

# 拟合稀疏信号的可穿戴设备混合现实视觉人机交互方法

张恒<sup>1</sup>, 任晓康<sup>1</sup>, 郝飞<sup>1</sup>, 高亚娟<sup>2</sup>, 王天宇<sup>3</sup>, 谢利德<sup>1†</sup>

(1. 承德医学院生物医学工程系, 河北承德 067000; 2. 北京大学第三医院放射科, 北京 100191;

3. 北京大学医学部医学技术研究院, 北京 100191)

**摘要:** 常规人机交互方法主要依托于深度自适应算法, 缺少对动作特征的重构, 无法对待检测样本进行动态对齐, 使得动作识别错误率较高。为此, 提出了拟合稀疏信号的可穿戴设备混合现实视觉人机交互方法。基于骨骼信息融合原理获取动作动态, 并将动作图像从深度图像中分割出来, 通过将动作图像进行超平面转换, 得到动作参数峰值。结合动作边缘方程提取动作特征值, 采用拟合稀疏信号算法对特征点进行空域离散化处理, 利用稀疏信号阵列模型对动作特征进行重构与动态对齐, 从而近似分类与识别动作类型, 借助鼠标控制 pynput 库对角色进行动作驱动, 由此实现人机交互。以可穿戴设备混合现实视觉数据集作为实验对象, 将所提方法应用于人机交互的动作识别, 结果表明, 所提方法对于不同类别的动作识别具有更低的识别错误率。

**关键词:** 拟合稀疏信号; 可穿戴设备; 混合现实视觉; 人机交互

**中图分类号:** TP212

**文献标识码:** A

## Wearable Device Hybrid Reality Visual Human-machine Interaction Method for Fitting Sparse Signals

ZHANG Heng<sup>1</sup>, REN Xiaokang<sup>1</sup>, HAO Fei<sup>1</sup>, GAO Yajuan<sup>2</sup>, WANG Tianyu<sup>3</sup>, XIE Lide<sup>1†</sup>

(1. Biomedical Engineering, Chengde Medical University, Chengde, Hebei 067000, China;

2. Department of Radiology, Peking University Third Hospital, Beijing 100191, China;

3. Institute of Medical Technology, Peking University Health Science Center, Beijing 100191, China)

**Abstract:** The conventional human-computer interaction methods mainly rely on deep adaptive algorithms. Due to the lack of reconstruction of action features, they cannot dynamically align the detected samples, resulting in a high error rate in action recognition. To this end, a wearable device hybrid reality visual human-machine interaction method that fits sparse signals is proposed. Based on the principle of bone information fusion, motion dynamics are obtained, and the motion image is segmented from the depth image. Through the hyperplane transformation of the motion image, the peak value of the motion parameters is obtained. Combined with the action edge equation, the action feature values are extracted, and the feature points are processed by spatial discretization using the sparse signal fitting algorithm. The action features are reconstructed and dynamically aligned using the sparse signal array model. Thus, approximate classification and recognition of action types can be achieved, and the pynput library can be controlled by the mouse to drive the actions of characters, thereby achieving human-machine interaction. Using a wearable device mixed reality visual dataset as the experimental object, the proposed method was applied to action recognition in human-computer interaction. The results showed that the proposed method has a lower recognition error rate for different types of action recognition.

**Key words:** fitting sparse signals; wearable devices; hybrid reality vision; human-machine interaction

计算机技术和虚拟现实技术正逐渐渗透到人们的生活中,推动人机交互领域的升级和演变。其中,可穿戴设备及其附带的各种运动传感器在人机交互过程中催生了各种运动识别应用。通过传感器数据收集和特定的反馈机制实现了人与虚拟机之间的运动匹配,从而达到人机交互的目的。

