

一种人体关键点检测方法及装置、电子设备和存储介质

本公开要求在2019年07月15日提交中国专利局、申请号为201910635763.6、申请名称为“一种人体关键点检测方法及装置、电子设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

5

技术领域

本公开涉及人体检测技术领域，尤其涉及一种人体关键点检测方法及装置、电子设备和存储介质。

10 **背景技术**

相关技术中，人体关键点检测技术是基于深度学习算法所开发的应用。在计算机算力不断提升、数据集不断扩大的背景下，深度学习算法作为机器学习的一个重要分支，被应用到了各个行业中。可对于诸如体感游戏、人体动态监控等应用场景中，对于人体处于运动状态时如何实现精确的人体关键点检测，目前未存在有效的解决方案。

15 **发明内容**

本公开提出了一种人体关键点检测技术方案。

根据本公开的一方面，提供了一种人体关键点检测方法，所述方法包括：

响应于检测到图像中包含人体，将所述图像中用于标识人体关键点位置的二维坐标数据提取出来，得到2D位姿数据；

20 将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。

采用本公开，将图像中用于标识人体关键点位置的二维坐标数据提取出来，可以得到2D位姿数据。将2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合，得到的3D位姿数据为用于标识人体关键点位置的三维坐标数据，通过标识人体
25 关键点位置的三维坐标数据能对于人体处于运动状态时实现精确的人体关键点检测。

可能的实现方式中，所述方法还包括：所述将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合之前，

将RGB图像数据流中每帧图像与对应同一图像的深度数据进行数据对齐的预处理，得到RGBD图像数据流。

30 采用本公开，将RGB数据和深度数据对齐，得到RGBD数据，可以实现对数据的预处理，之后分别对该RGB数据和RGBD数据进行相应的图像处理。

可能的实现方式中，所述检测到图像中包含人体，包括：

对于当前帧图像，经所述第一图像处理得到多个图像特征；

根据人体识别网络判断出所述多个图像特征为人体关键点特征的情况下，检测到所述当前帧图像中包含人体，直至对至少一帧图像完成检测。

采用本公开，根据人体识别网络判断出所述多个图像特征为人体关键点特征的情况下，可以检测到所述当前帧图像中包含人体。

可能的实现方式中，所述方法还包括：所述将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合之前，

对于当前帧图像，经所述第二图像处理得到多个深度数据，直至对至少一帧图像完成图像处理。

采用本公开，对于当前帧图像，经所述第二图像处理得到多个深度数据，直至对至少一帧图像完成图像处理，然后将多个深度数据与2D位姿数据相结合，以实现人体关键点的特征融合。

可能的实现方式中，所述方法还包括：

获取第一人体运动状态；

将所述第一人体运动状态对应的人体关键点位置变化通过第一3D位姿数据进行描述；根据所述第一3D位姿数据生成第一控制指令，将所述第一控制指令发送给接收侧设备，以在所述接收侧设备的显示屏上展示对应所述第一人体运动状态的动作模拟操作。

采用本公开，将所述第一人体运动状态对应的人体关键点位置变化通过第一3D位姿数据进行描述，通过发送第一控制指令给接收侧设备，实现了在接收侧设备的显示屏上展示对应第一人体运动状态的动作模拟操作。

可能的实现方式中，所述方法还包括：

获取第二人体运动状态；

将所述第二人体运动状态对应的人体关键点位置变化通过第二3D位姿数据进行描述；

将所述第二3D位姿数据与预配置的位姿数据进行比对，比对结果不一致的情况下生

成第二控制指令；

根据所述第二控制指令发出提示信息，以根据所述提示信息调整所述第二人体运动状态至符合目标状态。

采用本公开，将所述第二人体运动状态对应的人体关键点位置变化通过第二3D位姿数据进行描述，通过第二控制指令发出提示信息，实现了根据提示信息调整第二人体运动状态至符合目标状态。

