



(12) 发明专利

(10) 授权公告号 CN 120715913 B

(45) 授权公告日 2025. 11. 18

(21) 申请号 202511222403.5

G06T 7/00 (2017.01)

(22) 申请日 2025.08.29

G06V 20/70 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 120715913 A

(56) 对比文件

CN 118744426 A, 2024.10.08

CN 118348983 A, 2024.07.16

(43) 申请公布日 2025.09.30

审查员 崔清雨

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区粤海街
道高新技术产业园南区粤兴一道18号
香港理工大学产学研大楼205室

(72) 发明人 郑湃 陈泓芑

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205

专利代理师 周翀

(51) Int. Cl.

B25J 9/16 (2006.01)

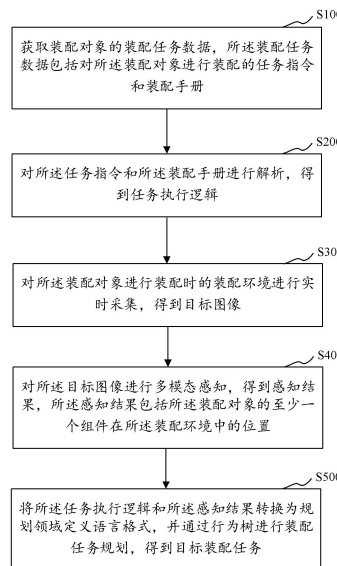
权利要求书3页 说明书19页 附图3页

(54) 发明名称

基于多智能体大模型的工业机器人自主装
配任务规划方法

(57) 摘要

本申请实施例提供了一种基于多智能体大模型的工业机器人自主装配任务规划方法,属于自动化装配技术领域。该方法包括:获取装配对象的装配任务数据,装配任务数据包括对装配对象进行装配的任务指令和装配手册;对任务指令和装配手册进行解析,得到任务执行逻辑;对装配对象进行装配时的装配环境进行实时采集,得到目标图像;对目标图像进行多模态感知,得到感知结果,感知结果包括装配对象的至少一个组件在装配环境中的位置;将任务执行逻辑和感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务。本申请实施例提升了工业机器人的装配效率和精度。



1. 一种基于多智能体大模型的工业机器人自主装配任务规划方法,其特征在于,所述方法包括:

获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册;

对所述任务指令和所述装配手册进行解析,得到任务执行逻辑;

对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像;

对所述目标图像进行多模态感知,得到感知结果,所述感知结果包括所述装配对象的至少一个组件在所述装配环境中的位置;

将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务;

所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务,包括:

将所述任务执行逻辑和所述感知结果分别进行数据格式转换,得到规划领域定义语言格式的新的任务执行逻辑和规划领域定义语言格式的新的感知结果;

根据新的任务执行逻辑和新的感知结果,生成行为树任务架构,所述行为树任务架构包括多个任务节点,所述任务节点包括顺序节点、条件节点和回退节点,所述顺序节点用于决定子任务的执行顺序,所述条件节点用于根据当前状态是否满足执行当前子任务的条件,所述回退节点用于在当前子任务执行失败时,重新执行当前子任务或重新规划所述装配任务;

根据所述行为树任务架构规划装配任务,得到所述目标装配任务。

2. 根据权利要求1所述的方法,其特征在于,所述对所述任务指令和所述装配手册进行解析,得到任务执行逻辑,包括:

通过视觉语言模型对所述任务指令和所述装配手册进行语义解析,得到各个组件的类别、各个组件的功能角色以及各个组件之间的空间结构关系;

根据各个所述组件的类别、各个所述组件的功能角色以及各个所述组件之间的空间结构关系,构建所述任务执行逻辑,所述任务执行逻辑包括多个子任务,每个所述子任务包括至少一个组件的安装描述信息,以及所述子任务的执行顺序。

3. 根据权利要求1所述的方法,其特征在于,所述对所述目标图像进行多模态感知,得到感知结果,包括:

对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像;

对每个所述装配对象的分割图像进行抓取位姿感知,得到每个所述装配对象的位姿信息,所述位姿信息包括各个所述装配对象的可抓取性评分和抓取位姿;

通过视觉语言模型对各个所述装配对象的分割图像和位姿信息进行语义分析,得到所述感知结果,所述感知结果包括各个所述装配对象之间的空间关系、各个所述装配对象的操作可行性和各个所述装配对象的属性。

4. 根据权利要求3所述的方法,其特征在于,所述对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像,包括:

对所述目标图像进行目标检测,得到所述目标图像中各个所述装配对象的边界框和初步类别标签;

根据各个所述装配对象的所述边界框和所述初步类别标签对所述目标图像进行图像分割,得到各个所述装配对象的分割图像,所述分割图像包括各个所述装配对象的轮廓边界和目标类别标签。

5. 根据权利要求1所述的方法,其特征在于,在所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务之后,所述方法还包括:

对所述装配对象进行装配时的装配环境进行拍摄,得到拍摄视频;

对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

6. 根据权利要求5所述的方法,其特征在于,所述对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行,包括:

通过视觉语言模型对所述拍摄视频进行图像分析,得到当前子任务执行结果;

获取对所述装配对象进行装配的装配设备的物理传感器数据,得到验证数据,所述验证数据包括抓手状态和力反馈数据;

通过所述验证数据对所述子任务执行结果进行验证,得到验证结果;

若所述验证结果指示所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

7. 一种基于多智能体大模型的工业机器人自主装配任务规划装置,其特征在于,所述装置包括:

获取模块,用于获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册;

解析模块,用于对所述任务指令和所述装配手册进行解析,得到任务执行逻辑;

采集模块,用于对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像;

感知模块,用于对所述目标图像进行多模态感知,得到感知结果,所述感知结果包括所述装配对象的至少一个组件在所述装配环境中的位置;

规划模块,用于将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务;将所述任务执行逻辑和所述感知结果分别进行数据格式转换,得到规划领域定义语言格式的新的任务执行逻辑和规划领域定义语言格式的新的感知结果;根据新的任务执行逻辑和新的感知结果,生成行为树任务架构,所述行为树任务架构包括多个任务节点,所述任务节点包括顺序节点、条件节点和回退节点,所述顺序节点用于决定子任务的执行顺序,所述条件节点用于根据当前状态是否满足执行当前子任务的条件,所述回退节点用于在当前子任务执行失败时,重新执行当前子任务或重新规划所述装配任务;根据所述行为树任务架构规划装配任务,得到所述目标装

配任务。

8. 一种电子设备,其特征在于,所述电子设备包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现权利要求1至6任一项所述的基于多智能体大模型的工业机器人自主装配任务规划方法。

9. 一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至6中任一项所述的基于多智能体大模型的工业机器人自主装配任务规划方法。

基于多智能体大模型的工业机器人自主装配任务规划方法

技术领域

[0001] 本申请涉及自动化装配技术领域,尤其涉及一种基于多智能体大模型的工业机器人自主装配任务规划方法。

背景技术

[0002] 随着工业4.0向工业5.0演进,制造模式正从以效率为核心的全自动化向以柔性协同为核心的人机共融系统转型,特别强调智能机器人在动态、多变环境中的自主任务执行与协作能力。在此背景下,工业机器人不仅需要具备基础的操作执行功能,还应具备语义理解、环境感知、任务推理与策略生成能力,以适应复杂、个性化和非结构化的生产需求。这种转变对机器人系统的智能水平提出了更高要求,尤其是在任务理解、多模态信息融合以及长时序规划等方面。

[0003] 然而,当前大多数工业机器人依然依赖静态编程与预定义流程驱动,缺乏对自然语言指令、图纸信息、任务说明书等非结构化输入的解析能力,导致任务的理解、分解与执行高度依赖人工配置,难以胜任复杂装配任务或变化频繁的生产流程。

发明内容

[0004] 本申请实施例的主要目的在于提出一种基于多智能体大模型的工业机器人自主装配任务规划方法,旨在解决现有的工业机器人装配依赖人工的技术问题,进而提升工业机器人的装配效率和精度。

[0005] 为实现上述目的,本申请实施例的第一方面提出了一种基于多智能体大模型的工业机器人自主装配任务规划方法,所述方法包括:

[0006] 获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册;

[0007] 对所述任务指令和所述装配手册进行解析,得到任务执行逻辑;

[0008] 对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像;

[0009] 对所述目标图像进行多模态感知,得到感知结果,所述感知结果包括所述装配对象的至少一个组件在所述装配环境中的位置;

[0010] 将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务。

[0011] 在一些实施例,所述对所述任务指令和所述装配手册进行解析,得到任务执行逻辑,包括:

[0012] 通过视觉语言模型对所述任务指令和所述装配手册进行语义解析,得到各个组件的类别、各个组件的功能角色以及各个组件之间的空间结构关系;

[0013] 根据各个所述组件的类别、各个所述组件的功能角色以及各个所述组件之间的空间结构关系,构建所述任务执行逻辑,所述任务执行逻辑包括多个子任务,每个所述子任务包括至少一个组件的安装描述信息,以及所述子任务的执行顺序。