目前已实现的人机交互动作识别算法主要包括文献[1]基于眼动方法和文献[2]基于语义本体方法。基于眼动方法的人机交互方法以眼电与脑电信号为基础,通过自适应阈值归一化方法进行眼电信号识别,结合特征匹配模板实现对角色的实时控制。但该方法为单通道交互方式,对于类不均衡的数据样本无法达到理想的交互识别效果。基于语义本体的人机交互方法将体感设备作为输入,通过三维医学影像的显示结果,构建自适应识别模型,由此完成人机交互。但此方法对于非接触式单手动态交互轨迹无法进行实时跟踪,且不适用于具有深度信息的动作识别。

针对以上所述方法存在的不足,提出了一种基于拟合稀疏信号的人机交互方法,适用于可穿戴设备混合现实视觉的人机交互。

## 1 可穿戴设备混合现实视觉人机交互方法设计

### 1.1 动作特征提取

为了确保动作识别不受环境光照条件和其他干扰因素的影响,应首先对操作员的动作进行分割,并根据分割参数提取动作特征,以实现更准确的人机交互<sup>[3]</sup>。

本文采用骨骼信息融合的原理,首先获得操作者的运动动力学,根据操作人员右手的骨骼特征数据,建立关键关节的空间坐标,并将其与深度坐标系进行深度转化,根据同一关节在不同坐标系下的映射坐标值,对动作图像设定相应的分割阈值,进而将动作从深度图像中分割出来<sup>[4]</sup>。一般情况下,建立以手的主要关节为质心的边界框,以及将边界框的高度作为分割阈值,以达到分割动作的目的。

定义动作分割的损失函数为:

$$f(x) = 1 - \frac{|A_c - U_c|}{A_c} - \alpha_c \quad (1)$$

式中,  $A_c$ 、 $U_c$  分别表示动作动态估计边框的质心点与模糊边缘点;  $\alpha_c$  表示提取框与图像边缘的距离。

度量动作包围框宽高比的相似性计算公式为:

$$v_v = \frac{4}{\pi^2} [\arctan \frac{\omega_c}{h_c} \times f(x)] \quad (2)$$

式中,  $\omega_c$  表示目标框与预测框的距离;  $h_c$  表示预测框的重叠率。

设包围框重采样点的离散序列为  $A$ ,  $A_k(x_k, y_k)$  为框中的第  $k$  个点,取当前采样点前后各 4 个点,利用最小二乘法分别对采样点进行拟合<sup>[5]</sup>,则得到的矩阵形式为:

$$M_k = \alpha_0 (A_k b + v_v) \quad (3)$$

式中,  $\alpha_0$  表示目标框的参量;  $b$  表示目标框与预测框的转换系数。

假设左右相机距离刚好为深度图像的长度  $l_s$ ,则视场角下的动作范围可由下式得到:

$$\omega_1 = \arctan \frac{4d \times l_f}{l_s - M_k} \quad (4)$$

式中,  $\omega_1$  表示动作范围视场角;  $d$  表示观察距离;  $l_f$  表示手部等效直径。

动作关节的三维坐标系与图像基准坐标系之间的转换方法为:

$$T_1 = \omega_1 (R_r * R_s^{-1}) \quad (5)$$

式中,  $T_1$  表示图像坐标系;  $R_r$  表示旋转矩阵;  $R_s$  表示平移矩阵。

则得到动作包围框的长、宽、高的像素值为:

$$\begin{cases} L_1 = T_l - T_1 \\ B_1 = N_k * 255 \\ H_1 = n_0 + p_0 \end{cases} \quad (6)$$

式中,  $T_l$  表示图像的转置矩阵;  $N_k$  表示  $k$  级灰度的像素数;  $n_0$  表示像素个数;  $p_0$  表示标定点数量。

