

DOI: 10.19650/j.cnki.cjsi.J2413574

基于辅助喂食全过程意图识别的多模态安全交互方法研究*

吕相谊¹, 赵东辉¹, 丁嘉辉¹, 杨俊友¹, 王硕玉²

(1. 沈阳工业大学电气工程学院 沈阳 110870; 2. 日本高知工科大学智能机械电子学院 高知 7828502)

摘要:为提高无法自主进食的弱机能人群面对被喂食过程中的安全性与灵活性,提出一种面向辅助喂食行为的多模态安全人机交互框架及其全过程意图识别算法。首先,结合用户特征及安全性与灵活性需求,提出一种结合视、触、力、位与语言融合的多模态人机交互框架并构建辅助喂食系统。其次,针对进食全过程包括对进食意图、菜品选择意图、动态喂食点估计与递送位姿计算、咀嚼意图等,提出一种基于视觉为主导的喂食意图全过程识别方法。选取有效刻画进食过程面部动态变化的特征点并设计融合口部纵横比与下颌纵横比识别算法,通过视线矢量分析用户菜品选择意图并基于面部实时位姿设计动态喂食点,从而形成全过程动态意图准确识别。同时,在辅助喂食虚拟映射系统中,结合大语言模型对交互全过程的模糊意图、临时变更意图进行提问,形成反馈机制进而提升交互安全性。最终通过仿真与综合性实验验证提出的方法,通过多模态交互框架可以有效提升辅助喂食过程的灵活性,同时结合大语言模型形成模糊意图与变更意图有效反馈,最终提升交互过程的安全性,该方法为弱机能人群的日常生活辅助喂食行为提供了新型护理方法。

关键词: 辅助喂食; 全过程意图识别; 多模态交互框架; 安全性与灵活性

中图分类号: TN911.72 TH711 TH9 **文献标识码:** A **国家标准学科分类代码:** 510.40

A multimodal safe interaction method based on intention recognition in the whole process of assisted feeding

Lyu Xiangyi¹, Zhao Donghui¹, Ding Jiahui¹, Yang Junyou¹, Wang Shuoyu²

(1. School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China;

2. Department of Intelligent Mechanical Systems Engineering, Kochi University of Engineering, Kochi 7828502, Japan)

Abstract: To improve the safety and flexibility of the feeding process for individuals with limited mobility who are unable to feed themselves, this study proposes a multimodal human-robot interaction framework for assisted feeding and a whole-process intention recognition algorithm. First, based on user characteristics and the requirements for safety and flexibility, a multimodal interaction framework integrating vision, touch, force, position, and language fusion is proposed, and an assisted feeding system is developed. Second, a vision-driven whole-process intention recognition method is introduced to address the entire feeding process, including feeding intention, dish selection intention, dynamic feeding point estimation, delivery pose calculation, and chewing intention. Key facial feature points that effectively capture dynamic changes during feeding are selected, and an algorithm combining the aspect ratio of the mouth and the mandible is designed. The user's dish selection intention is analyzed through gaze vector estimation, and dynamic feeding points are determined based on real-time facial pose tracking, enabling accurate recognition of dynamic intentions throughout the process. Furthermore, in the virtual mapping system for assisted feeding, a feedback mechanism is established by leveraging a large language model to clarify ambiguous intentions and adapt to temporary changes during the interaction, thereby enhancing safety. Finally, the proposed method is validated through simulations and comprehensive experiments. The results demonstrate that the multimodal interaction framework significantly improves the flexibility of the assisted feeding process, while the integration of the large language model provides

收稿日期: 2024-12-07 Received Date: 2024-12-07

* 基金项目: 国家自然科学基金项目(92248304)、辽宁省人工智能重点研发计划项目(2023JH26/10200018)、辽宁省教育厅基本科研项目(LJ212410142073)资助

effective feedback for ambiguous and changing intentions, ultimately enhancing the safety of the interaction. This approach offers a novel care solution for assisting feeding behaviors in the daily lives of individuals with limited mobility.

Keywords: assisted feeding; whole process intention recognition; multimodal interaction framework; safety and flexibility

0 引言

进食行为是日常生活频繁发生的重要基本行为之一,如何为无法自主进食的弱能人群提供灵活、安全的辅助喂食是养老护理机器人领域重要的研究内容之一^[1]。随着人口老龄化加剧和残疾人数量增多医护工作压力骤增^[2],进一步加剧了助餐机器人及其辅助喂食策略研究的重要性与紧迫性^[3]。亟待研究具有高护理质量与灵活护理策略的助餐机器人。

近年来,助餐机器人及其相关研究已经逐步被应用于部分生活场景^[4-5],但面向部分几乎无自主操作能力的弱能用户依然在交互智能性与灵活性存在局限性。在辅助喂食人机交互研究中,国内外学者^[6-7]提出使用操作开关、语音交互及生物电信号交互等意图识别方法并取得了研究进展。然而,针对于脊髓损伤等行动与操作能力较弱的人群来说,长进食时间、多重复咀嚼过程中,频繁的使用手动操作开关与语音,反复穿戴设备均具有一定的繁琐性,交互灵活性需要进一步加强。较多助餐机器人通过头部动作/四肢动作触发开关,驱动机械推送器完成取餐任务,系统应对复杂情况或不确定条件时还难以完成用户的个性化需求。为提高用户进餐多样化操作需求,文献^[8]设定手动、半自动与自动3种模式,并且配备两种力度的操纵杆适配于用户不同部位进行操作,用户控制操纵杆加按钮可实现用户选餐需求。系统预设固定喂食点或固定轨迹,需由用户主动伸脖迎取食物并进食,这种驱动方式对于弱能用户的灵活交互依然存在挑战。一种流食智能喂食器通过二自由度导轨滑台调整喂食口位置,结合图像识别技术,精准获取卧床人士面部和口部的信息,从而提供最优的进食频率与角度,然而对辅助进食的全过程意图并未有效识别。Ha等^[9]开发一种基于脑电的混合脑机接口系统,尽管该项研究能够根据快速捕捉用户意图进行自适应操作。Chio等^[10]提出了一种基于视觉的人脸识别辅助机器人系统,涵盖了食品获取点估计及其非接触式用户交互接口。相关研究采用的交互方式对于弱能用户依然存在模糊意图无法精准识别,临时更改意图无法执行,紧急情况难以快速停止,会为助餐行为带来不确定性甚至危险。因此,面相真实日常生活助餐场景,安全性与灵活性需要进一步提升。

如何结合用户的基本情况并拟合自然进食过程中的柔顺性,建立兼具安全性与灵活性的交互策略是辅助喂食的基础^[11-13]。其中,在辅助喂食交互全过程中,对人体

进食意图识别包括进食事件意图、菜品选择意图、动态喂食点、咀嚼过程与结束意图识别。同时,需要为交互策略提供安全保障。为此,提出以弱能人群为应用对象的一种多模态辅助喂食交互策略及其意图识别方法,旨在引导系统准确辅助用户进食。利用大语言模型构建动态询问反馈机制,实现用户个性化用餐需求及时响应并做出适应性调整,强化人机交互的灵活性与安全性,创新点如下:

1) 提出一种以视觉为主导,融合力、触觉、位置和语言安全反馈机制的多模态辅助喂食交互框架,该框架以拟合自然进食行为的柔顺性为出发点,提升辅助喂食过程中的安全性。

2) 提出一种以视觉为主导的非接触式进食全过程意图识别方法,包括进食事件意图、菜品选择意图、动态喂食点、咀嚼过程与结束等意图识别,通过融合计算提升用户在进食过程中的交互灵活性与舒适性。

1 助餐机器人系统及多模态交互框架

为满足弱能力用户辅助喂食行为,建立一种融合真实场景操作与虚拟场景安全反馈的助餐机器人系统,主要由意图识别系统、虚拟映射系统以及助餐机器人系统组成。其中,助餐机器人系统如图1所示,它集成了七自由度协作机器人、末端盛取器与六维力传感器进行喂食任务。意图识别系统采用RGB-D与RGB相机进行用户全过程进食意图识别、辅助机械臂盛取食物以及动态喂食点的位姿计算。基于Coppeliassim软件建立与真实场景相同的虚拟映射场景,如图2所示。本研究将一次助餐过程任务规划为包括意图识别,递送与交互的多个步骤,每个步骤都结合任务特性设定

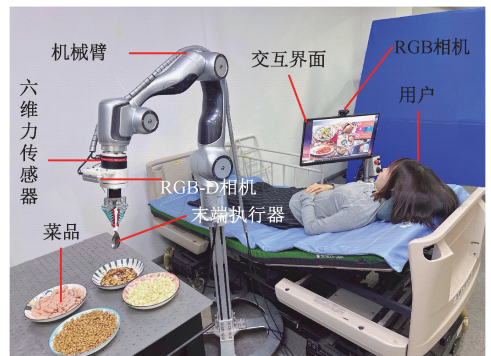


图1 助餐机器人系统

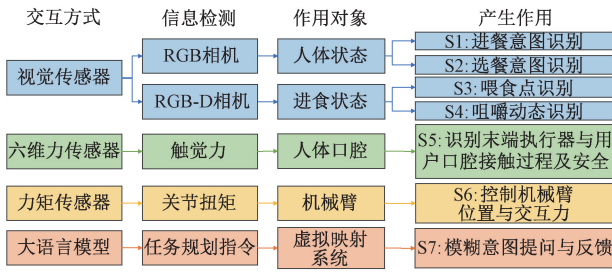
Fig. 1 Assisted feeding robot system

便捷的交互模式,同时在任务规划层以多线程形式嵌入大语言模型,对模糊意图与用户发出的语音进行交互,建立安全反馈机制,从而形成意图闭环系统,即通过视、触、力、位与语言多模态信息实现辅助喂食全过程的安全意图识别,如图 3 所示。



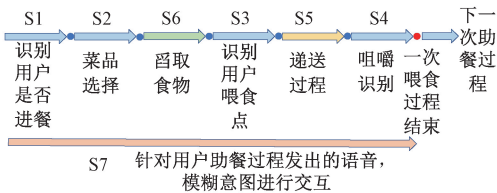
图 2 助餐机器人虚拟映射系统

Fig. 2 Virtual mapping system of assisted feeding robot system



(a) 基于多模态信息的助餐过程意图识别

(a) Intention recognition for assisted feeding based on multimodal information



(b) 多模态信息流

(b) Multimodal information flow

图 3 多模态交互系统信息流

Fig. 3 Information flow diagram of multimodal assisted feeding robot system

2 方 法

2.1 全流程喂食意图识别方法

为提升辅助喂食过程中的交互安全性、灵活性与舒适性,且需满足用户个性化用餐诉求,提出了全流程喂食意图识别方法。首先用户通过语言发出用餐需求,助餐机器人触发辅助进餐任务并初始化位姿。然后,结合图 3 全流程喂食识别系统主要包括进餐意图识别、菜品选

择意图识别、动态喂食点识别、咀嚼过程意图识别。为了准确识别用户进餐意图,采用口腔纵横比 (mouth aspect ratio, MAR) 识别用户嘴部状态进而分析用户进餐意图。当进餐意图被确定后,利用眼动追踪选餐意图识别算法,分析用户眼动偏航数据推断菜品所在位置,从而实现菜品选择意图识别。由于用户在每次进食中,进食点空间位置存在动态变化,因此基于动态喂食点识别确保系统实现精准的喂食行为。机械臂末端执行器将按照规划路径舀取食物并向用户递食物。同时机械臂末端执行器配有六维力传感器,在喂食过程中,可检测机械臂将食物送入口部并产生口部触觉交互过程的安全性。咀嚼过程采用下颌纵横比 (jaw aspect ratio, JAR) 可以有效刻画进食过程面部动态变化,通过数据分析判断用户是否完成咀嚼动作。当用户咀嚼完成后,系统再次检测用户的用餐意图,从而反复辅助喂食行为。如果未检测到嘴部动作,系统将判定用户无需继续进餐,从而结束整个辅助喂食流程。

1) 基于 MAR 动态表征的进餐意图识别

鉴于本框架服务于行动受限的半失能与失能老人,因此人机交互方式设计简单直观,避免复杂操作。考虑到张嘴和闭嘴是进食过程中最基本且自然的动作,因此采用这些日常动作作为传递进食意图的信号。采用基于面部特征点的嘴部状态检测算法,利用 FaceMesh 算法实现人脸及人脸 468 个关键点检测。通过获取嘴唇轮廓 12 个关键点信息计算口部纵横比 MAR 值,精确量化嘴部的张合状态,如图 4 所示。

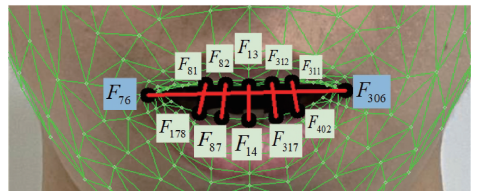


图 4 嘴唇外轮廓关键点

Fig. 4 Schematic diagram of key points on the outer lip contour

分析受试者嘴部的实时运动特征,即嘴部从闭合过渡到张开的状态,上、下唇之间的间距出现了明显的增加,但是两嘴角特征点关键点间距相对稳定,未见明显变化。因此,用户采用张嘴动作触发主动进食意图,结合嘴部特征点纵横比设计 MAR 的计算公式可表达为:

$$MAR = \frac{\|F_{13} - F_{14}\| + \|F_{82} - F_{87}\| + \|F_{312} - F_{317}\|}{5 \times \|F_{76} - F_{306}\|} + \frac{\|F_{81} - F_{178}\| + \|F_{311} - F_{402}\|}{5 \times \|F_{76} - F_{306}\|} \quad (1)$$