- [0014] 在一些实施例,所述对所述目标图像进行多模态感知,得到感知结果,包括:
- [0015] 对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像;
- [0016] 对每个所述装配对象的分割图像进行抓取位姿感知,得到每个所述装配对象的位姿信息,所述位姿信息包括各个所述装配对象的可抓取性评分和抓取位姿;
- [0017] 通过视觉语言模型对各个所述装配对象的分割图像和位姿信息进行语义分析,得到所述感知结果,所述感知结果包括各个所述装配对象之间的空间关系、各个所述装配对象的操作可行性和各个所述装配对象的属性。
- [0018] 在一些实施例,所述对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像,包括:
- [0019] 对所述目标图像进行目标检测,得到所述目标图像中各个所述装配对象的边界框和初步类别标签;
- [0020] 根据各个所述装配对象的所述边界框和所述初步类别标签对所述目标图像进行图像分割,得到各个所述装配对象的分割图像,所述分割图像包括各个所述装配对象的轮廓边界和目标类别标签。
- [0021] 在一些实施例,所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务,包括:
- [0022] 将所述任务执行逻辑和所述感知结果分别进行数据格式转换,得到规划领域定义语言格式的新的任务执行逻辑和规划领域定义语言格式的新的感知结果;
- [0023] 根据新的任务执行逻辑和新的感知结果,生成行为树任务架构,所述行为树任务架构包括多个任务节点,所述任务节点包括顺序节点、条件节点和回退节点,所述顺序节点用于决定子任务的执行顺序,所述条件节点用于根据当前状态是否满足执行当前子任务的条件,所述回退节点用于在当前子任务执行失败时,重新执行当前子任务或重新规划所述装配任务;
- [0024] 根据所述行为树任务架构规划装配任务,得到所述目标装配任务。
- [0025] 在一些实施例,在所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务之后,所述方法还包括:
- [0026] 对所述装配对象进行装配时的装配环境进行拍摄,得到拍摄视频;
- [0027] 对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。
- [0028] 在一些实施例,所述对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行,包括:
- [0029] 通过视觉语言模型对所述拍摄视频进行图像分析,得到当前子任务执行结果;
- [0030] 获取对所述装配对象进行装配的装配设备的物理传感器数据,得到验证数据,所述验证数据包括抓手状态和力反馈数据;
- [0031] 通过所述验证数据对所述子任务执行结果进行验证,得到验证结果;

[0032] 若所述验证结果指示所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

[0033] 为实现上述目的,本申请实施例的第二方面提出了一种基于多智能体大模型的工业机器人自主装配任务规划装置,所述装置包括:

[0034] 获取模块,用于获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册;

[0035] 解析模块,用于对所述任务指令和所述装配手册进行解析,得到任务执行逻辑;

[0036] 采集模块,用于对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像;

[0037] 感知模块,用于对所述目标图像进行多模态感知,得到感知结果,所述感知结果包括所述装配对象的至少一个组件在所述装配环境中的位置;

[0038] 规划模块,用于将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务。

[0039] 为实现上述目的,本申请实施例的第三方面提出了一种电子设备,所述电子设备包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现上述第一方面所述的方法。

[0040] 为实现上述目的,本申请实施例的第四方面提出了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现上述第一方面所述的方法。

[0041] 本申请提出的基于多智能体大模型的工业机器人自主装配任务规划方法,其通过视觉语言模型对装配任务指令和操作手册等信息进行解析,提取任务对象、角色及空间关系,构建任务执行逻辑;通过多模态感知系统集对装配环境的三维图像进行感知,确保机器人对当前场景具备实时认知能力;利用大语言模型生成具备执行条件和回退策略的符号化行为树,每个叶节点对应机器人技能库中定义的技能,并根据节点返回状态动态调度;在执行过程中融合视觉和传感器数据进行多模态验证,通过视觉问答和物理传感器数据进行状态判断,实现执行监控和反馈调整。本申请通过构建任务层级图结构、多模态感知与符号映射、动态行为树规划及多模态验证机制,实现了非结构化任务指令的语义解析、环境状态的符号化表征以及动态任务调整能力,提升了自主决策能力,实现多模态信息融合闭环,增强了复杂装配任务动态调整能力。

附图说明

[0042] 图1是本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法的流程示意图;

[0043] 图2是本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法的逻辑示意图;

[0044] 图3是本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划装置的结构示意图;

[0045] 图4是本申请实施例提供的电子设备的硬件结构示意图。

具体实施方式

[0046] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本申请,并不用于限定本申请。

[0047] 需要说明的是,虽然在装置示意图中进行了功能模块划分,在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于装置中的模块划分,或流程图中的顺序执行所示出或描述的步骤。说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。

[0048] 除非另有定义,本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的,不是旨在限制本申请。

[0049] 随着工业4.0向工业5.0演进,制造模式正从以效率为核心的全自动化向以柔性协同为核心的人机共融系统转型,特别强调智能机器人在动态、多变环境中的自主任务执行与协作能力。在此背景下,工业机器人不仅需要具备基础的操作执行功能,还应具备语义理解、环境感知、任务推理与策略生成能力,以适应复杂、个性化和非结构化的生产需求。这种转变对机器人系统的智能水平提出了更高要求,尤其是在任务理解、多模态信息融合以及长时序规划等方面。

[0050] 然而,当前大多数工业机器人依然依赖静态编程与预定义流程驱动,缺乏对自然语言指令、图纸信息、任务说明书等非结构化输入的解析能力,导致任务的理解、分解与执行高度依赖人工配置,难以胜任复杂装配或变化频繁的生产任务。

[0051] 此外,现有系统中感知与决策模块割裂,难以将视觉、力觉等多源传感信息转换为可用的符号表示,用于构建任务模型或生成动态规划策略,严重制约了机器人在开放环境下的适应性与鲁棒性。

[0052] 尤其在面对多步骤、强依赖的长时序任务时,传统方法在任务结构建模、执行中断恢复与多任务切换等方面表现出明显不足,缺乏基于任务语义与环境状态动态调度的能力。

[0053] 基于此,本申请实施例提供了一种基于多智能体大模型的工业机器人自主装配任务规划方法,旨在通过多模态感知、符号任务建模与行为树执行机制实现复杂工业装配任务的结构化规划与稳健执行,提升机器人在复杂动态环境下的任务完成率与执行效率。

[0054] 本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法,具体通过如下实施例进行说明,首先描述本申请实施例中的基于多智能体大模型的工业机器人自主装配任务规划方法。

[0055] 本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法,涉及自动化装配技术领域。本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法可应用于终端中,也可应用于服务器端中,还可以是运行于终端或服务器端中的软件。在一些实施例中,终端可以是智能手机、平板电脑、笔记本电脑、台式计算机等;服务器端可以配置成独立的物理服务器,也可以配置成多个物理服务器构成的服务器

集群或者分布式系统,还可以配置成提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN以及大数据和人工智能平台等基础云计算服务的云服务器;软件可以是实现基于多智能体大模型的工业机器人自主装配任务规划方法的应用等,但并不局限于以上形式。

[0056] 本申请可用于众多通用或专用的计算机系统环境或配置中。例如:个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0057] 需要说明的是,在本申请的各个具体实施方式中,当涉及到需要根据用户信息、用户行为数据,用户历史数据以及用户位置信息等与用户身份或特性相关的数据进行相关处理时,都会先获得用户的许可或者同意,而且,对这些数据的收集、使用和处理等,都会遵守相关法律法规和标准。此外,当本申请实施例需要获取用户的敏感个人信息时,会通过弹窗或者跳转到确认页面等方式获得用户的单独许可或者单独同意,在明确获得用户的单独许可或者单独同意之后,再获取用于使本申请实施例能够正常运行的必要的用户相关数据。

[0058] 图1是本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法的一个可选的流程图,图1中的方法可以包括但不限于包括步骤S100至步骤S500。

[0059] 步骤S100,获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册。

[0060] 在本实施例中,任务指令可以指明需要装配哪些零件、大致的装配顺序或关键步骤。装配手册则可能包含更详细的装配流程、技术要求、注意事项、零件信息等。具体地,系统首先可以从数据库、本地文件或通过网络接口获取待装配产品的装配任务数据。装配任务数据可以包括任务指令、装配手册等。其中,任务指令可以是一个简单的文本指令,也可以是结构化的数据,例如:将零件A安装到基座B上,然后将零件C安装到零件A上);装配手册可以是一个包含详细步骤、零件图、技术要求的电子文档或结构化数据库,例如:零件A有三个安装孔,需与基座B上对应的螺纹孔对准;安装零件C时,需确保其上的凹槽与零件A上的凸起完全啮合。