利用上式对操作人员关节坐标点单位进行换算,并从深度图像中二值分割出动作,以手势动作为例,具体如图 1 所示。

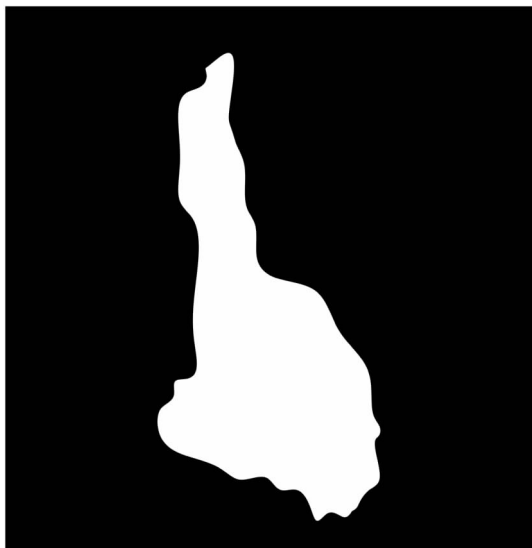


图1 手势动作分割图

将操作人员的动作动态分割出来后,需要提取动作的关键特征,为人机交互的实现奠定基础<sup>[6]</sup>。假设动作关节点  $x$  的邻域为高斯近似,则可以利用  $x$  的邻域信息根据动作周边像素点坐标估计出极值,具体如下:

$$y = \frac{1}{2\delta_0\pi} \times \exp\left(-\frac{L_1 B_1 H_1}{4}\right)^2 \quad (7)$$

式中,  $y$  表示一维高斯函数;  $\delta_0$  表示拟合系数。

利用动作分割图的等效质心点的坐标对高斯函数进行超平面转换,获得动作参数  $\mu_0$  即为峰值位置<sup>[7]</sup>,由此得到动作边缘的椭圆方程表达式为:

$$f_1(x) = by + f_{p_0} \quad (8)$$

式中,  $b$  表示方程系数;  $f_{p_0}$  表示椭圆中心点坐标。

将在椭圆方程上均匀的生成点作为特征点集合,则得到的动作特征向量集合为:

$$f_n(p) = \sum f_1(x) \times 2\tau(p; x_0, y_0) \quad (9)$$

式中,  $\tau$  表示特征点描述子;  $p(x_0)$ 、 $p(y_0)$  分别是图像中  $x_0$ 、 $y_0$  的像素灰度值;  $f_n(p)$  表示动作特征点向量集合。

根据动作关节坐标系与深度图像坐标系之间的关系,采用骨骼融合原理对动作轮廓进行分割,结合动作边缘椭圆方程,提取动作关键特征点,并构造特征向量集合。该过程可为后续动作识别做准备。

## 1.2 基于拟合稀疏信号的动作类型识别

在对动作进行特征提取和表示之后,需要使用

提取的动作特征来训练分类器,以完成动作类型的分类和识别<sup>[8]</sup>。本文采用稀疏信号拟合方法,通过最大区间超平面将两类数据尽可能分离,对输入样本进行近似可分,进而实现动作类型分类。

首先,基于动作特征值构建稀疏阵列信号模型,并根据特征点在整个空间域中的稀疏特性对其进行离散化<sup>[9]</sup>。动作特征点的空域稀疏性示意图如图2所示。

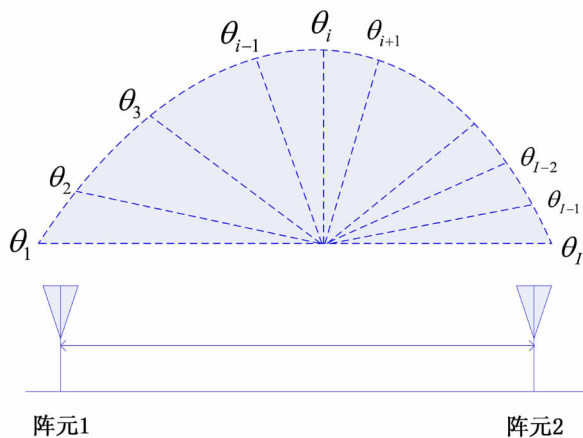


图2 动作特征点的空域稀疏性示意图

图中,  $\theta_1$ 、 $\theta_2$ 、 $\theta_3$  分别表示特征点对应的空域参数;  $\theta_{i-1}$ 、 $\theta_i$ 、 $\theta_{i+1}$  分别表示第  $i-1$ 、 $i$  和  $i+1$  个特征点的信号矢量;  $\theta_i$ 、 $\theta_{i-1}$ 、 $\theta_{i-2}$  分别表示对应特征点的高斯白噪声。

