(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114663917 A
(43)申请公布日 2022.06.24

(21)申请号 202210247791.2

(22)申请日 2022.03.14

(71)申请人 清华大学
地址 100084 北京市海淀区清华园

(72)发明人 季向阳 余杭 连晓聪

(74)专利代理机构 北京清亦华知识产权代理事务所(普通合伙) 11201
专利代理师 尚伟净

(51)Int.Cl.
G06V 40/10(2022.01)
G06V 20/64(2022.01)
G06V 20/52(2022.01)
G06V 10/774(2022.01)
G06K 9/62(2022.01)

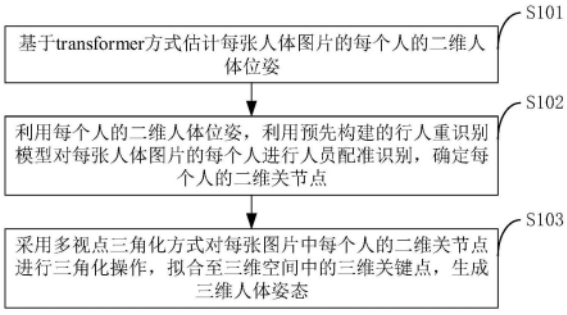
权利要求书2页 说明书9页 附图2页

(54)发明名称

基于多视角的多人三维人体位姿估计方法及装置

(57)摘要

本申请公开了一种基于多视角的多人三维人体位姿估计方法及装置,其中,方法包括:基于transformer方式估计每张人体图片的每个人的二维人体位姿;利用每个人的二维人体位姿,利用预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别,确定每个人的二维关节点;以及采用多视点三角化方式对每张图片中每个人的二维关节点进行三角化操作,拟合至三维空间中的三维关键点,生成三维人体姿态。由此,解决了相关技术中由于占用运算资源较多、运行时间较慢难以广泛应用于运算硬件条件较差的场景,适用性较低的技术问题。



CN 114663917 A

1.一种基于多视角的多人三维人体位姿估计方法,其特征在于,包括以下步骤:

基于transformer方式估计每张人体图片的每个人的二维人体位姿;

利用所述每个人的二维人体位姿,利用预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别,确定所述每个人的二维关节点;以及

采用多视点三角化方式对所述每张图片中每个人的二维关节点进行三角化操作,拟合至三维空间中的三维关键点,生成三维人体姿态。

2.根据权利要求1所述的方法,其特征在于,所述基于transformer方式估计每张人体图片的每个人的二维人体位姿,包括:

利用swin-transformer骨架检测所述每张人体图片的每个人的二维人体位姿;

或者,利用所述swin-transformer骨架估计所述每个人的关节点位置,确定所述二维人体位姿。

3.根据权利要求1所述的方法,其特征在于,在利用所述预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别之前,还包括:

获取用于训练模型的公用数据集;

利用所述公用数据集在基于深度学习构建的行人重识别模型上进行训练,生成所述预先构建的行人重识别模型。

4.根据权利要求1所述的方法,其特征在于,所述基于transformer方式估计每张人体图片的每个人的二维人体位姿,包括:

基于ViT的变形框架获取所述每张图片二维人体姿态估计。

5.一种基于多视角的多人三维人体位姿估计装置,其特征在于,包括:

位姿估计模块,用于基于transformer方式估计每张人体图片的每个人的二维人体位姿;

识别模块,用于利用所述每个人的二维人体位姿,利用预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别,确定所述每个人的二维关节点;以及

生成模块,用于采用多视点三角化方式对所述每张图片中每个人的二维关节点进行三角化操作,拟合至三维空间中的三维关键点,生成三维人体姿态。

6.根据权利要求5所述的装置,其特征在于,所述位姿估计模块进一步用于,利用swin-transformer骨架检测所述每张人体图片的每个人的二维人体位姿;或者,利用所述swin-transformer骨架估计所述每个人的关节点位置,确定所述二维人体位姿。

7.根据权利要求5所述的装置,其特征在于,所述识别模块包括:

获取单元,用于获取用于训练模型的公用数据集;