可能的实现方式中，所述方法还包括：

获取第三人体运动状态；

将所述第三人体运动状态对应的人体关键点位置变化通过第三3D位姿数据进行描述；
将所述第三3D位姿数据发送给接收侧设备，以在所述接收侧设备的显示屏上展示由
虚拟形象采样所述第三3D位姿数据执行的操作。

采用本公开，将所述第三人体运动状态对应的人体关键点位置变化通过第三3D位姿
5 数据进行描述，将所述第三3D位姿数据发送给接收侧设备，实现了在所述接收侧设备的
显示屏上展示由虚拟形象采样所述第三3D位姿数据执行的操作。

可能的实现方式中，所述人体识别网络的训练过程包括：

将预先标注好的人体关键点特征作为训练样本数据，将所述训练样本数据输入待训
练的人体识别网络进行训练，直至输出结果满足网络训练条件，训练后得到所述人体识
10 别网络。

采用本公开，将预先标注好的人体关键点特征作为训练样本数据，将所述训练样本
数据输入待训练的人体识别网络进行训练，可以将训练后得到人体识别网络用于人体关
键点检测，且确保检测的高效和准确性。

根据本公开的一方面，提供了一种人体关键点检测装置，所述装置包括：

15 检测模块，用于响应于检测到图像中包含人体，将所述图像中用于标识人体关键点
位置的二维坐标数据提取出来，得到2D位姿数据；

融合模块，用于将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体
关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。

可能的实现方式中，所述装置还包括：预处理模块，用于：

20 将RGB图像数据流中每帧图像与对应同一图像的深度数据进行数据对齐的预处理，
得到RGBD图像数据流。

可能的实现方式中，所述检测模块，进一步用于：

对于当前帧图像，经所述第一图像处理得到多个图像特征；

根据人体识别网络判断出所述多个图像特征为人体关键点特征的情况下，检测到所
25 述当前帧图像中包含人体，直至对至少一帧图像完成检测。

可能的实现方式中，所述装置还包括：图像处理模块，用于：

对于当前帧图像，经所述第二图像处理得到多个深度数据，直至对至少一帧图像
完成图像处理。

可能的实现方式中，所述装置还包括：

30 第一姿态获取模块，用于获取第一人体运动状态；

第一数据描述模块，用于将所述第一人体运动状态对应的人体关键点位置变化通过
第一3D位姿数据进行描述；

第一指令发送模块，用于根据所述第一3D位姿数据生成第一控制指令，将所述第一控制指令发送给接收侧设备，以在所述接收侧设备的显示屏上展示对应所述第一人体运动状态的动作模拟操作。

可能的实现方式中，所述装置还包括：

5 第二姿态获取模块，用于获取第二人体运动状态；

第二数据描述模块，用于将所述第二人体运动状态对应的人体关键点位置变化通过第二3D位姿数据进行描述；

数据比对模块，用于将所述第二3D位姿数据与预配置的位姿数据进行比对，比对结果不一致的情况下生成第二控制指令；

10 提示信息发送模块，用于根据所述第二控制指令发出提示信息，以根据所述提示信息调整所述第二人体运动状态至符合目标状态。

可能的实现方式中，所述装置还包括：

第三姿态获取模块，用于获取第三人体运动状态；

15 第三数据描述模块，用于将所述第三人体运动状态对应的人体关键点位置变化通过第三3D位姿数据进行描述；

第二指令发送模块，用于将所述第三3D位姿数据发送给接收侧设备，以在所述接收侧设备的显示屏上展示由虚拟形象采样所述第三3D位姿数据执行的操作。

可能的实现方式中，所述装置还包括：网络训练模块，用于：

20 在所述人体识别网络的训练过程中，将预先标注好的人体关键点特征作为训练样本数据，将所述训练样本数据输入待训练的人体识别网络进行训练，直至输出结果满足网络训练条件，训练后得到所述人体识别网络

根据本公开的一方面，提供了一种电子设备，包括：

处理器；

用于存储处理器可执行指令的存储器；

25 其中，所述处理器被配置为：执行上述人体关键点检测方法。

根据本公开的一方面，提供了一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现上述人体关键点检测方法。

30 根据本公开的一方面，提供一种计算机程序，其中，所述计算机程序包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于实现上述人体关键点检测方法。