通过将特征点在空域进行离散化以后,可以将特征向量的这列流行矩阵  $A(\theta)$  扩展为一个完备的导向矢量字典  $D(\theta)$ ,则有:

$$D(\theta) = A(\theta)(f_n(p) + \theta_i) \quad (10)$$

式中,  $f_n(p)$  表示特征点向量集合。

此时,阵列特征点的处理模型可表示为:

$$x(t) = m \times D(\theta) \times \theta_i \quad (11)$$

式中,  $m$  表示阵元个数。

基于上述模型,将特征向量进行稀疏信号重构,可得到:

$$x'(t) = 2n/s_0 \cdot x(t) \quad (12)$$

式中,  $n$  表示特征点个数;  $s_0$  表示重构范数。

采用动态规划方式将两个序列进行动态对齐<sup>[10]</sup>,将两段序列的特征匹配路径定义为  $R$ ,则输入特征序列  $T_1$  和待匹配序列  $T_2$  之间的规整距离为:

$$d_1(T, T_2) = \min \sqrt{\sum r_k / x'(t)} \quad (13)$$

式中,  $r_k$  表示匹配系数。

参考人体运动的运动学规律,初步设计了上点击、下点击、左点击和右点击 4 个基本动作,根据可穿戴设备的功能人机交互开关,扩展了上滑动、下滑动、左滑动和右滑动 4 个控制动作以及另外 2 个辅助动作<sup>[11]</sup>,并制定了数字人的动作规范,具体如表 1 所示。

表 1 数字人交互动作规范

动作名称	动作规范内容
上点击	双脚平行站立间隔 20 cm 左右,左脚不动,右脚抬起移动到原来位置的上方 15 cm 左右落下
下点击	双脚平行站立间隔 20 cm 左右,左脚不动,右脚抬起移动到原来位置的下方 15 cm 左右落下
左点击	双脚平行站立间隔 20 cm 左右,右脚不动,左脚抬起移动到原来位置的左方 15 cm 左右落下
右点击	双脚平行站立间隔 20 cm 左右,左脚不动,右脚抬起移动到原来位置的右方 15 cm 左右落下
上滑动	双脚平行站立间隔 20 cm 左右,左脚不动,右脚向上滑动移动 25 cm 左右停下
下滑动	双脚平行站立间隔 20 cm 左右,左脚不动,右脚向下滑动移动 25 cm 左右停下
左滑动	双脚平行站立间隔 20 cm 左右,右脚不动,左脚向左滑动移动 25 cm 左右停下
右滑动	双脚平行站立间隔 20 cm 左右,左脚不动,右脚向右滑动移动 25 cm 左右停下
启动动作	双脚平行站立状态,左右脚相距 25 cm 左右
停止动作	左右脚依次移动离开初始位置并且未返回原位

使用基尼指数作为分类器的分割指标,每次选择最优属性,在每次迭代中,分类器选择基尼指数最低的特征及其对应的分割点进行分类<sup>[12]</sup>,则第  $j$  个样本的类别被标记为第  $l$  类动作类型的概率为:

$$p_{jl} = \frac{\exp(\alpha_l)}{\sum_{i=1}^C d_i(T, T_2)} \quad (14)$$

式中,  $\alpha_l$  表示分类器最后一层的输出;  $C$  表示样本数。

根据求得的  $p_{jl}$  的大小,即可识别数字人当前的动作规范类型,为接下来实现人机交互提供有利条件。

### 1.3 实现人机交互

在上述对于操作人员肢体动作分类与识别的基础上,文中采用鼠标控制 pynput 库进行人机交互。具体实现过程如下。

(1)通过 pynput.mouse 设置和控制鼠标与触摸板的监控输入设备的后端参数<sup>[13]</sup>。

(2)将运动捕捉数据库中角色的所有关节坐标和  $X$ 、 $Y$ 、 $Z$  轴方向设置为初始姿势,即右侧坐标系是具有  $X$ 、 $Y$  和  $Z$  轴方向的标准世界坐标系。

