

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 10 月 8 日 (08.10.2020)



(10) 国际公布号
WO 2020/200082 A1

(51) 国际专利分类号:
H04N 21/478 (2011.01)

(21) 国际申请号: PCT/CN2020/081627

(22) 国际申请日: 2020 年 3 月 27 日 (27.03.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910251306.7 2019年3月29日 (29.03.2019) CN
201910252787.3 2019年3月29日 (29.03.2019) CN

(71) 申请人: 广州虎牙信息科技有限公司 (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市番禺区钟村街 (汉溪商业中心) 泽溪街13号1401, Guangdong 511400 (CN)。

(72) 发明人: 徐子豪(XU, Zihao); 中国广东省广州市番禺区钟村街 (汉溪商业中心) 泽溪街13号1401,

Guangdong 511400 (CN)。 吴昊(WU, Hao); 中国广东省广州市番禺区钟村街 (汉溪商业中心) 泽溪街13号1401, Guangdong 511400 (CN)。

(74) 代理人: 北京超凡宏宇专利代理事务所 (特殊普通合伙) (CHOFN INTELLECTUAL PROPERTY); 中国北京市海淀区北四环西路68号左岸工社1215-1218室, Beijing 100080 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) **Title:** LIVE BROADCAST INTERACTION METHOD AND APPARATUS, LIVE BROADCAST SYSTEM AND ELECTRONIC DEVICE

(54) 发明名称: 直播互动方法、装置、直播系统及电子设备

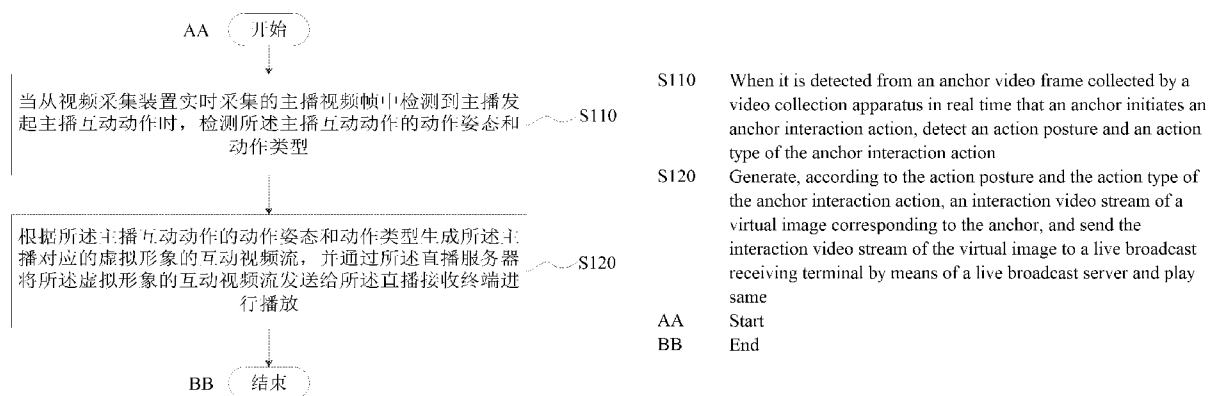


图 2

(57) **Abstract:** Provided are a live broadcast interaction method and apparatus, a live broadcast system and an electronic device. The method comprises: when it is detected from an anchor video frame collected by a video collection apparatus in real time that an anchor initiates an anchor interaction action, detecting an action posture and an action type of the anchor interaction action, wherein the anchor interaction action comprises a target prop wearing action and/or a target limb action; and then, generating, according to the action posture and the action type of the anchor interaction action, an interaction video stream of a virtual image corresponding to the anchor, and sending the interaction video stream of the virtual image to a live broadcast receiving terminal by means of a live broadcast server and playing same. Thus, by means of associating interaction content of a virtual image of an anchor with an action posture and an action type of an anchor interaction action, the interaction effect in a live broadcast process can be improved, manual operations when the anchor initiates virtual image interaction are reduced, and automatic interaction of the virtual image is achieved.

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本申请实施例提供了一种直播互动方法、装置、直播系统及电子设备, 当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时, 检测主播互动动作的动作姿态和动作类型, 其中, 主播互动动作包括佩戴目标道具和/或目标肢体动作。而后, 根据主播互动动作的动作姿态和动作类型生成主播对应的虚拟形象的互动视频流, 并通过直播服务器将虚拟形象的互动视频流发送给直播接收终端进行播放。如此, 通过将主播的虚拟形象的互动内容与主播互动动作的动作姿态和动作类型产生关联, 可以提高直播过程中的互动效果, 减少主播发起虚拟形象互动时的人为操作, 实现虚拟形象的自动互动。

直播互动方法、装置、直播系统及电子设备

相关申请的交叉引用

本申请要求于2019年3月29日提交中国专利局的申请号为2019102513067、名称为“直播互动方法、装置、直播系统及电子设备”的中国专利申请的优先权，以及于2019年3月29日提交中国专利局的申请号为2019102527873、名称为“虚拟形象控制方法、装置及电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及互联网技术领域，具体而言，涉及一种直播互动方法、装置、直播系统及电子设备。

背景技术

为了丰富主播和观众之间的互动，在网络直播过程中，在一些实施方式中，可以在直播界面展示虚拟形象，以通过该虚拟形象与观众进行互动。然而，该方案中虚拟形象仅仅只是单纯地演示某个互动动作，难以与主播产生动作关联，导致实际互动效果不佳。

发明内容

本申请提供一种电子设备，可以包括一个或多个存储介质和一个或多个与存储介质通信的处理器。一个或多个存储介质存储有处理器可执行的机器可执行指令。当电子设备运行时，处理器执行所述机器可执行指令，以执行直播互动方法。

本申请提供一种直播互动方法，应用于直播提供终端，所述方法包括：

当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型；

其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

根据所述主播互动动作的动作姿态和动作类型生成所述主播对应的虚拟形象的互动视频流，并通过所述直播服务器将所述虚拟形象的互动视频流发送给所述直播接收终端进行播放。

在一些可能的实现方式中，所述检测所述主播互动动作的动作姿态和动作类型的步骤，包括：

在检测到所述主播佩戴目标道具时，检测所述目标道具的道具属性和参考点位置向量，并根据所述道具属性查找所述目标肢体动作的动作类型；

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

在一些可能的实现方式中，所述检测所述主播互动动作的动作姿态和动作类型的步骤，包括：

在检测到所述主播发起目标肢体动作时，检测所述目标肢体动作的参考点位置向量，并采用互动动作识别模型识别所述目标肢体动作的动作类型；

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

在一些可能的实现方式中，所述根据所述参考点位置向量，采用逆向运动学算法预测所述主播互动动作的动作姿态的步骤，包括：

根据所述参考点位置向量计算所述主播的互动肢体的中心点高度以及所述主播的互动肢体相对于所述视频采集装置的姿态旋转矩阵；

根据所述姿态旋转矩阵、所述参考点位置向量和所述中心点高度，计算所述主播的互动肢体的各个肢体关节的位置向量，所述位置向量包括所述主播的互动肢体在各个参考轴方向上的分量；

根据计算得到的所述各个肢体关节的位置向量得到所述主播互动动作的动作姿态。

在一些可能的实现方式中，所述直播提供终端中预先存储有预设互动内容库，所述预设互动内容库包括各个动作类型对应的虚拟形象互动内容，所述虚拟形象互动内容包括对话互动内容、特效互动内容以及肢体互动内容中的一种或者多种组合；

所述根据所述主播互动动作的动作姿态和动作类型生成所述虚拟形象的互动视频流的

步骤, 包括:

从所述预设互动内容库中获取所述动作类型对应的虚拟形象互动内容;

根据所述动作姿态和所述虚拟形象互动内容生成所述虚拟形象的互动视频流。

在一些可能的实现方式中, 所述根据所述动作姿态和所述虚拟形象互动内容生成所述虚拟形象的互动视频流的步骤, 包括:

按照所述动作姿态关联的各个目标关节的位移坐标, 控制所述虚拟形象的各个目标关节沿对应的位移坐标移动, 并按照所述虚拟形象互动内容控制所述虚拟形象执行对应的互动动作, 以生成对应的互动视频流。

在一些可能的实现方式中, 当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时, 检测所述主播互动动作的动作姿态和动作类型的步骤, 包括:

将视频采集装置实时采集的主播视频帧输入到预先训练的互动动作识别模型中, 识别所述主播视频帧中是否包含目标肢体动作;

在检测到所述主播发起目标肢体动作时, 获得所述目标肢体动作的动作类型及所述目标肢体动作的参考点位置向量;

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

在一些可能的实现方式中, 所述互动动作识别模型包括输入层、至少一个卷积提取层、全连接层以及分类层, 每个卷积提取层包括依次设置的第一点卷积层、深度卷积层以及第二点卷积层, 所述卷积提取层内的每个卷积层之后设置一个激活函数层和池化层, 所述全连接层位于最后一个池化层之后, 所述分类层位于全连接层之后。

在一些可能的实现方式中, 所述互动动作识别模型还包括多个残差网络层, 每个残差网络层配置成将所述互动动作识别模型中任意相邻的两层的输出部分与该相邻的两层之后一层的输入部分串接。

在一些可能的实现方式中, 所述方法还包括预先训练所述互动动作识别模型的步骤, 具体包括:

建立神经网络模型;

采用公开数据集对所述神经网络模型进行预训练, 得到预训练神经网络模型;

采用收集数据集对所述预训练神经网络模型进行迭代训练, 得到所述互动动作识别模型, 其中, 所述收集数据集包括标记有不同主播互动动作的实际目标的训练样本图像集, 所述实际目标为所述主播互动动作在训练样本图像中的实际图像区域。

在一些可能的实现方式中, 所述采用收集数据集对所述预训练神经网络模型进行迭代训练, 得到所述互动动作识别模型的步骤, 包括:

将所述训练样本图像集中的每个训练样本图像输入到所述预训练神经网络模型的输入层进行预处理, 得到预处理图像;