在本公开中，响应于检测到图像中包含人体，将所述图像中用于标识人体关键点位置的二维坐标数据提取出来，得到2D位姿数据。将所述2D位姿数据和对应所述人体关键

点位置的深度数据进行人体关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。采用本公开，将图像中用于标识人体关键点位置的二维坐标数据提取出来，可以得到2D位姿数据。将2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合，得到的3D位姿数据为用于标识人体关键点位置的三维坐标数据，通过标识人体
5 关键点位置的三维坐标数据能对于人体处于运动状态时实现精确的人体关键点检测。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。

根据下面参考附图对示例性实施例的详细说明，本公开的其它特征及方面将变得清楚。

10

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

图1示出根据本公开实施例的人体关键点检测方法的流程图。

15

图2示出根据本公开实施例的人体关键点检测方法的流程图。

图3示出根据本公开实施例的人体骨架关键点的示意图。

图4示出根据本公开实施例手持手机终端的用户与电视等大屏幕设备互动的场景图。

图5示出根据本公开实施例生成虚拟形象的场景图。

图6示出根据本公开实施例的人体检测方案的示意图。

20

图7示出根据本公开实施例的人体关键点检测装置的框图。

图8示出根据本公开实施例的电子设备的框图。

图9示出根据本公开实施例的电子设备的框图。

具体实施方式

25

以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面，但是除非特别指出，不必按比例绘制附图。

在这里专用的词“示例性”意为“用作例子、实施例或说明性”。这里作为“示例性”所说明的任何实施例不必解释为优于或好于其它实施例。

30

本文中术语“和/或”，仅仅是一种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：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中术语“至少一种”表示多种中的任意一种或多种中的至少两种的任意组合，

例如，包括A、B、C中的至少一种，可以表示包括从A、B和C构成的集合中选择的任意一个或多个元素。

另外，为了更好的说明本公开，在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解，没有某些具体细节，本公开同样可以实施。在一些实例中，
5 对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述，以便于凸显本公开的主旨。

人体关键点检测技术是基于深度学习算法所开发的应用。深度学习算法发展迅速，得到了广泛的关注。在计算机算力不断提升、数据集不断扩大的背景下，深度学习作为机器学习的一个重要分支，被应用到了各个行业中。在计算机视觉领域中，深度学习凭借优异的计算效果，较高的鲁棒性更是成为了行业的关键技术。通过深度学习的卷积神经网络，可以有效提取图像中的关键特征。传统的全连接神经网络存在参数数量多、没有利用像素间的位置信息、网络深度受限（越深的网络表达能力越强，但随之而来的训练参数也会大量增加）等问题。而卷积神经网络（CNN）则很好的解决了这些问题。首先，CNN中的连接为局部连接，每个神经元不再和上一层的至少一神经元相连，而只和
10 一小部分神经元相连。同时，一组连接可以共享同一个权重参数，加上下采样的策略，很大程度上减少了参数数量。不同与全连接网络的一维排列，CNN的神经元结构为三维排列。通过去除大量不重要的参数，保留重要的权重值，让一个深度的神经网络得以实现。从而能够处理更多复杂度更高的信息。

考虑到2D人体关键点对于应用层面的限制，是无法获取到三维坐标。虽然通过3D方案可以弥补此缺陷，可是，如果只是基于RGB数据输出预测的3D关键点，该检测方案的精度较低。考虑到3D方案具备深度数据的输入与合并，可以输出更精准的三维坐标点。然而，如果只基于深度数据，输出对应的3D关键点，可能极大的受限于硬件的深度图质量。
20

采用本公开的人体关键点检测方案，将通过RGB数据预测的3D坐标与深度数据进行整合，可以有效降低对于3D硬件模组采集深度数据准确度的依赖，从而实现更好的检测精度和鲁棒性。
25

