145/3338533.3366569]
- Yang J, Barde A and Billingham M. 2022. Audio augmented reality: a systematic review of technologies, applications, and future research directions. *Journal of the Audio Engineering Society*, 70(10): 788-809 [DOI: 10.17743/jaes.2022.0048]
- Yang P, Shi Y X, Tao X L, Liu Z Q, Li S Y, Chen X Y, et al. 2023. Self-powered virtual olfactory generation system based on bionic fibrous membrane and electrostatic field accelerated evaporation. *EcoMat*, 5(2): #12298 [DOI: 10.1002/eom2.12298]
- Yang T H, Kim J R, Jin H, Gil H, Koo J H and Kim H J. 2021a. Recent advances and opportunities of active materials for haptic technologies in virtual and augmented reality. *Advanced Functional Materials*, 31(39): #2008831 [DOI: 10.1002/adfm.202008831]
- Yang T H, Son H, Byeon S, Gil H, Hwang I, Jo G, et al. 2021b. Magnetorheological fluid haptic shoes for walking in VR. *IEEE Transactions on Haptics*, 14(1): 83-94 [DOI: 10.1109/TOH.2020.3017099]
- Yang Z J, Wei Y L, Shen S and Choudhury R R. 2020. Ear-AR: indoor acoustic augmented reality on earphones//*Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. London, UK: Association for Computing Machinery [DOI: 10.1145/3372224.3419213]
- Yi S H, Qin Z R, Novak E, Yin Y F and Li Q. 2016. GlassGesture: exploring head gesture interface of smart glasses//*Proceedings of the 35th Annual IEEE International Conference on Computer Communications*. San Francisco, USA: IEEE: 1-9 [DOI: 10.1109/INFOCOM.2016.7524542]
- Yin J, Hinchet R, Shea H and Majidi C. 2021. Wearable soft technologies for haptic sensing and feedback. *Advanced Functional Materials*, 31(39): #2007428 [DOI: 10.1002/adfm.202007428]
- Yu X G, Xie Z Q, Yu Y, Lee J, Vazquez-Guardado A, Luan H W, et al. 2019. Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature*, 575(7783): 473-479 [DOI: 10.1038/s41586-019-1687-0]
- Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang C L, et al. 2020a. MediaPipe hands: on-device real-time hand tracking [EB/OL]. [2025-06-27]. <https://arxiv.org/pdf/2006.10214.pdf>
- Zhang G F, Huang G, Xie W J, Chen D P, Wang N, Liu H M, et al. 2024. A review of monocular visual-inertial SLAM. *Journal of Image and Graphics*, 29(10): 2839-2858 (章国锋, 黄赣, 谢卫健, 陈丹鹏, 王楠, 刘浩敏, 等. 2024. 基于单目视觉惯性的同步定位与地图构建方法综述. *中国图象图形学报*, 29(10): 2839-2858) [DOI: 10.11834/jig.230863]
- Zhang G T and Hansen J P. 2019. A virtual reality simulator for training gaze control of wheeled tele-robots//*The 25th ACM Symposium on Virtual Reality Software and Technology*. Parramatta, Australia: Association for Computing Machinery: #49 [DOI: 10.1145/3359996.3364707]
- Zhang Q, Lu H, Sak H, Tripathi A, McDermott E, Koo S, et al. 2020b. Transformer transducer: a streamable speech recognition model with transformer encoders and RNN-T loss//*Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE: 7829-7833 [DOI: 10.1109/ICASSP40776.2020.9053896]
- Zhang X C, Sugano Y, Fritz M and Bulling A. 2019. MPIIGaze: real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1): 162-175 [DOI: 10.1109/TPAMI.2017.2778103]
- Zheng J K, Liu X C, Liu W, He L X, Yan C G and Mei T. 2022. Gait recognition in the wild with dense 3D representations and a benchmark//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, USA: IEEE: 20196-20205 [DOI: 10.1109/CVPR52688.2022.01959]
- Zhou C, Li Q, Li C, Yu J, Liu Y X, Wang G J, et al. 2024. A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT [J/OL]. [2025-11-17]. *International Journal of Machine Learning and Cybernetics*: 1-65.

<https://link.springer.com/article/10.1007/s13042-024-02443-6>

Zhou Y T, Chen J J and Wang M H. 2022. A meta-analytic review on incorporating virtual and augmented reality in museum learning. *Educational Research Review*, 36: #100454 [DOI: 10.1016/j.edurev.2022.100454]

Zollmann S, Langlotz T, Grasset R, Lo W H, Mori S and Regenbrecht H. 2021. Visualization techniques in augmented reality: a taxonomy, methods and patterns. *IEEE Transactions on Visualization and Computer Graphics*, 27 (9) : 3808-3825 [DOI: 10.1109/

TVCG.2020.2986247]

作者简介

王怡雯,女,博士研究生,主要研究方向为三维视觉、增强现实和人机交互技术。E-mail: junzi521@stu.xjtu.edu.cn

赵玺,通信作者,女,副教授,主要研究方向为计算机图形学、计算机视觉、三维数据的分析、处理及合成。

E-mail: xi.zhao@mail.xjtu.edu.cn