针对所述预训练神经网络模型的每个卷积提取层, 通过该卷积提取层的第一点卷积层、深度卷积层以及第二点卷积层分别提取预处理图像的多维特征图像, 并将提取得到的多维特征图像输入到所连接的激活函数层中进行非线性映射, 而后将非线性映射后的多维特征图像输入到所连接的池化层中进行池化处理, 并将池化处理得到的池化特征图输入到下一层卷积层进行特征提取;

将最后一层池化层输出的池化特征图输入到全连接层, 得到全连接特征输出值;

将所述全连接特征输出值输入到分类层中进行预测目标分类, 得到每个训练样本图像的预测目标;

计算各个训练样本图像的预测目标与实际目标之间的损失函数值;

根据所述损失函数值进行反向传播训练, 并计算所述预训练神经网络模型的网络参数的梯度;

根据计算得到的所述梯度, 采用随机梯度下降法更新所述预训练神经网络模型的网络参数后继续训练, 直到所述预训练神经网络模型满足训练终止条件时, 输出训练得到的互

动动作识别模型。

在一些可能的实现方式中，所述根据所述损失函数值进行反向传播训练，并计算所述预训练神经网络模型的网络参数的梯度的步骤，包括：

根据所述损失函数值确定反向传播训练的反向传播路径；

通过所述预训练神经网络模型的残差网络层选择与所述反向传播路径对应的串接节点进行反向传播训练，并在到达所述反向传播路径对应的串接节点时，计算所述预训练神经网络模型的网络参数的梯度。

在一些可能的实现方式中，在采用收集数据集对所述预训练神经网络模型进行迭代训练，得到所述互动动作识别模型的步骤之前，所述方法还包括：

调整所述训练样本图像集中每个训练样本图像的图像参数，以对所述训练样本图像集进行样本扩展。

在一些可能的实现方式中，所述将视频采集装置实时采集的主播视频帧输入到预先训练的互动动作识别模型中，识别所述主播视频帧中是否包含主播互动动作的步骤，包括：

将所述主播视频帧输入到所述互动动作识别模型中，得到识别结果图，其中，所述识别结果图中包含至少一个目标框，所述目标框为标记所述识别结果图中的主播互动动作的几何框；

根据所述主播视频帧的识别结果图确定所述主播视频帧中是否包含主播互动动作。

在一些可能的实现方式中，所述将所述主播视频帧输入到所述互动动作识别模型中，得到识别结果图的步骤，包括：

通过所述互动动作识别模型将所述主播视频帧分割为多个网格；

针对每个网格，在该网格内生成多个不同属性参数的几何预测框，其中，每个几何预测框对应一个基准框，每个几何预测框的属性参数包括相对于所述基准框的中心点坐标、宽度、高度以及类别；

计算每个几何预测框的置信度得分，并根据计算结果剔除置信度得分低于预设得分阈值的几何预测框；

按照置信度得分由大到小的顺序对该网格内剩余的几何框进行排序，并根据排序结果将置信度得分最大的几何框确定为所述目标框，以得到识别结果图。

在一些可能的实现方式中，所述计算每个几何预测框的置信度得分的步骤，包括：

针对每个几何预测框，判断该几何预测框的区域内是否存在主播互动动作；

若不存在主播互动动作，则判定该几何预测框的置信度得分为 0；

若存在主播互动动作，则计算该几何预测框的区域属于主播互动动作的后验概率，并计算该几何预测框的检测评价函数值，其中，所述检测评价函数值用于表征主播互动动作与该几何预测框的交集与主播互动动作与该几何预测框的并集之间的比值；

根据所述后验概率与所述检测评价函数值得到该几何预测框的置信度得分。本申请一种直播互动装置，应用于直播提供终端，所述装置包括：

检测模块，配置成当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型，其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

生成模块，配置成根据所述主播互动动作的动作姿态和动作类型生成所述主播对应的虚拟形象的互动视频流，并通过所述直播服务器将所述虚拟形象的互动视频流发送给所述直播接收终端进行播放。

本申请提供一种直播系统，所述直播系统包括直播提供终端、直播接收终端以及分别与所述直播提供终端和所述直播接收终端通信连接的直播服务器；

所述直播提供终端配置成当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型，其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

所述直播服务器配置成将所述虚拟形象的互动视频流发送给所述直播接收终端；

所述直播接收终端配置成在直播界面中播放所述虚拟形象的互动视频流。

本申请提供一种可读存储介质，该可读存储介质上存储有机器可执行指令，该计算机程序被处理器运行时可以执行上述的直播互动方法的步骤。

本申请实施例在从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测主播互动动作的动作姿态和动作类型，其中，主播互动动作包括佩戴目标道具和/或目标肢体动作。而后，根据主播互动动作的动作姿态和动作类型生成主播对应的虚拟形象的互动视频流，并通过直播服务器将虚拟形象的互动视频流发送给直播接收终端进行播放。如此，通过将主播的虚拟形象的互动内容与主播互动动作的动作姿态和动作类型产生关联，可以提高直播过程中的互动效果，减少主播发起虚拟形象互动时的人为操作，实现虚拟形象的自动互动。

为使本申请实施例的上述目的、特征和优点能更明显易懂，下面将结合实施例，并配合所附附图，作详细说明。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本申请的某些实施例，因此不应被看作是对范围的限定，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相关的附图。

图 1 示出了本申请实施例所提供的直播系统的应用场景示意框图；

图 2 示出了本申请实施例所提供的直播互动方法的流程示意图；

图 3 示出了步骤 S110 的一种可能子步骤流程示意图；

图 4 示出了本申请实施例所提供的神经网络模型的网络结构示意图；

图 5 示出了本申请实施例所提供的神经网络模型的训练流程示意图；

图 6 示出了本申请实施例所提供的直播提供终端的一种直播界面示意图；

图 7 示出了本申请实施例所提供的直播提供终端的另一种直播界面示意图；

图 8 示出了本申请实施例所提供的图 1 所示的直播提供终端的示例性组件示意图。

具体实施方式

为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，应当理解，本申请中附图仅起到说明和描述的目的，并不用于限定本申请的保护范围。另外，应当理解，示意性的附图并未按实物比例绘制。本申请中使用的流程图示出了根据本申请实施例的一些实施例实现的操作。应该理解，流程图的操作可以不按顺序实现，没有逻辑的上下文关系的步骤可以反转顺序或者同时实施。此外，本领域技术人员在本申请内容的指引下，可以向流程图添加一个或多个其他操作，也可以从流程图中移除一个或多个操作。

另外，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此，以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围，而是仅仅表示本申请的选定实施例。基于本申请的实施例，本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本申请保护的范围。

图 1 是本申请实施例提供的直播系统 10 的应用场景示意图。例如，直播系统 10 可以是配置成诸如互联网直播之类的服务平台。参照图 1 所示，直播系统 10 可以包括直播服务器 200、直播提供终端 100 以及直播接收终端 300，直播服务器 200 分别与直播提供终端 100 以及直播接收终端 300 通信连接，配置成为直播提供终端 100 以及直播接收终端 300 提供直播服务。例如，直播提供终端 100 可以将直播间的直播视频流发送给直播服务器 200，观众可以通过直播接收终端 300 访问直播服务器 200 以观看直播间的直播视频。又例如，直播服务器也可以在观众订阅的直播间开播时向该观众的直播接收终端 300 发送通知消息。

直播视频流可以是当前正在直播平台中直播的视频流或者直播完成后形成的完整视频流。

可以理解，图 1 所示的直播系统 10 仅为一种可行的示例，在其它可行的实施例中，该直播系统 10 也可以仅包括图 1 所示组成部分的其中一部分或者还可以包括其它的组成部分。例如，在一些可能的实现方式中，直播提供终端 100 也可以直接与直播接收终端 300 通信，直播提供终端 100 可以直接将直播视频流数据发送给直播接收终端 300。

在一些实施场景中，直播提供终端 100 和直播接收终端 300 可以互换使用。例如，直播提供终端 100 的主播可以使用直播提供终端 100 来为观众提供直播视频服务，或者作为观众查看其它主播提供的直播视频。又例如，直播接收终端 300 的观众也可以使用直播接收终端 300 观看所关注的主播提供的直播视频，或者作为主播为其它观众提供直播视频服务。

本实施例中，直播系统 10 还可以包括配置成采集主播的主播视频帧的视频采集装置 400，视频采集装置 400 可以直接安装或者集成于直播提供终端 100，也可以独立于直播提供终端 100 并与直播提供终端 100 连接。

图 2 示出了本申请实施例提供的直播互动方法的流程示意图，该直播互动方法可由图 1 中所示的直播提供终端 100 执行。应当理解，在其它实施例中，本实施例的直播互动方法其中部分步骤的顺序可以根据实际需要相互交换，或者其中的部分步骤也可以省略或删除。该直播互动方法的详细步骤介绍如下。

步骤 S110，当从视频采集装置 400 实时采集的主播视频帧中检测到主播发起主播互动动作时，检测主播互动动作的动作姿态和动作类型。

作为一种可能的实施方式，视频采集装置 400 可以根据预设的实时主播视频帧采集速率，采集主播的主播视频帧。前述实时主播视频帧采集速率可以根据实际网络带宽、直播提供终端 100 的处理性能以及网络传输协议而设定。通常三维引擎可提供 60 帧/s 或者 30 帧/s 等不同的渲染速率，本实施例可以根据实际网络带宽、直播提供终端的处理性能以及目标传输协议等客观因素来确定所需的实时主播视频帧采集速率，由此可以确保后续渲染虚拟形象的视频流的实时性及流畅性。

本实施例中，主播互动动作可包括佩戴目标道具和/或目标肢体动作。

以根据目标道具确定动作类型和动作姿态为例，在从主播视频帧中检测到主播佩戴目标道具时，可检测目标道具的道具属性和参考点位置向量，并根据道具属性查找目标肢体动作的动作类型，而后根据参考点位置向量采用逆向运动学（Inverse Kinematic, IK）算法预测主播互动动作的动作姿态。

