(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113524175 B
(45)授权公告日 2022.08.12

(21)申请号 202110692988.2

(22)申请日 2021.06.22

(65)同一申请的已公布的文献号
申请公布号 CN 113524175 A

(43)申请公布日 2021.10.22

(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 李树飞 郑湃 范峻铭

(74)专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268
专利代理师 朱阳波 刘文求

(51)Int.Cl.
B25J 9/16(2006.01)
G06V 40/20(2022.01)
G06V 40/10(2022.01)
G06V 20/40(2022.01)
G06V 10/764(2022.01)
G06V 10/80(2022.01)
G06V 10/82(2022.01)
G06K 9/62(2022.01)
G06N 3/04(2006.01)
G06F 3/01(2006.01)

(56)对比文件
CN 112276944 A,2021.01.29
CN 108527370 A,2018.09.14
CN 107097227 A,2017.08.29
CN 111950485 A,2020.11.17
CN 104715493 A,2015.06.17
US 2020057935 A1,2020.02.20
LI SHUFEI,et al..An AR-assisted deep learning-based approach for automatic inspection of aviation connectors.《IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS》.2021,

审查员 方海林

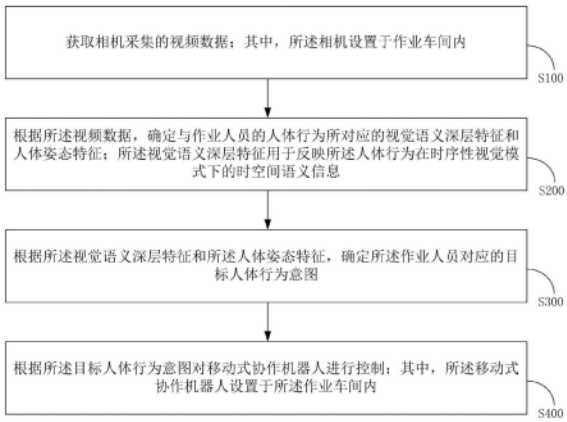
权利要求书2页 说明书10页 附图6页

(54)发明名称

一种基于多模态行为在线预测的人机协作方法和系统

(57)摘要

本发明公开了一种基于多模态行为在线预测的人机协作方法和系统，所述方法包括：获取视频数据；根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；根据所述目标人体行为意图确定移动式协作机器人对应的执行操作和移动路径。解决了现有技术中手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式的问题。



CN 113524175 B

1.一种基于多模态行为在线预测的人机协作方法,其特征在于,所述方法包括:

获取视频数据;

根据所述视频数据,确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征,所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息;

根据所述视觉语义深层特征和所述人体姿态特征,确定所述作业人员对应的目标人体行为意图;

根据所述目标人体行为意图对移动式协作机器人进行控制。

2.根据权利要求1所述的人机协作方法,其特征在于,所述根据所述视频数据,确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征,包括:

根据所述视频数据输出基础三原色视频流和三维人体姿态数据流;

根据所述基础三原色视频流提取所述视觉语义深层特征;

根据所述三维人体姿态数据流提取所述人体姿态特征。

3.根据权利要求2所述的人机协作方法,其特征在于,所述根据所述基础三原色视频流提取所述视觉语义深层特征,包括:

对所述基础三原色视频流中每一视频帧对应的人体行为区域进行剪裁,得到若干人体行为区域视频帧;

根据所述若干人体行为区域视频帧确定所述人体行为对应的视觉模态浅层特征,并根据所述视觉模态浅层特征提取所述视觉语义深层特征,其中,所述视觉模态浅层特征用于反映所述若干人体行为区域视频帧中的视觉几何特征。

4.根据权利要求3所述的人机协作方法,其特征在于,所述根据所述人体行为区域视频帧确定所述人体行为对应的视觉模态浅层特征,并根据所述视觉模态浅层特征提取所述视觉语义深层特征,包括:

将所述若干人体行为区域视频帧输入预先经过训练的二维卷积神经网络中,得到所述视觉模态浅层特征;

将所述视觉模态浅层特征输入预先经过训练的三维卷积神经网络中,得到所述视觉语义深层特征。

5.根据权利要求2所述的人机协作方法,其特征在于,所述根据所述三维人体姿态数据流提取所述人体姿态特征,包括:

获取所述三维人体姿态数据流中的人体姿态关节点坐标数据;

根据所述人体姿态关节点坐标数据构建人体姿态关节点拓扑图;

根据所述人体姿态关节点拓扑图提取所述人体姿态特征。

6.根据权利要求5所述的人机协作方法,其特征在于,所述根据所述人体姿态关节点拓扑图提取所述人体姿态特征,包括:

将所述人体姿态关节点拓扑图输入预先经过训练的图神经网络中,得到所述图神经网络基于所述人体姿态关节点拓扑图输出的所述人体姿态特征。

7.根据权利要求1所述的人机协作方法,其特征在于,所述根据所述视觉语义深层特征和所述人体姿态特征,确定所述作业人员对应的目标人体行为意图,包括:

根据所述视觉语义深层特征,确定所述人体行为对应的视觉特征类别;

根据所述人体姿态特征，确定所述人体行为对应的姿态特征类别；

根据所述视觉特征类别和所述姿态特征类别，确定所述目标人体行为意图。

8.根据权利要求7所述的人机协作方法，其特征在于，所述根据所述视觉特征类别和所述姿态特征类别，确定所述目标人体行为意图，包括：

获取若干种权重分配规则，其中，所述若干种权重分配规则中每一权重分配规则对应的人体行为意图不同；

确定所述视觉特征类别和所述姿态特征类别在每一所述权重分配规则下的加权和，得到所述若干种权重分配规则分别对应的加权和；

根据所述若干种权重分配规则分别对应的加权和，确定目标人体行为意图，其中，所述目标人体行为意图所对应的权重分配规则的加权和最大。

9.根据权利要求1所述的人机协作方法，其特征在于，所述根据所述目标人体行为意图对移动式协作机器人进行控制，包括：

根据所述目标人体行为意图确定所述移动式协作机器人对应的执行操作和移动路径；

根据所述执行操作和所述移动路径对所述移动式协作机器人进行控制。

10.一种基于多模态行为在线预测的人机协作系统，其特征在于，所述系统包括：

视频采集单元，用于获取视频数据；

特征确定单元，用于根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征，所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；

意图确定单元，用于根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；

人机协作单元，用于根据所述目标人体行为意图对移动式协作机器人进行控制。

一种基于多模态行为在线预测的人机协作方法和系统

技术领域

[0001] 本发明涉及人机协同智能制造装配领域，尤其涉及的是一种基于多模态行为在线预测的人机协作方法和系统。

背景技术

[0002] 现有产品制造模式中，产品装配是整个制造生命周期中时间和精力耗费量最大的环节之一。据统计，在工业化国家的产品生产过程中，大约1/3左右人力从事于有关产品装配的活动，该阶段占用超过40％的生产成本。同时，由于产品的复杂性或个性化发展趋势，极大制约了现有装配的自动化和智能化水平，使得手工装配仍然是现有的主流装配方式之一，繁重紧张的装配任务会增加人员的疲劳程度，进而影响整个产品的生产装配质量，不科学的装配工艺以及工作环境影响着员工的工作状态甚至危害人的健康，降低了工作效率。因此，现有的手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式。