(3)将角色与操作人员的不同动作类型与运动数据加载到运动模型的资源库中,其中不同的手势类型对应不同的肢体动作<sup>[14]</sup>。根据手势类型选择要调用的运动类型。

(4)对特定的动作片段进行虚拟预演分析,获得分解后的动作信息,并将其映射到当前视野中场景的水平面。

(5)修改运动数据库中角色的水平坐标,对运动轨迹数据进行归一化,并根据更新后的运动类型对数据进行校正,以驱动角色运动<sup>[15]</sup>。

通过以上分析,借助 pynput 数据库,即可实现对虚拟角色的交互控制。至此,完成基于拟合稀疏信号的可穿戴设备混合现实视觉人机交互方法的设计。

## 2 实验论证

### 2.1 实验准备

由于缺乏具体的动作控制数据集,当前开源数据库中针对人机交互动作识别的数据样本较少,容易导致样本与总体之间出现偏离,使得样本结果不能够泛化到总体,从而影响实验结论的可靠性。为了降低实验样本选取的偏差性影响,在本实验中,采用数据依赖项采集的方式收集大量角色的动作视频,增加样本量,使得样本更具代表性,减少选取偏差,提高实验结果的可靠性。并通过区域标记与动作帧截取构造实验数据集。对于利用双目摄像机与虚拟现实技术捕捉的动态彩色人物动作图像,可进行去噪与二值化等预处理,得到最终用于实验的像素级灰度图像。并以 10 帧/s 的速度将其从深度图像中分割出来,之后采用拟合稀疏信号算法进行手动标记,标记内容包括动作的类别信息与方位信息。

实验中主要针对的交互类型是可穿戴设备的虚拟界面交互。将智能设备的心智映射到可穿戴设备上,从交互逻辑的角度分析可穿戴设备和智能设备在显示界面上的差异,从而深入用户的实际应用场景,并根据交互信息的水平完成与界面的直接触屏交互。

实验中主要对可穿戴设备的手势进行识别,手

势交互模式如图3所示。

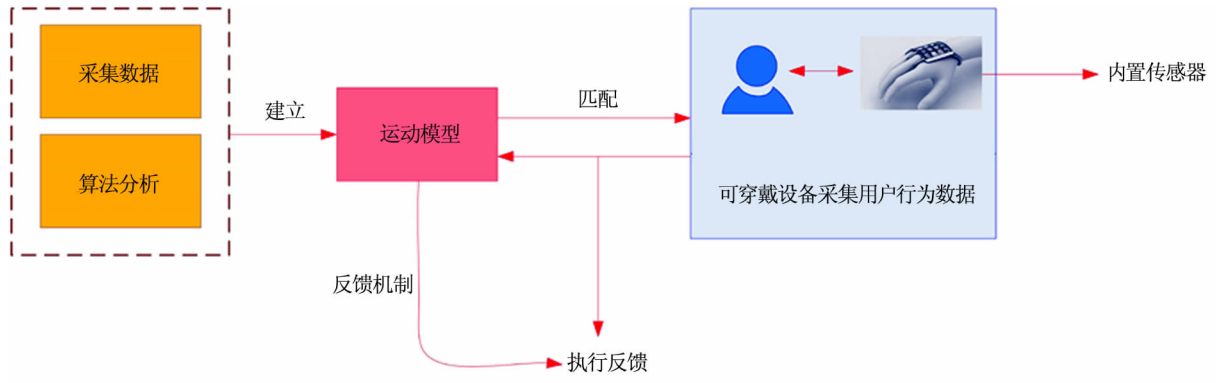


图3 手势交互模式

如图3所示,利用可穿戴设备内置的三轴加速度传感器收集用户当前的行为和运动数据,通过算法对数据进行分析,并建立运动模型。当用户的动作与相应的运动模型相匹配时,可以通过特定的反馈机制触发相应的反馈,完成人机交互动作。