其中，目标道具可以是直播平台能够识别的用于指示主播互动动作的动作类型的各种互动道具，这些互动道具的属性可以包括形状信息。在此情况下，互动道具可以为依据具体主播互动动作的动作类型进行设计的。例如，若互动道具 A 用于指示“剪刀手的卖萌动作”，则该互动道具 A 可以设计为剪刀手的形状。又例如，若互动道具 B 用于指示“手比爱心的温馨动作”，则该互动道具 B 可以设计为手比爱心的形状。

或者，这些互动道具的道具属性还可以包括颜色信息，在此情况下，互动道具的颜色可以依据具体主播互动动作的动作类型进行设计，例如，若互动道具 A 用于指示“剪刀手的卖萌动作”，则该互动道具 A 可以设计为红色，又例如，若互动道具 A 用于指示“手比爱心的温馨动作”，则该互动道具 B 可以设计为蓝色。如此设计，直播提供终端 100 可以通过识别互动道具的属性快速识别到目标肢体动作的动作类型，无需进行深度神经网络算法识别，从而大大减少计算量，提高识别速度和识别精度。

在另一种实施方式中，在从主播视频帧中检测到主播发起目标肢体动作时，可以检测目标肢体动作的参考点位置向量，并采用深度神经网络模型识别目标肢体动作的动作类型。而后，根据参考点位置向量采用逆向运动学(Inverse Kinematic, IK) 算法预测主播互动动作的动作姿态。

换句话说，在本实施例中，还可以将视频采集装置实时采集的主播视频帧输入到预先

训练的互动动作识别模型中，识别所述主播视频帧中是否包含目标肢体动作；在检测到所述主播发起目标肢体动作时，获得所述目标肢体动作的动作类型及所述目标肢体动作的参考点位置向量；根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

可选地，目标肢体动作的类型可包括但不限于起立、坐下、转圈、倒立、身体晃动、挥手、剪刀手、握拳、手比爱心、托手、鼓掌、手掌张开、手掌闭合、竖大拇指、手枪姿势、V 手势和 OK 手势等直播常用的肢体动作。

在此例子中，直播提供终端 100 可以在步骤 S110 中将主播视频帧输入到互动动作识别模型，得到识别结果图，并根据识别结果图确定主播视频帧中包含的目标肢体动作的动作类型。其中，上述识别结果图中包含至少一个目标框，目标框为标记识别结果图中的目标肢体动作的动作类型的几何框。请参照图 3，在此例子中，步骤 S110 可以包括以下子步骤。

子步骤 S111，通过互动动作识别模型将主播视频帧分割为多个网格。

子步骤 S112，针对每个网格，可以在该网格内生成多个不同属性参数的几何预测框，其中，每个几何预测框对应一个基准框，每个几何预测框的属性参数包括相对于基准框的中心点坐标、宽度、高度以及类别，从而可以适应直播场景的多样性。

子步骤 S113，计算每个几何预测框的置信度得分，并根据计算结果剔除置信度得分低于预设得分阈值的几何预测框。

例如，可以针对每个几何预测框，判断该几何预测框的区域内是否存在目标肢体动作：

若不存在目标肢体动作，则判定该几何预测框的置信度得分为 0；

若存在目标肢体动作，则计算该几何预测框的区域属于目标肢体动作的后验概率，并计算该几何预测框的检测评价函数值，其中，检测评价函数值用于表征目标肢体动作与该几何预测框的交集与目标肢体动作与该几何预测框的并集之间的比值。

最后，可以根据后验概率与检测评价函数值的乘积得到该几何预测框的置信度得分。

在此基础上，可以预先设定一个预设得分阈值，若该几何预测框的置信度得分低于该预设得分阈值的几何预测框，表示该几何预测框中的目标不可能是直播互动动作的预测目标；若该几何预测框的置信度得分大于等于该预设得分阈值的几何预测框，表示该几何预测框中的目标有可能是直播互动动作的预测目标。

由此，可以选择性地剔除置信度得分低于该预设得分阈值的几何预测框，从而一次性剔除大量的不可能存在直播互动动作的目标的几何预测框，只对有可能存在直播互动动作的目标的几何预测框进行后续的处理，从而大大减少后续计算量，进一步提高识别速度。

子步骤 S114，按照置信度得分由大到小的顺序对该网格内剩余的几何框进行排序，并根据排序结果将置信度得分最大的几何框确定为目标框，以得到识别结果图。

由此，通过直播图像的识别结果图，若存在标记有目标肢体动作的目标框，则确定主播视频帧中包含目标肢体动作，并可以确定该目标肢体动作的互动动作类型。

在从主播视频帧中检测到主播发起目标肢体动作时，直播提供终端 100 还可以根据目标肢体动作的参考点位置向量或者目标道具的参考点位置向量，采用逆向运动学算法预测主播互动动作的动作姿态，为后续实现虚拟形象与主播之间的整体动作同步提供数据基础。

例如，直播提供终端 100 可以根据参考点位置向量计算主播的互动肢体的中心点高度以及主播的互动肢体相对于视频采集装置 400 的姿态旋转矩阵。接着，根据姿态旋转矩阵、参考点位置向量和中心点高度，计算主播的互动肢体的各个肢体关节的位置向量，其中，位置向量包括主播的互动肢体在各个参考轴方向上的分量。最后，根据计算得到的各个肢体关节的位置向量得到主播互动动作的动作姿态。

其中，参考轴方向可以预先进行配置，以二维空间为例，参考轴方向可以包括相互垂直的 X 轴方向和 Y 轴方向；以三维空间为例，参考轴方向可以包括相互垂直的 X 轴方向、Y 轴方向以及 Z 轴方向。

主播的互动肢体相对于视频采集装置 400 的姿态旋转矩阵主要是指互动肢体相对于视

频采集装置 400 在二维空间或者三维空间的位置和姿态。以三维空间为例，该位置可以用一个位置矩阵来描述，该姿态可以记录为用坐标系三个坐标轴两两夹角的余弦值组成的姿态矩阵。

在本实施例中，互动动作识别模型可以基于神经网络模型训练获得，作为一种可能的实施方式，请参阅图 4，上述互动动作识别模型可包括输入层、至少一个卷积提取层、全连接层以及分类层。每个卷积提取层包括依次设置的第一点卷积层、深度卷积层以及第二点卷积层等多个按照第一点卷积层、深度卷积层以及第二点卷积层的顺序设置的卷积层。该卷积提取层内的每个卷积层之后设置一个激活函数层和池化层，全连接层位于最后一个池化层之后，分类层位于全连接层之后。接下来关于该互动动作识别模型的训练过程会在后文进行阐述，此处暂不介绍。

下面对前述神经网络模型训练得到互动动作识别模型的过程进行详细阐述。

首先，建立神经网络模型。可选地，该神经网络模型可以采用，但不限于 YOLOv2 网络模型。该 YOLOv2 网络采用计算量小的单元，以适应于直播提供终端，如手机或者用户终端等计算能力较弱的电子设备，具体可以采用 PointwiseDepthwise+Pointwise 卷积结构，或者三个普通卷积层结构，在训练过程中采用梯度下降法进行反向传播训练，在训练过程中采用残差网络改变训练时梯度的方向。

接着，采用公开数据集对神经网络模型进行预训练，得到预训练神经网络模型。其中，公开数据集可以采用 COCO 数据集，COCO 数据集是一个大型图像数据集，专为对象检测、分割、人体关键点检测、语义分割和字幕生成而设计，主要从复杂的日常场景中截取，图像中的检测目标通过精确的分割进行位置的标定，从而使得神经网络模型具有初步的目标检测、目标之间的上下文关系识别、目标在二维上的精确定位的能力。

而后，采用收集数据集对预训练神经网络模型进行迭代训练，得到互动动作识别模型。

其中，收集数据集包括标记有不同主播互动动作的实际目标的训练样本图像集，实际目标为主播互动动作在训练样本图像中的实际图像区域。例如，收集数据集可以包括但不限于直播过程中采集的不同主播互动动作对应的主播图像，或者主播自己做不同主播互动动作后上传的图像等。主播互动动作可以包括直播过程中的常用互动动作，例如剪刀手的卖萌动作、手比爱心的温馨动作等，本实施例对此不作具体限定。

可选地，为了使得互动动作识别模型能够在不同环境下识别主播互动动作，本实施例还可以调整训练样本图像集中每个训练样本图像的图像参数，以对训练样本图像集进行样本扩展。例如，为了适应主播在直播过程中与视频采集装置 400 之间相隔不同距离的环境，可以对初始收集数据集进行多种不同比例的等比例裁剪，从而得到与初始收集数据集相关的等比例裁剪数据集。又例如，为了适应直播在不同光线强度下的直播环境，可以对初始收集数据集进行曝光度调整处理，从而得到与初始收集数据集相关的曝光度调整数据集。又例如，为了适应直播在不同噪点环境下的直播环境，还可以对初始收集数据集添加不同程度的噪声，从而得到与初始收集数据集相关的噪声数据集。如此，通过对训练样本图像集进行样本扩展，可以有效提高后续互动动作识别模型在不同直播场景下的识别能力。

由于整个互动动作的识别过程都发生在直播提供终端 100，为了有效降低直播提供终端 100 的计算量，提高识别速度，通过上述网络结构设计，每个卷积提取层采用可分离卷积结构，即由第一点卷积层、深度卷积层以及第二点卷积层的级联结构组成，采用这种级联结构与采用普通的三个卷积层的结构相比，计算量和网络参数量更小。

下面结合图 4 所示的神经网络模型，对前述采用收集数据集对预训练神经网络模型进行迭代训练的过程进行示例性说明，请参阅图 5，在步骤 S110 之前还包括步骤 S101、步骤 S102、步骤 S103、步骤 S104、步骤 S105、步骤 S106 以及步骤 S107，下面分别对步骤 S101、步骤 S102、步骤 S103、步骤 S104、步骤 S105、步骤 S106 以及步骤 S107 进行介绍。