[0061] 步骤S200,对所述任务指令和所述装配手册进行解析,得到任务执行逻辑。

[0062] 在本实施例中,系统可以利用自然语言处理(NLP)技术或视觉语言模型(VLM)等方式对任务指令和装配手册文本进行分析。例如,可以识别出“安装”、“对准”、“啮合”等动作关键词,提取出涉及的零件名称(A、B、C),并解析出它们之间的装配关系和顺序约束。对于结构化的手册数据,则可以直接提取相应的规则和参数。解析的结果,即任务执行逻辑,通过识别其中的对象类别、功能角色与空间结构关系,输出组件清单及其关系属性,并构建任务执行逻辑,任务执行逻辑可以通过任务层级图结构表示。

[0063] 具体地,利用视觉语言模型对输入的装配图纸和操作手册进行语义理解,具体可以采用预训练的视觉语言模型对图像中的对象类别、功能角色及空间关系进行提取,进而

构建任务层级图结构。通过将非结构化视觉信息转化为结构化任务图,解决了传统方法无法解析图纸和自然语言指令的问题。

[0064] 步骤S300,对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像。

[0065] 在本实施例中,利用机器人搭载的三维相机实时采集作业场景图像,包括彩色图像与深度信息,用于支持后续目标检测、分割与姿态估计。可以通过安装在机器人臂端、固定在工位上方或侧面的摄像头、深度相机等多传感器来实现对装配环境的采集,以获取装配区域内的实时视觉信息。具体地,在装配工作区域安装有多个传感器,包括至少一个深度相机。在机器人执行装配任务的过程中,这些传感器持续工作,实时采集装配区域的图像数据,包括彩色图像和深度图。例如,当机器人准备执行抓取零件A的操作时,传感器会捕捉到当前工作台上零件A的实际位置、姿态以及周围环境(如其他零件、工具、夹具)的信息。

[0066] 步骤S400,对所述目标图像进行多模态感知,得到感知结果,所述感知结果包括所述装配对象的至少一个组件在所述装配环境中的位置。

[0067] 在本实施例中,系统对接收到的目标图像(RGB图像和深度图像)进行处理。首先,利用目标检测和识别算法在RGB图像中识别出装配对象的各个组件(零件A、B、C等)。然后,结合深度图像信息,精确计算出这些组件在机器人坐标系下的三维位置和姿态(位姿)。感知结果不仅包括零件A的位置坐标和朝向,还可能包括组件的状态(如是否已被安装、是否正确摆放、是否处于可抓取状态、是否有遮挡等信息),以及与其他组件的相对关系等。

[0068] 步骤S500,将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务。

[0069] 在本实施例中,系统将任务执行逻辑与感知结果进行融合。具体可以将自然语言指令、任务执行逻辑及感知结果作为输入,利用大语言模型生成具备执行条件与回退策略的符号化行为树。行为树任务规划是指利用大语言模型将自然语言指令、任务层级图及感知状态生成符号化行为树,具体可以采用大语言模型生成包含执行条件与回退策略的行为树,其叶节点绑定规划领域定义语言(Planning Domain Definition Language,PDDL)定义的原子技能。行为树中的每个叶节点对应一个机器人技能库定义的机器人技能(如抓取、插入、搬运等),并在执行过程中根据节点返回的Success、Failure或Running状态动态调度。

[0070] 具体地,系统将任务执行逻辑与当前的感知结果进行匹配和融合。如果任务逻辑要求安装零件B,而感知结果发现零件B不在预期的位置,或者其姿态不正确,系统可以根据当前感知到的零件B的实际位置和姿态,以及任务逻辑中关于安装顺序和约束的要求,动态地规划出一条新的、适应当前环境的装配路径或策略。目标装配任务可以是一系列具体的动作指令,如“移动到X坐标,以Y姿态抓取零件B,然后移动到Z坐标进行安装”。

[0071] 本实施例的逻辑示意图如图2所示,首先,利用视觉语言模型对装配图纸和操作手册等视觉信息进行解析,提取任务对象、角色及空间关系,构建任务层级图结构。这一步骤实现了非结构化输入到结构化任务模型的转换。其次,通过多模态感知系统进行对象识别和姿态估计,结合视觉语言模型提取对象类别、空间关系和任务状态,生成符号谓词并映射为PDDL格式的初始状态和动作前提。这一步骤解决了多模态感知与符号规划模块之间的语义断层问题。第三,利用大语言模型生成具备执行条件和回退策略的符号化行为树,每个叶节点对应机器人技能库中定义的技能,并根据节点返回状态动态调度。这一步骤实现了基于实时感知状态的执行逻辑调整。最后,在执行过程中融合视觉和传感器数据进行多模态

验证,通过视觉问答和物理传感器数据进行状态判断,实现执行监控和反馈调整。这一步骤确保了状态判断的准确性,并支持失败时的回退或重规划。各步骤之间的协同工作体现在:视觉语言模型提取的任务结构为行为树生成提供了框架;多模态感知结果被映射为PDDL格式,为行为树的执行条件判断提供了依据;行为树的动态调度机制与多模态验证相结合,实现了任务执行的实时监控和调整。这种协同工作方式使得系统能够适应复杂、动态的装配环境,提高了任务执行的灵活性和鲁棒性。

[0072] 本实施例通过实时采集装配环境图像并进行多模态感知,能够及时发现装配环境中的变化,并根据任务执行逻辑动态调整装配计划,提高了装配系统对复杂、动态环境的适应能力;能够根据实际环境最优地规划装配路径和策略,避免不必要的移动或调整,从而可能缩短装配时间,提高生产效率;结合了对任务本身的逻辑理解,使得规划出的任务不仅仅是对环境的简单反应,而是基于装配目标和规则的智能决策,减少了因环境微小扰动导致的装配失败。

[0073] 在一些实施例中,步骤S200可以包括但不限于包括步骤S210至步骤S220:

[0074] 步骤S210,通过视觉语言模型对所述任务指令和所述装配手册进行语义解析,得到各个组件的类别、各个组件的功能角色以及各个组件之间的空间结构关系;

[0075] 步骤S220,根据各个所述组件的类别、各个所述组件的功能角色以及各个所述组件之间的空间结构关系,构建所述任务执行逻辑,所述任务执行逻辑包括多个子任务,每个所述子任务包括至少一个组件的安装描述信息,以及所述子任务的执行顺序。

[0076] 在本实施例中,通过视觉语言模型(Vision-Language Model, VLM)对任务指令和装配手册进行语义解析,能够理解自然语言文本和图像之间的语义关联,从而提取出各个组件的类别、功能角色以及它们之间的空间结构关系,进而构建任务执行逻辑,即任务层级图结构 $G=(O,H,R)$ 。其中 O 为对象集合,包含所有识别出的零件。 H 为子任务的有向组合关系,表示组装顺序与层次。 R 为可替换组件的等价关系,标识功能相同的可互换零件。最后,通过结构映射自动生成PDDL格式的任务问题定义。PDDL是人工智能规划领域广泛使用的标准表示语言,这种转换使得后续的规划工具能够理解和处理。将对象映射为PDDL对象,关系映射为谓词,组装步骤映射为动作。生成的PDDL文件包含:domain、objects、init和goal部分,分别对应问题域、对象集、初始状态和目标状态。

[0077] 具体地,视觉语言模型通过多模态输入解析生成结构化组件清单,对象集合 O 包含装配任务中涉及的物理实体及其类别标签,例如螺丝、轴承、基座等;子任务的有向组合关系 H 通过解析操作手册中的步骤顺序与条件依赖关系生成,例如“安装基座”需在“固定螺丝”之前执行;可替换组件的等价关系 R 通过识别功能等效的不同零件建立,例如同一种规格的两种螺丝可互换。通过结构映射自动生成PDDL格式的任务问题定义,为后续的符号规划和行为树生成提供结构化输入。

[0078] 具体地,视觉语言模型首先对装配图纸中的图形符号与文本标注进行联合解析,提取对象类别与空间关系,例如识别图纸中标注的“零件A需插入零件B的卡槽”。随后,模型根据操作手册中的步骤描述构建子任务间的有向组合关系,例如手册中“步骤1:放置基座;步骤2:安装支架”生成 H 中的顺序关系。对于可替换组件,模型通过比对零件参数表与功能描述建立等价关系,例如参数相同的两种螺母在 R 中被标记为可互换。完成图结构构建后,系统将 O 中的对象映射为PDDL中的对象实例, H 中的顺序关系转换为PDDL动作的前后置条

件,R中的等价关系生成替代动作选项。通过此过程,原本非结构化的任务输入被转化为标准化的符号描述,使得后续规划模块可直接调用PDDL定义的初始状态与目标约束,避免人工配置导致的语义偏差与效率损失。

[0079] 在本实施例的一个实现方式中,视觉语言模型可以分析装配手册中的图片或任务指令中引用的图片,识别出其中包含的各个组件,并结合文本描述,确定每个组件的类别,例如“螺栓”、“螺母”、“垫片”、“齿轮”、“轴承”等。视觉语言模型能结合文本描述(如“将电机安装到支架上”、“用密封圈防止漏油”)和图像中的组件位置、与其他组件的连接方式,推断出每个组件在装配体中的功能角色,例如“被安装件”、“安装基座”、“紧固件”、“密封件”、“传动件”等。视觉语言模型还能够分析装配手册中的示意图或3D模型截图,理解组件之间的空间布局 and 连接关系。例如,识别出“组件A位于组件B的上方”、“组件C通过螺纹连接到组件D”、“组件E包围着组件F”等空间结构关系。视觉语言模型不仅能识别表面信息,还能理解组件的功能角色和空间关系,这使得生成的任务执行逻辑更符合实际装配需求,减少了因理解偏差导致的错误。