[0003] 因此，现有技术还有待改进和发展。

发明内容

[0004] 本发明要解决的技术问题在于，针对现有技术的上述缺陷，提供一种基于多模态行为在线预测的人机协作方法和系统，旨在解决现有技术中手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式的问题。

[0005] 本发明解决问题所采用的技术方案如下：

[0006] 第一方面，本发明实施例提供一种基于多模态行为在线预测的人机协作方法，其中，所述方法包括：

[0007] 获取视频数据；

[0008] 根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；

[0009] 根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；

[0010] 根据所述目标人体行为意图对移动式协作机器人进行控制。

[0011] 在一种实施方法中，所述根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征，包括：

[0012] 根据所述视频数据输出基础三原色视频流和三维人体姿态数据流；

[0013] 根据所述基础三原色视频流提取所述视觉语义深层特征；

[0014] 根据所述三维人体姿态数据流提取所述人体姿态特征。

[0015] 在一种实施方法中，所述根据所述基础三原色视频流提取所述视觉语义深层特

征，包括：

[0016] 对所述基础三原色视频流中每一视频帧对应的人体行为区域进行剪裁，得到若干人体行为区域视频帧；

[0017] 根据所述若干人体行为区域视频帧确定所述人体行为对应的视觉模态浅层特征，并根据所述视觉模态浅层特征提取所述视觉语义深层特征，其中，所述视觉模态浅层特征用于反映所述若干人体行为区域视频帧中的视觉几何特征。

[0018] 在一种实施方法中，所述根据所述人体行为区域视频帧确定所述人体行为对应的视觉模态浅层特征，并根据所述视觉模态浅层特征提取所述视觉语义深层特征，包括：

[0019] 将所述若干人体行为区域视频帧输入预先经过训练的二维卷积神经网络中，得到所述视觉模态浅层特征；

[0020] 将所述视觉模态浅层特征输入预先经过训练的三维卷积神经网络中，得到所述视觉语义深层特征。

[0021] 在一种实施方法中，所述根据所述三维人体姿态数据流提取所述人体姿态特征，包括：

[0022] 获取所述三维人体姿态数据流中的人体姿态关节点坐标数据；

[0023] 根据所述人体姿态关节点坐标数据构建人体姿态关节点拓扑图；

[0024] 根据所述人体姿态关节点拓扑图提取所述人体姿态特征。

[0025] 在一种实施方法中，所述根据所述人体姿态关节点拓扑图提取所述人体姿态特征，包括：

[0026] 将所述人体姿态关节点拓扑图输入预先经过训练的图神经网络中，得到所述图神经网络基于所述人体姿态关节点拓扑图输出的所述人体姿态特征。

[0027] 在一种实施方法中，所述根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图，包括：

[0028] 根据所述视觉语义深层特征，确定所述人体行为对应的视觉特征类别；

[0029] 根据所述人体姿态特征，确定所述人体行为对应的姿态特征类别；

[0030] 根据所述视觉特征类别和所述姿态特征类别，确定所述目标人体行为意图。

[0031] 在一种实施方法中，所述根据所述视觉特征类别和所述姿态特征类别，确定所述目标人体行为意图，包括：

[0032] 获取若干种权重分配规则，其中，所述若干种权重分配规则中每一权重分配规则对应的人体行为意图不同；

[0033] 确定所述视觉特征类别和所述姿态特征类别在每一所述权重分配规则下的加权和，得到所述若干种权重分配规则分别对应的加权和；

[0034] 根据所述若干种权重分配规则分别对应的加权和，确定目标人体行为意图，其中，所述目标人体行为意图所对应的权重分配规则的加权和最大。

[0035] 在一种实施方法中，所述根据所述目标人体行为意图对移动式协作机器人进行控制，包括：

[0036] 根据所述目标人体行为意图确定所述移动式协作机器人对应的执行操作和移动路径；

[0037] 根据所述执行操作和所述移动路径对所述移动式协作机器人进行控制。

[0038] 第二方面，本发明实施例还提供一种基于多模态行为在线预测的人机协作系统，其中，所述系统包括：

[0039] 视频采集单元，用于获取视频数据；

[0040] 特征确定单元，用于根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征，所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；

[0041] 意图确定单元，用于根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；

[0042] 人机协作单元，用于根据所述目标人体行为意图对移动式协作机器人进行控制。

[0043] 第三方面，本发明实施例还提供一种终端，其中，所述终端包括有存储器和一个或者一个以上处理器；所述存储器存储有一个或者一个以上的程序；所述程序包含用于执行如上述任一所述的人机协作筛选方法的指令；所述处理器用于执行所述程序。

[0044] 第四方面，本发明实施例还提供一种计算机可读存储介质，其上存储有多条指令，其中，由处理器加载并执行所述指令，以实现上述任一所述的人机协作方法的步骤。

[0045] 本发明的有益效果：本发明实施例通过获取视频数据；根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；根据所述目标人体行为意图确定移动式协作机器人对应的执行操作和移动路径。解决了现有技术中手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式的问题。

附图说明

[0046] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0047] 图1是本发明实施例提供的人机协作方法的流程示意图。

[0048] 图2是本发明实施例提供的人机协作方法的详细模块流程图。

[0049] 图3是本发明实施例提供的作业车间内部示意图。

[0050] 图4是本发明实施例提供的视频关键帧的抽取方法。

[0051] 图5是本发明实施例提供的人体姿态拓扑图的示意图。

[0052] 图6是本发明实施例提供的图神经网络、二维卷积神经网络、三维卷积神经网络之间的连接示意图。

[0053] 图7是本发明实施例提供的人机协作系统的内部模块示意图。

[0054] 图8是本发明实施例提供的终端的原理框图。

具体实施方式

[0055] 为使本发明的目的、技术方案及优点更加清楚、明确，以下参照附图并举实施例对

本发明进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0056] 需要说明，若本发明实施例中有涉及方向性指示(诸如上、下、左、右、前、后……)，则该方向性指示仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等，如果该特定姿态发生改变时，则该方向性指示也相应地随之改变。

[0057] 现有产品制造模式中，产品装配是整个制造生命周期中时间和精力耗费量最大的环节之一。据统计，在工业化国家的产品生产过程中，大约1/3左右人力从事于有关产品装配的活动，该阶段占用超过40％的生产成本。同时，由于产品的复杂性或个性化发展趋势，极大制约了现有装配的自动化和智能化水平，使得手工装配仍然是现有的主流装配方式之一，繁重紧张的装配任务会增加人员的疲劳程度，进而影响整个产品的生产装配质量，不科学的装配工艺以及工作环境影响着员工的工作状态甚至危害人的健康，降低了工作效率。因此，现有的手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式。