步骤 S101，将训练样本图像集中的每个训练样本图像输入到预训练神经网络模型的输入层进行预处理，得到预处理图像。详细地，由于后续需要使用随机梯度下降法进行训练，

因此输入每个训练样本图像需要进行标准化处理。

例如，可以将每个训练样本图像进行均值化，详细地，可以将每个训练样本图像的各个维度都中心化到 0，所有训练样本图像求和后再求平均值得到均值样本，然后将所有的训练样本图像减去这个均值样本，得到预处理图像。

又例如，还可以将每个训练样本图像的数据幅度归一化到同样的范围，比如对于每个特征而言，范围是 $[-1, 1]$ ，从而得到预处理图像。

又例如，还可以将每个训练样本图像进行 PCA 降维，让每个维度的相关度取消，特征和特征之间是相互独立的，然后再对每个训练样本图像在每个特征轴上的幅度归一化，得到预处理图像。

步骤 S102，针对每个卷积提取层，通过该卷积提取层的第一点卷积层、深度卷积层以及第二点卷积层分别提取预处理图像的多维特征图像，并将提取得到的多维特征图像输入到所连接的激活函数层中进行非线性映射，而后将非线性映射后的多维特征图像输入到所连接的池化层中进行池化处理，并将池化处理得到的池化特征图输入到下一层卷积层进行特征提取。

第一点卷积层、深度卷积层以及第二点卷积层的功能是对输入的图像数据进行特征提取，其内部包含多个卷积核，组成卷积核的每个元素都对应一个权重系数和一个偏差量，也即一个神经元。对于每个预处理图像的多维特征图像，有一个性质称作局部关联性质，一个预处理图像的像素点影响最大的是该预处理图像周边的像素点，而与距离这个像素点比较远的像素点二者之间关系不大。如此，每一个神经元只需要和上一层局部连接，相当于每一个神经元扫描一小区域，然后许多神经元（这些神经元权值共享）合起来就相当于扫描了全局的特征图，这样就构成一个一维特征图，多维特征图像也即提取了这个预处理图像的多维特征得到。

在此基础上，将提取得到的多维特征图像输入到所连接的激活函数层中进行非线性映射，以协助表达多维特征图像中的复杂特征。可选地，激活函数层可以采用但不限于线性整流单元(Rectified Linear Unit, ReLU)、Sigmoid 函数和双曲正切函数(Hyperbolic Tangent)等。

而后将非线性映射后的多维特征图像输入到所连接的池化层中进行池化处理，也即，非线性映射后的多维特征图像会被传递至池化层进行特征选择和信息过滤，池化层可以包含预设定的池化函数，从而将非线性映射后的多维特征图像单个点的结果替换为其相邻区域的特征图统计量。接着，将池化处理得到的池化特征图输入到下一层卷积层继续进行特征提取。

步骤 S103，将最后一层池化层输出的池化特征图输入到全连接层，得到全连接特征输出值。详细地，在全连接层中所有神经元都有权重连接，当前面的所有卷积层（也即第一点卷积层、深度卷积层以及第二点卷积层）提取到足以用来识别待处理图像的特征图像后，接下来需要通过全连接层进行分类，得到全连接特征输出值。

步骤 S104，将全连接特征输出值输入到分类层中进行预测目标分类，得到每个训练样本图像的预测目标。

步骤 S105，计算各个训练样本图像的预测目标与实际目标之间的损失函数(Loss Function)值。

步骤 S106，根据损失函数值进行反向传播训练，并计算预训练神经网络模型的网络参数的梯度。

可选地，在本实施例中，互动动作识别模型还可以包括多个残差网络层（图未示出），每个残差网络层配置成将互动动作识别模型中任意相邻的两层的输出部分与该相邻的两层之后一层的输入部分串接。由此，可以使得梯度在反向传播训练时可以选择不同的反向传播路径，增强训练效果。

详细地，当确定出损失函数值后，可以根据损失函数值确定反向传播训练的反向传播

路径，然后通过预训练神经网络模型的残差网络层选择与反向传播路径对应的串接节点进行反向传播训练，并在到达反向传播路径对应的串接节点时，计算预训练神经网络模型的网络参数的梯度。

步骤 S107，根据计算得到的梯度，采用随机梯度下降法更新预训练神经网络模型的网络参数后继续训练，直到预训练神经网络模型满足训练终止条件时，输出训练得到的互动动作识别模型。

其中，上述的训练终止条件可以包括以下条件中的至少一种：

1) 迭代训练次数达到设定次数；2) 损失函数值低于设定阈值；3) 损失函数值不再下降。

其中，在条件 1) 中，为了节省运算量，可以设置迭代次数的最大值，如果迭代次数达到设定次数，可以停止本迭代周期的迭代，将最后得到的预训练神经网络模型作为互动动作识别模型。在条件 2) 中，如果损失函数值低于设定阈值，说明当前的互动动作识别模型已经基本可以满足条件，此时可以停止迭代。在条件 3) 中，损失函数值不再下降，表明已经形成了最佳的互动动作识别模型，可以停止迭代。

需要说明的是，上述迭代停止条件可以结合使用，也可以择一使用，例如，可以在损失函数值不再下降时停止迭代，或者，在迭代次数达到设定次数时停止迭代，或者，在损失函数值不再下降时停止迭代。或者，还可以在损失函数值低于设定阈值，并且损失函数值不再下降时，停止迭代。

此外，在实际实施过程中，也可以不限于采用上述示例作为训练终止条件，本领域技术人员可以根据实际需求设计与上述示例不同的训练终止条件。

步骤 S120，根据主播互动动作的动作姿态和动作类型生成主播对应的虚拟形象的互动视频流，并通过直播服务器 200 将虚拟形象的互动视频流发送给直播接收终端 300 进行播放。

其中，虚拟形象可以采用与主播的外貌、体态、行动方式等相符的虚拟人物形象，并可以以二维虚拟形象、三维虚拟形象、VR 虚拟形象、AR 虚拟形象等显示在直播界面中，从而与观众进行直播互动。

本实施例中，直播提供终端 100 中可以预先存储有预设互动内容库，预设互动内容库包括各个动作类型对应的虚拟形象互动内容，虚拟形象互动内容包括对话互动内容、特效互动内容以及肢体互动内容中的一种或者多种组合。可选地，直播提供终端 100 本地可以预先配置预设互动内容库，直播提供终端 100 也可以从直播服务器 200 中下载预设互动内容库，本实施例对此不作具体限制。

可选地，对话互动内容可以包括字幕图片、字幕特效等互动信息，特效互动内容可以包括静态特效图片、动态特效图片等图像信息，肢体互动内容可以包括面部表情（诸如开心、愤怒、激动、痛苦和悲伤等）特效图片等图像信息。

由此，在确定主播互动动作的动作姿态和动作类型后，可以从预设互动内容库中获取动作类型对应的虚拟形象互动内容，然后根据动作姿态和虚拟形象互动内容生成虚拟形象的互动视频流。详细地，可以按照动作姿态关联的各个目标关节的位移坐标，控制虚拟形象的各个目标关节沿对应的位移坐标移动，并按照虚拟形象互动内容控制虚拟形象执行对应的互动动作，以生成对应的互动视频流。如此，可以使虚拟形象的互动动作与主播的动作相似，从而提高主播和观众的交互度。

作为一种可能的实施方式，在上述过程中，可以通过使用的图形图像绘制或渲染方法等生成虚拟形象的互动视频流。可选地，可以基于 OpenGL 图形绘制引擎或 Unity 3D 渲染引擎等进行 2D 图形形象或 3D 图形形象的绘制，生成虚拟形象的互动视频流，以使带有虚拟形象的互动效果的互动视频流得到展现。OpenGL 定义了一个跨编程语言、跨平台的编程接口规格的专业的图形程序接口，其与硬件无关，可以方便地进行 2D 或 3D 图形图像的绘制。通过 OpenGL 和/或 Unity 3D 渲染引擎等，不仅可以实现 2D 效果如 2D 贴纸或特效

的绘制，还可以实现 3D 特效的绘制及粒子特效的绘制等等。

仅作为示例，请参阅图 6，示出了直播提供终端 100 的一种直播界面示例图，在该直播界面中，可以包括直播界面显示框、主播视频帧显示框以及虚拟形象区域。其中，直播界面显示框用于显示当前正在直播平台中直播的视频流或者直播完成后形成的完整视频流，主播视频帧显示框用于显示视频采集装置 400 实时采集到的主播视频帧，虚拟形象区域用于展示主播的虚拟形象。

当主播发起主播互动动作时，该主播视频帧显示框中会显示主播的发起的主播互动动作，同时可以检测到主播互动动作的动作姿态和动作类型，然后获取动作类型对应的虚拟形象互动内容，并控制虚拟形象区域中的虚拟形象执行对应的互动动作。例如，若识别到的主播互动动作为手比爱心的温馨动作，此时控制虚拟形象执行对应的手比爱心的温馨动作，并且显示对话互动内容“比心”以及特效互动内容“比心”的特效，然后生成虚拟形象的互动视频流，并通过直播服务器 200 将互动视频流发送给直播接收终端 300 进行播放。

如此，本实施例通过将主播的虚拟形象的互动内容与主播互动动作的动作姿态和动作类型产生关联，可以提高直播过程中的互动效果，减少主播发起虚拟形象互动时的人为操作，实现虚拟形象的自动互动。

在其他一些实现方式中，在视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作后也可以直接根据主播互动动作确定虚拟形象互动内容，并将虚拟形象的互动视频流发送给直播接收终端 300。

例如，可以先将视频采集装置实时采集的主播视频帧输入到预先训练的互动动作识别模型中，识别所述主播视频帧中是否包含主播互动动作。接着，当在预设数量的主播视频帧中均识别到主播互动动作时，获取预先配置的与该主播互动动作对应的虚拟形象互动内容。然后根据所述虚拟形象互动内容，控制所述直播提供终端的直播界面中的虚拟形象执行对应的互动动作，以生成所述虚拟形象的互动视频流，并通过直播服务器将所述互动视频流发送给直播接收终端进行播放。其中，为了避免主播互动动作的误识别，可以在预设数量的主播视频帧中均识别到主播互动动作时，获取预先配置的与该主播互动动作对应的虚拟形象互动内容。