生成单元,用于利用所述公用数据集在基于深度学习构建的行人重识别模型上进行训练,生成所述预先构建的行人重识别模型。

8.根据权利要求5所述的装置,其特征在于,所述位姿估计模块包括:

姿态估计单元,用于基于ViT的变形框架获取所述每张图片二维人体姿态估计。

9.一种电子设备,其特征在于,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述程序,以实现如权利要求1-4任一项所述的基于多视角的多人三维人体位姿估计方法。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器

执行,以用于实现如权利要求1-4任一项所述的基于多视角的多人三维人体位姿估计方法。

基于多视角的多人三维人体位姿估计方法及装置

技术领域

[0001] 本申请涉及计算机视觉技术领域，特别涉及一种基于多视角的多人三维人体位姿估计方法及装置。

背景技术

[0002] 人体姿态估计任务是当前计算机视觉领域中的一个重要研究分支，也是当下基于应用和产业需求的研究热点。人体姿态估计任务常用的划分方式有三种：根据提供输入视角数量划分，可以分为单视角估计任务和多视角估计任务；根据检测人数划分可分为单人场景任务和多人场景任务；根据目的信息可分为二维估计任务和三维估计任务。根据其基本分类和应用场景，人体姿态估计任务在全息现实、人体仿真、视频监控、无人机群等领域有着广泛的应用，并且还有着巨大的开发潜力。人体姿态估计也是诸多计算机视觉任务的研究基础，其估计精度对下游任务的效果有着重要的影响。所以，研究人体姿态估计问题，有着愈发重要的意义。

[0003] 相关技术遵循单图内二维人体姿态估计、多图内人员配准识别，三维人体姿态拟合三个步骤，采用了一种多路匹配算法，该匹配算法首先通过分析外观信息本身建立交叉匹配矩阵，寻找多个视图中检测到的二维姿态的周期一致性对应，从而匹配多视角图片中不同的人，该匹配算法能够在不知道场景中真实人数的情况下，修剪错误检测并处理视图之间的部分重叠，在跨链接二维视图人员匹配问题上达到了很好的效果。同时，相关技术改良了常规的3DPS(3D pictorial structure，沉浸式投影显示系统)方法，通过在多个视图之间匹配检测到的二维姿态，生成二维姿态簇，每个簇包含不同视图下同一人的二维姿态，解决了身体层次上的对应问题。

[0004] 然而，相关技术中，仍存在不少缺陷：

[0005] 一、二维人体位姿估计方法较差。相关技术设计的二维人体位姿是在MSCOCO二维人体姿态数据集上训练的CPN(Cascaded Pyramid Network，堆叠金字塔网络)，该CPN系统有两个子网络，一个全局网络进行全局的人定位和粗略估计关键点，一个局部精调网络用来估计关节点精细位置。该方法精度较差，并且运行时间慢、占用运算资源较多。

[0006] 二、整体系统运行速率慢。该系统与其他类似系统处理步骤基本一致，都遵循单帧间二维姿态估计、人员匹配、三维拟合这个管道过程。相关技术在每个系统中设计了过多的环节来确保精确性，比如人员匹配环节中设计了相似度估计、行人重识别、循环一致性等多种技术来确保拟合精度，三维拟合环节使用了运算量很大的3DPS方法来确保拟合精度，大量占用运算资源，并且延长了许多运算时间。

[0007] 综上所述，相关技术中由于占用运算资源较多、运行时间较慢难以广泛应用于运算硬件条件较差的场景，适用性较低，亟需改善。

发明内容

[0008] 本申请提供一种基于多视角的多人三维人体位姿估计方法及装置，以解决相关技

术中由于占用运算资源较多、运行时间较慢难以广泛应用于运算硬件条件较差的场景，适用性较低，且人体位姿估计的精度较差等技术问题。

[0009] 本申请第一方面实施例提供一种基于多视角的多人三维人体位姿估计方法，包括以下步骤：基于transformer方式估计每张人体图片的每个人的二维人体位姿；利用所述每个人的二维人体位姿，利用预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别，确定所述每个人的二维关节点；以及采用多视点三角化方式对所述每张图片中每个人的二维关节点进行三角化操作，拟合至三维空间中的三维关键点，生成三维人体姿态。