[0058] 针对现有技术的上述缺陷，本发明提供了一种基于多模态行为在线预测的人机协作方法，通过获取视频数据；根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；根据所述目标人体行为意图确定移动式协作机器人对应的执行操作和移动路径。解决了现有技术中手工装配模式需要耗费大量的装配时间，难以适应工业技术体系中生命周期逐渐缩短、产品创新日益加快的发展模式的问题。

[0059] 如图1所示，所述方法包括如下步骤：

[0060] 步骤S100、获取视频数据。

[0061] 具体地，本实施例首先需要获取作业车间内的视频数据，由于该视频数据可以反映作业人员的人体行为，因此可以基于该视频数据调控移动式协作机器人，从而实现人机协作。

[0062] 在一种实现方式中，如图图3所示，可以在作业车间1内设置一个相机4，通过该相机采集作业车间的视频数据，以观测作业人员3的人体行为。为了采集到准确的视频数据，本实施例启动该相机4后，还需要对该相机4进行调试，调试过程包括但不限于：对该相机进行初始化操作、设置该相机的分辨率和采样帧率、调整该相机的视场和深度模式。调试完毕以后，即可通过该相机采集作业车间的视频数据，后续可以通过该视频数据调控移动式协作机器人5，活动于零部件装配区域2和零件工具存储区6协助作业人员3进行作业任务。

[0063] 在一种实现方式中，如图2所示，所述相机可以采用Azure Kinect相机，其中，Azure Kinect相机包含有RGB-D相机和IR红外相机。

[0064] 如图1所示，所述方法还包括如下步骤：

[0065] 步骤S200、根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息。

[0066] 具体地，为了使终端准确地获知作业人员的人体行为的具体类别，本实施例需要通过该视频数据获取两种信息，即作业人员的人体行为所对应的视觉语义深层特征和人体

姿态特征。其中，视觉语义深层特征可以反映人体行为在时序性视觉模式下的时空间语义信息，而人体姿态特征可以反映人体骨架对应的姿态信息，例如人体关节点的相对位置的变化可以反映不同的人体姿态。由于本实施例是采用结合视觉语义深层特征和人体姿态特征两种信息确定作业人员当前发生的人体行为，因此相较于仅通过一种信息判定人体行为类别的方法，可以更准确地确定作业人员当前发生的人体行为，降低判定错误的概率。

[0067] 在一种实现方式中，所述根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征，具体包括如下步骤：

[0068] 步骤S201、根据所述视频数据输出基础三原色视频流和三维人体姿态数据流；

[0069] 步骤S202、根据所述基础三原色视频流提取所述视觉语义深层特征；

[0070] 步骤S203、根据所述三维人体姿态数据流提取所述人体姿态特征。

[0071] 具体地，由于该视频数据中捕获了作业车间内的作业人员的行为数据，因此可以根据该视频数据输出基础三原色视频流(即RGB视频流)和三维人体姿态数据流。其中，基础三原色视频流中包含每一视频帧的图像信息，例如颜色、纹理、轮廓等信息，因此本实施例可以通过基础三原色视频流提取视觉语义深层特征。而三维人体姿态数据流中包含有每一视频帧中人体运动的姿态的跟踪信息，因此本实施例可以通过三维人体姿态数据流提取人体姿态特征。

[0072] 在一种实现方式中，所述步骤S202具体包括如下步骤：

[0073] 步骤S2021、对所述基础三原色视频流中每一视频帧对应的人体行为区域进行剪裁，得到若干人体行为区域视频帧；

[0074] 步骤S2022、根据所述若干人体行为区域视频帧确定所述人体行为对应的视觉模态浅层特征，并根据所述视觉模态浅层特征提取所述视觉语义深层特征，其中，所述视觉模态浅层特征用于反映所述若干人体行为区域视频帧中的视觉几何特征。

[0075] 具体地，由于每一视频帧中包含有许多冗余的图像信息，例如作业车间的环境背景产生的图像信息，这些图像信息对于确定作业人员的人体行为都是不必要的，因此本实施例首先需要对获得的基础三原色视频流中每一视频帧对应的人体行为区域进行剪裁，得到若干人体行为区域视频帧。可以理解的是，人体行为区域即为作业人员的骨架所运动的最大区域。本实施例首先对人体行为区域视频帧进行第一次特征提取，得到视觉模态浅层特征，所述视觉模态浅层特征用于反映每一视频帧中的视觉几何特征，例如图像的颜色、形状等信息。再对视觉模态浅层特征进行第二次特征提取，得到视觉语义深层特征，所述视觉语义深层特征用于反映每一视频帧中的人体行为的抽象语义特征，即相当于人体行为在时序性视觉模式下的时空间语义信息。

[0076] 为了得到人体行为区域视频帧，同时减少内存消耗，在一种实现方式中，本实施例可以将所述基础三原色视频流等长分割为若干视频流，针对所述若干视频流中每一视频流，从该视频流中按照预设规则抽取一帧视频帧作为视频关键帧，得到所述若干视频流对应的若干视频关键帧。再对所述若干视频关键帧中的人体行为区域进行剪裁，得到所述若干人体区域视频帧。

[0077] 具体地，首先通过所述三维人体姿态数据流获取三维人体姿态关节点坐标信息。然后，获取相机的内参数和外参数，针对每一视频关键帧，根据所述三维人体姿态关节点坐标信息、所述内参数以及所述外参数计算该视频关键帧中的人体关节点坐标，根据所述人

体关节点坐标计算出人体行为区域的四个端点坐标，根据四个端点坐标计算出人体行为区域宽度和人体行为区域高度。并根据所述人体关节点坐标确定人体脖颈关节点坐标和人体髋关节点坐标，根据所述人体脖颈关节点坐标和所述人体髋关节点坐标，计算出尺度参考阈值。根据所述人体行为区域宽度、所述人体行为区域高度以及所述尺度参考阈值，确定裁剪区域宽度和裁剪区域高度。最后，根据所述裁剪区域宽度和所述裁剪区域高度对该视频关键帧进行裁剪，即得到该视频关键帧对应的人体区域视频帧。

