

数据处理方法及装置、电子设备及存储介质

相关申请的交叉引用

本申请基于申请号为 201811089872.4、申请日为 2018 年 09 月 18 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此引入本申请作为参考。

技术领域

本申请涉及信息技术领域但不限于信息技术领域，尤其涉及一种数据处理方法及装置、电子设备及存储介质。

10 背景技术

在体感游戏等体感场景中，一般都需由人体佩戴体感设备，体感设备采集人体的 3D 姿态传输给受控设备，进行受控设备的控制。这种受控设备的控制，通常需要使用到体感设备。

发明内容

15 有鉴于此，本申请实施例期望提供一种数据处理方法及装置、电子设备及存储介质。

本申请的技术方案是这样实现的：

一种数据处理方法，包括：

20 根据参考深度值及图像中目标的关键点的实际深度值，将所述关键点的第一 2D 坐标的转换为第二 2D 坐标，其中，所述第二 2D 坐标和所述参考深度值，构成了所述关键点的第一 3D 特征；

基于所述第一 3D 特征，获得所述目标的 3D 姿态。

一种数据处理装置，包括：

25 第一转换模块，配置为根据参考深度值及图像中目标的关键点的实际深度值，将所述关键点的第一 2D 坐标的转换为第二 2D 坐标，其中，所述第二 2D 坐标和所述参考深度值，构成了所述关键点的第一 3D 特征；

第一获得模块，配置为基于所述第一 3D 特征，获得所述目标的 3D 姿态。

30 一种计算机存储介质，所述计算机存储介质存储有计算机可执行代码；所述计算机可执行代码被执行后，能够实现一个或多个技术方案提供的数

据处理方法。

一种电子设备，其中，包括：

存储器，用于存储信息；

5 处理器，与所述存储器连接，用于通过执行存储在所述存储器上的计算机可执行指令，能够实现一个或多个技术方案提供的数据处理方法。

本申请实施例提供的技术方案，可以对目标采集 3D 图像。该 3D 图像包括 2D 图像和深度图像组成，深度图像提供标识目标与相机之间距离的深度值；3D 图像提供 RGB 或 YUV 等目标在 2D 成像平面内的姿势。如此，3D 图像可以用于目标的三维姿态获取。例如，通过神经网络等深度学习模块对 3D 图像的处理，可以提取出目标在三维空间内的三维姿态。但是在图像采集的过程中，目标相对于相机是前后移动的；如此，使得相机采集的深度图像中的深度值发生变化。若深度学习模块能够识别不同远近目标的三维姿态，一方面需要利用特定远近的训练样本进行训练，训练难度大，且训练周期长；另一方面，即便利用各种远近的训练样本进行了训练之后，深度学习模块的训练效果未必足够好，故对于某些样本比较少的远近的 3D 图像的三维姿态提取精度依然会不够。在本申请实施例中，在将 3D 图像目标关键点的第一 3D 特征输入到深度模型之前，通过 2D 成像平面内目标的平移，实际深度值向参考深度值的转换，可以获得目标在参考深度值时的第一 3D 特征，将该第一 3D 特征输入到深度学习模块中进行处理，由于采用的参考深度值是深度学习模块的训练时使用的深度值，从而可以精准的提取出目标的 3D 姿态；同时减少了深度学习模块训练所需的样本和时间。

附图说明

图 1A 为本申请实施例提供的一种数据处理方法的流程示意图；

图 1B 为本申请实施例提供的一种数据处理方法的流程示意图；

25 图 2 为本申请实施例提供的一种第一 2D 坐标转换为第二 2D 坐标的转换示意图；

图 3 为本申请实施例提供的一种关键点的示意图；

图 4 为本申请实施例提供的一种 2D 图像中目标平移的效果示意图；

图 5 为本申请实施例提供的得到关键点的第一 2D 坐标的示意图；

30 图 6A 为本申请实施例提供的一种关键点和基准点的示意图；

图 6B 为本申请实施例提供的另一种关键点的示意图；

图 7 为本申请实施例提供的一种数据处理装置的结构示意图；

图 8 为本申请实施例提供的一种电子设备的结构示意图；

图 9 为本申请实施例提供的一种神经网络的网络示意图。

具体实施方式

以下结合说明书附图及具体实施例对本申请的技术方案做进一步的详细阐述。

如图 1A 所示, 本实施例提供一种数据处理方法, 包括:

5 步骤 S110: 根据参考深度值及图像中目标的关键点的实际深度值, 将所述关键点的第一 2D 坐标的转换为第二 2D 坐标, 其中, 所述第二 2D 坐标和所述参考深度值, 构成了所述关键点的第一 3D 特征;

步骤 S120: 基于所述第一 3D 特征, 获得所述目标的 3D 姿态。

10 本实施例中提供基于数据处理方法, 该基于数据处理方法可以应用于一台或多台电子设备中。所述电子设备可包括: 处理器, 该处理器通过计算机程序等可执行指令的执行, 可以实现基于数据处理方法中一个或多个步骤的执行。在一些实施例中, 可以单一电子设备进行集中式的数据处理, 也可以运用多台电子设备进行分布式的数据处理。

所述图像可为三维图像; 所述三维图像包括: 2D 图像和深度图像。所述 2D 图像可为 RGB 图像或 YUV 图像等。所述深度图像可为利用深度采集模组采集的深度信息。所述深度信息的像素值为深度值。所述深度值可为: 图像采集模组距离目标之间的距离。此处, 本申请实施例中所述实际深度值就来自深度图像

20 通过第一 2D 坐标到第二 2D 坐标的转换, 就得到了将目标的实际深度值转换为参考深度值之后, 在相机坐标系上目标的关键点的 2D 坐标。

在步骤 S110 中, 就相当于得到了可供深度学习模块能够精准提取出目标的 3D 姿态的第一 3D 特征。向深度学习模块输入所述第一 3D 特征, 深度学习模型会自动输出目标的 3D 姿态; 所述 3D 姿态可用于位于三维空间坐标系内的第一 3D 特征之间的相对位置表示。

25 此处的所述第一 3D 特征可包括: 图像坐标系内的坐标和参考深度值构成。

例如, 基于所述关键点得到了目标的基准点的 2D 坐标; 所述 3D 姿态可为表征采集对象在三维空间内的姿势体态的信息。具体而言, 所述 3D 姿态可由: 各个关键点与所述基准点之间的相对位置表示。假设, 在 3D 空间内, 若所述基准点的 2D 坐标为: (0, 0, 0), 且目标为人体, 则可以通过表征人体骨骼的多个关键点相对于 (0, 0, 0) 的相对位置或相对坐标。

30 所述基准点可为: 人体髌部两端之间的中心点。例如, 所述关键点可为: 表示头、颈、肘、腕、髌、膝及踝的坐标点。如此, 根据这些关键点相对于基准点的相对位置, 就知道当前人体前、后、左、右的平移距离, 还可以根据人脸的关键点与基准点之间的相对位置, 知道人脸朝向, 从而知道人体头部的转动量和/或转动方向等转动参数; 根据躯干的关键点与基准点之间的相对位置, 可以知道躯干的转动量和/或转动方向等转动参数。

所述人脸的关键点可取位于鼻子上的一个点，例如，鼻尖坐标点。所述躯干的关键点可取胸部中心点坐标。当然以上仅是关键点举例，具体实现不局限于此。

进一步地，若所述目标为人体，则在步骤 S110 中可以根据 N 个关键点的坐标，得到 N+M 个第二 2D 坐标。其中增加的 M 个第二 2D 坐标，可为根据 N 个关键点的第一 2D 坐标生成的。例如，所述 M 可为 1，增加的 1 个第二 2D 坐标可对应于人体的基准点的 2D 坐标。所述 N 可为 14。

在步骤 S120 中可以根据 N+1 个关键点的第二 2D 坐标和参考深度值输入到深度学习模块中，可能会得到 N+S 个关键点的三维 (3D) 坐标，作为所述 3D 姿态输出。其中，N+S 个关键点中的 N 个关键点与第一 2D 坐标的 N 个关键点一一对应；S 个关键点是基于 N 个关键点生成的。