[0010] 可选地，在本申请的一个实施例中，所述基于transformer方式估计每张人体图片的每个人的二维人体位姿，包括：利用swin-transformer骨架检测所述每张人体图片的每个人的二维人体位姿；或者，利用所述swin-transformer骨架估计所述每个人的关节点位置，确定所述二维人体位姿。

[0011] 可选地，在本申请的一个实施例中，在利用所述预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别之前，还包括：获取用于训练模型的公用数据集；利用所述公用数据集在基于深度学习构建的行人重识别模型上进行训练，生成所述预先构建的行人重识别模型。

[0012] 可选地，在本申请的一个实施例中，所述基于transformer方式估计每张人体图片的每个人的二维人体位姿，包括：基于ViT的变形框架获取所述每张图片二维人体姿态估计。

[0013] 本申请第二方面实施例提供一种基于多视角的多人三维人体位姿估计装置，包括：位姿估计模块，用于基于transformer方式估计每张人体图片的每个人的二维人体位姿；识别模块，用于利用所述每个人的二维人体位姿，利用预先构建的行人重识别模型对所述每张人体图片的每个人进行人员配准识别，确定所述每个人的二维关节点；以及生成模块，用于采用多视点三角化方式对所述每张图片中每个人的二维关节点进行三角化操作，拟合至三维空间中的三维关键点，生成三维人体姿态。

[0014] 可选地，在本申请的一个实施例中，所述位姿估计模块进一步用于，利用swin-transformer骨架检测所述每张人体图片的每个人的二维人体位姿；或者，利用所述swin-transformer骨架估计所述每个人的关节点位置，确定所述二维人体位姿。

[0015] 可选地，在本申请的一个实施例中，所述识别模块包括：获取单元，用于获取用于训练模型的公用数据集；生成单元，用于利用所述公用数据集在基于深度学习构建的行人重识别模型上进行训练，生成所述预先构建的行人重识别模型。

[0016] 可选地，在本申请的一个实施例中，所述位姿估计模块包括：姿态估计单元，用于基于ViT的变形框架获取所述每张图片二维人体姿态估计。

[0017] 本申请第三方面实施例提供一种电子设备，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序，以实现如上述实施例所述的基于多视角的多人三维人体位姿估计方法。

[0018] 本申请第四方面实施例提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行，以用于实现如权利要求1-4任一项所述的基于多视角的多人三维人体位姿估计方法。

[0019] 本申请实施例可以基于transformer方式估计每张人体图片的每个人的二维人体位姿，并利用行人重识别模型进行人员配准识别，提取每个人的二维关节点，通过多视点三角化方式对二维关键点进行三角化操作，进而生成三维人体姿态，无需苛刻的硬件设备运行条件，即可在保证一定精度的前提下，完成多人三维人体位姿估计，适用性更强。由此，解决了相关技术中为实现多人三维人体位姿估计，导致运行时占用运算资源较多、运行时间较慢，难以广泛应用于运算硬件条件较差的场景，适用性较低的技术问题。

[0020] 本申请附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本申请的实践了解到。

附图说明

[0021] 本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

[0022] 图1为根据本申请实施例提供的一种基于多视角的多人三维人体位姿估计方法的流程图；

[0023] 图2为根据本申请一个实施例的基于多视角的多人三维人体位姿估计方法的流程图；

[0024] 图3为根据本申请实施例提供的一种基于多视角的多人三维人体位姿估计装置的结构示意图；

[0025] 图4为根据本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0026] 下面详细描述本申请的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本申请，而不能理解为对本申请的限制。

[0027] 下面参考附图描述本申请实施例的基于多视角的多人三维人体位姿估计方法及装置。针对上述背景技术中心提到的相关技术中为实现多人三维人体位姿估计，导致运行时占用运算资源较多、运行时间较慢，难以广泛应用于运算硬件条件较差的场景，适用性较低的技术问题，本申请提供了一种基于多视角的多人三维人体位姿估计方法，在该方法中，可以基于transformer方式估计每张人体图片的每个人的二维人体位姿，并利用行人重识别模型进行人员配准识别，提取每个人的二维关节点，通过多视点三角化方式对二维关键点进行三角化操作，进而生成三维人体姿态，无需苛刻的硬件设备运行条件，即可在保证一定精度的前提下，完成多人三维人体位姿估计，适用性更强。由此，解决了相关技术中为实现多人三维人体位姿估计，导致运行时占用运算资源较多、运行时间较慢，难以广泛应用于运算硬件条件较差的场景，适用性较低的技术问题。