其中, $F_i (i = 0, 17, 37, \dots)$ 表示选取 468 个关键点中嘴部外廓的特征点, i 为特征点标号,分母表示 5 倍嘴角

之间的距离。分子表示上下嘴唇5对特征点垂直距离。为确保在辅助喂食过程中,准确地识别用户进食意图,传达出用户主动进食想法,既要准确识别嘴部张开状态,又要避免因说话而导致的误检。为此,设定一个恰当的MAR阈值尤为重要。针对MAR阈值设定的过程需要结合不同用户的张口习惯,因此需提前采集一段用户主动张口视频,并根据面部特征值设定MAR阈值,同时匹配固定的持续时间窗,具体分析过程见3.1节第1部分。系统预设好用户的MAR阈值,受助者只需张嘴就能够触发用户主动用餐意图。当捕捉到用餐意图后,选餐界面中“开始选餐”按流程被激活并且会触发食物检测环节,对桌面食物进行实时检测通过选餐界面进行显示。进食意图识别方法如算法1所示。

算法1 进食意图识别算法

```

Img_color=get_aligned_images(),获取相机捕捉图像
Results=faceMesh.process(Img_color),分析图片,获得分析结果。
if results:
    Img,Faces=findFaceMesh(Img_color),在图上绘制检测信息,并收集全部的面部特征点。
    mouth_point,MAR=get_mouth_info(Faces),提取12个嘴部关键点信息,并计算MAR值。
    if MAR>=MAR_threshold:如果满足条件则说明用户有用餐意图。
        Order_food_QPushButton(purple),激活选餐界面中“开始选餐”按钮。
        start_foodmouth(),触发食物检测。
end

```

2) 菜品选择意图识别

系统检测到用户用餐意图就会触发食物检测,通过机械臂搭载的深度相机获取食物信息显示在选餐界面中,让卧床受助者可以更方便的选择食物。为了实现精确且简单的选餐,采用实时追踪眼动的方法实现餐食的选定的意图识别。通过智能床上方摄像头捕捉受助者眼球运动直接操控界面光标。该光标在某个菜单选项上静止一定时间后,系统将自动识别出受助者的选择意图,并相应地激活选中的食物项,并结束选餐程序。这种方法为受助者提供了明确直观的反馈,允许在选择过程中通过语音随时自由中断,确保了操作灵活性和控制自主性。在本研究中,在输出凝视角度的基础上结合的头部姿态,进一步精确计算视线方向,从而实现在屏幕上准确指示光标位置,即“眼睛看到哪,光标就指到哪”的效果^[14-15]。具体而言, \mathbf{R}_{head} 是头部综合旋转矩阵,如式(2)所示。 \mathbf{V}_{head} 是头部综合旋转

矩阵应用得到的视线方向,如式(3)所示。对视线矢量 \mathbf{V}_{head} ,进一步应用眼球的旋转角度,得到眼部视线矢量 \mathbf{V}_{eye} ,如式(4)所示。

$$\mathbf{R}_{head} = \begin{bmatrix} \cos(Ry_{head}) & 0 & \sin(Ry_{head}) \\ 0 & 1 & 0 \\ -\sin(Ry_{head}) & 0 & \cos(Ry_{head}) \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(Rx) & -\sin(Rx) \\ 0 & \sin(Rx) & \cos(Rx) \end{bmatrix} \times \begin{bmatrix} \cos(Rz_{head}) & -\sin(Rz_{head}) & 0 \\ \sin(Rz_{head}) & \cos(Rz_{head}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$\mathbf{V}_{head} = \mathbf{R}_{head} \times \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \quad (3)$$

$$\mathbf{V}_{eye} = \begin{bmatrix} \cos(Ry_{eye}) & 0 & \sin(Ry_{eye}) \\ 0 & 1 & 0 \\ -\sin(Ry_{eye}) & 0 & \cos(Ry_{eye}) \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(Rx_{eye}) & -\sin(Rx_{eye}) \\ 0 & \sin(Rx_{eye}) & \cos(Rx_{eye}) \end{bmatrix} \times \mathbf{V}_{head} \quad (4)$$

通过测量卧床时头部到屏幕的视距 $D=0.7\text{ m}$,并保证选餐的准确性将显示食物窗口定为光标移动区。在物理空间中,计算视线与屏幕的交点坐标 (C_x, C_y) ,如式(5)所示,最后将物理坐标映射到屏幕的像素坐标 (P_x, P_y) ,如式6所示。这样就可以通过捕捉受助者眼球运动直接操控界面光标进行选餐。

$$(C_x, C_y) = \left(V_{eye_x} \times \frac{-D}{V_{eye_z}}, V_{eye_y} \times \frac{-D}{V_{eye_z}} \right) \quad (5)$$

$$\begin{cases} P_x = \left(\frac{C_x}{width} + 0.5 \right) \times W_{pixels} \\ P_y = \left(0.5 - \frac{C_y}{height} \right) \times H_{pixels} \end{cases} \quad (6)$$

此外,为了实现食物的选定引入 $\rho(W)$ 密度,在一定时间内计算注视在餐食上点的密度,密度最高的食物即为受试者可能选择的项。密度计算公式为:

$$\rho(W) = \frac{N(W)}{m_{00}} \quad (7)$$

其中, $\rho(W)$ 为食物分割掩膜轮廓中注视点的密度, $N(W)$ 表示点 W 的分割掩膜内所有注视点的集合,点 W 为餐食分割掩膜轮廓中心点 (m_x, m_y) , m_{00} 为餐食分割掩膜面积。

3) 动态喂食点估计与递送位姿计算

大部分现有助餐机器人通过预设固定喂食点的方式需要用户主动迎取食物进而影响用户使用舒适度,

对于自主摆动头部能力较弱的用户来说并不友好,为此提出根据用户动态喂食点估计与跟踪的方式为用户提供舒适的喂食策略。动态喂食点位置估计与跟踪方法通过对用户嘴部中心点内侧安全距离有效估计同时,根据面部朝向姿态进行实时位姿补偿,进而提供一种便捷的随动式辅助喂食策略。系统基于 FaceMesh 算法输出面部特征图,识别并对齐 468 个关键点。为了实现能够精准的将食物喂至嘴中,选取上嘴唇中心关键点 F_{13} 与下嘴唇关键点 F_{14} ,求解加权平均坐标 F_{mouth} ,经过多次试验,权重分别为 0.75 和 0.25,贴近嘴唇的中心点,如式(8)所示。

$$\begin{cases} F_{mouth}^x = F_{14,x} \times 0.75 + F_{13,x} \times 0.25 \\ F_{mouth}^y = F_{14,y} \times 0.75 + F_{13,y} \times 0.25 \end{cases} \quad (8)$$