图1示出根据本公开实施例的人体关键点检测方法的流程图，该人体关键点检测方法应用于人体关键点检测装置，例如，人体关键点检测装置可以由终端设备或服务器或其它处理设备执行，其中，终端设备可以为用户设备（UE，User Equipment）、移动设备、
30 蜂窝电话、无绳电话、个人数字处理（PDA，Personal Digital Assistant）、手持设备、计算设备、车载设备、可穿戴设备等。在一些可能的实现方式中，该人体关键点检测方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。如图1所示，该流程

包括：

步骤S101、响应于检测到图像中包含人体，将所述图像中用于标识人体关键点位置的二维坐标数据提取出来，得到2D位姿数据。

5 步骤S102、将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。

采用本公开，通过2D位姿数据+深度数据，可以得到3D位姿数据。其中，2D位姿数据是人体关键点在RGB图像中的二维坐标，3D位姿数据是3D的人体关键点，通过3D位姿数据，可以对人体处于运动状态进行精确的人体关键点检测，如将某一运动状态分解为：抬手、踢腿、摆头、弯腰中的至少一种节点位姿，从而实时跟踪这些节点位姿对应的人
10 体关键点检测。

图2示出根据本公开实施例的人体关键点检测方法的流程图，该人体关键点检测方法应用于人体关键点检测装置，例如，人体关键点检测装置可以由终端设备或服务器或其它处理设备执行，其中，终端设备可以为用户设备（UE，User Equipment）、移动设备、蜂窝电话、无绳电话、个人数字处理（PDA，Personal Digital Assistant）、手持设备、计
15 算设备、车载设备、可穿戴设备等。在一些可能的实现方式中，该人体关键点检测方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。如图2所示，该流程包括：

步骤S201、将RGB图像数据流中每帧图像与对应同一图像的深度数据进行数据对齐的预处理，得到RGBD图像数据流。

20 一个示例中，对于数据预处理，需要将RGB数据和深度数据对齐，得到RGBD数据，然后才可以执行本方法流程中分别对该RGB数据和RGBD数据的处理。

步骤S202、从RGB图像数据流中检测到图像中包含人体，将图像中用于标识人体关键点位置的二维坐标数据提取出来，得到2D位姿数据。

步骤S203、从RGBD图像数据流中获取深度数据，将2D位姿数据和深度数据（对应
25 人体关键点位置的深度数据）进行人体关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。

采用本公开，预处理后，每一个RGB和RGBD所构成的数据对，是对应同一视角的图像帧。是将RGB图像数据流中每帧图像的人体关键点与对应同一图像中人体关键点的深度数据进行对齐，从而，对于图像中任一个人体关键点而言，既有了表征该人体关键点位置的二维坐标，又有了表征该人体关键点运动变化的距离值，则得到了针对该人体
30 关键点位置的三维信息。其中，深度数据从深度图（DepthMap）中获取，在一个包含摄像机或摄像模块的采集场景中，DepthMap可以认为是：包含与该场景中所采集目标对象

其表面的距离有关的信息所构成的图像（或称为图像通道）。该场景中至少一个点相对于摄像机或摄像模块的距离用深度图来表示时，深度图中的每一个像素值可以表示场景中某一点与摄像机之间的距离。

本公开可能的实现方式中，所述检测到图像中包含人体，包括：获取所述RGB图像数据流，对所述RGB图像数据流中的每帧图像进行第一图像处理。对于当前帧图像，经所述第一图像处理得到多个图像特征。根据人体识别网络判断出所述多个图像特征为人体关键点特征的情况下，检测到所述当前帧图像中包含人体，直至对至少一帧图像完成检测。

本公开可能的实现方式中，所述方法还包括：所述将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合之前，获取所述RGBD图像数据流，对所述RGBD图像数据流中的每帧图像进行第二图像处理。对于当前帧图像，经所述第二图像处理得到多个深度数据，直至对至少一帧图像完成图像处理。