[0028] 具体而言，图1为本申请实施例所提供的一种基于多视角的多人三维人体位姿估计方法的流程示意图。

[0029] 如图1所示，该基于多视角的多人三维人体位姿估计方法包括以下步骤：

[0030] 在步骤S101中，基于transformer方式估计每张人体图片的每个人的二维人体位姿。

[0031] 可以理解的是，二维人体姿态估计，可以由图像人数区分为单人人体姿态估计和多人人体姿态估计，单人人体姿态估计可以看作是，在给定人数、确定关节点的条件下，对图像中给定的单人关节点进行定位；而多人姿态估计通常无法提前确定人数及每个人在输入(图片或视频)中的位置，且出现人体遮挡、自遮挡的概率较大，相较于单人人体姿态估计任务，多人人体姿态估计任务难度更大。

[0032] 在实际执行过程中，本申请实施例可以使用transformer框架，基于数据流转与变换的方式，对每张人体图片的每个人的二维人体位姿进行估计，进而实现对单人人体姿态或多人人体姿态的估计，并为后续进行三维人体姿态转化奠定基础。

[0033] 可选地，在本申请的一个实施例中，基于transformer方式估计每张人体图片的每个人的二维人体位姿，包括：基于ViT的变形框架获取每张图片二维人体姿态估计。

[0034] 在一些情况下，本申请实施例可以基于ViT的变形框架获取每张图片二维人体姿态估计，为后续进行每个人的二维关节点确定奠定基础，可以在保证较高的精度的前提下，缩减运算消耗和运算时间，有利于本申请实施例可以广泛应用于硬件设备较低的场景。

[0035] 可选地，在本申请的一个实施例中，基于transformer方式估计每张人体图片的每个人的二维人体位姿，包括：利用swin-transformer骨架检测每张人体图片的每个人的二维人体位姿；或者，利用swin-transformer骨架估计每个人的关节点位置，确定二维人体位姿。

[0036] 在另一些情况下，本申请实施例可以利用ViT模型的变体，即改进后的swin-transformer骨架替换从上至下的CPN中的人体检测的环节，或者利用改进后的swin-transformer骨架完全代替整个CPN环节(人体检测+关键点估计环节)。

[0037] 在步骤S102中，利用每个人的二维人体位姿，利用预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别，确定每个人的二维关节点。

[0038] 本领域技术人员可以理解到的是，行人重识别技术，是一种利用计算机视觉技术来检索图像或者视频序列中是否存在特定行人的人工智能技术，可应用于城市监控等场景。本申请实施例预先构建行人重识别模型的流程，可以分为特征学习，度量学习和排序优化三个部分。

[0039] 进一步地，本申请实施例可以基于预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别，实现对每个人的二维关节点的提取，并为后续进行三角化技术提供数据基础，有助于实现基于多视角的多人三维人体的位姿估计。

[0040] 可选地，在本申请的一个实施例中，在利用预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别之前，还包括：获取用于训练模型的公用数据集；利用公用数据集在基于深度学习构建的行人重识别模型上进行训练，生成预先构建的行人重识别模型。

[0041] 具体地，行人重识别模型的构建步骤包括：

[0042] 1、数据采集。通常数据的来源包括：从视频中截取的原始数据，或同一场景下的监控摄像头的数据，进而获取构建行人重识别模型的数据集。

[0043] 2、行人框生成。行人重识别模型的构建可以从视频数据中，通过人工方式、行人检测、跟踪方式等将行人从途中剪切出来，从而生成行人框。

[0044] 3、标注训练数据。其中，标注的数据包括相机标签和行人标签等信息。

[0045] 4、训练模块。

[0046] 5、在测试环境下进行模型测试。

[0047] 在本申请实施例中，可以使用公用数据集，在深度学习构建的行人重识别模型上进行训练，进而可以省略数据采集、行人框生成和标注训练数据三个步骤，实现简化构建行人重识别模型过程的目的，从而放宽对硬件设备的要求，使得本申请实施例可以在保证较高精度的同时，广泛适用于硬件设备条件较低的场景。