通过机械臂末端 RGB-D 深度相机对失能者面部关键点信息捕捉,将二维像素坐标 $(F_{mouth}^x, F_{mouth}^y)$ 映射到三维空间坐标系,得到喂食点基于相机的三维坐标 $(F_{camera,x}, F_{camera,y}, F_{camera,z})$ 。由于喂食执行机构是机械臂,因此通过式(9),基于相机坐标系的喂食点三维坐标转换为机械臂基坐标系的三维坐标 $(x_{mouth}, y_{mouth}, z_{mouth})$ 作为喂食点。

$$P_{base} = {}_B^C T \cdot P_c \quad (9)$$

其中, P_c 是基于相机坐标系的喂食点坐标, ${}_B^C T$ 是相机坐标到机械臂基坐标转换矩阵, P_{base} 是将经过坐标变换后基于机械臂基坐标系的喂食点。常规状态下,弱机能人体由于躺卧在智能床上,头部位姿变化较小。当用户在喂食过程中自主调整头部位姿发生较大的动态变化,会导致喂食点位置也会发生相应的动态变化。因此选择合适的递送位姿尤为重要,需要考虑人体脸部朝向,以适当的偏移角度控制勺子递送食物并递送口中。基于面部轮廓特征点,通过 PnP (perspective-n-point, PnP) 算法确定面部位姿。找到映射关系来匹配面部三维模型和面部关键点,为机械臂末端提供精准的喂食点与喂食姿态,其映射关系表示为:

$$p_i = K[R|t]P_i \quad (10)$$

其中, P_i 为面部模型中的三维点, p_i 为图像坐标系下的面部关键点, K 为相机内参矩阵, R 和 t 为摄像机的旋转矩阵与平移矩阵。从而得到面部在三维空间中的姿态。为保障受试者安全,在 Z 轴方向增设一定安全距离为预喂食点,使饭勺先以到达预喂食点为目标再根据用户喂食点与面部朝向进行动态跟进。当喂食点稳定 1 s 后,系统将根据面部朝向自动调整饭勺姿态并缓慢靠近喂食点。将头部正面方向矢量旋转矩阵 R_{head} ,即以面部朝向作为基准,并基于该基准沿 Y 轴旋转一定角度作为控制勺子递送食物的实际角度,例如可以根据用户要求旋转 30° 将勺子以侧向方式递送至口部。这种方式可以避免食物以朝向脸部正方向的方式直接递送给用户带来

的不适性并提升用户心理舒适性^[16]。

4) 咀嚼行为与咀嚼完毕识别

在受助者开始咀嚼用餐的同时,系统实时监测受助者的用餐状态,判断是否完成了一次进食。咀嚼这一动作是由嘴部肌肉收缩产生的力量推动下颌骨运动而形成的,表现为嘴巴轻微张开和闭合,从而改变面部形状。通过综合分析嘴部和面部轮廓的动态特征,可以准确判断受助者的咀嚼动作。本研究采用特征量化指标 JAR ,这一参数能够有效地捕捉到咀嚼活动中的关键细微动作。首先定义关键点来捕捉面部结构的核心变化,包括鼻尖和下颌等位置,以确保能够全面而精确地捕捉脸部的整体形态和下颌骨的运动轨迹。它们在不同个体间显示出一致性,而且不易受到面部表情肌肉活动的干扰,从而保证了咀嚼检测的普遍适用性和稳定性。这一指数是监测下颌运动情况的有效工具,进一步提升了对咀嚼动作判断的准确度。其中, JAR 值计算公式为:

$$JAR = \frac{\|F_1 - F_{175}\| + \|F_{218} - F_{140}\| + \|F_{438} - F_{369}\|}{3 \times \|F_{192} - F_{416}\|} \quad (11)$$

式中: F_{192} 、 F_{416} 表示嘴角外的 2 个关键点; F_1 、 F_{218} 、 F_{438} 表示鼻尖的 3 个关键点; F_{175} 、 F_{140} 、 F_{369} 表示下颌的 3 个关键点。面部外轮廓关键点示意图如图 5 所示。

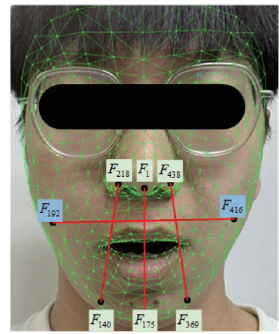


图 5 面部外轮廓关键点

Fig. 5 Schematic diagram of key points on the facial contour

式(11)中,分母表示嘴角关键点宽度的 3 倍,分子是鼻尖 3 个关键点到下颌 3 个关键点的距离之和,该公式可以描述了咀嚼过程中下颌的开合程度。前期实验通过对多人咀嚼过程中 JAR 值发现,由于不同用户的面部肌肉分布和咀嚼习惯不尽相同, JAR 值的也具有一定差异。因此,不能以固定阈值来判断咀嚼动作。为了能够准确识别咀嚼动作,提出了一种结合动态阈值的咀嚼识别算法。在咀嚼过程中 JAR 值从“谷值 - 峰值 - 谷值”的波动较大,因此引入滑动窗口,计算窗内标准差来判断 JAR 值在窗口内的波动,每当有新数据到达时,会在一个固定大小的窗口 w 内计算均值 μ 和标准差 σ ,如式(12)所示,并设定一个合理的阈值来判断是否为咀嚼状态。

阈值设定如式(13)所示,如满足该式则说明发生了咀嚼动作,其中阈值使用均值加 k 倍标准差来设定。

$$\begin{cases} \mu = \frac{1}{w} \sum_{i=1}^w JAR_i \\ \sigma = \sqrt{\frac{\sum_{i=1}^w (JAR_i - \mu)^2}{w}} \end{cases} \quad (12)$$

$$\begin{cases} threshold = \mu + k \times \sigma \\ JAR > threshold \end{cases} \quad (13)$$

2.2 辅助喂食模糊意图反馈机制

在喂食全过程中,针对用户个体差异性产生的临时意图更改、模糊意图等不确定意图和紧急呼叫诉求会导致交互系统无法正常工作,进而为辅助喂食过程带了中断甚至是风险性。针对用户的个性化需求或基于视觉指令机器人技术具有局限性,为提供更具备安全性和灵活性的服务,将辅助喂食流程系统结合大语言模型对交互过程中的临时意图、模糊意图以及紧急状态进行提问,并采用 Coppeliassim 软件搭建喂食流程虚拟映射系统,进而实现喂食系统在识别用户临时意图、模糊意图时可通过语言交互反馈提高效率,在通过辅助喂食系统与映射系统协同下提高辅助喂食全过程的安全性。辅助喂食系统结合大语言模型工作框架图如图6所示。

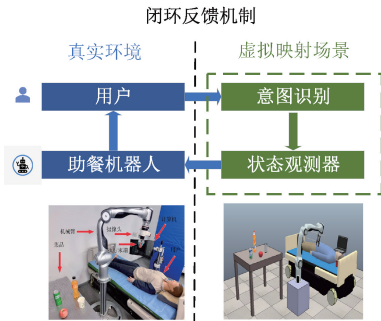


图6 模糊意图反馈框架

Fig. 6 Framework diagram of fuzzy intention feedback

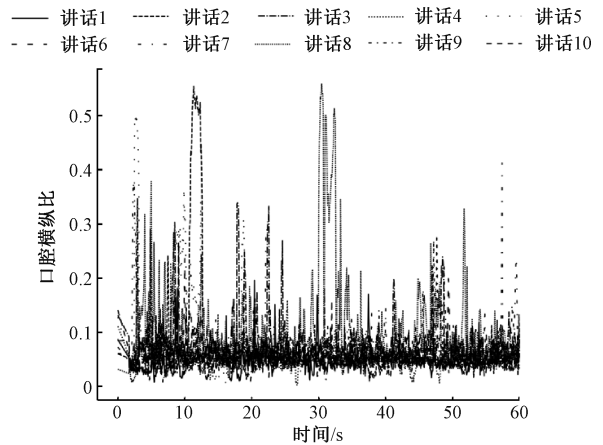
当机器人与用户交互时,系统通过不断地将来自各种用户意图作为反馈源的信息注入机器人规划系统。当模糊意图、临时更改意图出现时,采用大型语言模型 (large language model, LLM) 规划语言提示中来制定反馈机制。使用启示功能,结合用户语言的清晰捕捉,对 LLM 进行预测至关重要,生成可由机器人执行的高效计划^[17-18]。然而,在这种情况下,单一的使用视觉感知或单一的采用 LLM 方法仍然是单向的提供一系列技能,而不进行纠正或相应地重新规划。相比之下,反馈机制研究的是以闭环方式直接向 LLM 提供用户模糊意图反馈与临时诉求。这促进了在复杂的长期环境中改进任务规划。