例如，以人体为例，N 个第一 2D 坐标可为：14 个关键点；S 可等于 3；即最终会得到 17 个关键点的第一 3D 特征。在一些实施例中，17 个关键点中有一个为基准点。该基准点可为人体髋部两个端点（对应于两个关键点）的中心点。另外两个关键点可为人脸的鼻头坐标及胸部的中心点坐标；当然此处仅是举例，具体实现不局限于此。

图 6A 可为相对于图 3 所示的 14 个关键点增加了关键点 0 的示意图；图 6B 可为一种基于图 3 所示的 14 个关键点生成的 17 个关键点的示意图。图 6B 中 17 个关键点，相当于图 3 所示的关键点，增加了关键点 0、关键点 15 及关键点 16；其中，关键点 16 的 2D 坐标可基于关键点 1 及关键点 2 的 2D 坐标的初步确定；关键点 15 的 2D 坐标可根据关键点 2 的 2D 坐标及关键点 0 的 2D 坐标的确定的。关键点 0 可为本申请实施例提供的基准点。

在本申请实施例中，一方面，神经网络等深度学习模块在训练过程中，若为了能够直接检测不同实际深度的目标的 3D 姿态，则需要将利用不同实际深度值的训练样本训练神经网络；若这样，训练所需训练样本多，训练样本多则神经网络等深度学习模块的收敛速度慢，导致训练周期长。若采用本实施例中的方法，神经网络等深度学习模块可以仅采用为同一个深度值的训练样本进行训练，从而训练样本的数据量小；神经网络等深度学习模块的收敛速度快，训练周期短等；如此，可以简化神经网络等深度学习模块。

另一方面，若采用单一深度值（即参考深度值），可以使得神经网络等深度学习模块不会因为需要顾及到不同深度值牺牲掉单一深度值所对应的 3D 坐标的 3D 姿态提取的精准度；从而具有 3D 姿态提取的精准度高的特点。

第一 3D 特征深度学习模块第一 3D 特征深度学习模块深度学习模块深度学习模块在一些实施例中，所述步骤 S110 可包括：根据所述实际深度值与所述参考深度值的比值，及所述第一 2D 坐标，得到所述第二 2D 坐标。

进一步地，例如，所述步骤 S110 可包括：利用如下函数关系确定所述

第二 2D 坐标;

$$X2 = (X1 * d) / D,$$

$$Y2 = (Y1 * d) / D,$$

其中, $X2$ 为所述第二 2D 坐标在第一方向上的坐标值; $X1$ 为所述第
5 一 2D 坐标在所述第一方向上的坐标值;

$Y2$ 为第二 2D 坐标在第二方向上的坐标值, $Y1$ 为所述第一 2D 坐标在
所述第二方向上的坐标值; 其中, 所述第二方向垂直于所述第一方向;

d 为所述实际深度值; D 为所述参考深度值。

所述 D 可为空间距离, 单位可为: 毫米、厘米或分米等。

10 参考如图 2 所示, of 为图像采集的焦距 (可简写成 f), 是可以通过查
询相机参数知道的, 通过三角函数变换可以得到第二 2D 坐标和参考深度
值。第二 2D 坐标和参考深度值, 构成了所述第一 3D 特征, 将标准的深度
值的第一 3D 特征输入到深度学习模块, 能够实现目标 3D 姿态的精准提取。
故在一些实施例中 od 表示的距离为所述实际深度值, 可简写成 d ; oD 表示
15 的距离为参考深度值。根据三角函数关系可以指 $y0/y1 = f/d$; $y2/y1 = f/D$; $y0$
表示的第一 2D 坐标, $y2$ 表示的第二 2D 坐标。如此, $y2 = (d * y0) / D$ 。

在本实施例中, 根据图像关键点在采集的图像中的 2D 坐标, 在本实施
例中实际采集得到的图像的 2D 坐标称之为第三 2D 坐标。如图 1B 所示,
所述方法还包括:

20 步骤 S100: 光心位置根据所述关键点的第二 3D 特征及所述图像对应的
光心位置, 得到所述第一 2D 坐标。

实际采集的 2D 图像上目标人体的关键点。例如, 人体骨骼上的关键点。

所述关键点的个数可为多个。例如, 若目标为人体, 则所述关键点可
包括: 14 个 2D 关键点。此处, 可以利用 2D 图像的深度学习模块, 通过对
25 2D 图像的处理, 获得所述第三 2D 坐标。

如图 3 所示为一种人体骨骼的关键点的 2D 坐标。在图 3 中用 14 个黑
色圆点表示 14 个关键点。

在一些实施例中, 可以利用深度学习模块对所述 2D 图像进行处理, 从
而获得所述第三 2D 坐标。该第三 3D 坐标 和从深度图像中提取的实际深
30 度值, 可以组成第二 3D 特征。

利用深度学习模块基于目标的第一 3D 特征进行目标的 3D 姿态估计
时, 当前目标距离图像采集模组之间的距离可近可远, 若深度学习模块之
前缺少对应的训练样本, 则会导致无法精确的预估目标 3D 姿态。而一个深
度学习模块需要尽可能对不同远近的 3D 图像中目标 3D 姿态的精准提取,
35 则需要引入更多的训练样本对深度学习模块进行处理; 如此, 深度学习模
块的训练难度大且训练周期长。所述深度学习模块可为: 各种神经网络,
例如, 可包括: 全连接网络及残差网络的残差模块等具有 3D 姿态识别的网
络。故在本实施例中, 为了提升目标 3D 姿态的精准度, 会将目标的关键点

的第一 3D 特征中深度值转换为参考深度值。

为了实现成功将所述关键点的第一 3D 特征中的深度值转换为参考深度值，首先需要将第三 2D 坐标转换为第一 2D 坐标，使得转换后的 2D 坐标位于所述图像的光轴上。

- 5 图 4 为一个 2D 图像，拍摄的人物原本位于照片的非中间位置，通过基于光心位置之间的坐标平移，可以将图 4 中实线表示的人物从第三 2D 坐标所在的位置，移动到虚线表示的第一 2D 坐标所在的位置上。通过将关键点中的基准点在相机平面上的平移，使得基准点移动到相机平面的光轴上；相对于直接向深度学习模块输入第三 2D 坐标可以减少干扰，从而提升 3D 姿态的精准度，同时减少 3D 姿态提取的深度学习模块训练所需的数据和/或时长，再次简化深度学习模块的训练及提升训练的速率。

第三 2D 坐标到第一 2D 坐标的转换方式有多种，以下提供一种可选方式：

如图 5 所示，所述步骤 S100 可包括：

- 15 步骤 S101：移动所述关键点的第二 3D 特征，使得所述关键点中基准点的 3D 特征平移到所述光心位置，并得到各所述关键点的第三 3D 特征；
步骤 S102：将所述第三 3D 特征投影到 2D 成像平面，得到所述第一 2D 坐标。

- 在本实施例中，若神经网络等深度学习模型从 3D 图像中的关键点的第二 3D 特征不包括基准点的第二 3D 特征时，可以根据其他关键点的第三 2D 坐标，得到基准点的 2D 坐标，然后基于基准点的 2D 坐标查找深度图像得到基准点对应位置的实际深度值；从而获得基准点的第二 3D 特征。然后在步骤 100 中整体移动所有关键点，在移动的过程中，使得基准点的第一 3D 特征移动到光心位置上。例如，光心位置 $(0, 0, 0)$ ；可以根据基准点的第二 3D 特征移动到光心位置的移动向量，求解出其他关键点的第二 3D 特征与基准点采用同样移动向量移动之后的第三 3D 特征。

在得到所有关键点的第三 3D 特征之后，可以将第三 3D 特征投影到 2D 成像平面就得到了前述的第一 2D 坐标。

- 30 通过第二 3D 坐标的移动，至少使得目标的基准点移动到了图像的相机坐标系的光轴上。神经网络等深度学习模块对位于光轴上的目标的 3D 姿态的提取精准度是更高的；减少因为目标的基准点位于非光轴位置上引入了误差的现象，提升 3D 姿态的精准度。在还有一些实施例中，若所述目标为人体骨骼，所述基准点的第一 3D 特征是基于所述关键点中的两个胯部关键点的第二 3D 特征确定的。