[0048] 在步骤S103中，采用多视点三角化方式对每张图片中每个人的二维关节点进行三角化操作，拟合至三维空间中的三维关键点，生成三维人体姿态。

[0049] 作为一种可能实现的方式，本申请实施例可以采用多视点三角化技术，将每张图片中每个人确定的二维关节点通过三角化操作，直接将每张图片中每个人确定的二维关节点拟合至三维空间中的三维关键点，从而生成三维人体姿态。本申请实施例采用三角化技术，计算规模较小，可以有效提高运算速率，可以广泛应用于对实际精度需求不高且硬件设备条件较差的场景。

[0050] 下面结合图2所示，以一个实例对本申请实施例的基于多视角的多人三维人体位姿估计方法进行详细阐述。

[0051] 如图2所示，本申请实施例包括以下步骤：

[0052] 步骤S201：使用transformer框架进行二维人体姿态估计。在实际执行过程中，本申请实施例可以使用transformer框架，基于数据流转与变换的方式，对每张人体图片的每个人的二维人体位姿进行估计，进而实现对单人人体姿态或多人人体姿态的估计，并为后续进行三维人体姿态转化奠定基础。

[0053] 具体地，本申请实施例可以通过以下两种方式进行二维人体姿态估计：

[0054] 1、本申请实施例可以利用ViT模型的变体，即改进后的swin-transformer骨架替换从上至下的CPN中的人体检测的环节；

[0055] 2、利用改进后的swin-transformer骨架完全代替整个CPN环节(人体检测+关键点估计环节)。

[0056] 可以理解的是，ViT模型是transformer技术应用于计算机视觉领域的一大突破口，标志着自然语言处理和计算机视觉两大领域开始出现可以统一的表达模型。ViT模块采取了固定大小分割的图像块作为输入，送入transformer板块，最终得到如图像分割、图像检测灯图像任务的结果，且基本可以取得良好的效果。

[0057] 然而，ViT模型存在一定的缺陷，首先由于视觉实体差别很大，并且图像中有效信息聚合不均匀，导致ViT模型会接受输入过剩的信息；其次图像形式的输入导致ViT模型运行时的计算复杂度是图像尺度的平方，从而导致计算量过于庞大。综上，这两大缺点导致ViT技术并不能很好的适应大多数实际环境，并且对运算设备要求较高。

[0058] 为了解决上述问题，swin-transformer模型采用了滑动窗口操作，从而能够将自适应的信息输入模型中，该滑动窗口操作方案会通过模型中的多头自注意力机制得到不重叠的局部窗口。同时，该模块还允许信息的跨窗口、跨尺度链接，既能够中和不同尺度中间过程的有效信息聚合，提高了感受野，还能得到更高的效率。上述结构能够将给予transformer的基本图像方法的运算复杂度由图像尺度的平方降低至线性度，进而拓展了其实际的应用范围。

[0059] 步骤S202:使用行人重识别技术进行人员配准。进一步地,本申请实施例可以基于预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别,实现对每个人的二维关节点的提取,并为后续进行三角化技术提供数据基础,有助于实现基于多视角的多人三维人体的位姿估计。

[0060] 其中,行人重识别模型的构建步骤包括:

[0061] 1、数据采集。通常数据的来源包括:从视频中截取的原始数据,或同一场景下的监控摄像头的数据,进而获取构建行人重识别模型的数据集。

[0062] 2、行人框生成。行人重识别模型的构建可以从视频数据中,通过人工方式、行人检测、跟踪方式等将行人从途中剪切出来,从而生成行人框。

[0063] 3、标注训练数据。其中,标注的数据包括相机标签和行人标签等信息。

[0064] 4、训练模块。

[0065] 5、在测试环境下进行模型测试。

[0066] 在本申请实施例中,可以使用公用数据集,在深度学习构建的行人重识别模型上进行训练,进而可以省略数据采集、行人框生成和标注训练数据三个步骤,实现简化构建行人重识别模型过程的目的,从而放宽对硬件设备的要求,使得本申请实施例可以在保证较高精度的同时,广泛适用于硬件设备条件较低的场景。