[0078] 举例说明，如图4所示，将RGB视频流等长分割为前次视频流41和当前视频流42，然后将前次视频帧41和当前视频流42等间距划分至不同的视频子集411，从前次视频帧41的后三分之一视频子集411中随机选取单个视频帧4111作为视频关键帧43，从当前视频流42的全部视频子集411中随机选取单个视频帧4111作为视频关键帧43。由于所述的视频关键帧43每次更新三分之一的视频帧数，因此可以实现减少内存消耗。同时，根据获取的3D人体姿态关节点坐标$P_w=(X_i,Y_i,Z_i)$，结合Azure Kinect相机4(如图3所示)的RGB相机内参K_c、RGB相机外参T_c和深度相机外参T_d，计算视频关键帧43中的人体关节点坐标$P_{uv-c}=K_cT_cT_d^{-1}P_d=(x_i,y_i)$；根据所述视频关键帧43中的人体关节点坐标$P_{uv-c}=(x_i,y_i)$，计算人体行为区域坐标31(如图5所示)的数值$(x_1,y_1,x_2,y_2)=(minx_i,miny_i,maxx_i,maxy_i)$，根据所述视频关键帧43中的人体脖颈关节点坐标$P_{uv-c}^3=(x_3,y_3)$和人体髋关节点坐标$P_{uv-c}^4=(x_4,y_4)$，计算尺度参考阈值32(如图5所示)的数值$d_c=\left\|P_{uv-c}^4-P_{uv-c}^3\right\|$；根据所述的人体行为区域坐标31的数值$(x_1,y_1,x_2,y_2)$，计算人体行为区域宽度$w_c=x_4-x_3$和高度$h_c=y_4-y_3$，结合尺度参考阈值32的数值$d_c$，计算裁剪区域宽度$w_m=max\{\lfloor w_c,h_c/2\rfloor\}+d_c/2$和高度$h_m=2\times w_m$，以此尺寸从视频关键帧43中裁剪人体行为区域，避免了对Azure Kinect相机4采集的视频流进行额外的人体目标检测任务。

[0079] 在一种实现方式中，为了得到视觉模态浅层特征和视觉语义深层特征，本实施例可以将所述若干人体行为区域视频帧输入预先经过训练的二维卷积神经网络中，得到所述视觉模态浅层特征；将所述视觉模态浅层特征输入预先经过训练的三维卷积神经网络中，得到所述视觉语义深层特征(如图2所示)。

[0080] 具体地，本实施例预先训练了一个二维卷积神经网络和一个三维卷积神经网络，由于两个卷积神经网络的维度不同，因此提取的特征维度也不相同。将得到的多个人体行为区域视频帧输入该二维卷积神经网络中，该二维神经网络即可提取出每一帧人体行为区域视频帧中的图像特征，得到视觉模态浅层特征，由于该视觉模态浅层特征反映的是人体行为区域视频帧中的视觉几何信息，例如形状，位置，边缘，感兴趣区域等信息，是低级的视觉特征，因此还需要将视觉模态浅层特征输入该三维卷积神经网络中，得到视觉语义深层特征，由于该视觉语义深层特征反映的是作业人员的人体行为在时序性视觉模式下的时空间语义信息，因此根据该视觉语义深层特征辅助终端正确理解人体行为区域视频帧中的人体行为。

[0081] 在一种实现方式中，所述步骤S203具体包括如下步骤：

[0082] 步骤S2031、获取所述三维人体姿态数据流中的人体姿态关节点坐标数据；

[0083] 步骤S2032、根据所述人体姿态关节点坐标数据构建人体姿态关节点拓扑图；

[0084] 步骤S2033、根据所述人体姿态关节点拓扑图提取所述人体姿态特征。

[0085] 具体地，为了得到人体姿态特征，本实施例首先需要获取三维人体姿态数据流中的人体姿态关节点坐标数据，例如可以通过相机中的人体姿态跟踪API获取三维人体姿态数据流中的人体姿态关节点坐标数据。然后根据构建人体姿态关节点拓扑图，其中，在所述人体姿态关节点拓扑图中，同一时序数据下相邻关节点互相连接，连续时序数据下同一关节点互相连接。由于该人体姿态关节点拓扑图中各关节点的相对位置关系可以反映作业人员的人体行为的姿态信息，因此对该人体姿态关节点拓扑图进行特征提取，即可得到人体姿态特征。

[0086] 在一种实现方式中，为了得到人体姿态特征，本实施例可以将所述人体姿态关节点拓扑图输入预先经过训练的图神经网络中，得到所述图神经网络基于所述人体姿态关节点拓扑图输出的所述人体姿态特征。

[0087] 具体地，本实施例预先训练一个用于提取图像特征的图神经网络，获得人体姿态关节点拓扑图以后，将其作为图神经网络的输入图像，图神经网络即可对其进行图像特征提取，并输出与该人体姿态关节点拓扑图对应的人体姿态特征，该人体姿态特征用于反映作业人员的人体行为的姿态信息。

[0088] 在一种实现方式中，如图6所示，图神经网络包含九层图卷积层和一层极值特征组，所述极值特征组连接在第六层图卷积层后，用于划分不同人体的姿态关节拓补图特征至不同小组，并将每组中的姿态关节拓补图的最大特征值作为该组对应的最大值。二维卷积神经网络包含2D卷积层和2D池化层，三维卷积神经网络包含四层3D卷积层和一层3D池化层。本实施例可以通过两次多模态人体行为特征融合，从而实现学习融合的可用数据流的全部知识信息。其中，第一次融合为：三维卷积神经网络中第三层3D卷积层的输出数据和图神经网络中极值特征组的输出数据进行融合，得到第一融合数据，并将所述第一融合数据作为三维卷积神经网络中第四层3D卷积层的输入数据；第二次融合为：三维卷积神经网络中第四层3D卷积层的输出数据和图神经网络中第九层图卷积层的输出数据进行融合，得到第二融合数据，并将所述第二融合数据作为三维卷积神经网络中3D池化层的输入数据和图神经网络中最后一层卷积层的输入数据。

[0089] 如图1所示，所述方法还包括如下步骤：

[0090] 步骤S300、根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图。

[0091] 具体地，虽然基于视觉语义深层特征和人体姿态特征中任意一种特征，也可以确定作业人员的人体行为意图，即预测作业人员的下一步行为，但是这样的确定方法的准确性不高。因此本实施例选择采用两种特征，即视觉语义深层特征和人体姿态特征相结合来确定作业人员对应的目标人体行为意图，该目标人体行为意图即可准确地反映作业人员的下一步行为。

[0092] 在一种实现方式中，所述步骤S300具体包括如下步骤：

[0093] 步骤S301、根据所述视觉语义深层特征，确定所述人体行为对应的视觉特征类别；

[0094] 步骤S302、根据所述人体姿态特征，确定所述人体行为对应的姿态特征类别；

[0095] 步骤S303、根据所述视觉特征类别和所述姿态特征类别，确定所述目标人体行为意图。

[0096] 具体地，本实施例可以将所述视觉语义深层特征输入预先训练好的第一分类器中，通过第一分类器确定所述视觉语义深层特征对应的视觉特征类别；同理，将所述人体姿态特征输入预先训练好的第二分类器中，通过第二分类器确定所述人体姿态特征对应的姿态特征类别，然后结合所述视觉特征类别和所述姿态特征类别，确定目标人体行为意图。

[0097] 在一种实现方式中，为了确定目标人体行为意图，本实施例可以获取若干种权重分配规则，其中，所述若干种权重分配规则中每一权重分配规则对应的人体行为意图不同；确定所述视觉特征类别和所述姿态特征类别在每一所述权重分配规则下的加权和，得到所述若干种权重分配规则分别对应的加权和；根据所述若干种权重分配规则分别对应的加权和，确定目标人体行为意图，其中，所述目标人体行为意图所对应的权重分配规则的加权和最大。