如图 6B 所示的关键点 9 及关键点 10 的第三 2D 坐标，可以计算这两个关键点的基准点的 2D 坐标。而该点的坐标即为所述基准点的 2D 坐标。

在一些实施例中，所述基准点的 2D 坐标可以称之为根节点的 2D 坐标。

在一些实施例中，所述基准点可为目标的基准点或者靠近中心的位置的一个点。在本实施例中针对人体而言，采用的两个跨部关键点的基准点作为基准点的 2D 坐标，是与人体的具体结构相适合的。

在一些实施例中，所述方法还包括：所述基于所述第一 3D 特征，获得
5 所述目标的 3D 姿态，包括：对所述关键点的第二 2D 坐标对应的深度值减去所述基准点的深度值，得到第四 2D 坐标及所述第四 2D 坐标对应的深度值；

对所述第四 2D 坐标及所述第四 2D 坐标对应的深度值进行归一化处理，得到归一化后的所述第四 2D 坐标及归一化后的所述第四 2D 坐标对应的
10 深度值；

利用深度学习模型对归一化后的所述第四 2D 坐标及归一化后的所述第四 2D 坐标对应的深度值进行处理，得到所述目标的 3D 姿态。

例如，将归一化后的第四 2D 坐标及其对应的深度值，分别输入神经网络；神经网络可直接输出所述 3D 姿态；或者，神经网络可以输出一个能够
15 求解出所述 3D 姿态的第四 3D 特征；基于第四 3D 特征的转换可以得到所述 3D 姿态。

在本实施例中，通过归一化处理，可以消除不同相机参数的相机采集导致的差异，从而消除神经网络等深度学习模型因为不同相机参数导致的 3D 姿态提取的精度低的问题，从而可以进一步提升目标的 3D 姿态提取精
20 确度。

在一些实施例中，所述对所述第四 2D 坐标及所述第四 2D 坐标对应的深度值进行归一化处理，得到归一化后的所述第四 2D 坐标及归一化后的所述第四 2D 坐标对应的深度值，包括：

基于第四 2D 坐标及所述第四 2D 坐标对应的深度值，得到所述关键点的坐标均值及方差；
25

根据所述坐标均值及方差及所述第四 2D 坐标及所述第四 2D 坐标对应的深度值，得到归一化后的第四 2D 坐标。

具体地，所述均值利用 Mean 表示，方差利用 Std 表示；则所述第四 2D 坐标可以用于如下函数关系计算：

30 $X4' = (X4 - \text{Mean}) / \text{Std}_x$;

$Y4' = (Y4 - \text{Mean}) / \text{Std}_y$ 。

X4 为第四 2D 坐标在第一方向上的坐标值；Y4 为第四 2D 坐标在第二方向上的坐标值；X4' 为归一化后的第四 2D 坐标在第一方向上的坐标值；Y4' 为归一化后的第四 2D 坐标在第二方向上的坐标值；Std_x 为第一方向上的坐标值的方差。Std_y 为第二方向上的坐标值的方差。
35

在一些实施例中，所述方法还包括：

基于所述实际深度值，进行将所述 3D 姿态投影到二维平面内的迭代运算，以得到与所述第三 2D 坐标之间距离最小的第五 2D 坐标；

根据所述第五 2D 坐标及所述第一 3D 特征，获得所述目标的转动参数和平移参数。

在本实施例中，将 3D 姿态投影到二维平面内，可包括：将表征 3D 姿态的第一 3D 特征投影到 2D 成像平面内，从而获得 2D 成像平面内的 2D 投影图像。

投影的方式有多种，以下提供两种可选方式：

可选方式一：根据所述 3D 姿势及投影矩阵，得到投影到 2D 成像平面内的 2D 坐标；例如，将 3D 姿势左乘投影矩阵，得到所述投影到 2D 成像平面内的坐标。此处的投影矩阵可以为根据相机参数和/或者投影的经验值确定的。