一个示例中，数据对齐预处理：是根据多路RGB数据流得到多路RGBD数据流。可以将RGB图像数据流中每帧图像的人体关键点与对应同一图像中人体关键点的深度数据进行对齐，如果将RGB和RGBD当做数据对来看，则每一个RGB和RGBD构成的数据对，二者都是对应同一视角的图像帧。数据对齐预处理后，可以输入多个RGB和RGBD数据对，比如，可以分两路输入运行本公开人体关键点检测流程的逻辑模型，对于第一路数据（RGB数据），经第一图像处理，通过已经训练得到的人体追踪网络来判断当前图像帧中是否检测到人体，如果检测到人体，将对应当前图像帧中的目标RGB数据交由后续步骤处理。在后续步骤中，对于第二路数据（RGBD数据），经第二图像处理，将RGBD数据与目标RGB数据结合，以根据RGBD数据与目标RGB数据得到3D位姿数据（3D坐标的人体骨架关键点）。动态跟踪：用3D坐标表征人体骨架关键点的3D位姿数据，实现对人体处于运动状态的追踪，如跟踪节点位姿对的变化，支持抬手、踢腿、摆头、弯腰等至少一种人体动作。

一个示例中，运行本公开人体关键点检测流程的处理逻辑，可以采用离线端软件开发工具包（SDK，Software Development Kit）形式集成到手机上。基于手机作为移动端的算法优化，可以加快上述处理逻辑的运行速度，区别于现有技术C/S在线模式是将处理逻辑放置在服务器，这样，若终端发起请求到服务器，二者间容易有传输时延，或者网络故障等导致不能及时得到终端所请求的处理结果。而本公开将处理逻辑直接以SDK的离线模式放置在终端，大大加快了检测方法的处理效率。

图3示出根据本公开实施例的人体骨架关键点的示意图，包括人体骨架中17个关键点，通过检测这17个关键点，可以实时跟踪用户的动态姿态变化，如抬手、踢腿、摆头、弯

腰等至少一种人体动作。

如手持手机终端的用户如何与电视等大屏幕设备互动的场景中，获取第一人体运动状态（比如打网球时的挥拍动作等），将所述第一人体运动状态对应的人体关键点位置变化通过第一3D位姿数据进行描述。根据所述第一3D位姿数据生成第一控制指令，将所述
5 第一控制指令发送给接收侧设备，以在所述接收侧设备的显示屏上展示对应所述第一人体运动状态的动作模拟操作。这只是一个示例，其他互动场景也适用本公开。对于这类手机端的体感游戏而言，采用相关技术中的体感游戏机，如Xbox、Switch等，都需要另外购买设备，使用成本和空间成本都相对手机终端更高。采用本公开，用户只要拥有飞行时间（ToF，Time of Flight）手机，便能连接屏幕进行游戏。同时加上3D人体坐标检测
10 的技术，能够带来更丰富的游戏内容。ToF手机可以设置TOF模块，其3D成像方案可以通过给目标物体连续发送光脉冲，然后用传感器接收从目标物体返回的光，通过探测光脉冲的飞行（往返）时间来得到目标物体基于采集位置的距离。

图4示出根据本公开实施例手持手机终端的用户与电视等大屏幕设备互动的场景图，为一个打羽毛球的互动场景，可以通过检测用户的人体骨架关键点跟踪用户当前的姿态
15 变化，将得到的姿态变化回传回电视等电子设备中，并在该电子设备中呈现对应的姿态变化。

如手持手机的用户在锻炼，如何纠正锻炼姿态的场景中，获取第二人体运动状态（如尝试抬起双手至与水平面呈现90度），将所述第二人体运动状态对应的人体关键点位置变化通过第二3D位姿数据进行描述。将所述第二3D位姿数据与预配置的位姿数据进行比对，
20 比对结果不一致的情况下生成第二控制指令（比如，用户抬起双手仅仅达到85度，未能与预配置的位姿数据“90度”一致），根据所述第二控制指令发出提示信息，以根据所述提示信息调整所述第二人体运动状态至符合目标状态。其中，所述提示信息包括：语音，文字、声光电等信息，提示用户注意当前运动姿态完全不对或者姿势不到位等。这只是一个示例，其他锻炼姿态纠正场景也适用本公开。比如，可以基于本公开开发健身行业的
25 虚拟教练软件，通过手机或类似3D模组检测到用户的健身动作，并给予指导。