[0080] 然后,系统将上一步解析得到的组件类别、功能角色和空间结构关系进行整合。例如,知道“螺栓”是紧固件,“支架”是安装基座,“电机”是被安装件,并且知道电机需要安装在支架上,螺栓用于固定电机和支架。基于整合后的信息,系统根据装配的基本原则(如先安装基础结构,再安装上层组件;先安装大件,再安装小件;先安装内部件,再安装外部件等)以及解析出的空间结构关系,将整个装配过程分解为多个逻辑上独立的子任务。每个子任务都聚焦于完成一个特定的安装动作。系统根据组件之间的空间结构关系和装配的物理约束,确定这些子任务的合理执行顺序。最终,构建出的任务执行逻辑是一个包含多个子任务的有序列表。每个子任务都明确包含了至少一个组件的安装描述信息(例如,需要安装哪个组件、安装到哪个位置、与其他哪些组件发生关系),以及该子任务在整个序列中的执行顺序。

[0081] 本实施例通过引入视觉语言模型,自动解析复杂的图文装配信息,大大减少了人工理解、标注和输入装配步骤的工作量,提高了效率;视觉语言模型不仅能识别表面信息,还能理解组件的功能角色和空间关系,这使得生成的任务执行逻辑更符合实际装配需求,减少了因理解偏差导致的错误;生成的PDDL定义为后续规划提供了结构化输入,提升了整体系统的智能化水平和任务适应性。

[0082] 在一些实施例中,步骤S400可以包括但不限于包括步骤S410至步骤S430:

[0083] 步骤S410,对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像;

[0084] 步骤S420,对每个所述装配对象的分割图像进行抓取位姿感知,得到每个所述装配对象的位姿信息,所述位姿信息包括各个所述装配对象的可抓取性评分和抓取位姿;

[0085] 步骤S430,通过视觉语言模型对各个所述装配对象的分割图像和位姿信息进行语义分析,得到所述感知结果,所述感知结果包括各个所述装配对象之间的空间关系、各个所述装配对象的操作可行性和各个所述装配对象的属性。

[0086] 在本实施例中,可以采用图像分割算法(例如Grounded SAM模型等),对实时采集的目标图像进行处理。算法能够识别并区分出图像中的不同装配对象(如待安装的零件、已安装的部件、工具等),并为每个对象生成一个精确的边界框或像素级掩码,从而得到各个

装配对象的分割图像。例如,如果目标图像中包含一个螺栓和一个螺母,分割算法将分别生成代表螺栓和螺母的两个分割图像,清晰地界定它们的轮廓。

[0087] 在本实施例中,对每个由分割图像界定的装配对象,进一步进行抓取位姿感知。这通常通过专门的位姿估计算法或抓取检测算法来实现。这些算法分析分割图像中对象的形状、边缘、纹理等特征,估计出对象在三维空间中的姿态(即位姿,包括位置和方向),并评估从不同角度抓取该对象的可行性和稳定性。位姿信息通常包括可抓取性评分和抓取位姿。其中,可抓取性评分是一个量化指标,表示当前状态下抓取该对象的难易程度或成功率;抓取位姿是一个或多个建议的抓取点及其对应的机械臂末端执行器的姿态(例如,夹爪的中心位置和开合方向),这些位姿被认为是抓取该对象的理想位置和角度。

[0088] 在本实施例中,利用视觉语言模型(VLM),将每个装配对象的分割图像(视觉信息)及其对应的位姿信息作为输入。视觉语言模型能够理解图像内容并结合位姿的描述,进行更深层次的语义分析。例如,视觉语言模型(如GPT-4V模型)可以理解分割图像中的物体是什么(需要结合装配手册中的组件信息进行关联),分析出这些物体之间的空间关系(如“螺栓在螺母的左侧”、“垫片位于螺栓下方”),判断当前位姿下对物体的操作是否可行(如“当前位姿下螺栓可以被夹爪稳定抓取”、“螺母因遮挡暂时不可操作”),并提取或推断出物体的某些属性(可能包括材质、形状、尺寸、功能角色,如“该物体是金属材质”、“该物体需要被拧紧”)等。最终得到的感知结果是一个综合了视觉、空间、操作可行性和属性信息的丰富集合。

[0089] 这些信息通过预定义的谓词模板转换为标准符号格式,例如可抓取性映射为 `graspable(obj)`,空间位置映射为 `at(obj,loc)`。感知-谓词映射函数将上述符号集合编码为初始状态,供任务规划模块调用。该过程通过多模型协同处理,将原始感知数据转化为结构化符号描述,解决了传统方法中感知与符号映射精度不足的问题,为后续规划提供可靠的环境状态输入。

[0090] 本方案的具体实施如下:

[0091] 首先,机器人搭载的三维相机实时采集作业场景图像I,包括彩色图像与深度信息,用于支持后续目标检测、分割与姿态估计。

[0092] 接着,采用Grounding DINO模型对图像进行目标检测,获取所有候选对象的边界框及初步类别标签。然后使用Grounded SAM对检测框内区域进行精细图像分割,获取各对象的语义掩膜与轮廓边界。将上述分割结果输入AnyGrasp模型,对每个物体计算可抓取性评分、抓取位姿。

[0093] 进一步地,进行语义信息抽取:将分割图像、物体候选框及姿态结果输入视觉语言模型GPT-4V,结合场景上下文进行语言引导的语义分析,提取物体间空间关系(如“靠近”“位于…上方”“部分插入”)、操作可行性(如“可抓取”“可插入”)及任务相关属性(如“零件A属于组件X”)。

[0094] 最后,根据预定义的PDDL谓词模板,将上述感知结果统一映射为符号谓词集合 $L_x = \{x_1, x_2, \dots, x_M\}$,包括 `graspable(obj)`、`at(obj,loc)`、`inserted(obj1,obj2)`、`type(obj,class)` 等格式,用以描述当前环境状态。通过感知-谓词映射函数 $f_{pred}(K(I))$,将符号化感知结果编码为标准PDDL初始状态格式,供行为树任务规划模块在后续步骤中调用。

[0095] 本实施例通过图像分割和位姿感知,能够精确识别装配环境中的各个对象及其状

态；结合位姿信息和语义分析,能更智能地判断对象是否可以被当前工具(如机械臂夹爪)操作,避免了无效或危险的尝试;将视觉信息(图像、位姿)与语言描述信息相结合,提供了对装配环境更全面、更智能的感知,为后续根据任务逻辑和实时感知结果进行动态规划提供了高质量的数据输入;通过将感知结果映射为标准PDDL格式,实现了感知模块与规划模块的无缝对接,增强了系统的模块化程度和可扩展性。

[0096] 在一些实施例中,步骤S410可以包括但不限于包括步骤S411至步骤412:

[0097] 步骤S411,对所述目标图像进行目标检测,得到所述目标图像中各个所述装配对象的边界框和初步类别标签;

[0098] 步骤S412,根据各个所述装配对象的所述边界框和所述初步类别标签对所述目标图像进行图像分割,得到各个所述装配对象的分割图像,所述分割图像包括各个所述装配对象的轮廓边界和目标类别标签。

[0099] 在本实施例中,为实现对目标图像中装配对象的高精度识别与定位,本实施例采用了一种结合目标检测与引导式分割的技术方案。首先,利用Grounding DINO模型,结合预设的文本提示对目标图像进行目标检测。该模型能够输出图像中所有候选装配对象的边界框及其对应的初步类别标签。

[0100] 具体地,Grounding DINO模型是一个结合了自然语言指导能力的目标检测模型。与传统仅依赖视觉特征的目标检测器不同,Grounding DINO可以利用文本提示(例如,装配任务中可能涉及的“螺丝”、“螺母”、“支架”等词语)来引导检测过程。Grounding DINO模型通过扫描整个目标图像,识别出所有与给定文本提示相关的潜在对象。对于图像中每一个被识别出的候选对象,模型会输出一个边界框,框定该对象在图像中的大致位置。同时,模型还会给出一个初步类别标签,这个标签通常是对应于输入文本提示的类别名称(例如,“螺丝”)。

[0101] 随后,将上述边界框作为输入,引导Grounded SAM模型对边界框所覆盖的区域进行精细图像分割。Grounded SAM模型依据边界框内的视觉特征生成高精度的语义掩膜,从而精确获取各个装配对象的轮廓边界。