3 实验与分析

3.1 辅助喂食意图识别实验

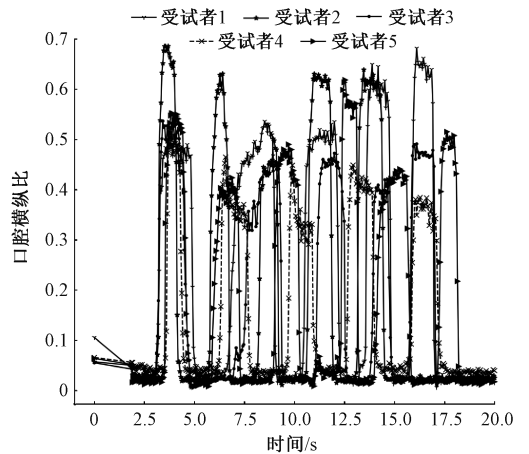
1) 进食意图识别实验

在进食过程中,用户主动张嘴代表想要进食,同时结合不同用户差异,设定一个恰当的 MAR 阈值尤为重要。首先,在分析相关阈值设定方面,提前采集并分析了不同用户日常生活常规说话数据与张口数据,分析相应场景下的 MAR 数据值。其中,列举 10 段短话 MAR 值及其分析结果显示,在说话时 MAR 值以波动较大的方式呈无规则状态,但在绝大多数情况下, MAR 值未超过 0.2,仅在极少数个别发音时出现短时帧数据超过 0.23 的情况,如图 7(a)所示。同时受试者进行一系列嘴部从完全闭合到稍大幅度张开的连续动作,以采集一段连续的“张开-闭合”状态下的 MAR 数据,如图 7(b)所示,可以发现 MAR 峰值变化范围大致在 0.26~0.54 之间。这一对比



(a) 说话时MAR变化值

(a) MAR variation value during speaking



(b) 连续张嘴MAR变化值

(b) MAR variation during continuous mouth opening

图7 MAR 变化曲线对比分析

Fig. 7 Comparative analysis of MAR variation curves

数据为设定阈值提供了重要参考,不同用户可根据相应的数值设定 MAR 阈值,从而更准确区分说话和进食意图动作,保证实时主动意图捕捉并避免误识别。

由于受试者说话时的 MAR 值波动较大且考虑到用户长时间保持嘴部大幅度张开可能会导致嘴部不适感,因此不宜将 MAR 阈值设置过高,为此结合滑动窗内的连续帧检测机制,只有在用户张嘴过程窗中的连续帧中全部满足条件时才判定为“张嘴”,从而有效提升进食意图识别精确性。列举 5 名受试者说一段话,字数 N_w 。在启动进食意图识别程序后,受试者开始发音讲述,同时计算并记录不同检测帧数时识别出的代表进食意图的张嘴次数 C_M 。最终,计算误检率(false detection rate, FDR),如式(14)所示:

$$FDR = \frac{C_M}{N_w} \times 100\% \quad (14)$$

实验数据如表 1 所示,由实验数据可知当 MAR 阈值设为 0.264 并且检测窗口设置为 10 帧时,实验未发生误检,其中部分志愿者意图识别效果如图 8 所示。

表 1 测试数据的误检率

Table 1 Error detection rate of test data

帧数	MAR 值	指标	受试者 1	受试者 2	受试者 3	
9 帧	0.262	N_w	68	124	178	
		C_M	5	8	12	
		$FDR/\%$	7.35	6.45	6.74	
	0.263	N_w	68	124	178	
		C_M	3	7	11	
		$FDR/\%$	4.41	5.64	6.17	
	0.264	N_w	68	124	178	
		C_M	1	2	8	
		$FDR/\%$	1.47	1.61	4.49	
	10 帧	0.262	N_w	68	124	178
			C_M	3	6	10
			$FDR/\%$	4.41	4.83	5.61
0.263		N_w	68	124	178	
		C_M	0	1	2	
		$FDR/\%$	0.00	0.80	0.11	
0.264	N_w	68	124	178		
	C_M	0	0	0		
	$FDR/\%$	0.00	0.00	0.00		

将连续检测窗口设为 10 帧, MAR 阈值设为 0.264, 在一定时间段内,对 10 名受试者进行以张嘴为信号的进食意图检测。实验数据如表 2 所示,其中识别的平均准确率达到 97.248%。



图 8 用餐意图检测

Fig. 8 Eating intention detection

表 2 10 名志愿者进食意图识别结果

Table 2 Eating intention recognition results for 10 volunteers

受试者	张嘴次数	识别次数	准确率/%
1	124	122	98.38
2	147	142	96.59
3	79	79	100.00
4	125	124	99.20
5	164	161	98.17
6	187	181	96.79
7	147	142	96.59
8	321	312	97.19
9	109	102	93.57
10	100	96	96.00

2) 选餐意图识别实验

本部分对基于眼动跟踪的菜品选择意图检测方案进行实验。系统检测到用户用餐意图后,控制机械臂到达捕捉桌面食物的基础位置,通过机械臂末端的深度相机获取食物图像信息并显示在选餐界面中,最后通过智能床上方的显示屏实时显示桌面食物,让卧床用户可以更方便直观地选择食物。为了实现精确且直观的选餐与结束选餐意图的识别,采用实时眼动追踪方法,通过智能床上方摄像头捕捉受助者眼神凝视点运动直接操控界面光标,其中用户凝视点识别方向如图 9 所示。选择菜品场景与选择菜品交互界面如图 10 所示。为了增强交互性并提升受助者体验,设计了一个交互式光标。该光标在某个菜单选项上静止一定时间后,系统将自动识别出用户的选餐意图,并相应地激活选中的食物项或是结束选餐程序。这种方法为用户提供了明确且直观的反馈,允许在选择过程中随时自由中断,确保操作的灵活性和控制的自主性。