直播提供终端 100 中预先存储有预设互动内容库，预设互动内容库包括预先配置的各个主播互动动作对应的虚拟形象互动内容，虚拟形象互动内容可以包括对话互动内容、特效互动内容以及肢体互动内容中的一种或者多种组合。可选地，直播提供终端 100 可以在本地配置预设互动内容库，也可以从直播服务器 200 中下载预设互动内容库，本实施例对此不作具体限制。

仅作为示例，请参阅图 7，示出了直播提供终端 100 的一种直播界面示例图，在该直播界面中，可以包括直播界面显示框、主播视频帧显示框以及虚拟形象区域。其中，直播界面显示框用于显示当前正在直播平台中直播的视频流或者直播完成后形成的完整视频流，主播视频帧显示框用于显示视频采集装置 400 实时采集到的主播视频帧，虚拟形象区域用于展示主播的虚拟形象。

当主播发起主播互动动作时，该主播视频帧显示框中会显示主播的发起主播互动动作，同时可以获取到主播互动动作对应的虚拟形象互动内容，然后控制虚拟形象区域中的虚拟形象执行对应的互动动作。例如，若识别到的主播互动动作为手比爱心的温馨动作，此时则可以控制虚拟形象执行对应的手比爱心的温馨动作，并且显示对话互动内容“比心”以及“爱你哟”的特效。由此可以生成虚拟形象的互动视频流，并通过直播服务器 200 将互动视频流发送给直播接收终端 300 进行播放。

如此，本实施例通过将主播的虚拟形象的互动内容与主播互动动作产生关联，可以提高直播过程中的互动效果，减少主播发起虚拟形象互动时的人为操作，实现虚拟形象的自动互动。

图 8 示出了本申请实施例提供的图 1 中所示的直播提供终端 100 的示例性组件示意图，

直播提供终端 100 可包括存储介质 110、处理器 120 以及直播互动装置 500。本实施例中，存储介质 110 与处理器 120 均位于直播提供终端 100 中且二者分离设置。然而，应当理解的是，存储介质 110 也可以是独立于直播提供终端 100 之外，且可以由处理器 120 通过总线接口来访问。可替换地，存储介质 110 也可以集成到处理器 120 中，例如，可以是高速缓存和/或通用寄存器。

直播互动装置 500 可以理解为上述直播提供终端 100，或直播提供终端 100 的处理器 120，也可以理解为独立于上述直播提供终端 100 或处理器 120 之外的在直播提供终端 100 控制下实现上述直播互动方法的软件功能模块。如图 7 所示，该直播互动装置 500 可以包括检测模块 510 以及生成模块 520，下面分别对该直播互动装置 500 的各个功能模块的功能进行详细阐述。

检测模块 510，配置成当从视频采集装置 400 实时采集的主播视频帧中检测到主播发起主播互动动作时，检测主播互动动作的动作姿态和动作类型，其中，主播互动动作包括佩戴目标道具和/或目标肢体动作。可以理解，该检测模块 510 可以配置成执行上述步骤 S110，关于该检测模块 510 的详细实现方式可以参照上述对步骤 S110 有关的内容。

生成模块 520，配置成根据主播互动动作的动作姿态和动作类型生成主播对应的虚拟形象的互动视频流，并通过直播服务器 200 将虚拟形象的互动视频流发送给直播接收终端 300 进行播放。可以理解，该生成模块 520 可以配置成执行上述步骤 S120，关于该生成模块 520 的详细实现方式可以参照上述对步骤 S120 有关的内容。

进一步地，本申请实施例还提供一种计算机可读存储介质，计算机可读存储介质存储有机器可执行指令，机器可执行指令被执行时实现上述实施例提供的直播互动方法。

以上仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

工业实用性

本申请实施例在从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测主播互动动作的动作姿态和动作类型，其中，主播互动动作包括佩戴目标道具和/或目标肢体动作。而后，根据主播互动动作的动作姿态和动作类型生成主播对应的虚拟形象的互动视频流，并通过直播服务器将虚拟形象的互动视频流发送给直播接收终端进行播放。如此，通过将主播的虚拟形象的互动内容与主播互动动作的动作姿态和动作类型产生关联，可以提高直播过程中的互动效果，减少主播发起虚拟形象互动时的人为操作，实现虚拟形象的自动互动。

权利要求书

1.一种直播互动方法，其特征在于，应用于直播提供终端，所述方法包括：

当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型；

其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

根据所述主播互动动作的动作姿态和动作类型生成所述主播对应的虚拟形象的互动视频流，并通过直播服务器将所述虚拟形象的互动视频流发送给直播接收终端进行播放。

2.根据权利要求1所述的直播互动方法，其特征在于，所述检测所述主播互动动作的动作姿态和动作类型的步骤，包括：

在检测到所述主播佩戴目标道具时，检测所述目标道具的道具属性和参考点位置向量，并根据所述道具属性查找所述目标肢体动作的动作类型；

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

3.根据权利要求1所述的直播互动方法，其特征在于，所述检测所述主播互动动作的动作姿态和动作类型的步骤，包括：

在检测到所述主播发起目标肢体动作时，检测所述目标肢体动作的参考点位置向量，并采用深度神经网络模型识别所述目标肢体动作的动作类型；

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

4.根据权利要求2或3所述的直播互动方法，其特征在于，所述根据所述参考点位置向量，采用逆向运动学算法预测所述主播互动动作的动作姿态的步骤，包括：

根据所述参考点位置向量计算所述主播的互动肢体的中心点高度以及所述主播的互动肢体相对于所述视频采集装置的姿态旋转矩阵；

根据所述姿态旋转矩阵、所述参考点位置向量和所述中心点高度，计算所述主播的互动肢体的各个肢体关节的位置向量，所述位置向量包括所述主播的互动肢体在各个参考轴方向上的分量；

根据计算得到的所述各个肢体关节的位置向量得到所述主播互动动作的动作姿态。

5.根据权利要求1-4任意一项所述的直播互动方法，其特征在于，所述直播提供终端中预先存储有预设互动内容库，所述预设互动内容库包括各个动作类型对应的虚拟形象互动内容，所述虚拟形象互动内容包括对话互动内容、特效互动内容以及肢体互动内容中的一种或者多种组合；

所述根据所述主播互动动作的动作姿态和动作类型生成所述虚拟形象的互动视频流的步骤，包括：

从所述预设互动内容库中获取所述动作类型对应的虚拟形象互动内容；

根据所述动作姿态和所述虚拟形象互动内容生成所述虚拟形象的互动视频流。

6.根据权利要求5所述的直播互动方法，其特征在于，所述根据所述动作姿态和所述虚拟形象互动内容生成所述虚拟形象的互动视频流的步骤，包括：

按照所述动作姿态关联的各个目标关节的位移坐标，控制所述虚拟形象的各个目标关节沿对应的位移坐标移动，并按照所述虚拟形象互动内容控制所述虚拟形象执行对应的互动动作，以生成对应的互动视频流。

7.根据权利要求1所述的直播互动方法，其特征在于，当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型的步骤，包括：

将视频采集装置实时采集的主播视频帧输入到预先训练的互动动作识别模型中，识别所述主播视频帧中是否包含目标肢体动作；

在检测到所述主播发起目标肢体动作时，获得所述目标肢体动作的动作类型及所述目标肢体动作的参考点位置向量；

根据所述参考点位置向量采用逆向运动学算法预测所述主播互动动作的动作姿态。

8.根据权利要求7所述的直播互动方法,其特征在于,所述互动动作识别模型包括输入层、至少一个卷积提取层、全连接层以及分类层,每个卷积提取层包括依次设置的第一点卷积层、深度卷积层以及第二点卷积层,所述卷积提取层内的每个卷积层之后设置一个激活函数层和池化层,所述全连接层位于最后一个池化层之后,所述分类层位于全连接层之后。

9.根据权利要求8所述的直播互动方法,其特征在于,所述互动动作识别模型还包括多个残差网络层,每个残差网络层配置成将所述互动动作识别模型中任意相邻的两层的输出部分与该相邻的两层之后一层的输入部分串接。

10.根据权利要求7-9中任意一项所述的直播互动方法,其特征在于,所述方法还包括预先训练所述互动动作识别模型的步骤,具体包括:

建立神经网络模型;

采用公开数据集对所述神经网络模型进行预训练,得到预训练神经网络模型;

采用收集数据集对所述预训练神经网络模型进行迭代训练,得到所述互动动作识别模型,其中,所述收集数据集包括标记有不同主播互动动作的实际目标的训练样本图像集,所述实际目标为所述主播互动动作在训练样本图像中的实际图像区域。

11.根据权利要求10所述的直播互动方法,其特征在于,所述采用收集数据集对所述预训练神经网络模型进行迭代训练,得到所述互动动作识别模型的步骤,包括:

将所述训练样本图像集中的每个训练样本图像输入到所述预训练神经网络模型的输入层进行预处理,得到预处理图像;

针对所述预训练神经网络模型的每个卷积提取层,通过该卷积提取层的第一点卷积层、深度卷积层以及第二点卷积层分别提取预处理图像的多维特征图像,并将提取得到的多维特征图像输入到所连接的激活函数层中进行非线性映射,而后将非线性映射后的多维特征图像输入到所连接的池化层中进行池化处理,并将池化处理得到的池化特征图输入到下一层卷积层进行特征提取;

将最后一层池化层输出的池化特征图输入到全连接层,得到全连接特征输出值;

将所述全连接特征输出值输入到分类层中进行预测目标分类,得到每个训练样本图像的预测目标;

计算各个训练样本图像的预测目标与实际目标之间的损失函数值;

根据所述损失函数值进行反向传播训练,并计算所述预训练神经网络模型的网络参数的梯度;

根据计算得到的所述梯度,采用随机梯度下降法更新所述预训练神经网络模型的网络参数后继续训练,直到所述预训练神经网络模型满足训练终止条件时,输出训练得到的互动动作识别模型。

12.根据权利要求11所述的直播互动方法,其特征在于,所述根据所述损失函数值进行反向传播训练,并计算所述预训练神经网络模型的网络参数的梯度的步骤,包括:

根据所述损失函数值确定反向传播训练的反向传播路径;

通过所述预训练神经网络模型的残差网络层选择与所述反向传播路径对应的串接节点进行反向传播训练,并在到达所述反向传播路径对应的串接节点时,计算所述预训练神经网络模型的网络参数的梯度。

13.根据权利要求10所述的直播互动方法,其特征在于,在采用收集数据集对所述预训练神经网络模型进行迭代训练,得到所述互动动作识别模型的步骤之前,所述方法还包括:

调整所述训练样本图像集中每个训练样本图像的图像参数,以对所述训练样本图像集进行样本扩展。

14.根据权利要求7-9中任意一项所述的直播互动方法,其特征在于,所述将视频采集装置实时采集的主播视频帧输入到预先训练的互动动作识别模型中,识别所述主播视频帧

中是否包含主播互动动作的步骤，包括：

将所述主播视频帧输入到所述互动动作识别模型中，得到识别结果图，其中，所述识别结果图中包含至少一个目标框，所述目标框为标记所述识别结果图中的主播互动动作的几何框；

根据所述主播视频帧的识别结果图确定所述主播视频帧中是否包含主播互动动作。

15.根据权利要求 14 所述的直播互动方法，其特征在于，所述将所述主播视频帧输入到所述互动动作识别模型中，得到识别结果图的步骤，包括：

通过所述互动动作识别模型将所述主播视频帧分割为多个网格；

针对每个网格，在该网格内生成多个不同属性参数的几何预测框，其中，每个几何预测框对应一个基准框，每个几何预测框的属性参数包括相对于所述基准框的中心点坐标、宽度、高度以及类别；

计算每个几何预测框的置信度得分，并根据计算结果剔除置信度得分低于预设得分阈值的几何预测框；

按照置信度得分由大到小的顺序对该网格内剩余的几何框进行排序，并根据排序结果将置信度得分最大的几何框确定为所述目标框，以得到识别结果图。

16.根据权利要求 15 所述的直播互动方法，其特征在于，所述计算每个几何预测框的置信度得分的步骤，包括：

针对每个几何预测框，判断该几何预测框的区域内是否存在主播互动动作；

若不存在主播互动动作，则判定该几何预测框的置信度得分为 0；

若存在主播互动动作，则计算该几何预测框的区域属于主播互动动作的后验概率，并计算该几何预测框的检测评价函数值，其中，所述检测评价函数值用于表征主播互动动作与该几何预测框的交集与主播互动动作与该几何预测框的并集之间的比值；

根据所述后验概率与所述检测评价函数值得到该几何预测框的置信度得分。

17.一种直播互动装置，其特征在于，应用于直播提供终端，所述装置包括：

检测模块，配置成当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型，其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

生成模块，配置成根据所述主播互动动作的动作姿态和动作类型生成所述主播对应的虚拟形象的互动视频流，并通过直播服务器将所述虚拟形象的互动视频流发送给直播接收终端进行播放。

18.一种直播系统，其特征在于，所述直播系统包括直播提供终端、直播接收终端以及分别与所述直播提供终端和所述直播接收终端通信连接的直播服务器；

所述直播提供终端配置成当从视频采集装置实时采集的主播视频帧中检测到主播发起主播互动动作时，检测所述主播互动动作的动作姿态和动作类型，并根据所述主播互动动作的动作姿态和动作类型生成所述主播对应的虚拟形象的互动视频流，将所述虚拟形象的互动视频流发送给直播服务器，其中，所述主播互动动作包括佩戴目标道具和/或目标肢体动作；

所述直播服务器配置成将所述虚拟形象的互动视频流发送给所述直播接收终端；

所述直播接收终端配置成在直播界面中播放所述虚拟形象的互动视频流。

19.一种电子设备，其特征在于，所述电子设备包括一个或多个存储介质和一个或多个与存储介质通信的处理器，一个或多个存储介质存储有处理器可执行的机器可执行指令，当电子设备运行时，处理器执行所述机器可执行指令，以执行权利要求 1-6 中任意一项所述的直播互动方法。

20.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有机器可执行指令，所述机器可执行指令被执行时实现权利要求 1-16 中任意一项所述的直播互动方法。