[0102] 具体地,Grounded SAM模型是基于Segment Anything Model (SAM)的一个变体,特别设计用于根据给定的“grounding”信息(即Grounding DINO提供的边界框和类别标签)来进行精确的图像分割。将边界框作为输入,Grounded SAM会利用边界框内的视觉特征,并结合可能的类别信息(或仅依赖边界框区域),来精确地勾勒出对象的轮廓。对于每一个输入的边界框区域,Grounded SAM会生成一个语义掩膜。语义掩膜是一个二值图像(或带有置信度的图像),其中属于该对象的像素被标记为1(或高置信度),不属于的像素被标记为0(或低置信度)。语义掩膜精确地描绘了对象的轮廓边界。

[0103] 本实施例通过有效结合Grounding DINO的快速定位与语言指导能力以及Grounded SAM的高精度分割能力,能够在保证效率的同时,为后续的位姿感知和语义分析提供精确的对象形状与位置信息。

[0104] 在一些实施例中,步骤S500可以包括但不限于包括步骤S510至步骤S530:

[0105] 步骤S510,将所述任务执行逻辑和所述感知结果分别进行数据格式转换,得到规划领域定义语言格式的新的任务执行逻辑和规划领域定义语言格式的新的感知结果;

[0106] 步骤S520,根据新的任务执行逻辑和新的感知结果,生成行为树任务架构,所述行

为树任务架构包括多个任务节点,所述任务节点包括顺序节点、条件节点和回退节点,所述顺序节点用于决定子任务的执行顺序,所述条件节点用于根据当前状态是否满足执行当前子任务的条件,所述回退节点用于在当前子任务执行失败时,重新执行当前子任务或重新规划所述装配任务;

[0107] 步骤S530,根据所述行为树任务架构规划装配任务,得到所述目标装配任务。

[0108] 在本实施例中,通过结构映射将任务执行逻辑(任务层级图结构)转换为PDDL格式的任务问题定义,为后续的符号规划和行为树生成提供结构化输入。结构映射过程将图结构中的节点与边转换为PDDL中的对象、初始状态及目标状态,例如将“基座位于装配台”映射为PDDL谓词(at base station)。具体来说,完成图结构构建后,系统将O中的对象映射为PDDL中的对象实例,H中的顺序关系转换为PDDL动作的前后置条件,R中的等价关系生成替代动作选项。通过此过程,原本非结构化的任务输入被转化为标准化的符号描述,使得后续规划模块可直接调用PDDL定义的初始状态与目标约束,避免人工配置导致的语义偏差与效率损失。

[0109] 在本实施例中,针对从实时感知(如摄像头、传感器)获取的关于装配环境的信息(即感知结果),系统使用预定义的PDDL谓词模板,将这些感知数据统一映射为一个符号谓词集合。这些谓词以形式化的方式描述了当前环境的各种状态,例如(在…之上)(在…旁边)(抓取)(空闲)等。通过“感知-谓词映射函数”,将这个符号化的感知结果集合编码成标准的PDDL初始状态格式。

[0110] 在本实施例中,通过对任务执行逻辑和感知结果进行数据格式转换,确保了任务层面的高层规划指令(任务执行逻辑)与感知层面的底层环境信息(感知结果)能够以兼容的方式被后续的行为树任务架构生成模块所利用。

[0111] 在本实施例中,基于转换后的新任务执行逻辑和新感知结果,系统会生成一个任务架构,即行为树。具体地,系统会根据自然语言任务指令、PDDL知识域与感知状态,通过大语言模型结合任务技能库,生成行为树结构。

[0112] 具体地,自然语言任务指令为用户或系统以自然语言形式给出的装配任务要求(例如“请将蓝色螺母安装到红色支架上”),作为任务目标的高层语义输入;PDDL知识域定义了任务域内的对象类型、谓词及可用技能集合;感知状态通过符号谓词集合描述当前环境状态。

[0113] 具体地,行为树包括以下几种节点:顺序节点、条件节点和回退节点。其中,顺序节点用于体现任务执行逻辑中预设的子任务执行顺序。一个顺序节点包含多个子节点(通常是其他任务节点或具体的执行动作),它按预定顺序依次执行这些子节点。例如,如果任务逻辑规定先安装A组件,再安装B组件,则顺序节点会确保A组件的安装任务在B组件之前执行。条件节点用于在执行特定子任务之前进行状态检查,它会根据当前感知结果中提供的信息,判断是否满足执行该子任务的前提条件。例如,在尝试安装一个螺母之前,条件节点会检查感知结果中对应的螺栓是否已经就位且处于可安装状态。如果条件满足,则执行后续的子任务;如果不满足,则可能跳过该子任务或触发其他处理逻辑(如等待、调整)。这体现了感知结果对任务执行的动态约束和引导。回退节点用于处理任务执行过程中可能出现的失败情况,增强了规划方法的鲁棒性。当某个子任务执行失败时(例如,抓取失败、安装位置错误等,这些失败可能通过传感器反馈或执行结果检测到),回退节点会被激活。它可以

执行预设的恢复动作,例如重试当前子任务(可能需要先调整位姿或清除干扰)、执行清理操作,或者在必要时触发更复杂的重新规划过程,调用更高层级的规划器根据最新的感知结果重新制定后续任务步骤。这确保了即使在非理想环境下,装配过程也能继续进行或安全中止。

[0114] 在本实施例中,基于生成的行为树任务架构,系统进行具体的装配任务规划,输出最终的目标装配任务。对于行为树任务架构中确定的每个具体子任务(例如,抓取某个组件、将其移动到指定位置、执行装配动作),系统结合最新的感知结果(组件的精确位姿、可达性等),计算出机器人执行该动作所需的具体路径、末端执行器的姿态、力度等控制参数。系统会沿着行为树任务架构的节点顺序执行,在遇到条件节点时进行实时判断,在遇到回退节点时处理异常。整个过程是一个动态的、迭代优化的过程,不断将最新的感知信息反馈到规划中。最终,规划结果被转化为机器人控制器能够理解的具体指令序列,即目标装配任务。这个任务指令序列不仅包含了要执行的动作,还包含了执行这些动作的时序、条件判断逻辑和异常处理预案。

[0115] 本方案的具体实施如下:

[0116] 首先,输入自然语言任务指令 i 、PDDL知识域 D 与感知状态 P 。例如,任务指令 i 为“将零件A插入零件B”,PDDL知识域 D 定义了可用的机器人技能集合,感知状态 P 包含当前场景中零件A和B的位置信息。

[0117] 其次,通过大语言模型结合任务技能库 L_{ψ} ,生成行为树结构 B 。具体地,将任务指令 i 、知识域 D 和状态 P 作为输入,通过GPT-4等大语言模型进行解析和推理,输出包含顺序执行、条件判断与回退节点的行为树结构 B 。

[0118] 然后,为行为树的每个叶子节点绑定PDDL定义的原子技能 ψ_i 。原子技能 ψ_i 的格式为 $(name, params, \psi_p, \psi_e, \phi)$,其中 $name$ 为技能名称, $params$ 为参数列表, ψ_p 表示前提条件, ψ_e 为技能效果, ϕ 为底层控制接口。例如,“抓取”技能可定义为 $(grasp, [obj], \{at(robot, obj_loc), graspable(obj)\}, \{holding(obj)\}, grasp_control)$ 。

[0119] 最后,行为树执行时根据节点状态返回值动态驱动机器人任务执行与失败恢复机制。节点状态包括Success、Failure和Running三种。当子节点返回Failure时,触发父节点的回退策略;当所有子节点均返回Success时,任务完成。

[0120] 本实施例通过数据格式转换实现信息融合,通过行为树任务架构引入逻辑与容错机制,最终生成具体可行的目标装配任务,提升了基于多智能体大模型的工业机器人自主装配任务规划的智能化水平和适应性;不仅能够遵循预设的装配流程,还能通过感知结果实时响应环境变化;通过行为树任务架构中的条件节点和回退节点进行条件判断和异常处理,从而提高了装配成功的概率,降低了人工干预的需求,尤其适用于环境复杂、对象状态多变的自动化装配场景。

[0121] 在一些实施例中,在步骤S500之后还可以包括但不限于包括步骤S600至步骤S700:

[0122] 步骤S600,对所述装配对象进行装配时的装配环境进行拍摄,得到拍摄视频;

[0123] 步骤S700,对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规

划,得到目标装配任务的步骤执行。

[0124] 在本实施例中,系统会在装配任务执行过程中对装配现场进行监控。利用装配环境中的摄像头或其他视觉传感器,对正在或即将进行装配操作的装配对象及其周围环境进行拍摄,从而获取一段或多段拍摄视频。获取到拍摄视频后,对这些视频进行图像分析,判断目标装配任务(或当前正在执行的子任务)是否按照预期在进行。具体而言,分析系统会检查视频中装配设备(如机械臂、抓手等)的动作是否符合规划路径,装配对象上的组件是否被正确抓取、移动或安装,以及是否存在任何意外的干扰或错误(例如,组件位置偏移、抓取失败、与其他物体发生碰撞等)。