[0098] 具体地，本实施例预先设置了多种权重分配规则，每一种权重分配规则对应一个人体行为意图，不同权重分配规则中所述视觉特征类别和所述姿态特征类别分别对应的权重值不同，即本实施例在确定不同人体行为意图时，会为视觉特征类别和姿态特征类别设置不同的权重值，以区分视觉特征类别和姿态特征类别的重要程度。计算出视觉特征类别和姿态特征类别在每一所述权重分配规则下的加权和，并将加权和的数值最大的一种权重分配规则所对应的人体行为意图作为目标人体行为意图。

[0099] 在一种实现方式中，得到目标人体行为意图后，还需要判断目标人体行为意图是否符合作业种类的预定义约束条件，即是否属于预设的若干个装配工艺类别中的一种，若不属于则说明该目标人体行为意图无效，需要重新获取人体行为意图。

[0100] 如图1所示，所述方法还包括如下步骤：

[0101] 步骤S400、根据所述目标人体行为意图对移动式协作机器人进行控制。

[0102] 具体地，为了实现人机协作装配，在确定目标人体行为意图之后，就可以基于该目标人体行为意图准确地获知作业人员的下一步操作，因此可以根据该目标人体行为意图对作业车间内的移动式协作机器人进行控制，使移动式协作机器人可以做出相应的操作或者移动到相应的地方以配合作业人员执行其下一步的操作，从而实现人机协作装配。

[0103] 在一种实现方式中，当Azure Kinect相机完成初始化操作以后，需要进行RGB-D相机标定获取世界坐标系，根据得到的世界坐标系判断Azure Kinect相机的标定精度是否满足预设的尺寸约束条件，当满足时启动移动式协作机器人进行手眼标定，手眼标定完毕以后即可统一Azure Kinect相机和移动式协作机器人的空间坐标系。在一种实现方式中，为满足作业车间的实际生产需求和作业质量要求，所述尺寸约束条件为Azure Kinect相机4标定精度误差小于0.5cm。

[0104] 在一种实现方式中，所述步骤S400具体包括如下步骤：

[0105] 步骤S401、根据所述目标人体行为意图确定所述移动式协作机器人对应的执行操作和移动路径；

[0106] 步骤S402、根据所述执行操作和所述移动路径对所述移动式协作机器人进行控制。

[0107] 具体地，本实施例可以根据得到的目标人体行为意图使移动式协作机器人做出相应的操作和移动到相应的地方以配合作业人员执行其下一步的操作，从而实现人机协作装配。

[0108] 在一种实现方式中，在确定所述移动式协作机器人对应的执行操作和移动路径后，还需要确定所述执行操作和所述移动路径是否满足安全约束条件，其中，所述安全约束条件为所述作业人员与所述移动式协作机器人之间的距离大于预设距离阈值，例如20cm。

[0109] 根据所述目标人体行为意图对移动式协作机器人进行控制。

[0110] 基于上述实施例，本发明还提供了一种基于多模态行为在线预测的人机协作系统，如图7所示，该系统包括：

[0111] 视频采集单元01，用于获取视频数据；

[0112] 特征确定单元02，用于根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征，所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息；

[0113] 意图确定单元03，用于根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图；

[0114] 人机协作单元04，用于根据所述目标人体行为意图对移动式协作机器人进行控制。

[0115] 基于上述实施例，本发明还提供了一种终端，其原理框图可以如图8所示。该终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏。其中，该终端的处理器用于提供计算和控制能力。该终端的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现人机协作方法。该终端的显示屏可以是液晶显示屏或者电子墨水显示屏。

[0116] 本领域技术人员可以理解，图8中示出的原理框图，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的终端的限定，具体的终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0117] 在一种实现方式中，所述终端的存储器中存储有一个或者一个以上的程序，且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行人机协作方法的指令。

[0118] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读取存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM

(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0119] 综上所述,本发明公开了一种基于多模态行为在线预测的人机协作方法和系统,所述方法包括:获取视频数据;根据所述视频数据,确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征;所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息;根据所述视觉语义深层特征和所述人体姿态特征,确定所述作业人员对应的目标人体行为意图;根据所述目标人体行为意图确定移动式协作机器人对应的执行操作和移动路径。