可选方式二：利用可以将 3D 姿势投影到 2D 成像平面内的投影模型，例如，投影神经网络，以所述 3D 姿势为输入，以投影到 2D 成像平面内的 2D 坐标为输出。

得到投影到 2D 成像平面内的输出 2D 坐标（即所述第五 2D 坐标）就可以计算与第三 2D 坐标之前的距离，选择距离最小的一组，计算所述转动参数及所述平移参数。总之在投影的过程中，相当于将深度值去除，仅保留 2D 成像平面内的 2D 坐标。但是在本实施例中，实质上 3D 姿势是基于参考深度值计算的，如此，可能利用图 2 所示的三角函数关系，将 3D 姿势平移回到实际深度值的位置上。但是考虑到深度学习模块的处理误差及相机的处理误差等，故可以基于实际深度值及其近似值，进行 3D 姿势向 2D 成像平面内的投影。投影的过程中，需要使得投影到二维平面内的 2D 坐标，与实际的第三 2D 坐标距离最小。例如，可用如下函数表示第五 2D 坐标与第三 2D 坐标的最小化值为： $\min\{(X5-X3)^2+(Y5-Y3)^2\}$ ；

$(X5, Y5)$ 为所述第五 2D 坐标； $(X3, Y3)$ 为所述第三 2D 坐标。

接着利用如下函数关系，可以求解出所述转动参数 R 和平移参数 T， $\arg \min_{R,T} K[R|T] \cdot S_3 - S_2$ 。S₃ 表示关键点的第一 3D 特征；S₂ 表示关键点的 2D 坐标。

由于实际深度值给出了迭代计算的深度范围。例如，根据实际深度值加上一个偏移量得到所述深度范围的最大值；实际深度值减去一个偏移量得到了所述深度范围的最小值。在进行所述 3D 姿态向 2D 成像平面内的投影时，可以在该深度范围内选择实际深度值。之所以基于实际深度值选择深度范围，一方面是考虑了深度摄像头采集深度图像的有偏差，另一方面是考虑了网络的误差，基于上述两个方面的考虑通过深度范围来进行容错处理，从而实现 3D 姿态向 2D 成像平面内投影，以获得最优的第五 2D 坐标，从而估计出转动参数和/或平动参数。

所述平动参数可表征了目标平移状况，所述转动参数表征了目标的转动状况。所述平动参数可包括：在各个方向上平动位移量；所述转动参数可包括：在各个方向上的转动位移量。

在本申请实施例中，由于迭代的过程中，是预先知道了实际深度值，可以以所述实际深度值为参考深度值，在包含有所述实际深度值的深度范围内，进行所述 3D 姿态到二维平面内的投影；相对于没有实际深度值提供深度范围的情况下迭代计算，大大的减少了迭代次数，节省了计算量及提升了计算速率。

如图 7 所示，本实施例提供一种数据处理装置，包括：

第一转换模块 110，配置为根据参考深度值及图像中目标的关键点的实际深度值，将所述关键点的第一 2D 坐标的转换为第二 2D 坐标，其中，所述第二 2D 坐标和所述参考深度值，构成了所述关键点的第一 3D 特征；

第一获得模块 120，配置为基于所述第一 3D 特征，获得所述目标的 3D 姿态。

在一些实施例中，所述第一转换模块 110 及第一获得模块 120 可为程序模块，该程序模块被处理器执行之后，能够实现第一 2D 坐标向第二 2D 坐标的转换，及 3D 姿态的获得。

在另一些实施例中，所述第一转换模块 110 及第一获得模块 120 还可为硬件模块及程序模块的组合，例如，复杂可编程阵列或者现场可编程阵列。

在还有一些实施例中，所述第一转换模块 110 及第一获得模块 120 可对应于硬件模块，例如，所述第一转换模块 110 及第一获得模块 120 可为专用集成电路。