如将用户的人体数据应用于虚拟形象的场景中，获取第三人体运动状态（如用户奔跑的姿态），将所述第三人体运动状态对应的人体关键点位置变化通过第三3D位姿数据进行描述。将所述第三3D位姿数据发送给接收侧设备，以在所述接收侧设备的显示屏上展示由虚拟形象采样所述第三3D位姿数据执行的操作（虚拟形象可以是一个小动物、一个
30 男孩或一个女孩正在游戏场景中进行奔跑）。这只是一个示例，其他虚拟形象场景也适用本公开。比如，可以基于本公开开发虚拟游戏，通过实时的用户动作捕捉，驱动一个虚拟的形象，代替真人在游戏场景，为一种跨越触屏的交互方式。图5示出根据本公开实施

例生成虚拟形象的场景图，为一个跑酷的场景，可以通过检测用户的人体骨架关键点生成电视等电子设备中虚拟形象对应的姿态变化数据，并在该电子设备中呈现对应的姿态变化。

本公开可能的实现方式中，所述人体识别网络的训练过程包括：将预先标注好的人体关键点特征作为训练样本数据，将所述训练样本数据输入待训练的人体识别网络（如CNN）进行训练，直至输出结果满足网络训练条件，训练后得到所述人体识别网络。通过CNN可以提取图像中的人体关键点特征，根据人体的骨架关键点标注数据集训练的算法模型用于识别图像中是否包括人体。借助深度学习算法的优异性能，并利用深度数据，可以得到准确的节点位姿，并实时跟踪节点位姿对的变化，支持抬手、踢腿、摆头、弯腰等至少一种人体动作。

应用示例：

图6示出根据本公开实施例的人体检测方案的示意图。图6中，对于两路图像数据流，如RGB图像数据流和RGBD图像数据流分别进行图像处理。对于RGB图像数据流这一路处理中，经图像处理判断当前RGB图像帧中是否检测到人体，如果检测到人体，将对应当前RGB图像帧中的目标RGB数据交由后续对于RGBD图像数据流这一路处理中予以处理。对于RGBD图像数据流这一路处理中，经图像处理后将得到的目标RGBD数据（深度数据）与目标RGB数据（2D位姿数据）结合，以根据2D位姿数据和深度数据得到3D位姿数据，即3D坐标的人体骨架关键点，将3D位姿数据进行数据转化后得到数据转化结果，以用于至少一种场景的检测处理。

本领域技术人员可以理解，在具体实施方式的上述方法中，各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定，各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

本公开提及的上述各个方法实施例，在不违背原理逻辑的情况下，均可以彼此相互结合形成结合后的实施例，限于篇幅，本公开不再赘述。

此外，本公开还提供了人体关键点检测装置、电子设备、计算机可读存储介质、程序，上述均可用来实现本公开提供的任一种人体关键点检测方法，相应技术方案和描述和参见方法部分的相应记载，不再赘述。

图7示出根据本公开实施例的人体关键点检测装置的框图，如图7所示，本公开实施例的人体关键点检测装置，包括：检测模块31，用于响应于检测到图像中包含人体，将所述图像中用于标识人体关键点位置的二维坐标数据提取出来，得到2D位姿数据；融合模块32，用于将所述2D位姿数据和对应所述人体关键点位置的深度数据进行人体关键点特征融合，得到用于标识人体关键点位置的3D位姿数据。

可能的实现方式中，所述装置还包括：预处理模块，用于：将RGB图像数据流中每帧图像与对应同一图像的深度数据进行数据对齐的预处理，得到RGBD图像数据流。

可能的实现方式中，所述检测模块，进一步用于：对于当前帧图像，经所述第一图像处理得到多个图像特征；根据人体识别网络判断出所述多个图像特征为人体关键点特征的情况下，检测到所述当前帧图像中包含人体，直至对至少一帧图像完成检测。