图9 凝视点识别

Fig. 9 Gaze point recognition

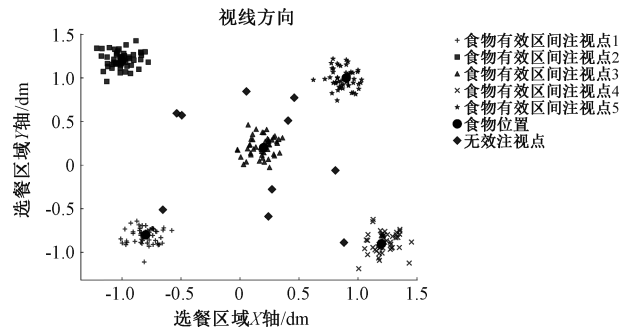


图11 选餐意图识别的注视数据

Fig. 11 Gaze data of meal selection intention recognition



(a) 选择菜品场景
(a) Scene of selecting a dish



(b) 选择菜品交互界面
(b) Interactive interface of selecting a dish

图10 选餐意图识别

Fig. 10 Identification of meal selection intention

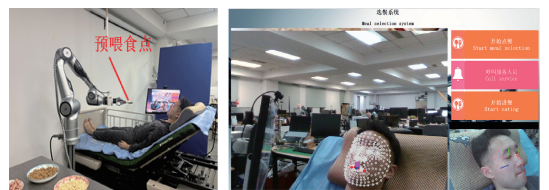
为确保选餐意图识别过程中信息的直观表达,采用了颜色编码方式记录用户实时动态视线数据,其中圆心点代表食物中心坐标,是交互系统中选餐位置参考值。

其他形状点代表用户在用餐全程中视线追踪点集。当用户注视点偏离食物选择的有效边界,动态点标记为无效注释点,表征潜在的误操作或系统的异常检测点,用户凝视选餐区域不同食物的数据示意图如图11所示。能够有效地捕捉用户选餐意图并提升交互过程的直观性,从用户感官角度也间接提升交互过程中意图识别准确性。

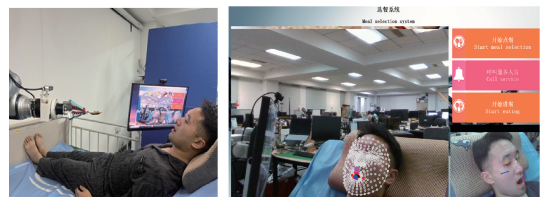
3) 喂食点识别实验

本部分对喂食过程中喂食点与喂食姿态进行计算。基于机械臂规划的舀取轨迹和递送轨迹,首先机械臂将食物递送至预喂食点如图12(a)所示。识别用户喂食点的同时准备以最佳进餐姿势提供给用户。此时,系统通

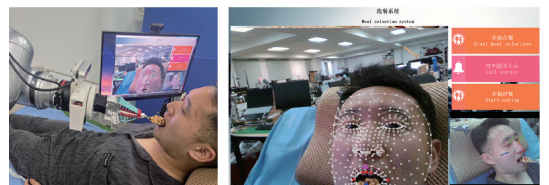
过深度相机实时捕捉受助者的用餐意图和喂食点位置如图12(b)所示。当准备好进餐时,仅需张嘴即可触发用餐意图识别。根据捕捉到的喂食点位置,结合用户舒适的便宜角度,机械臂通过自适应递送算法,将食物送至受助者嘴中,如图12(c)所示。结合2.1节第3部分,以面部朝向作为基准,并基于该基准沿Y轴旋转一定角度作为控制勺子递送食物的实际角度,在本实验中,可以根据用户要求旋转30°将勺子以侧向方式缓慢递送至口部,避免食物以朝向脸部正方向的方式直接递送给用户带来的压迫性。



(a) 预喂食点
(a) Pre-feeding points



(b) 食物递送中喂食点识别
(b) Identification of feeding points during food delivery



(c) 递送食物过程
(c) Process of delivering food

图12 喂食点识别

Fig. 12 Identification of feeding point

4) 咀嚼意图识别实验

根据前期咀嚼实验发现,由于不同用户脸部差异与咀嚼习惯导致咀嚼过程中 JAR 值的具有差异性。为此选择最优的 w 与 k 值,来刻画咀嚼过程中 JAR 值从“谷值 - 峰值 - 谷值”的变化。为了验证所提出的咀嚼意图识别算法,并确定 w 与 k 值,采集了多位受试者咀嚼食物的视频,并对咀嚼状态下的下颌宽高比 JAR 进行了详尽分析。其中列举一位测试者咀嚼识别数据如图 13(a)、(b) 所示,

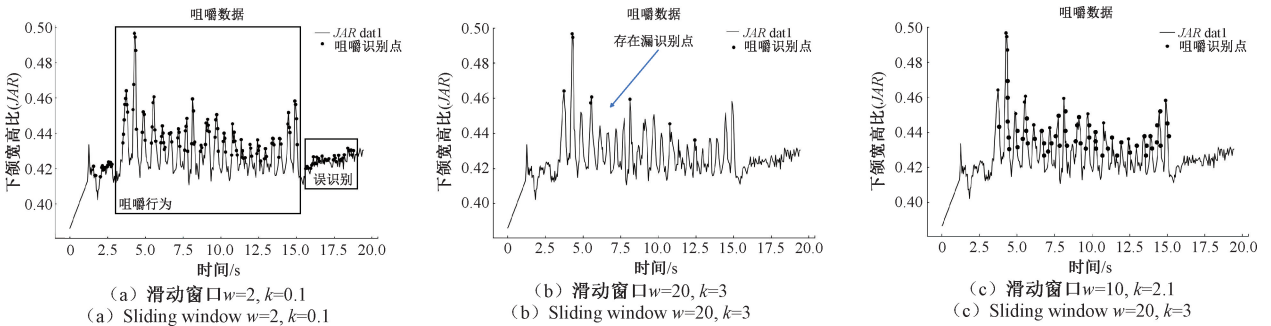


图 13 JAR 数据阈值分析

Fig. 13 Analysis of JAR data threshold

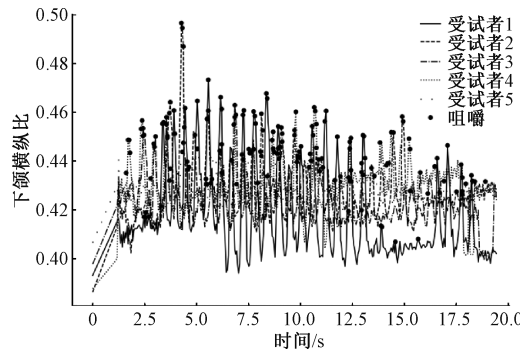


图 14 咀嚼检测

Fig. 14 Chewing detection

5) 嘴部交互力实验

为了增强用户在进餐过程中的安全性,保障喂食执行器与用户口腔接触力的柔性,对食物送入嘴中并咬取的嘴部触觉力交互实验。如图 15 所示记录了 2 次进食过程中,口部对末端执行器的触觉力。六维力传感器在喂食执行器未舀取食物时测量的力为 0 N。1.0~3.2 s 间,在喂食执行器舀取食物后,将食物送入用户口腔中并进行咬合,此时执行器与用户的口腔会产生短时间的交互触觉力。当六维力传感器检测到食物被纳入口腔并交互后,机械臂以 0.002 m/s 的速度缓慢抽离移动至预喂食点。最后,喂食执行器完成送餐任务,采集到六维力传感器的相关参数短暂骤减后恢复到 0 N。通过该实验发现,触觉力可以反应口部咬合实物的过程,虽然交互力信号呈现无规则运动,但该信号起到喂食过程开始至送