[0125] 在本实施例中,系统会将分析结果与预设的“成功”标准或“可接受偏差”范围进行比较。若确定目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值。动作失败可以指装配设备未能完成预定动作(如抓取失败、放置不准),或者装配对象的状态未达到预期(如螺丝未拧紧、零件未对齐)。动作偏差超出预设阈值指装配设备的实际运动轨迹、速度、力度等与规划值或理想值之间的差异超出了系统允许的范围,这个阈值是根据装配任务的精度要求和装配对象的物理特性预先设定的。一旦图像分析的结果表明确实存在“动作失败”或“超出阈值的动作偏差”,系统便不会继续按原计划执行后续任务,而是触发相应的纠正措施。如果判断认为当前子任务只是执行过程中出现了可纠正的、非致命性的错误或偏差,系统可以选择让装配设备停止当前动作,然后尝试重新执行同一个子任务。在重新执行前,系统可能会根据失败原因进行微小的参数调整或路径修正。如果图像分析或后续判断(可能结合其他传感器信息)认为当前的问题比较严重,或者简单的重试可能无效甚至导致更坏的结果,系统则会选择中断当前子任务的执行,并“跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行”。这意味着系统将重新回到任务规划的起点或当前子任务之前的某个节点,利用最新的感知结果(可能包括从失败中获取的新信息)和任务逻辑,重新进行一次或多次任务规划,生成一个修正后的、可能包含不同策略或路径的目标装配任务,然后再尝试执行。

[0126] 本实施例通过引入这一基于视频图像分析的实时监控和动态调整环节,能够显著提高装配系统对执行错误的响应能力,避免错误累积,确保即使在非理想条件下,装配任务也能尽可能按照预期完成,从而提升了整个装配过程的可靠性、鲁棒性和自动化水平。

[0127] 在一些实施例中,步骤S700可以包括但不限于包括步骤S710至步骤S740:

[0128] 步骤S710,通过视觉语言模型对所述拍摄视频进行图像分析,得到当前子任务执行结果;

[0129] 步骤S720,获取对所述装配对象进行装配的装配设备的物理传感器数据,得到验证数据,所述验证数据包括抓手状态和力反馈数据;

[0130] 步骤S730,通过所述验证数据对所述子任务执行结果进行验证,得到验证结果;

[0131] 步骤S740,若所述验证结果指示所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

[0132] 在本实施例中,对拍摄视频进行图像分析以确定当前子任务执行状态的过程,具

体是通过视觉语言模型来完成的。视觉语言模型是一种能够同时理解和处理图像与文本信息的先进人工智能模型。系统会将拍摄到的视频片段(或其中的关键帧)输入到预先训练好的视觉语言模型中。该模型不仅能够识别视频画面中的物体、位置、姿态等视觉信息,还能理解这些信息与装配任务指令、装配手册等文本知识之间的关联。通过这种跨模态的理解能力,视觉语言模型能够生成对当前子任务执行结果的描述性判断,即“当前子任务执行结果”。例如,模型可能判断“螺丝已成功拧入但未到位”、“零件A未能正确对准零件B的插槽”或“抓手已抓取到目标零件”。

[0133] 在得到视觉语言模型输出的当前子任务执行结果之后,系统还会同步获取装配设备(例如工业机器人、机械臂等)上安装的物理传感器所采集的数据。这些传感器数据构成了重要的“验证数据”,具体包括但不限于抓手的状态信息(如抓取/释放状态、当前抓取力度、抓取位置坐标等)以及力反馈数据(如抓手与物体接触时产生的力的大小、方向、分布等)。这些物理传感器数据直接反映了装配设备与装配对象之间的物理交互情况,用于验证任务执行是否成功。

[0134] 最后,基于综合了视觉语言模型判断和物理传感器数据验证的“验证结果”,系统做出最终决策。如果验证结果明确指示目标装配任务在执行过程中确实存在动作失败(如零件未抓取、放置错误等)或动作偏差(如零件位置略有偏移、拧紧力度不足等)超出了预设的安全或精度阈值,则系统将触发相应的纠正机制,指令装配设备重新执行当前未能成功的子任务,期望在重试中修正错误;若如果错误较为严重或重试可能无效,则跳转回任务规划的环节,即重新执行“将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务”的步骤。在重新规划时,系统可以利用最新的感知结果(包括从失败尝试中获取的信息)和任务逻辑调整策略、修正路径或参数,生成一个修正后的任务计划再进行执行。

[0135] 具体地,系统可以通过视觉问答方式判断操作是否达成预期效果,并结合抓手状态、力反馈等低层传感器数据进行交叉验证。若检测到动作失败或偏差超出阈值,系统将自动触发行为树中的回退策略或重新规划路径。视觉问答方式可以通过视觉语言模型对当前场景图像进行语义判断,提取操作结果与预期目标的匹配度;物理验证方式可以通过六维力传感器采集夹持力、接触力数据,与预设的安全阈值进行比对;交叉验证机制将视觉语义判断结果与物理传感器数据通过加权融合算法进行综合评估。

[0136] 本方案的具体实施如下:

[0137] 首先,通过机器人搭载的高分辨率相机采集当前场景图像。随后,将该图像输入预训练的视觉语言模型GPT-4V,通过视觉问答方式对操作结果进行语义判断。例如,系统可以询问“零件A是否已经完全插入到零件B中”或“组件C的朝向是否正确”等问题,由模型基于图像内容给出回答。

[0138] 同时,系统还结合机器人末端执行器的力矩传感器和位置编码器数据,对操作过程中的力反馈和位置偏差进行量化分析。例如,在插入操作中,若检测到异常的阻力或位置偏差,系统会将其视为潜在的执行失败。

[0139] 进一步地,系统将视觉语义判断结果与物理传感数据进行交叉验证。如果两者存在明显不一致,或者任一方面显示操作未达到预期效果,系统将自动触发失败处理机制。

[0140] 由此,若检测到动作执行失败或偏差超出预设阈值,系统会首先尝试启动行为树

中预定义的回退策略。例如,在零件插入失败时,系统可能会执行“抽出-调整姿态-重新插入”的回退序列。如果多次回退仍无法解决问题,系统将重新调用任务规划模块,基于当前状态生成新的执行路径。

[0141] 在整个执行过程中,系统持续记录各项操作的成功率、耗时、精度等关键指标。这些数据将用于后续的任务优化,如调整操作参数、优化规划策略等,从而不断提升系统的整体性能。

[0142] 本实施例通过融合视觉语义理解和物理传感数据,大幅提升了系统对复杂环境变化的适应能力;通过多模态验证和实时反馈机制,系统能够及时发现并纠正执行偏差,有效降低了任务失败率;动态重规划和回退策略的引入,增强了系统的容错能力和执行稳定性,使其能够应对各种意外情况。

[0143] 本申请实施例通过视觉语言模型(VLM)对装配图纸或自然语言指令进行解析,提取任务对象、语义角色和空间结构关系,构建任务层级图结构;结合三维图像、分割算法与抓取估计器,提取对象类别、位姿与操作属性,生成PDDL格式的谓词集以描述环境状态;通过大语言模型(LLM)结合指令、任务图与环境状态信息,自动生成行为树结构,组织任务执行顺序,行为树支持条件控制、回退机制与模块化执行;系统在任务执行过程中融合视觉验证与传感器反馈,自动判断动作成功与否,如遇失败则触发备用路径或重新规划,确保任务完成的稳定性与精确性。本申请实施例通过融合视觉语言模型与大语言模型,构建了一个支持复杂语义解析、多模态感知与行为树驱动执行的机器人长时序任务规划系统。该系统具备高度智能化、自适应和模块化特点,能够显著提升工业机器人在多任务、多步骤、动态环境下的执行稳定性与操作效率。依托预训练大模型的通用语义理解能力,无需针对每个任务进行特定训练,即可解析新的自然语言指令或装配图纸,自动生成可执行的任务计划。极大地拓宽了机器人在变化频繁或非结构化生产线中的适应范围。通过视觉语言模型提取任务对象、空间关系与操作顺序,系统可自动构建任务图结构,识别出任务的层次依赖与并发关系,适用于多零部件、多工序、高精度要求的装配流程,满足柔性化、定制化生产场景对任务建模与规划的复杂性需求。系统使用行为树作为任务表示载体,支持顺序执行、条件判断、失败回退等控制逻辑。在任务执行过程中,若因环境变化或感知误差导致执行失败,可自动触发备用策略或重新规划,保证任务不中断,有效提升系统的容错能力与执行稳定性。系统在任务完成率、执行耗时、重复稳定性等关键指标方面均显著优于传统基于状态机或模板驱动的机器人方法。尤其在复杂、多步骤任务中展现出更强的灵活性与鲁棒性,具备良好的工程推广前景。

[0144] 请参阅图3,本申请实施例还提供一种基于多智能体大模型的工业机器人自主装配任务规划装置800,可以实现上述基于多智能体大模型的工业机器人自主装配任务规划方法,该基于多智能体大模型的工业机器人自主装配任务规划装置800包括:

[0145] 获取模块10,用于获取装配对象的装配任务数据,所述装配任务数据包括对所述装配对象进行装配的任务指令和装配手册;