在一些实施例中，所述第一转换模块 110，配置为根据所述实际深度值与所述参考深度值的比值，及所述第一 2D 坐标，得到所述第二 2D 坐标。

在一些实施例中，所述第一转换模块 110，配置为利用如下函数关系确定所述第二 2D 坐标；

$$X2 = (X1 * d) / D,$$

$$Y2 = (Y1 * d) / D,$$

其中，X2 为所述第二 2D 坐标在第一方向上的坐标值；X1 为所述第一 2D 坐标在所述第一方向上的坐标值；

Y2 为第二 2D 坐标在第二方向上的坐标值，Y1 为所述第一 2D 坐标在所述第二方向上的坐标值；其中，所述第二方向垂直于所述第一方向；

d 为所述实际深度值；D 为所述参考深度值。

在一些实施例中，所述装置还包括：

第二转换模块，配置为光心位置根据所述关键点的第二 3D 特征及所述图像对应的光心位置，得到所述第一 2D 坐标；其中，所述第二 3D 特征包括：基于 2D 图像得到的第三 2D 坐标及基于深度图像得到的实际深度值。

在一些实施例中，所述第二转换模块，配置为移动所述关键点的第二 3D 特征，使得所述关键点中基准点的 3D 特征平移到所述光心位置，并得到各所述关键点的第三 3D 特征；将所述第三 3D 特征投影到 2D 成像平面，

得到所述第一 2D 坐标光心位置。

在一些实施例中，若所述目标为人体骨骼，所述基准点的第一 3D 特征是基于所述关键点中的两个胯部关键点的第二 3D 特征确定的。

在一些实施例中，所述第一获得模块，配置为对所述关键点的第二 2D 坐标对应的深度值减去所述基准点的深度值，得到第四 2D 坐标及所述第四 2D 坐标对应的深度值；对所述第四 2D 坐标及所述第四 2D 坐标对应的深度值进行归一化处理，得到归一化后的所述第四 2D 坐标及归一化后的所述第四 2D 坐标对应的深度值；利用深度学习模型对归一化后的所述第四 2D 坐标及归一化后的所述第四 2D 坐标对应的深度值进行处理，得到所述目标的 3D 姿态。

在一些实施例中，所述第一获得模块 120，配置为基于第四 2D 坐标及所述第四 2D 坐标对应的深度值，得到所述关键点的坐标均值及方差；根据所述坐标均值及方差及所述第四 2D 坐标及所述第四 2D 坐标对应的深度值，得到归一化后的第四 2D 坐标。

在一些实施例中，所述装置还包括：

迭代模块，配置为基于所述实际深度值，进行将所述 3D 姿态投影到二维平面内的迭代运算，以得到与所述第三 2D 坐标之间距离最小的第五 2D 坐标；

第二获得模块，配置为根据所述第五 2D 坐标及所述第一 3D 特征，获得所述目标的转动参数和平移参数。

如图 8 所示，本申请实施例提供了一种电子设备，包括：

存储器，配置为存储信息；

处理器，与所述存储器连接，配置为通过执行存储在所述存储器上的计算机可执行指令，能够实现前述一个或多个技术方案提供的数据处理方法，例如，如图 1A、图 1B 及图 5 所示的方法中的一个或多个。

该存储器可为各种类型的存储器，可为随机存储器、只读存储器、闪存等。所述存储器可用于信息存储，例如，存储计算机可执行指令等。所述计算机可执行指令可为各种程序指令，例如，目标程序指令和/或源程序指令等。

所述处理器可为各种类型的处理器，例如，中央处理器、微处理器、数字信号处理器、可编程阵列、数字信号处理器、专用集成电路或图像处理器等。

所述处理器可以通过总线与所述存储器连接。所述总线可为集成电路总线等。

在一些实施例中，所述终端设备还可包括：通信接口，该通信接口可包括：网络接口、例如，局域网接口、收发天线等。所述通信接口同样与所述处理器连接，能够用于信息收发。

在一些实施例中，所述终端设备还包括人机交互接口，例如，所述

人机交互接口可包括各种输入输出设备，例如，键盘、触摸屏等。

本申请实施例提供了一种计算机存储介质，所述计算机存储介质存储有计算机可执行代码；所述计算机可执行代码被执行后，能够实现前述一个或多个技术方案提供的数据处理方法，例如，例如，图 1A、图 1B 及图 5 所示的方法中的一个或多个。