当滑动窗口大小 w 和阈值系数 k 较小时,未咀嚼状态可能会被误判为咀嚼状态,而当 w 和 k 较大时,实际的咀嚼状态可能会被漏检。经过多次实验,在 30 Hz 的 JAR 采样频率下,滑动窗口大小 $w = 10$ 和阈值系数 $k = 2.1$ 时,咀嚼状态的误检率最低,如图 13(c) 所示。该算法适用于不同咀嚼习惯的用户,如图 14 所示。通过结合 w 与 k 值所设计 JAR 域值能够很好地适应每个受试者,减少误判的发生,从而实现高识别率与泛化性。

入口中的喂食事件过程描述效果,同时驱动喂食的下一步骤。当完成咬合实物后,机械臂缓慢退回至预喂食点,进一步优化喂食过程的安全性。

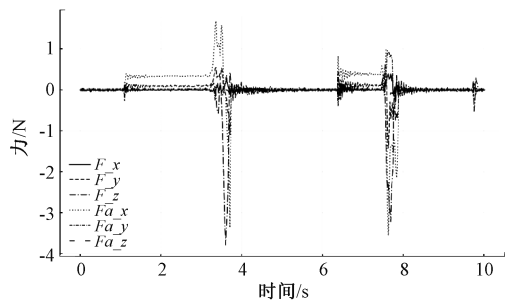


图 15 末端执行器受力变化曲线

Fig. 15 Force change curve of end effector

3.2 辅助喂食模糊意图反馈及交互评估实验

在实验中,机器人辅助喂食任务的人机交互安全的主要挑战之一在于应对用户与机器人的交互过程中依赖单一意图识别方法进行交互的局限性。当依靠视觉传感器进行进餐意图识别和选餐意图识别时,用户的个体差异性行为幅度不同,持续时间过短,临时改变意图等条件,可能会导致识别精度不高从而导致系统做出违背用户意图的行为。为解决上述喂食场景中存在的矛盾和挑战,采用 LLM 的深入理解与规划能力用于确认用户的真实意图,形成基于辅助喂食全过程意图识别的多模态安全交互方法。如图 16(a) 所示,为模糊意图识别,由于用户个体差异性导致几次张嘴的动作较小,无法直接触发

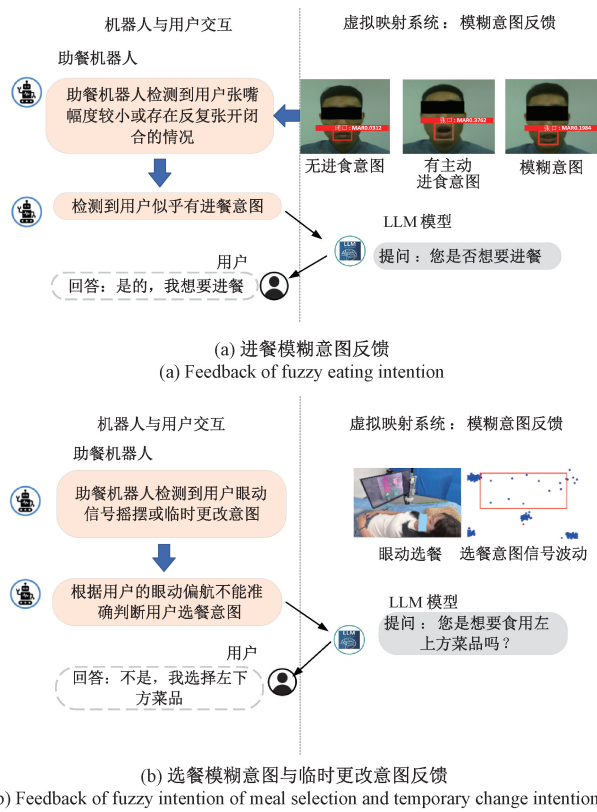


图 16 模糊意图与临时更改意图询问反馈交互场景
Fig. 16 Interactive scene of fuzzy intention and temporary change intention inquiry and feedback

喂食功能,因此结合大语言模型对用户收否有进食意图进行识别与确认。如图 16(b)所示,在用户进食过程中,存在临时意图更改,结合大语言模型对用户进行意图询问从而确定用户的最终诉求。图 17 展示了辅助喂食机器人与用户在交互场景中识别模糊意图时的交互过程,直观地展示了这种集成解决方案的优越性。对辅助喂食实验的人机交互性进行了评估,分别对 9 名行动不便者在辅助喂食过程中的成功性,完成时间,易用性,舒适性与安全性进行评估,采用 5 点李克特量表分别对上述指标进行打分;5 代表非常满意,4 代表满意,3 代表一般,2 代表不满意,1 代表很不满意。评估结果如表 3 所示。根据评估结果,发现受试者已经成功得到辅助进食服务并相信能够成功地用辅助喂食机器人系统完成进食行为的得分为 4.78,满分为 5。表明通过使用送餐机器人系统可以增强他们的感知能力并提供给用户长期使用的信心。在完成时间方面,受试者的平均得分为 3.89 分。在使用该机器人系统初期,用户需要进行一段时间的适应性训练,包括采用面部特征点驱动喂食各个环节的熟悉与训练,同时对交互界面各部分功能的熟悉使用。虽然受试者受限于手部灵活操作,但在视觉交互与语言表达方面不存在问题,因此可以通过阶段性训练完成送餐各环节的驱动,对于语言交互障碍或存在视觉交互障碍的用户并不适用。同时,在完成时间方面,机器人规划与运动速度起到了重要作用,然而对于受试者来说,过



图 17 辅助喂食实验过程
Fig. 17 Experimental process of assisted feeding

表 3 辅助喂食交互方法评估量表
Table 3 Evaluation scale for assisted feeding interaction methods

评估指标	分数	
	平均值	标准差
成功受助	4.56	0.89
完成时间	3.89	0.93
易用性	4.22	0.83
安全性	4.78	0.67
舒适性	4.33	0.71

以根据用户需求设定速度即可。在系统易用性与安全性方面,结果得分分别为 4.22 和 4.56,证明系统安全且易于使用。在安全性方面,受试者普遍认为设定通过语言的紧急报警、意图更改方式让用户主观感觉到安全保障,会提醒护理人员进一步辅助。相比于采用紧急按键方式来说更加便捷,特别对于手部无行动能力用户,语音交互触发安全保障机制是可靠性选择。然而,少量参与者认为被体积巨大的协作机器人辅助过程中,特别是当协作机器人靠近面部时感到了威胁与心里不适。为此,调整了喂食姿态,确保机械臂根据用户需求以倾斜一定角度缓慢靠近喂食点,避免给用户带了心理压力。在易用性方面,本系统采用独特的视觉主导 GUI 交互界面,表明视觉触发的方式对设备的输入是高效且简单的。方法具有

快的移动速度对用户主观舒适性较差,在实际应用中可

较低工作负荷,无需反复语言交互或操作进食按键,符合前期研究相关结论^[16]。在舒适性方面,用户给出平均得分为 4.22 分,其中辅助进食策略及其规划方法、人机交互方式甚至机器人外观均起到了影响,随着使用的熟练性增加,用户相应舒适性会有所提高。