[0146] 解析模块20,用于对所述任务指令和所述装配手册进行解析,得到任务执行逻辑;

[0147] 采集模块30,用于对所述装配对象进行装配时的装配环境进行实时采集,得到目标图像;

[0148] 感知模块40,用于对所述目标图像进行多模态感知,得到感知结果,所述感知结果

包括所述装配对象的至少一个组件在所述装配环境中的位置；

[0149] 规划模块50,用于将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务。

[0150] 在一些实施方式中,解析模块20可以包括:

[0151] 解析子模块,用于通过视觉语言模型对所述任务指令和所述装配手册进行语义解析,得到各个组件的类别、各个组件的功能角色以及各个组件之间的空间结构关系;

[0152] 构建子模块,用于根据各个所述组件的类别、各个所述组件的功能角色以及各个所述组件之间的空间结构关系,构建所述任务执行逻辑,所述任务执行逻辑包括多个子任务,每个所述子任务包括至少一个组件的安装描述信息,以及所述子任务的执行顺序。

[0153] 在一些实施方式中,感知模块40可以包括:

[0154] 图像分割子模块,用于对所述目标图像进行图像分割,得到所述目标图像中各个装配对象的分割图像;

[0155] 位姿感知子模块,用于对每个所述装配对象的分割图像进行抓取位姿感知,得到每个所述装配对象的位姿信息,所述位姿信息包括各个所述装配对象的可抓取性评分和抓取位姿;

[0156] 语义分析子模块,用于通过视觉语言模型对各个所述装配对象的分割图像和位姿信息进行语义分析,得到所述感知结果,所述感知结果包括各个所述装配对象之间的空间关系、各个所述装配对象的操作可行性和各个所述装配对象的属性。

[0157] 在一些实施方式中,图像分割子模块可以包括:

[0158] 目标检测单元,用于对所述目标图像进行目标检测,得到所述目标图像中各个所述装配对象的边界框和初步类别标签;

[0159] 图像分割单元,用于根据各个所述装配对象的所述边界框和所述初步类别标签对所述目标图像进行图像分割,得到各个所述装配对象的分割图像,所述分割图像包括各个所述装配对象的轮廓边界和目标类别标签。

[0160] 在一些实施方式中,规划模块50可以包括:

[0161] 转换子模块,用于将所述任务执行逻辑和所述感知结果分别进行数据格式转换,得到规划领域定义语言格式的新的任务执行逻辑和规划领域定义语言格式的新的感知结果;

[0162] 生成子模块,用于根据新的任务执行逻辑和新的感知结果,生成行为树任务架构,所述行为树任务架构包括多个任务节点,所述任务节点包括顺序节点、条件节点和回退节点,所述顺序节点用于决定子任务的执行顺序,所述条件节点用于根据当前状态是否满足执行当前子任务的条件,所述回退节点用于在当前子任务执行失败时,重新执行当前子任务或重新规划所述装配任务;

[0163] 规划子模块,用于根据所述行为树任务架构规划装配任务,得到所述目标装配任务。

[0164] 在一些实施方式中,该装置还包括:

[0165] 拍摄模块,用于对所述装配对象进行装配时的装配环境进行拍摄,得到拍摄视频;

[0166] 分析模块,用于对所述拍摄视频进行图像分析,若确定所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所

述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

[0167] 在一些实施方式中,分析模块可以包括:

[0168] 图像分析子模块,用于通过视觉语言模型对所述拍摄视频进行图像分析,得到当前子任务执行结果;

[0169] 获取子模块,用于获取对所述装配对象进行装配的装配设备的物理传感器数据,得到验证数据,所述验证数据包括抓手状态和力反馈数据;

[0170] 验证子模块,用于通过所述验证数据对所述子任务执行结果进行验证,得到验证结果;

[0171] 执行子模块,用于若所述验证结果指示所述目标装配任务在执行过程中存在动作失败或动作偏差超出预设阈值,则重新执行当前子任务或跳转至所述将所述任务执行逻辑和所述感知结果转换为规划领域定义语言格式,并通过行为树进行装配任务规划,得到目标装配任务的步骤执行。

[0172] 该基于多智能体大模型的工业机器人自主装配任务规划装置的具体实施方式与上述基于多智能体大模型的工业机器人自主装配任务规划方法的具体实施例基本相同,在此不再赘述。

[0173] 本申请实施例还提供了一种电子设备,电子设备包括存储器和处理器,存储器存储有计算机程序,处理器执行计算机程序时实现上述基于多智能体大模型的工业机器人自主装配任务规划方法。该电子设备可以为包括平板电脑、车载电脑等任意智能终端。

[0174] 请参阅图4,图4示意了另一实施例的电子设备的硬件结构,电子设备包括:

[0175] 处理器801,可以采用通用的中央处理器(CentralProcessingUnit,CPU)、微处理器、应用专用集成电路(ApplicationSpecificIntegratedCircuit,ASIC)、或者一个或多个集成电路等方式实现,用于执行相关程序,以实现本申请实施例所提供的技术方案;

[0176] 存储器802,可以采用只读存储器(ReadOnlyMemory,ROM)、静态存储设备、动态存储设备或者随机存取存储器(RandomAccessMemory,RAM)等形式实现。存储器802可以存储操作系统和其他应用程序,在通过软件或者固件来实现本说明书实施例所提供的技术方案时,相关的程序代码保存在存储器802中,并由处理器801来调用执行本申请实施例的基于多智能体大模型的工业机器人自主装配任务规划方法;

[0177] 输入/输出接口803,用于实现信息输入及输出;

[0178] 通信接口804,用于实现本设备与其他设备的通信交互,可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信;

[0179] 总线805,在设备的各个组件(例如处理器801、存储器802、输入/输出接口803和通信接口804)之间传输信息;

[0180] 其中处理器801、存储器802、输入/输出接口803和通信接口804通过总线805实现彼此之间在设备内部的通信连接。

[0181] 本申请实施例还提供了一种计算机可读存储介质,该计算机可读存储介质存储有计算机程序,该计算机程序被处理器执行时实现上述基于多智能体大模型的工业机器人自主装配任务规划方法。

[0182] 存储器作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序以及非

暂态性计算机可执行程序。此外,存储器可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施方式中,存储器可选包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至该处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0183] 本申请实施例提供的基于多智能体大模型的工业机器人自主装配任务规划方法、基于多智能体大模型的工业机器人自主装配任务规划装置、电子设备及存储介质,其通过视觉语言模型对装配任务指令和操作手册等信息进行解析,提取任务对象、角色及空间关系,构建任务执行逻辑;通过多模态感知系统集对装配环境的三维图像进行感知,确保机器人对当前场景具备实时认知能力;利用大语言模型生成具备执行条件和回退策略的符号化行为树,每个叶节点对应机器人技能库中定义的技能,并根据节点返回状态动态调度;在执行过程中融合视觉和传感器数据进行多模态验证,通过视觉问答和物理传感器数据进行状态判断,实现执行监控和反馈调整。本申请通过构建任务层级图结构、多模态感知与符号映射、动态行为树规划及多模态验证机制,实现了非结构化任务指令的语义解析、环境状态的符号化表征以及动态任务调整能力,提升了自主决策能力,实现多模态信息融合闭环,增强了复杂装配任务动态调整能力。

[0184] 本申请实施例描述的实施例是为了更加清楚的说明本申请实施例的技术方案,并不构成对于本申请实施例提供的技术方案的限定,本领域技术人员可知,随着技术的演变和新应用场景的出现,本申请实施例提供的技术方案对于类似的技术问题,同样适用。

[0185] 本领域技术人员可以理解的是,图中示出的技术方案并不构成对本申请实施例的限定,可以包括比图示更多或更少的步骤,或者组合某些步骤,或者不同的步骤。

[0186] 以上所描述的装置实施例仅仅是示意性的,其中作为分离部件说明的单元可以是或者也可以不是物理上分开的,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0187] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、设备中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。

[0188] 本申请的说明书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数字在适当情况下可以互换,以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0189] 应当理解,在本申请中,“至少一个(项)”是指一个或者多个,“多个”是指两个或两个以上。“和/或”,用于描述关联对象的关联关系,表示可以存在三种关系,例如,“A和/或B”可以表示:只存在A,只存在B以及同时存在A和B三种情况,其中A,B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达,是指这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b或c中的至少一项(个),可以表示:a,b,c,“a和b”,“a和c”,“b和c”,或“a和b和c”,其中a,b,c可以是单个,也可

以是多个。

[0190] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,上述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0191] 上述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0192] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0193] 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括多指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本申请各个实施例的方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等各种可以存储程序的介质。

[0194] 以上参照附图说明了本申请实施例的优选实施例,并非因此局限本申请实施例的权利范围。本领域技术人员不脱离本申请实施例的范围和实质内所作的任何修改、等同替换和改进,均应在本申请实施例的权利范围之内。