附图

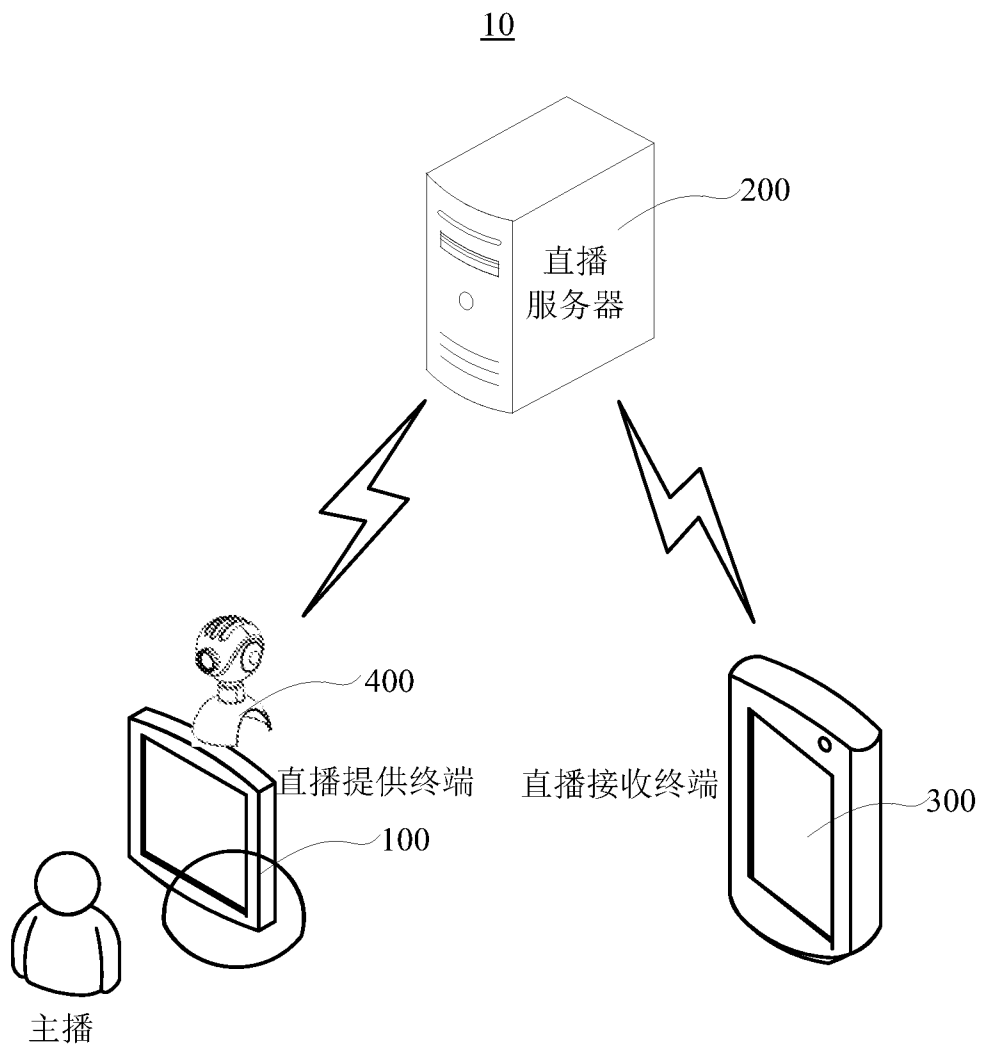


图 1

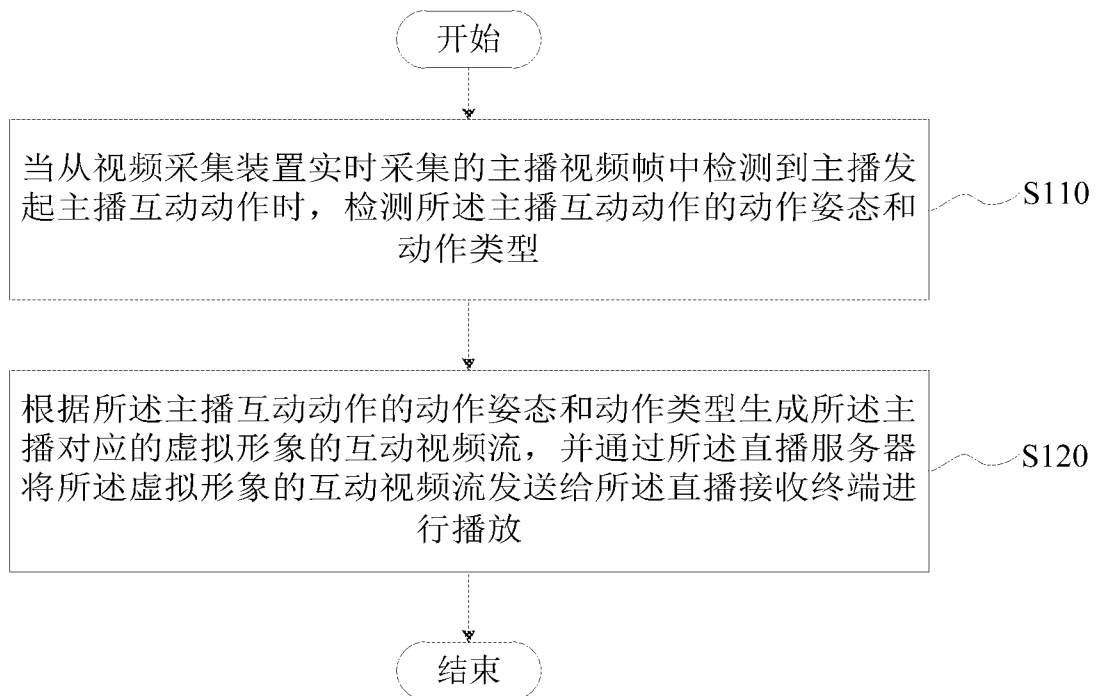


图 2

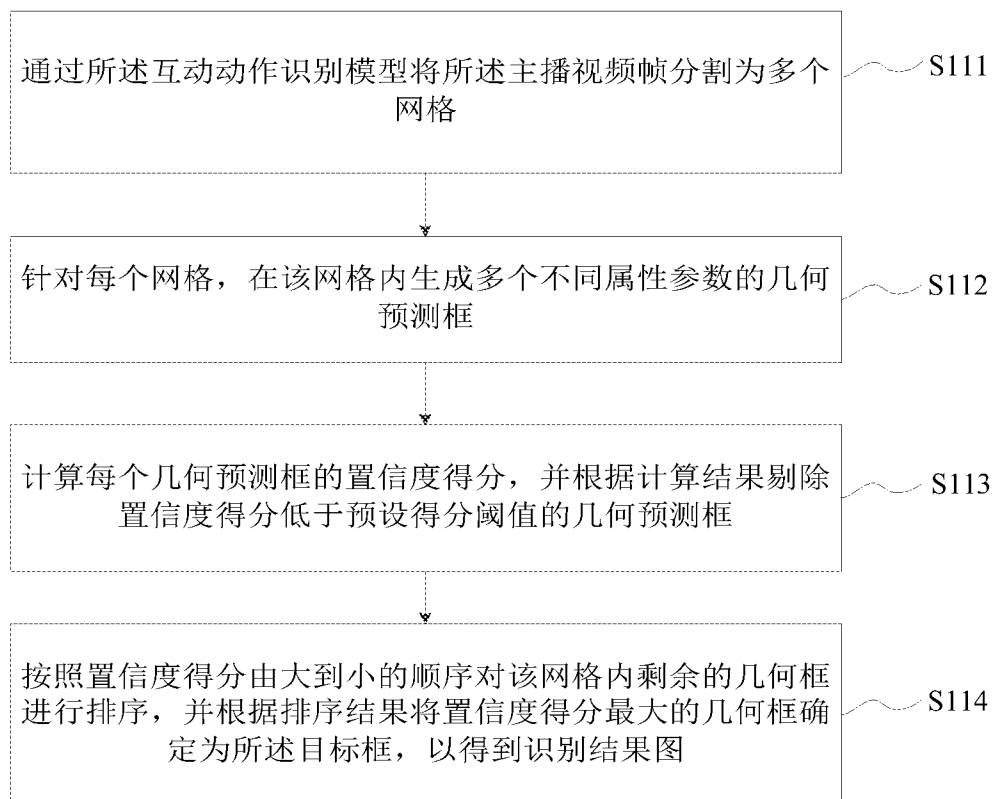


图 3

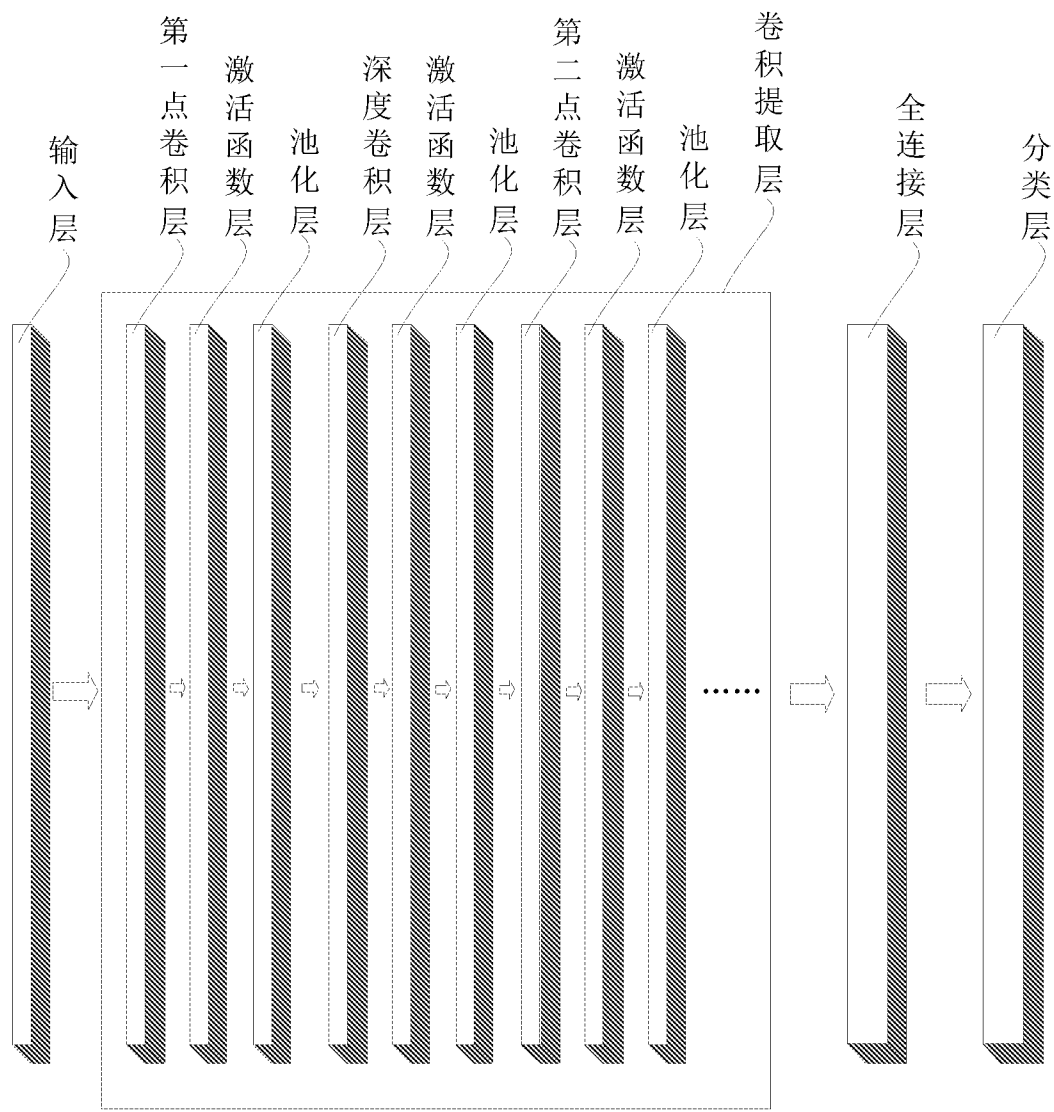


图 4



图 5

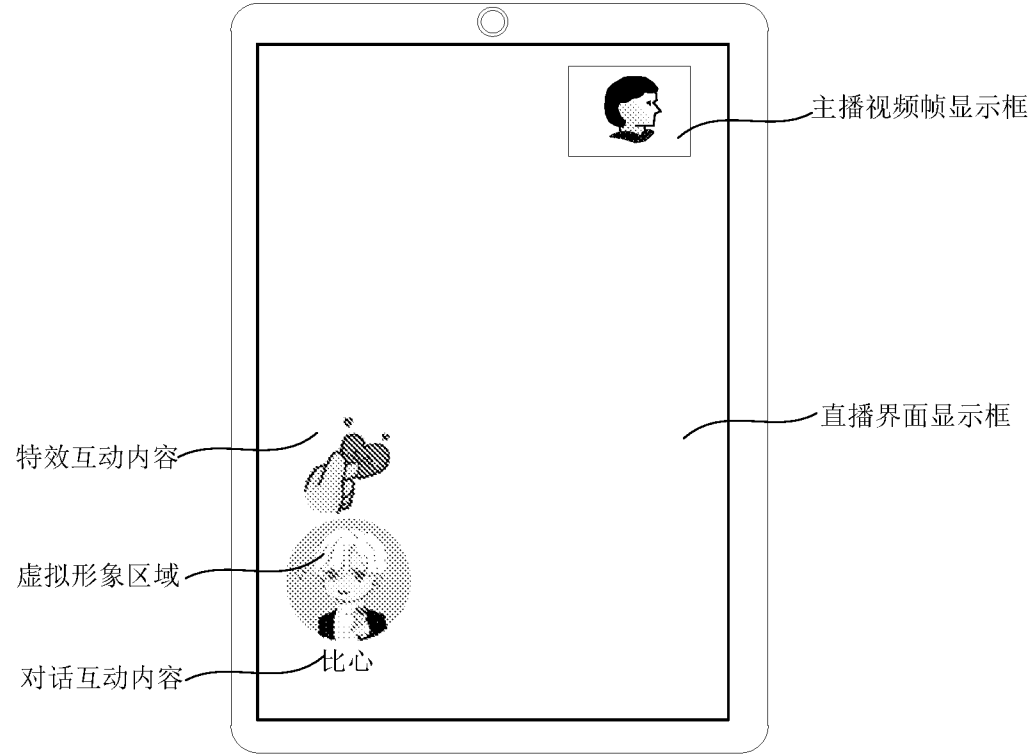


图 6

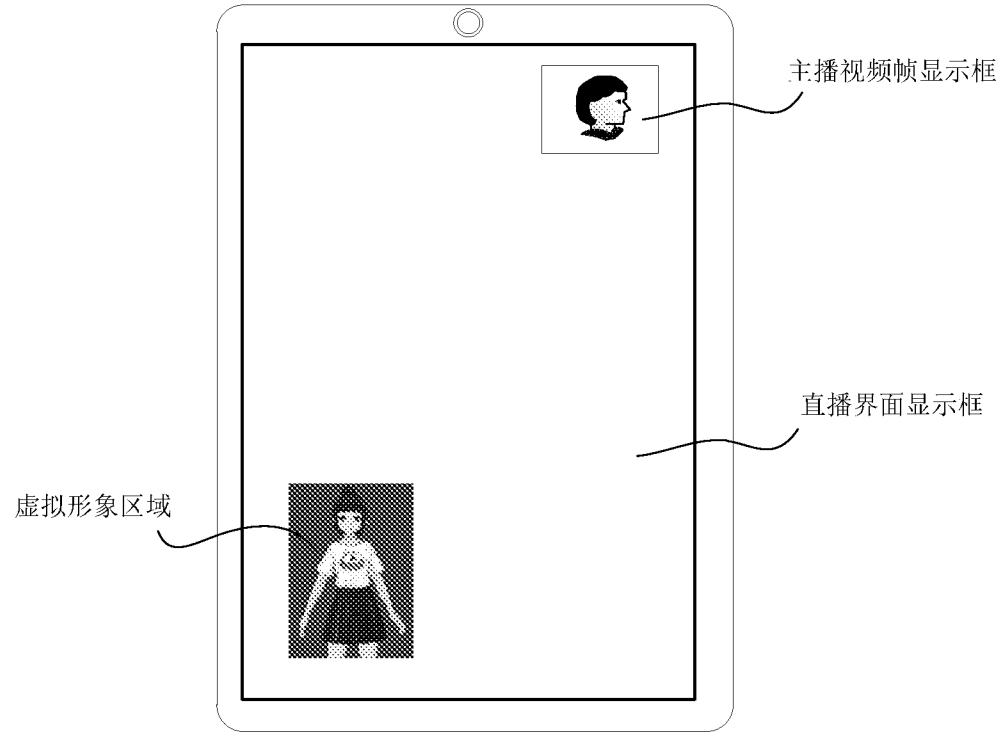


图 7

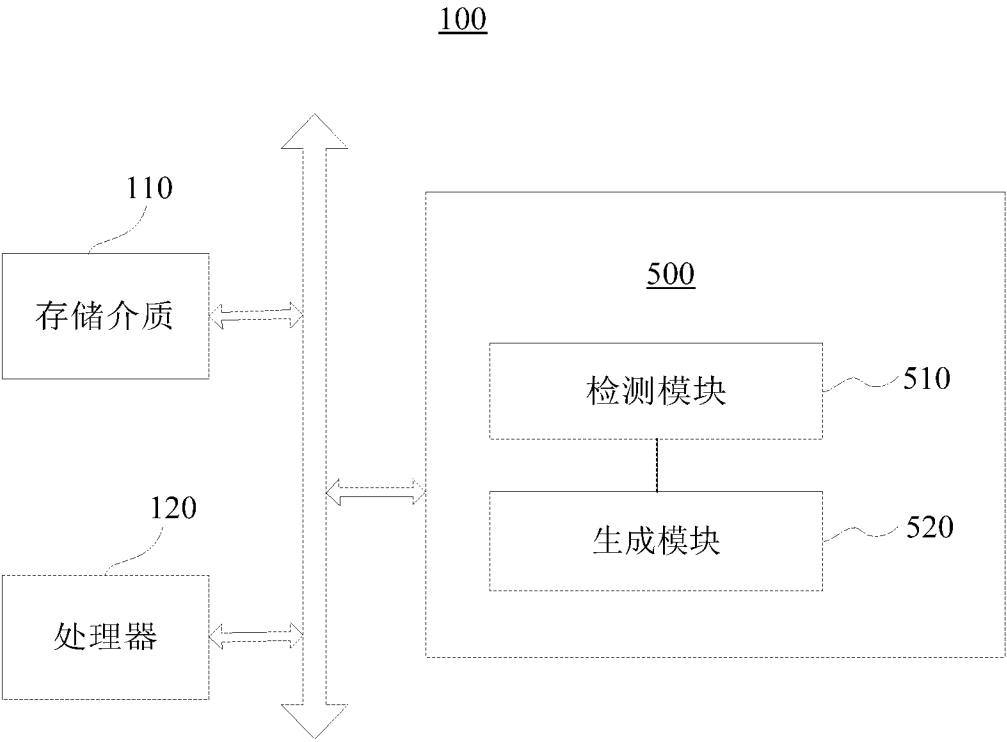


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081627

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/478(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI, VEN, WOTXT, USTXT, EPTXT: 直播, 主播, 即时通讯, 虚拟, 虚构, 角色, 形象, 人物, 对象, 表情, 动作, 行为, 神经网络, 逆向, 反向, 运动, 网格; live, broadcast, virtual, image, role, character, action

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106993195 A (GUANGZHOU CHUANGHUAN DIGITAL TECHNOLOGY CO., LTD.) 28 July 2017 (2017-07-28) description paragraphs [0017] -[0053] and figures 1-2	1, 5, 17-20
Y	CN 106993195 A (GUANGZHOU CHUANGHUAN DIGITAL TECHNOLOGY CO., LTD.) 28 July 2017 (2017-07-28) description paragraphs [0017]-[0053] and figures 1-2	2-4, 7-16
Y	CN 108681263 A (SHANGHAI HERAIN QISHEN MULTIMEDIA CO., LTD.) 19 October 2018 (2018-10-19) description paragraphs [0054]-[0081] and figures 1-2	2-4, 7-16
Y	CN 107423721 A (ZHUHAI XIYUE INFORMATION TECHNOLOGY CO., LTD.) 01 December 2017 (2017-12-01) description, paragraphs [0026]-[0034]	3, 7-16
Y	CN 108960185 A (TAIHUA WISDOM INDUSTRY GROUP CO., LTD.) 07 December 2018 (2018-12-07) description, paragraphs [0142]-[0158]	14-16



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

24 June 2020

Date of mailing of the international search report

30 June 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081627**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106804007 A (UNION NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 06 June 2017 (2017-06-06) entire document	1-20
PX	CN 109922354 A (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) 21 June 2019 (2019-06-21) entire document	1-6, 17-20
PX	CN 109936774 A (GUANGZHOU HUYA INFORMATION TECHNOLOGY CO., LTD.) 25 June 2019 (2019-06-25) entire document	7-16

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/081627

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	106993195	A	28 July 2017	None	
CN	108681263	A	19 October 2018	None	
CN	107423721	A	01 December 2017	None	
CN	108960185	A	07 December 2018	None	
CN	106804007	A	06 June 2017	None	
CN	109922354	A	21 June 2019	None	
CN	109936774	A	25 June 2019	None	

A. 主题的分类 H04N 21/478(2011.01) i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																										
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS, CNTXT, CNKI, VEN, WOTXT, USTXT, EPTXT; 直播, 主播, 即时通讯, 虚拟, 虚构, 角色, 形象, 人物, 对象, 表情, 动作, 行为, 神经网络, 逆向, 反向, 运动, 网格; live, broadcast, virtual, image, role, character, action																										