## 2.2 实验说明

选择健康的受试者进行实验,并按照前文所述方法进行动作特征提取以及动作类型分类与识别等操作。首先,令受试者双脚呈平行状态站立在交互区域,然后在鼠标控制 pynput 库中调用激活模块,对虚拟角色进行动作控制,并根据手势的实时运动轨迹,将所有动作类型归一化到平面场景窗口,然后利用转换指令控制角色按照指定的运动轨迹移动,并通过手势的上下移动来控制角色的动作速度。

设计实验如下:受试者随机做出5种待识别样本中的任意一种手势动作,交互界面从 pynput 运动数据库中检索手势图像并进行特征值提取,之后将特征信息反馈给监控计算机,计算机根据监测到的手势调用反馈程序来控制角色做出相应的动作。同时,监测模块通过实时监测人机交互界面的角色动作对其进行调整,从而实现人机交互。

## 2.3 人机交互结果分析

选取操作人员5种待检测手势(如图3)。将所提方法应用于实验设计的人机交互场景,引入虚拟角色反馈手势与操作人员的动作手势之间的匹配度来衡量人机交互效果。动作匹配度的计算公式如下。

$$M_s = \frac{1}{k} \sqrt{\frac{1}{J}(\alpha_1 - \alpha_2)} \quad (15)$$

式中,  $M_s$  表示动作匹配度;  $k$  表示动作类别样本数;  $J$  表示实验次数;  $\alpha_1$  表示动作的真实方位;

$\alpha_2$  表示动作的识别方位。

采用所设计的方法进行人机交互,利用上式求取动作匹配度,所得匹配度越高,表明动作识别越准确,人机交互效果越好。基于本文方法的人机交互结果如图4所示。

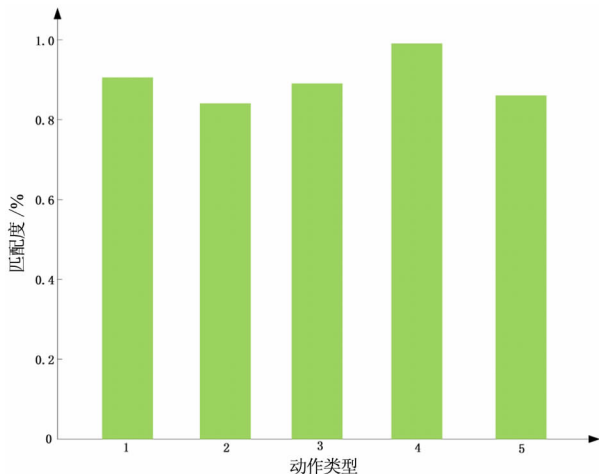


图4 人机交互动作识别结果

如图4所示,利用本文设计的人机交互方法对待检测的手势动作进行识别,对于5种不同的动作类型,本文方法均以较高准确率对其进行了分类与识别,在整个实验过程中,角色与操作人员的动作匹配度始终在80%以上。由此可以说明,本文所提的人机交互方法能够使角色按照操作人员的具体手势完成相应的动作,可以达到预期效果,人机交互效果良好。

## 2.4 动作识别准确性对比实验分析

为进一步证明本文设计的人机交互方法在动作识别方面的优越性能,采用文献[1]基于眼动(方法1)、文献[2]基于语义本体(方法2)作为本文方法的对比方法。以上述5种手势动作类别作为不同方法的输入样本数据,并统计三种人机交互方法