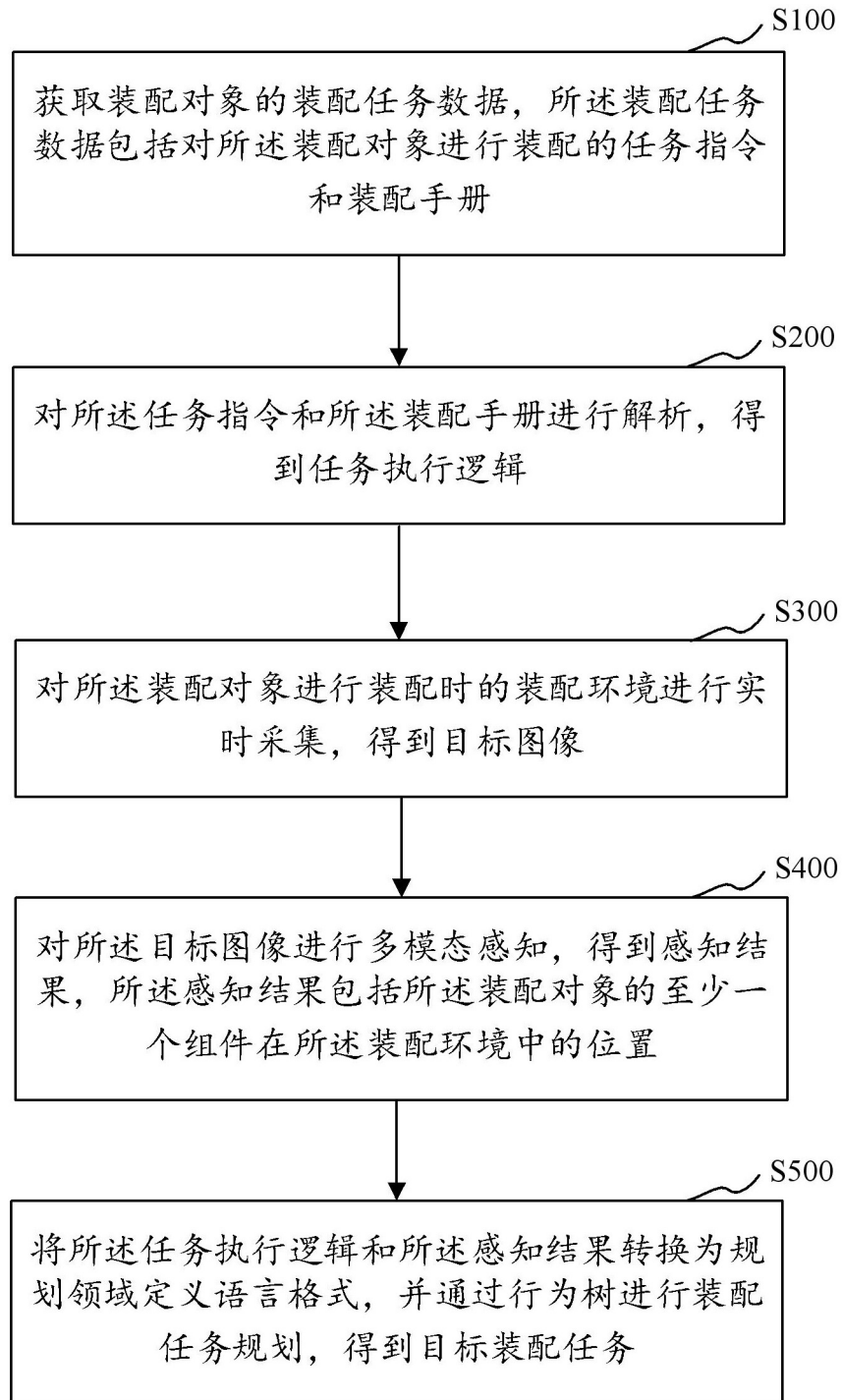


图1

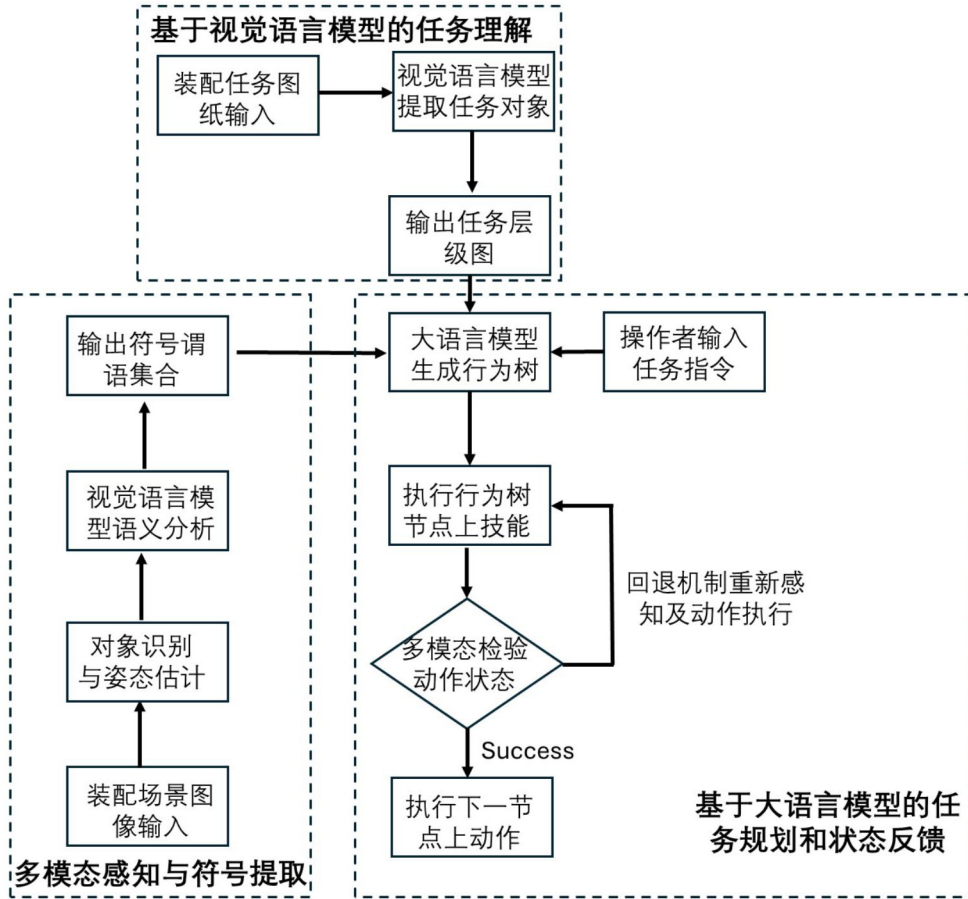


图2

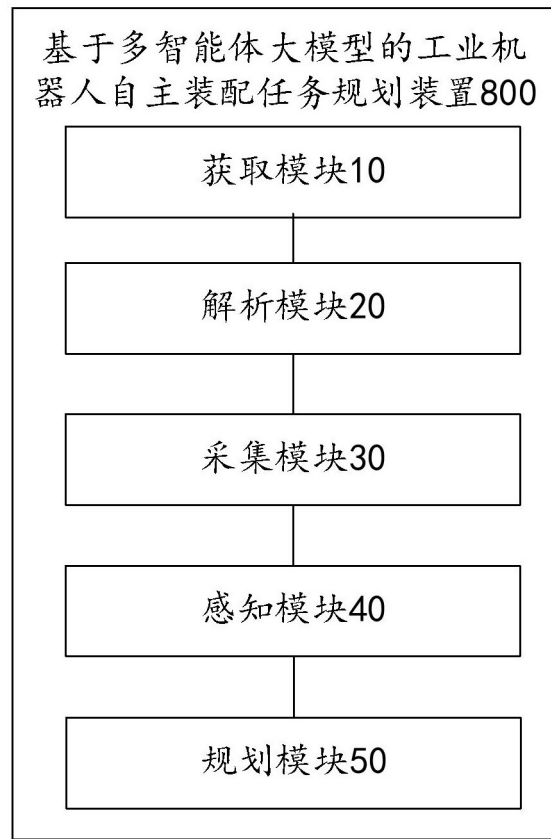


图3

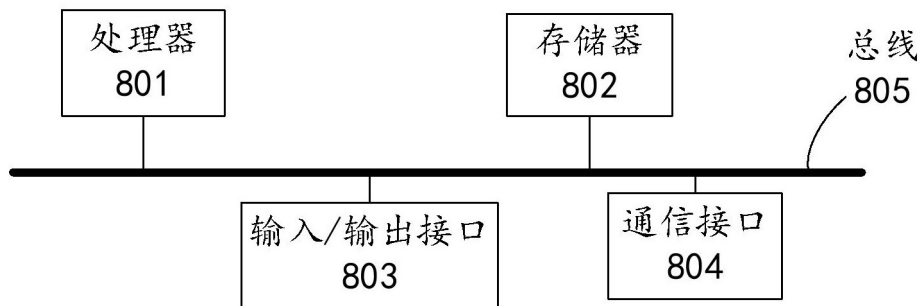


图4