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2</td> <td>1, 5, 17-20</td> </tr> <tr> <td>Y</td> <td>CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2</td> <td>2-4, 7-16</td> </tr> <tr> <td>Y</td> <td>CN 108681263 A (上海恒润申启多媒体有限公司) 2018年 10月 19日 (2018 - 10 - 19) 说明书第[0054] - [0081]段及图1-2</td> <td>2-4, 7-16</td> </tr> <tr> <td>Y</td> <td>CN 107423721 A (珠海习悦信息技术有限公司) 2017年 12月 1日 (2017 - 12 - 01) 说明书第[0026] - [0034]段</td> <td>3, 7-16</td> </tr> <tr> <td>Y</td> <td>CN 108960185 A (泰华智慧产业集团股份有限公司) 2018年 12月 7日 (2018 - 12 - 07) 说明书第[0142] - [0158]段</td> <td>14-16</td> </tr> <tr> <td>A</td> <td>CN 106804007 A (合一网络技术北京有限公司) 2017年 6月 6日 (2017 - 06 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>PX</td> <td>CN 109922354 A (广州虎牙信息科技有限公司) 2019年 6月 21日 (2019 - 06 - 21) 全文</td> <td>1-6, 17-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2	1, 5, 17-20	Y	CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2	2-4, 7-16	Y	CN 108681263 A (上海恒润申启多媒体有限公司) 2018年 10月 19日 (2018 - 10 - 19) 说明书第[0054] - [0081]段及图1-2	2-4, 7-16	Y	CN 107423721 A (珠海习悦信息技术有限公司) 2017年 12月 1日 (2017 - 12 - 01) 说明书第[0026] - [0034]段	3, 7-16	Y	CN 108960185 A (泰华智慧产业集团股份有限公司) 2018年 12月 7日 (2018 - 12 - 07) 说明书第[0142] - [0158]段	14-16	A	CN 106804007 A (合一网络技术北京有限公司) 2017年 6月 6日 (2017 - 06 - 06) 全文	1-20	PX	CN 109922354 A (广州虎牙信息科技有限公司) 2019年 6月 21日 (2019 - 06 - 21) 全文	1-6, 17-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2	1, 5, 17-20																								
Y	CN 106993195 A (广州创幻数码科技有限公司) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0017] - [0053]段及图1-2	2-4, 7-16																								
Y	CN 108681263 A (上海恒润申启多媒体有限公司) 2018年 10月 19日 (2018 - 10 - 19) 说明书第[0054] - [0081]段及图1-2	2-4, 7-16																								
Y	CN 107423721 A (珠海习悦信息技术有限公司) 2017年 12月 1日 (2017 - 12 - 01) 说明书第[0026] - [0034]段	3, 7-16																								
Y	CN 108960185 A (泰华智慧产业集团股份有限公司) 2018年 12月 7日 (2018 - 12 - 07) 说明书第[0142] - [0158]段	14-16																								
A	CN 106804007 A (合一网络技术北京有限公司) 2017年 6月 6日 (2017 - 06 - 06) 全文	1-20																								
PX	CN 109922354 A (广州虎牙信息科技有限公司) 2019年 6月 21日 (2019 - 06 - 21) 全文	1-6, 17-20																								
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																										
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																									
国际检索实际完成的日期 2020年 6月 24日		国际检索报告邮寄日期 2020年 6月 30日																								
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 陈德锋 电话号码 86-(010)-62411501																								

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109936774 A (广州虎牙信息科技有限公司) 2019年 6月 25日 (2019 - 06 - 25) 全文	7-16

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/081627

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106993195	A	2017年 7月 28日	无	
CN	108681263	A	2018年 10月 19日	无	
CN	107423721	A	2017年 12月 1日	无	
CN	108960185	A	2018年 12月 7日	无	
CN	106804007	A	2017年 6月 6日	无	
CN	109922354	A	2019年 6月 21日	无	
CN	109936774	A	2019年 6月 25日	无	