[0120] 应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

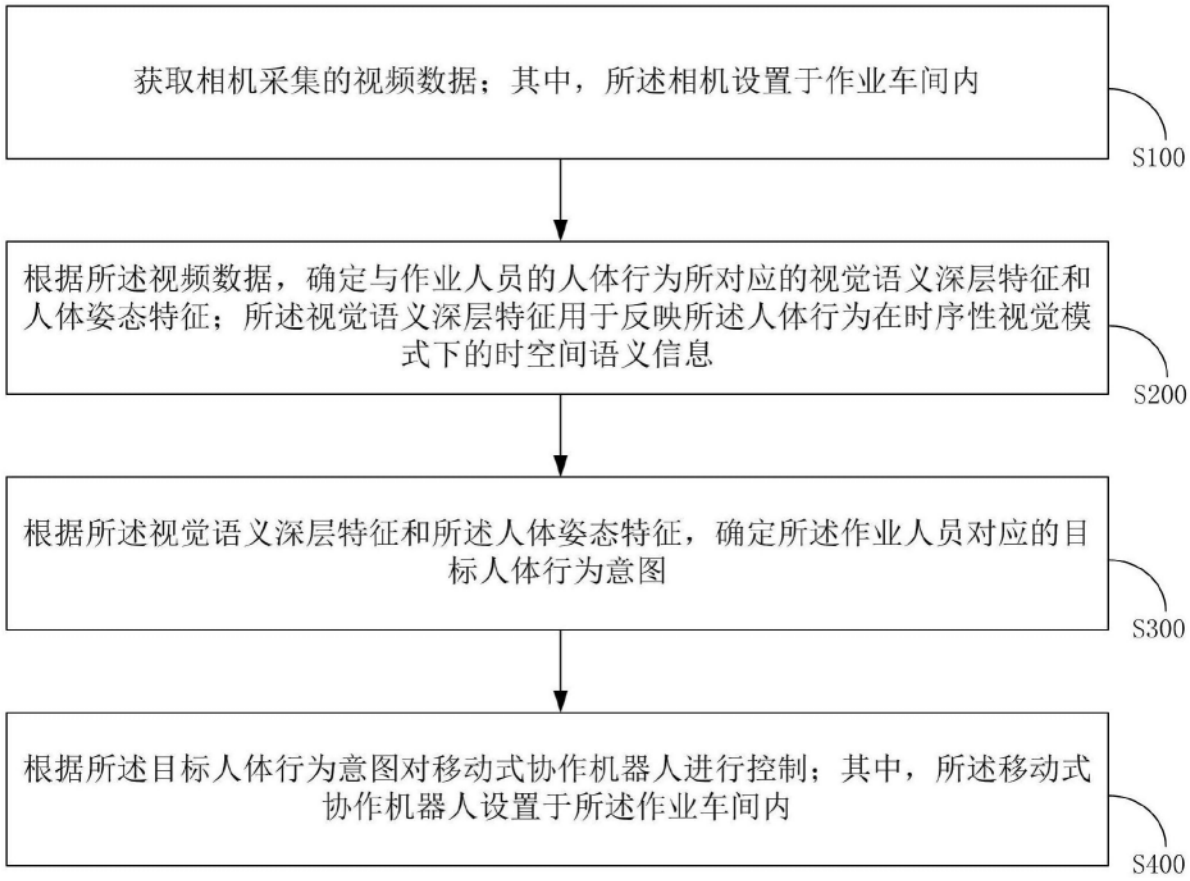

获取相机采集的视频数据；其中，所述相机设置于作业车间内
S100
根据所述视频数据，确定与作业人员的人体行为所对应的视觉语义深层特征和人体姿态特征；所述视觉语义深层特征用于反映所述人体行为在时序性视觉模式下的时空间语义信息
S200
根据所述视觉语义深层特征和所述人体姿态特征，确定所述作业人员对应的目标人体行为意图
S300
根据所述目标人体行为意图对移动式协作机器人进行控制；其中，所述移动式协作机器人设置于所述作业车间内
S400

图1

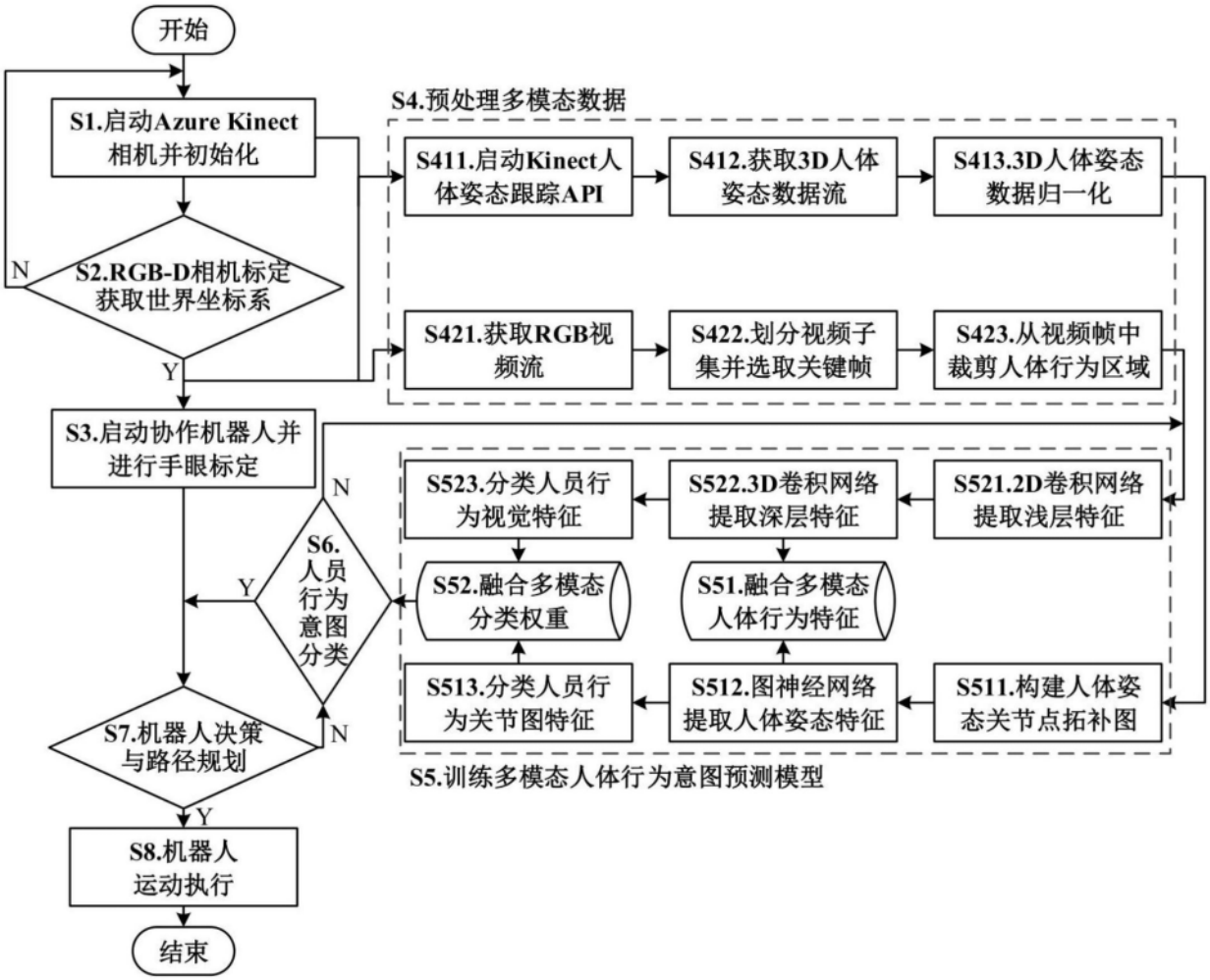

开始
S1.启动Azure Kinect相机并初始化
S2.RGB-D相机标定获取世界坐标系
N
Y
S3.启动协作机器人并进行手眼标定
S4.预处理多模态数据
S411.启动Kinect人体姿态跟踪API
S412.获取3D人体姿态数据流
S413.3D人体姿态数据归一化
S421.获取RGB视频流
S422.划分视频子集并选取关键帧
S423.从视频帧中裁剪人体行为区域
S523.分类人员行为视觉特征
S522.3D卷积网络提取深层特征
S521.2D卷积网络提取浅层特征
S52.融合多模态分类权重
S51.融合多模态人体行为特征
S513.分类人员行为关节图特征
S512.图神经网络提取人体姿态特征
S511.构建人体姿态关节点拓补图
S5.训练多模态人体行为意图预测模型
S6.人员行为意图分类
S7.机器人决策与路径规划
S8.机器人运动执行
结束