可能的实现方式中，所述装置还包括：图像处理模块，用于：对于当前帧图像，经所述第二图像处理得到多个深度数据，直至对至少一帧图像完成图像处理。

可能的实现方式中，所述装置还包括：第一姿态获取模块，用于获取第一人体运动状态；第一数据描述模块，用于将所述第一人体运动状态对应的人体关键点位置变化通过第一3D位姿数据进行描述；第一指令发送模块，用于根据所述第一3D位姿数据生成第一控制指令，将所述第一控制指令发送给接收侧设备，以在所述接收侧设备的显示屏上展示对应所述第一人体运动状态的动作模拟操作。

可能的实现方式中，所述装置还包括：第二姿态获取模块，用于获取第二人体运动状态；第二数据描述模块，用于将所述第二人体运动状态对应的人体关键点位置变化通过第二3D位姿数据进行描述；数据比对模块，用于将所述第二3D位姿数据与预配置的位姿数据进行比对，比对结果不一致的情况下生成第二控制指令；提示信息发送模块，用于根据所述第二控制指令发出提示信息，以根据所述提示信息调整所述第二人体运动状态至符合目标状态。

可能的实现方式中，所述装置还包括：第三姿态获取模块，用于获取第三人体运动状态；第三数据描述模块，用于将所述第三人体运动状态对应的人体关键点位置变化通过第三3D位姿数据进行描述；第二指令发送模块，用于将所述第三3D位姿数据发送给接收侧设备，以在所述接收侧设备的显示屏上展示由虚拟形象采样所述第三3D位姿数据执行的操作。

可能的实现方式中，所述装置还包括：网络训练模块，用于：在所述人体识别网络的训练过程中，将预先标注好的人体关键点特征作为训练样本数据，将所述训练样本数据输入待训练的人体识别网络进行训练，直至输出结果满足网络训练条件，训练后得到所述人体识别网络。

在一些实施例中，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述。

本公开实施例还提出一种计算机可读存储介质，其上存储有计算机程序指令，所述

计算机程序指令被处理器执行时实现上述人体关键点检测方法。计算机可读存储介质可以是非易失性计算机可读存储介质。

本公开实施例还提出一种电子设备，包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器被配置为上述人体关键点检测方法。

5 电子设备可以被提供为终端、服务器或其它形态的设备。

本公开实施例还提出一种计算机程序，其中，所述计算机程序包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于实现上述人体关键点检测方法。

图8是根据一示例性实施例示出的一种电子设备800的框图。例如，电子设备800可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等终端。

参照图8，电子设备800可以包括以下一个或多个组件：处理组件802，存储器804，电源组件806，多媒体组件808，音频组件810，输入/输出（I/O）的接口812，传感器组件814，以及通信组件816。

15 处理组件802通常控制电子设备800的整体操作，诸如与显示，电话呼叫，数据通信，相机操作和记录操作相关联的操作。处理组件802可以包括一个或多个处理器820来执行指令，以完成上述的方法的全部或部分步骤。此外，处理组件802可以包括一个或多个模块，便于处理组件802和其他组件之间的交互。例如，处理组件802可以包括多媒体模块，以方便多媒体组件808和处理组件802之间的交互。

20 存储器804被配置为存储各种类型的数据以支持在电子设备800的操作。这些数据的示例包括用于在电子设备800上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

电源组件806为电子设备800的各种组件提供电力。电源组件806可以包括电源管理系统，一个或多个电源，及其他与为电子设备800生成、管理和分配电力相关联的组件。

多媒体组件808包括在所述电子设备800和用户之间的提供一个输出接口的屏幕。在一些实施例中，屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一

些实施例中，多媒体组件808包括一个前置摄像头和/或后置摄像头。当电子设备800处于操作模式，如拍摄模式或视频模式时，前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