[0067] 步骤S203:使用三角化技术完成三维信息拟合。作为一种可能实现的方式,本申请实施例可以采用多视点三角化技术,将每张图片中每个人确定的二维关节点通过三角化操作,直接将每张图片中每个人确定的二维关节点拟合至三维空间中的三维关键点,从而生成三维人体姿态。本申请实施例采用三角化技术,计算规模较小,可以有效提高运算速率,可以广泛应用于对实际精度需求不高且硬件设备条件较差的场景。

[0068] 根据本申请实施例提出的基于多视角的多人三维人体位姿估计方法,可以基于transformer方式估计每张人体图片的每个人的二维人体位姿,并利用行人重识别模型进行人员配准识别,提取每个人的二维关节点,通过多视点三角化方式对二维关键点进行三角化操作,进而生成三维人体姿态,无需苛刻的硬件设备运行条件,即可在保证一定精度的前提下,完成多人三维人体位姿估计,适用性更强。由此,解决了相关技术中为实现多人三维人体位姿估计,导致运行时占用运算资源较多、运行时间较慢,难以广泛应用于运算硬件条件较差的场景,适用性较低的技术问题。

[0069] 其次参照附图描述根据本申请实施例提出的基于多视角的多人三维人体位姿估计装置。

[0070] 图3是本申请实施例的基于多视角的多人三维人体位姿估计装置的方框示意图。

[0071] 如图3所示,该基于多视角的多人三维人体位姿估计装置10包括:位姿估计模块100、识别模块200和生成模块300。

[0072] 具体地,位姿估计模块100,用于基于transformer方式估计每张人体图片的每个人的二维人体位姿。

[0073] 识别模块200,用于利用每个人的二维人体位姿,利用预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别,确定每个人的二维关节点。

[0074] 生成模块300,用于采用多视点三角化方式对每张图片中每个人的二维关节点进行三角化操作,拟合至三维空间中的三维关键点,生成三维人体姿态。

[0075] 可选地，在本申请的一个实施例中，位姿估计模块进一步用于，利用swin-transformer骨架检测每张人体图片的每个人的二维人体位姿；或者，利用swin-transformer骨架估计每个人的关节点位置，确定二维人体位姿。

[0076] 可选地，在本申请的一个实施例中，识别模块200包括：获取单元和生成单元。

[0077] 其中，获取单元，用于获取用于训练模型的公用数据集。

[0078] 生成单元，用于利用公用数据集在基于深度学习构建的行人重识别模型上进行训练，生成预先构建的行人重识别模型。

[0079] 可选地，在本申请的一个实施例中，位姿估计模块100包括：姿态估计单元。

[0080] 其中，姿态估计单元，用于基于ViT的变形框架获取每张图片二维人体姿态估计。

[0081] 需要说明的是，前述对基于多视角的多人三维人体位姿估计方法实施例的解释说明也适用于该实施例的基于多视角的多人三维人体位姿估计装置，此处不再赘述。

[0082] 根据本申请实施例提出的基于多视角的多人三维人体位姿估计装置，可以基于transformer方式估计每张人体图片的每个人的二维人体位姿，并利用行人重识别模型进行人员配准识别，提取每个人的二维关节点，通过多视点三角化方式对二维关键点进行三角化操作，进而生成三维人体姿态，无需苛刻的硬件设备运行条件，即可在保证一定精度的前提下，完成多人三维人体位姿估计，适用性更强。由此，解决了相关技术中为实现多人三维人体位姿估计，导致运行时占用运算资源较多、运行时间较慢，难以广泛应用于运算硬件条件较差的场景，适用性较低的技术问题。