图2

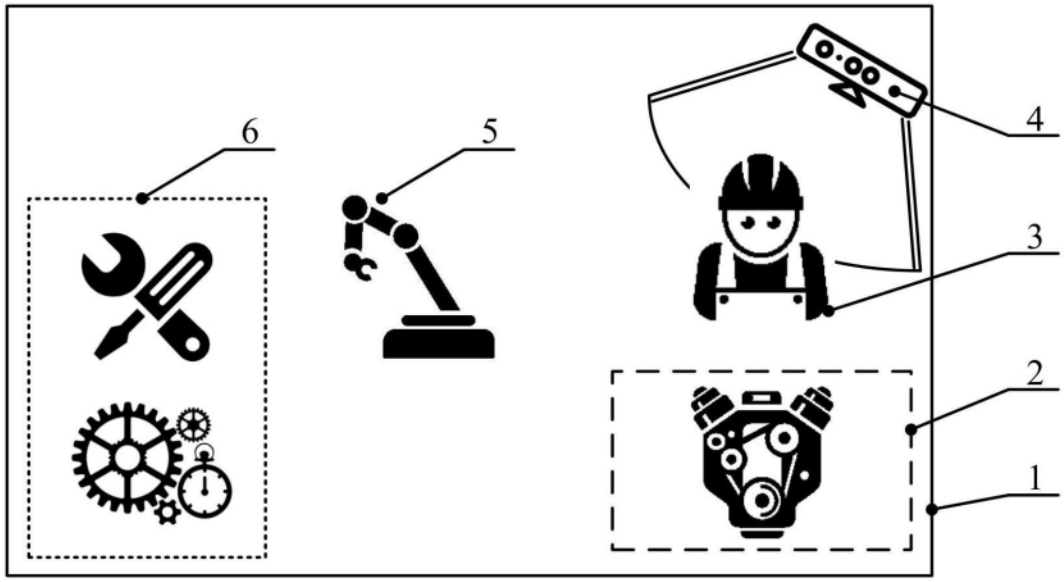

6
5
4
3
2
1

图3

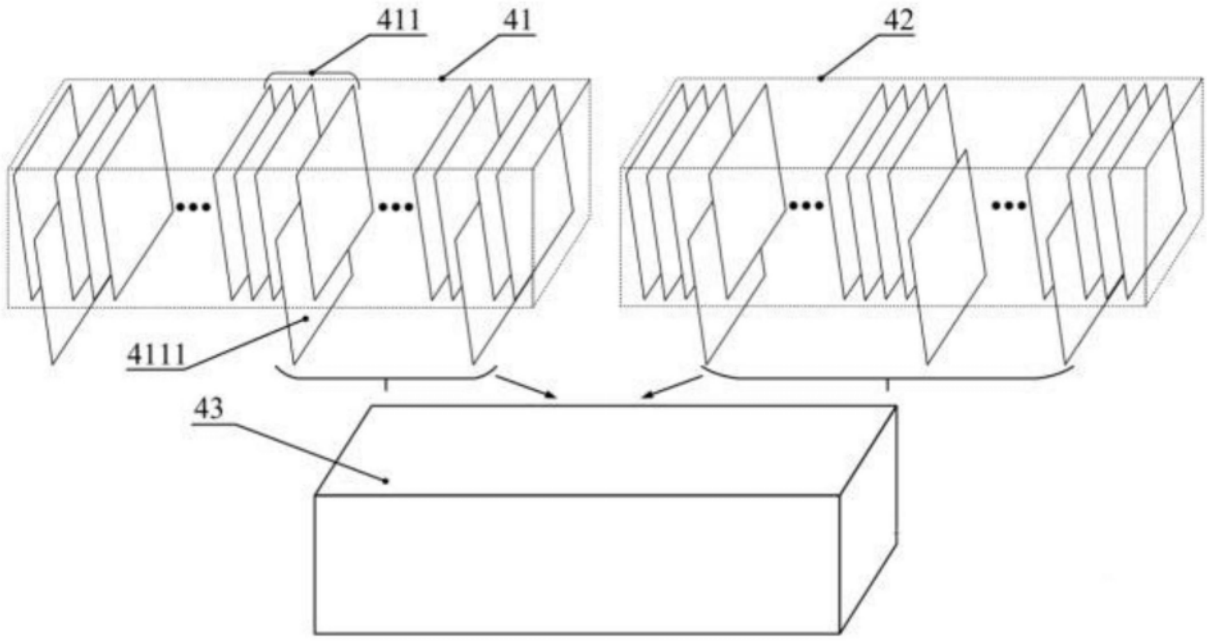


图4

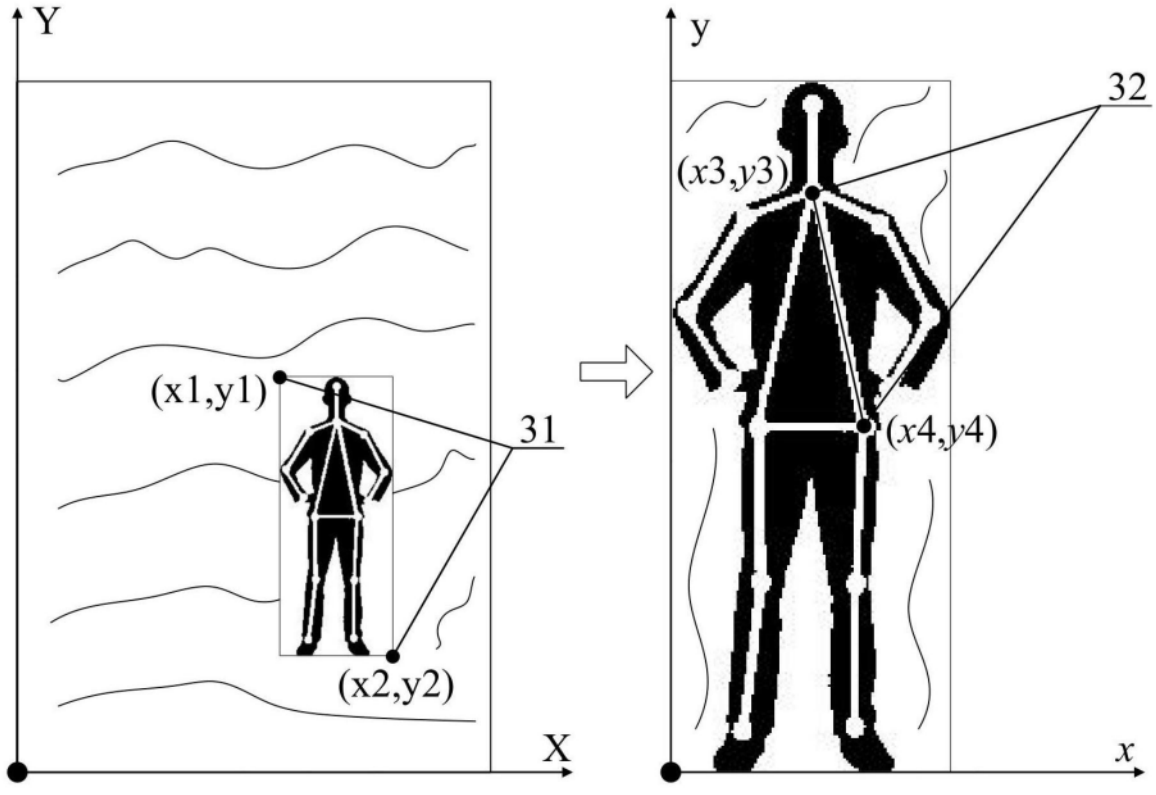


图5

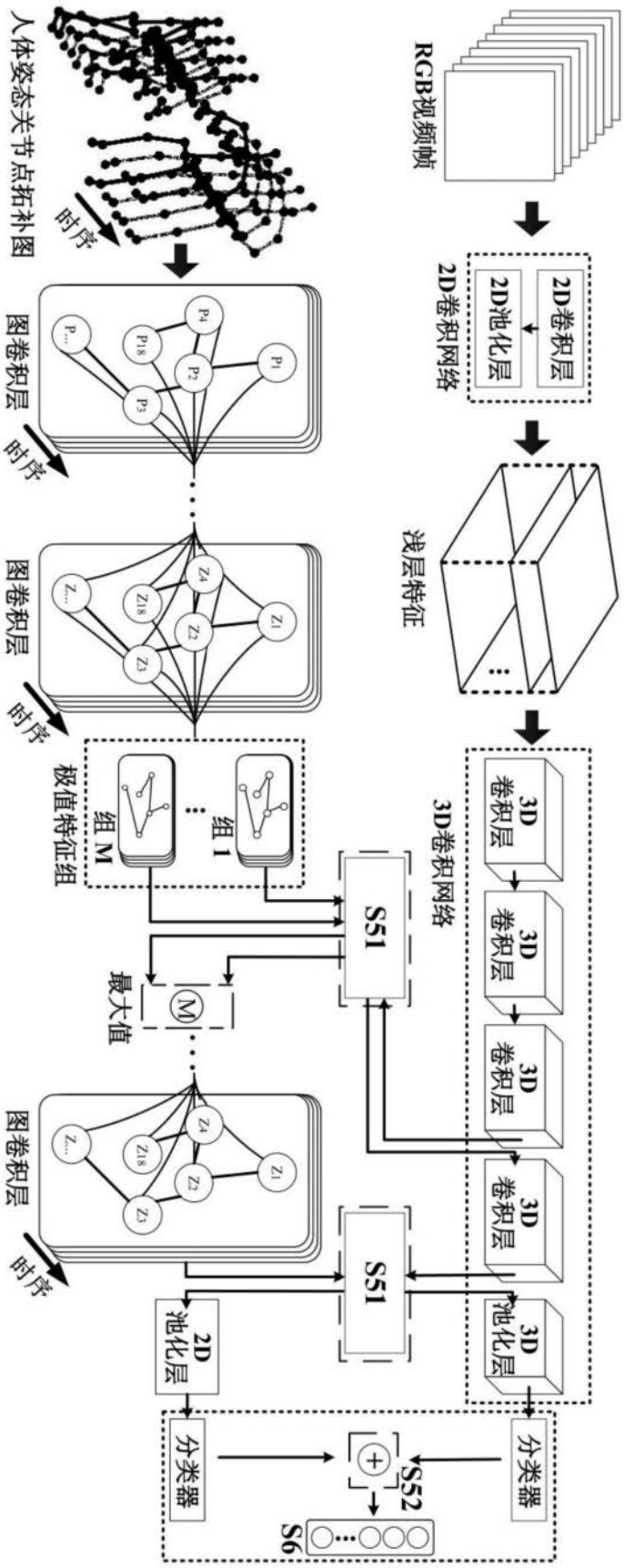

RGB视频帧
2D卷积层
2D池化层
2D卷积网络