本研究提出的多模态交互方法及其安全保障机制与其他送餐机器人对比如表 4 所示,目前很多研究都采用了基于视觉引导的方式,因为这种交互方式符合辅助进食过程中灵活、方便的自然交互规律。再此基础上,针对末端执行器与口部交互过程建立了触觉感知方法,针对模糊意图与风险行为建立了语言交互机制。同时,针对在床用户,分别采用两台摄像头实现人体状态、取餐与递送的识别,避免单一机械臂端摄像头反复摆动、目标离开视野等问题造成用户的心里不适。在整个实验过程中,送餐机器人系统成功地为用户提供一种易于使用的进食策略。

表 4 送餐机器人交互方式与安全保护机制对比

Table 4 Comparison of interaction modes and safety protection mechanisms for assisted feeding robots

助残机器人	交互方法	安全保护机制
Robot-assisted feeding system ^[5-6,10]	基于视觉交互的 GUI	执行监视器/停止按键
ADA assisted feeding robot ^[7]	视觉与触觉	三级安全保障策略: 机械限位/碰撞 停止/护理人员
My Spoon/Obi ^[8]	固定喂食位置	按键/操纵杆
Real-Life Meal-Assist Robot ^[9]	脑机接口	按键/操纵杆
I-feed ^[13]	语言交互	按键
本文提出的 送餐机器人	基于视觉、触觉、力、 位与语言多模态交互	基于语音交互的保障机制

4 结 论

本研究提出一种基于辅助喂食全过程意图识别的多模态安全交互方法。该方法是融合了视觉、触觉、力、位与语言的多模态人机交互框架,从而可柔顺地辅助用户进餐,提升辅助喂食过程的灵活性。其中,提出一种喂食意图全过程识别方法,该方法通过对进餐意图识别、菜品选择意图识别、动态喂食点估计、咀嚼意图进行识别,完成自然的意图识别交互过程。

同时结合大语言模型形成对不确定意图交互反馈闭环,最终提升整体系统框架的安全性。实验结果表

明,采用基于辅助喂食全过程意图识别的多模态安全交互方法可以确保上肢不便人群在日常生活辅助喂食行为安全性前提下,有效的提升助餐过程的灵活性,有望被应用于医院、家庭等弱机能用户的日常生活辅助进食行为中来。

参考文献

- [1] ZHAO D H, SUN X W, SHAN B, et al. Research status of elderly-care robots and safe human-robot interaction methods [J]. *Frontiers in Neuroscience*, 2023, 17: 1291682.
- [2] DING H, YANG X J, ZHENG N N, et al. Tri-Co Robot: A Chinese robotic research initiative for enhanced robot interaction capabilities [J]. *National Science Review*, 2018, 5(6):799-801.
- [3] MASHRUR T, GHULAM Z, FRENCH G, et al. Assistive feeding robot for upper limb impairment-testing and validation [J]. *International Journal of Advanced Robotic Systems*, 2023, 20(4):17298806231183571.
- [4] FUJII A, OKADA K, INABA M. Development of meal partner robot and applications towards enriching mealtime experience [C]. 2024 ACM/IEEE International Conference on Human-Robot Interaction, 2024:468-472.
- [5] PARK D, HOSHI Y, MAHAJAN H P, et al. Active robot-assisted feeding with a general-purpose mobile manipulator: Design, evaluation, and lessons learned [J]. *Robotics and Autonomous Systems*, 2020, 124:103344.
- [6] NANAVATI A, ALVES P, SCHRENK T, et al. Design principles for robot-assisted feeding in social contexts [C]. 2023 ACM/IEEE International Conference on Human-Robot Interaction, 2023: 24-33.
- [7] BHATTACHARJEE T, GORDON E K, SCALISE R, et al. Is more autonomy always better? Exploring preferences of users with mobility impairments in robot-assisted feeding [C]. 2020 15th ACM/IEEE International Conference on Human-robot Interaction, 2020:181-190.
- [8] ISHII S. Meal-assistance robot “my spoon” [J]. *Journal of the Robotics Society of Japan*, 21(4): 378-381.
- [9] HA J, PARK S, IM C H, et al. A hybrid brain-computer interface for real-life meal-assist robot control [J]. *Sensors*, 2021, 21(13):4578.
- [10] CHOI I, KO K, SONG H J, et al. A meal-assistance robot system for asian food and its food acquisition point estimation and user interface based on face recognition [J]. *Applied Sciences*, 2023, 13(5):3216.
- [11] 戴廷飞, 刘逸, 叶阳阳, 等. 人机共享控制机器人系统的应用与发展 [J]. *仪器仪表学报*, 2019, 40(3): 62-73.

DAI T F, LIU M, YE Y Y, et al. Application and

- development of human-machine shared control robot system[J]. Chinese Journal of Scientific Instrument, 2019, 40(3):62-73.
- [12] 唐心宇,宋爱国. 人体姿态估计及在康复训练情景交互中的应用[J]. 仪器仪表学报, 2018, 39(11):195-203.
TANG X Y, SONG A I G. Human pose estimation and its implementation in scenario interaction system of rehabilitation training[J]. Chinese Journal of Scientific Instrument, 2018, 39(11):195-203.
- [13] 刘策,刘小峰. 助老服务机器人多模式友好交互系统设计及实现[J]. 电子测量与仪器学报, 2019, 33(1):177-182.
LIU C, LIU X F. Design and implementation of multi-mode friendly interactive system for elderly assistance robot [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(1):177-182.
- [14] LEBLOND-MENARD C, ACHICHE S. Non-intrusive real time eye tracking using facial alignment for assistive technologies[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31:954-961.
- [15] KLAIB A F, ALSREHIN N O, MELHEM W Y, et al. Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies[J]. Expert Systems with Applications, 2021, 166:114037.
- [16] ZHOU L R, FENG ZH Q, WANG H Y, et al. MIUC: A human-computer collaborative multimodal intention-understanding algorithm incorporating comfort analysis[J]. International Journal of Human-Computer Interaction, 2024, 40(19): 6077-6090.
- [17] LIANG J, HUANG W L, XIA F, et al. Code as policies: Language model programs for embodied control[C]. 2023 IEEE International Conference on Robotics and Automation, 2023:9493-9500.
- [18] 张天骐,沈夕文,唐娟,等. 双分支复频谱下多特征聚合的轻量化语音增强方法[J]. 仪器仪表学报, 2024, 45(7):279-291.
ZHANG T Q, SHEN X W, TANG J, et al. A lightweight speech enhancement method based on dual branch complex spectrum with multiple feature aggregation[J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 279-291.

作者简介



吕相谊,沈阳工业大学硕士研究生,主要研究方向为养老护理机器人与人机交互。
E-mail:1437182545@qq.com

Lyu Xiangyi is a master student in Shenyang University of Technology. Her main research interests include elderly-care robot and human robot interaction.



赵东辉(通信作者),沈阳工业大学电气工程学院副教授、博士生导师,主要研究方向为智能机器人及其灵巧操作,养老护理机器人,具身智能等。
E-mail:putongdeyu@126.com

Zhao Donghui (Corresponding author) is an associate professor and Ph.D. supervisor at the School of Electrical Engineering, Shenyang University of Technology. His main research interests include intelligent robots and agile operations, elderly-care robots, embodied AI, etc.