的动作识别错误率,以定量、直观地比较不同方法的人机交互效果。对比结果如图5所示。

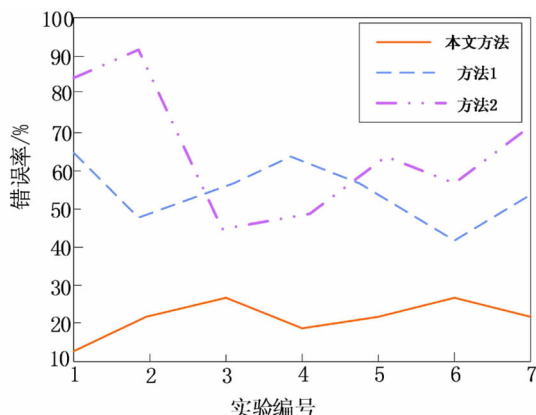


图5 基于不同方法的人机交互动作识别错误率对比

从图5可以看出,本文提出的方法对于不同类别的手势动作识别得到的识别错误率相对更低,始终控制在10%~30%范围内,而方法1与方法2的识别错误率较高,其中方法2的最高错误率达到90%以上,该方法在分类模型的训练过程中通过减慢动作损失的下降速度来提升模型的训练效率,因此影响了模型是识别效果;方法1虽然可以考虑不同动作尺度参数对回归损失的影响,但未考虑到待检测手势与反馈手势位置坐标的相关性,所以两个手势的重合程度较差。通过实验对比结果可以得知,本文方法能够准确识别人体动作,更好地实现人机交互。本文方法之所以能在对比实验中取得优势,是因为该方法基于骨骼信息融合原理获取全面的动作动态,并对动作图像进行分割和超平面转换,得到动作参数峰值,为后续人体动作识别提供了更为精准的基础数据;结合动作边缘方程提取动作特征值,采用拟合稀疏信号算法对特征点进行空域离散化处理,并对动作特征进行重构与动态对齐,提高了动作类型的识别精度,从而加强了人机交互效果。

### 3 结论

为提高人机交互动作识别的准确性,结合拟合稀疏信号算法研究并设计了针对可穿戴设备混合现实视觉的人机交互方法。通过提取动作特征向量与识别动作类型,借助运动数据库实现人机交互。但本研究在动作交互场景的实时识别方面还存在一定不足,接下来将对该方面进行深入研究。

### 参考文献

- [1] 蒋贵虎,刘仲,王亚卓.基于眼动的战机座舱人机交互方法[J].飞机设计,2022,42(3):11-14+19.
- [2] 徐巍峰,徐非非,王文军,等.基于语义本体的无人机人机交互控制方法研究[J].计算机应用与软件,2021,38(12):103-109+148.
- [3] 柯健,张量.用于虚拟现实/增强现实人机交互的实时动作识别方法的研究[J].现代计算机,2023,29(1):59-62.
- [4] 程瑞洲,黄攀峰,刘正雄,等.一种面向在轨服务的空间遥操作人机交互方法[J].宇航学报,2021,42(9):1187-1196.
- [5] 厉钰琪,巩森森.复杂驾驶任务下基于多资源理论的汽车人机交互设计方法[J].设计,2022,35(10):24-27.
- [6] 张玉兰,杜羽.多自由度包装机器人人机交互控制方法[J].包装工程,2021,42(15):239-244.
- [7] 周巡,黄保青,陈亚晨,等.智能眼镜人机交互功能中眼行为识别方法的设计与实现[J].激光杂志,2022,43(3):76-81.
- [8] 苏翎菲,化永朝,董希旺,等.人与无人集群多模态智能交互方法[J].航空学报,2022,43(S1):129-142.
- [9] 段文婷.智能翻译机器人人机交互模块自动优化方法[J].自动化与仪器仪表,2021(10):194-197.
- [10] 郭馨蔚,陶霖密.用于轻度认知障碍测试的自然交互方法[J].应用科技,2022,49(2):33-39.
- [11] 钱一润,王从庆,张闻锐,等.基于SSVEP的空间机械臂脑机交互方法[J].计算机测量与控制,2022,30(6):197-202.
- [12] 苏静,刘兆峰,王姬,等.基于改进的YOLOv3实现动作识别的人机交互方法[J].天津科技大学学报,2021,36(6):49-54.
- [13] 刘尚旺,王培哲,张翰林,等.人机交互系统多维语音信息识别方法[J].计算机仿真,2021,38(12):367-370+469.
- [14] 龙年,刘智慧.虚拟运动目标人机交互方法设计与仿真[J].计算机仿真,2022,39(6):201-205.
- [15] 任雅琪.新闻领域中的可穿戴交互技术应用[J].全媒体探索,2022(7):111-112.