[0083] 图4为本申请实施例提供的电子设备的结构示意图。该电子设备可以包括：

[0084] 存储器401、处理器402及存储在存储器401上并可在处理器402上运行的计算机程序。

[0085] 处理器402执行程序时实现上述实施例中提供的基于多视角的多人三维人体位姿估计方法。

[0086] 进一步地，电子设备还包括：

[0087] 通信接口403，用于存储器401和处理器402之间的通信。

[0088] 存储器401，用于存放可在处理器402上运行的计算机程序。

[0089] 存储器401可能包含高速RAM存储器，也可能还包括非易失性存储器(non-volatile memory)，例如至少一个磁盘存储器。

[0090] 如果存储器401、处理器402和通信接口403独立实现，则通信接口403、存储器401和处理器402可以通过总线相互连接并完成相互间的通信。总线可以是工业标准体系结构(Industry Standard Architecture，简称为ISA)总线、外部设备互连(Peripheral Component，简称为PCI)总线或扩展工业标准体系结构(Extended Industry Standard Architecture，简称为EISA)总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图4中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

[0091] 可选的，在具体实现上，如果存储器401、处理器402及通信接口403，集成在一块芯片上实现，则存储器401、处理器402及通信接口403可以通过内部接口完成相互间的通信。

[0092] 处理器402可能是一个中央处理器(Central Processing Unit，简称为CPU)，或者是特定集成电路(Application Specific Integrated Circuit，简称为ASIC)，或者是被配置成实施本申请实施例的一个或多个集成电路。

[0093] 本实施例还提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如上的基于多视角的多人三维人体位姿估计方法。

[0094] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或N个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0095] 此外，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本申请的描述中，“N个”的含义是至少两个，例如两个，三个等，除非另有明确具体的限定。

[0096] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更N个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本申请的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本申请的实施例所属技术领域的技术人员所理解。

[0097] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是用于实现逻辑功能的可执行指令的定序列表，可以具体实现在任何计算机可读介质中，以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，"计算机可读介质"可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下：具有一个或N个布线的电连接部(电子装置)，便携式计算机盘盒(磁装置)，随机存取存储器(RAM)，只读存储器(ROM)，可擦除可编辑只读存储器(EPROM或闪速存储器)，光纤装置，以及便携式光盘只读存储器(CDROM)。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

[0098] 应当理解，本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，N个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。如，如果用硬件来实现和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

[0099] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

[0100] 此外，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模块中，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模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。

[0101] 上述提到的存储介质可以是只读存储器，磁盘或光盘等。尽管上面已经示出和描述了本申请的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本申请的限制，本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

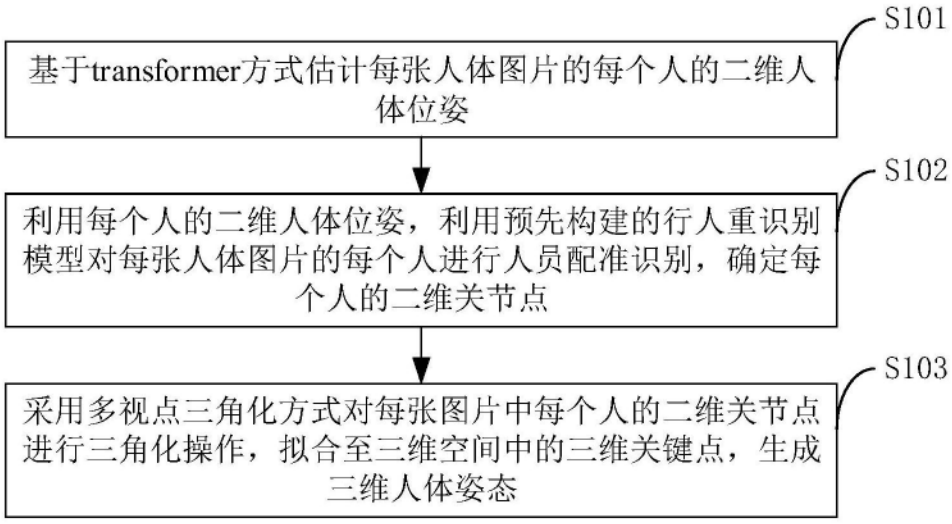

S101
基于transformer方式估计每张人体图片的每个人的二维人体位姿
S102
利用每个人的二维人体位姿，利用预先构建的行人重识别模型对每张人体图片的每个人进行人员配准识别，确定每个人的二维关节点
S103
采用多视点三角化方式对每张图片中每个人的二维关节点进行三角化操作，拟合至三维空间中的三维关键点，生成三维人体姿态