所述存储介质包括：移动存储设备、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。所述存储介质可为非瞬间存储介质。

本申请实施例提供一种计算机程序产品，所述程序产品包括计算机可执行指令；所述计算机可执行指令被执行后，能够实现前述任意实施提供的数据处理方法，例如，图 1A、图 1B 及图 5 所示的方法中的一个或多个。

以下结合上述实施例提供几个具体示例：

示例 1：

本示例使用深度神经网络预测人体的二维和三维的关键点，再利用三维视觉算法计算人体的三维姿态；具体可包括：

使用 2D 人体的关键点估计工具预测 14 个人体的关键点在 2D 图像中的 2D 位置；

提取出 2D 图像对应的深度图像中与 14 个人体的关键点的实际深度值；

通过三角函数等方式，将对应于实际深度值的 2D 坐标转换与参考深度值对应的 2D 坐标；

使用相机内参对所有关键点转换后的 2D 坐标进行内参归一化操作；

统计归一化后各关键点的均值与标准差，用于更进一步的坐标归一化操作；得到归一化后的 2D 坐标及参考深度值；

将归一化后的 2D 坐标及参考深度值，输入深度神经网络，由神经网络进行 2D 关键点到 3D 关键点的第一 3D 特征。基于该第一 3D 特征可以得到三维视觉算法等，得到 3D 姿态。例如，基于透视 n 点定位（PnP）优化基于所述第一 3D 特征得到 3D 姿态。

图 9 所示可为得到本示例提供的 3D 姿态的一种神经网络，包括：

全连接层（Fc）、批处理+ReLU 层及 Dropout 层；

其中，全连接层获得 14 个关键点的第一 3D 特征；输出的是 3D 姿态。

该神经网络可以用于提取出所述 3D 姿态。

示例 2：

本示例提供一种数据处理方法，包括：

利用深度神经网络获取输入 2D 图像的若干人体的 2D 关键点（对应

于 2D 坐标)；

将二维人体关键点进行相机内参归一化后输入到第二个深度神经网络得到相对于人体某一关键点（一般为骨盆处）的相对三维关键点；

最后将所得二维关键点和三维关键点的点序对齐，使用 PnP 算法，
5 求出三维人体空间位姿。

示例 3：对于每一帧 3D 图像，使用人体的二维关键点检测工具，得到图像上 14 点的关键点的坐标；

以第一步中得到的二维关键点坐标作为输入，3D 关键点的提取网络，获得得到对于的三维人体骨架（17 个关键点，其中骨盆处关键点位置
10 固定为 0）。

将得到的两个人体关键点模型进行对齐操作，使得每个关键点在物理意义上一致。

已知当前设备的内参 K ，计算目标人体在相机坐标系下的外参 R 和

T 。其中，
$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

15 f_x, f_y, c_x, c_y 可由张正友标定法标定当前设备而来。不妨设对齐后的二维人体骨架 S_2 和三维人体骨架 S_3 ，那么最优化下列公式即可

$$\arg \min_{R, T} K[R|T] \cdot S_3 - S_2$$

20 由于使用了一段连续的视频作为输入，所以前一帧的 R 和 T 可以用来作为后一帧的初始值。在本申请所提供的几个实施例中，应该理解到，所揭露的设备和方法，可以通过其它的方式实现。以上所描述的设备实施例仅仅是示意性的，例如，所述单元的划分，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，如：多个单元或组件可以结合，或可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以是通过一些接口，设备或单元
25 的间接耦合或通信连接，可以是电性的、机械的或其它形式的。

上述作为分离部件说明的单元可以是、或也可以不是物理上分开的，作为单元显示的部件可以是、或也可以不是物理单元，即可以位于一个地方，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全部单元来实现本实施例方案的目的。

30 另外，在本申请各实施例中的各功能单元可以全部集成在一个处理模块中，也可以是各单元分别单独作为一个单元，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，也可以采用硬件加软件功能单元的形式实现。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分
35 步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的

步骤；而前述的存储介质包括：移动存储设备、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

- 5 以上所述，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所述权利要求的保护范围为准。