浅层特征
3D
卷积层
3D
卷积层
3D
卷积层
3D
卷积层
3D
池化层
3D卷积网络
S51
S51
人体姿态关节点拓扑图
时序
图卷积层
时序
图卷积层
时序
极值特征组
组 M
组 1
最大值
图卷积层
时序
2D
池化层
分类器
分类器
S52
S6

图6

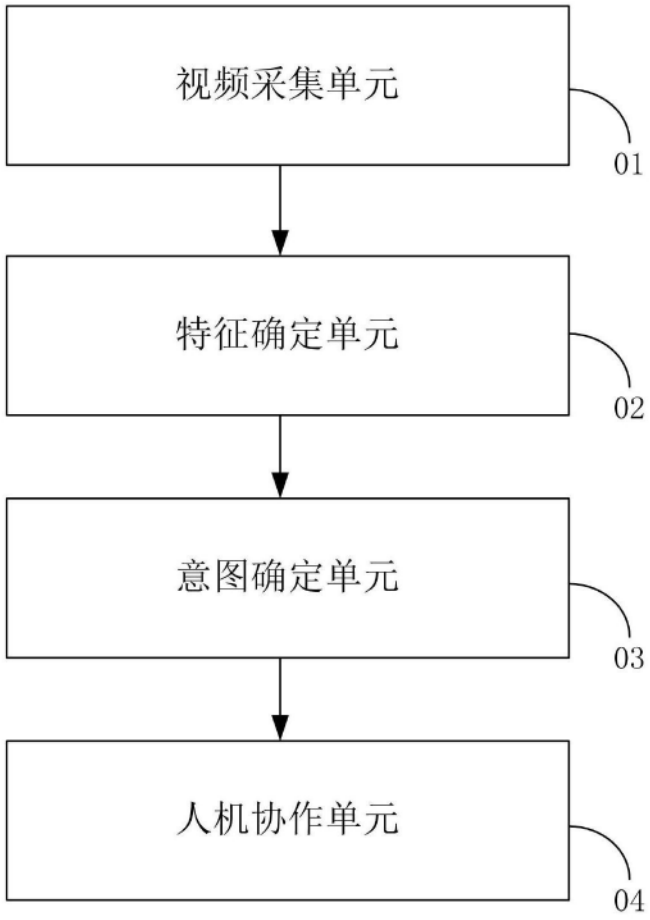


图7

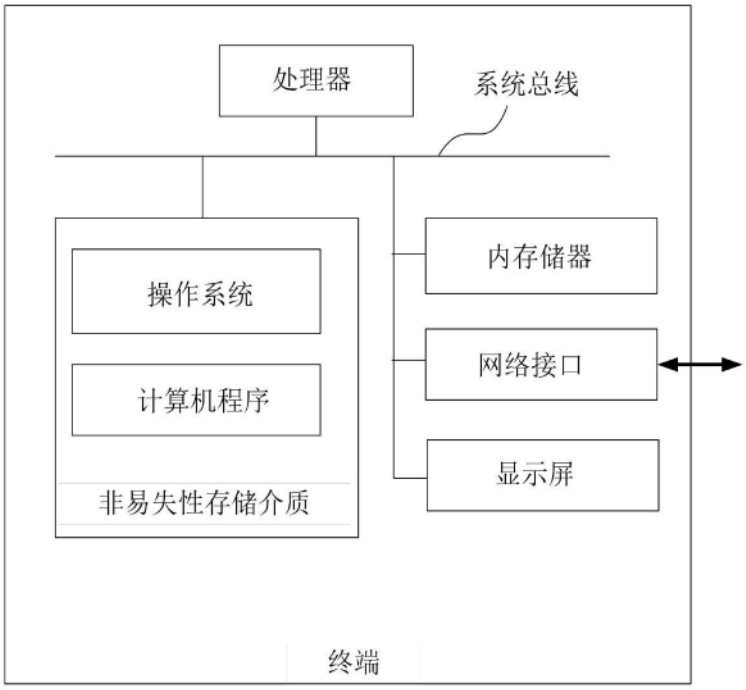


图8