图1

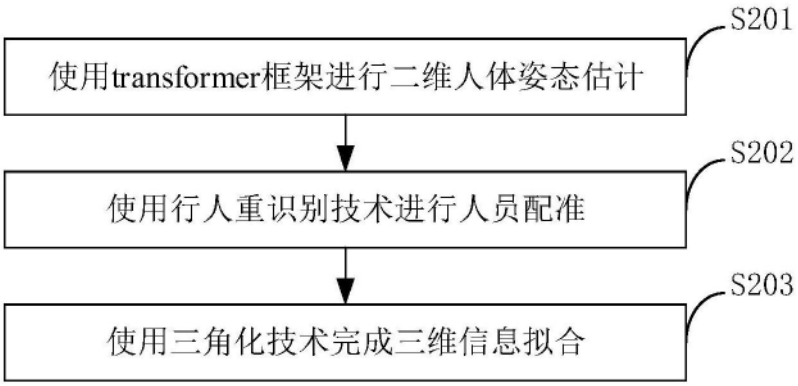

S201
使用transformer框架进行二维人体姿态估计
S202
使用行人重识别技术进行人员配准
S203
使用三角化技术完成三维信息拟合

图2

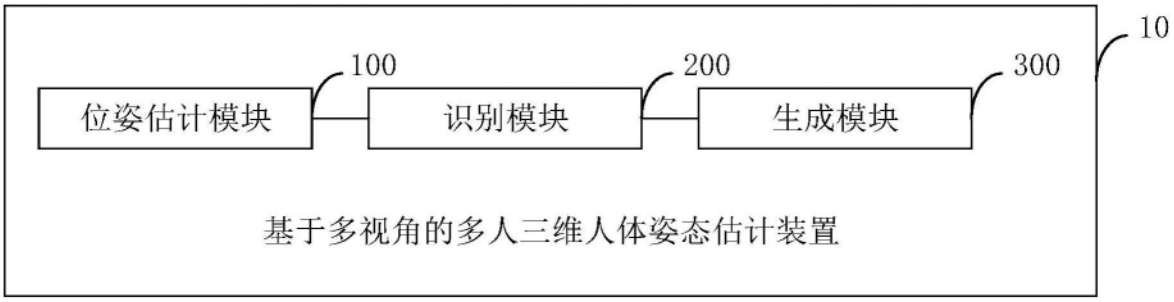

10
100
位姿估计模块
200
识别模块
300
生成模块
基于多视角的多人三维人体姿态估计装置

图3

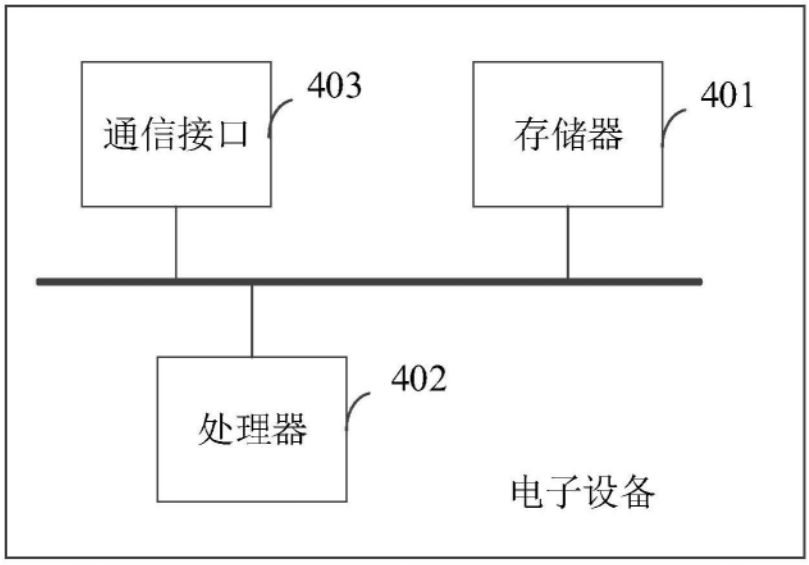


图4