5 音频组件810被配置为输出和/或输入音频信号。例如，音频组件810包括一个麦克风（MIC），当电子设备800处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中，音频组件810还包括一个扬声器，用于输出音频信号。

10 I/O接口812为处理组件802和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

传感器组件814包括一个或多个传感器，用于为电子设备800提供各个方面的状态评估。例如，传感器组件814可以检测到电子设备800的打开/关闭状态，组件的相对定位，
15 例如所述组件为电子设备800的显示器和小键盘，传感器组件814还可以检测电子设备800或电子设备800一个组件的位置改变，用户与电子设备800接触的存在或不存在，电子设备800方位或加速/减速和电子设备800的温度变化。传感器组件814可以包括接近传感器，被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器，如CMOS或CCD图像传感器，用于在成像应用中使用。在一些实施例中，该传
20 感器组件814还可以包括加速度传感器，陀螺仪传感器，磁传感器，压力传感器或温度传感器。

通信组件816被配置为便于电子设备800和其他设备之间有线或无线方式的通信。电子设备800可以接入基于通信标准的无线网络，如WiFi，2G或3G，或它们的组合。在一个示例性实施例中，通信组件816经由广播信道接收来自外部广播管理系统的广播信号或
25 广播相关信息。在一个示例性实施例中，所述通信组件816还包括近场通信（NFC）模块，以促进短程通信。例如，在NFC模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

在示例性实施例中，电子设备800可以被一个或多个应用专用集成电路（ASIC）、数字信号处理器（DSP）、数字信号处理设备（DSPD）、可编程逻辑器件（PLD）、现场可编程门阵列（FPGA）、控制器、微控制器、微处理器或其他电子元件实现，用于执行上述方法。
30 在示例性实施例中，还提供了一种非易失性计算机可读存储介质，例如包括计算机

程序指令的存储器804，上述计算机程序指令可由电子设备800的处理器820执行以完成上述方法。

图9是根据一示例性实施例示出的一种电子设备900的框图。例如，电子设备900可以被提供为一服务器。参照图9，电子设备900包括处理组件922，其进一步包括一个或多个
5 处理器，以及由存储器932所代表的存储器资源，用于存储可由处理组件922的执行的指令，例如应用程序。存储器932中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外，处理组件922被配置为执行指令，以执行上述方法。

电子设备900还可以包括一个电源组件926被配置为执行电子设备900的电源管理，一个有线或无线网络接口950被配置为将电子设备900连接到网络，和一个输入输出（I/O）
10 接口958。电子设备900可以操作基于存储在存储器932的操作系统，例如Windows Server™，Mac OS X™，Unix™，Linux™，FreeBSD™或类似。

在示例性实施例中，还提供了一种非易失性计算机可读存储介质，例如包括计算机程序指令的存储器932，上述计算机程序指令可由电子设备900的处理组件922执行以完成上述方法。

15 本公开可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于使处理器实现本公开的各个方面的计算机可读程序指令。

计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是——但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM或闪存）、静态随机存取存储器（SRAM）、便携式压缩盘只读存储器（CD-ROM）、数字多功能盘（DVD）、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身，诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波（例如，通过光纤电缆的光脉冲）、或者通过电线传输的电信号。
20 25

这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备，或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从
30 网络接收计算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构（ISA）指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等，以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。

- 5 计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。在一些实施例中，通过利用
- 10 计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列（FPGA）或可编程逻辑阵列（PLA），该电子电路可以执行计算机可读程序指令，从而实现本公开的各个方面。

- 这里参照根据本公开实施例的方法、装置（系统）和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/
- 15 或框图中各方框的组合，都可以由计算机可读程序指令实现。

- 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些
- 20 指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

- 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以
- 25 产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

- 附图中的流程图和框图显示了根据本公开的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，所述模块、程序段或指令的一部分包含一个或
- 30 多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意

的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

- 5 以上已经描述了本公开的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实施例的原理、实际应用或对市场中技术的技术改进，或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。