



(12) 发明专利

(10) 授权公告号 CN 113065506 B

(45) 授权公告日 2023.12.26

(21) 申请号 202110411237.9

CN 108876814 A, 2018.11.23

(22) 申请日 2021.04.16

CN 104715493 A, 2015.06.17

(65) 同一申请的已公布的文献号

CN 106910247 A, 2017.06.30

申请公布号 CN 113065506 A

CN 110008915 A, 2019.07.12

(43) 申请公布日 2021.07.02

彭帅. 基于卷积神经网络的人体姿态估计算法综述. 《北京信息科技大学学报(自然科学版)》. 2020, 第35卷(第03期), 全文.

(73) 专利权人 南京大学

Naigong Yu. Human body posture recognition algorithm for still images. 《JOURNAL OF ENGINEERING-JOE》. 2020, 全文.

地址 210046 江苏省南京市栖霞区仙林大道163号

(72) 发明人 江沛远 周余 于耀 都思丹

李昕昕; 龚勋. 三维人脸建模及在跨姿态人脸匹配中的有效性验证. 计算机应用. 2017, (01), 全文.

(74) 专利代理机构 北京高沃律师事务所 11569

专利代理师 杜阳阳

(51) Int. Cl.

G06F 30/27 (2020.01)

(56) 对比文件

CN 111428586 A, 2020.07.17

US 2021097759 A1, 2021.04.01

CN 109949368 A, 2019.06.28

CN 111753747 A, 2020.10.09

吴巍; 郭飞; 郭毓; 郭健. 一种基于全卷积神经网络的横担姿态测量方法. 华中科技大学学报(自然科学版). 2018, (12), 全文.

张广翻; 计忠平. 基于二维点云图的三维人体建模方法. 计算机工程与应用. 2019, (19), 全文.

审查员 张丽萍

权利要求书3页 说明书9页 附图3页

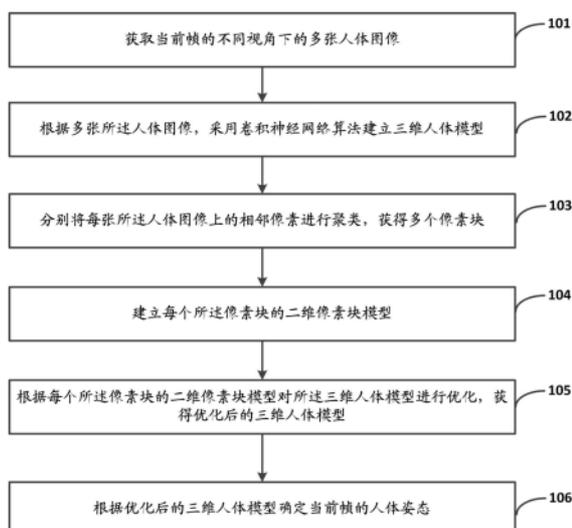
(54) 发明名称

一种人体姿态识别方法及系统

(57) 摘要

本发明公开了一种人体姿态识别方法, 所述识别方法包括如下步骤: 获取当前帧的不同视角下的多张人体图像; 根据多张所述人体图像, 采用卷积神经网络算法建立三维人体模型; 分别将每张所述人体图像上的相邻像素进行聚类, 获得多个像素块; 建立每个所述像素块的二维像素块模型; 根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化, 获得优化后的三维人体模型; 根据优化后的三维人体模型确定当前帧的人体姿态。本发明首先采用卷积神经网络算法, 简单快速的建立一个三维人体模型, 然后利用图像中人体姿态对所述三维人体模型进行优化, 利用优化后的三维人体模型进行姿态识别, 实现了在保证人体姿态估计精度的同时, 提高人

体姿态估计的速度。



CN 113065506 B

1. 一种人体姿态识别方法,其特征在于,所述识别方法包括如下步骤:
 获取当前帧的不同视角下的多张人体图像;
 根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型;
 分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块;
 建立每个所述像素块的二维像素块模型;
 根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型;
 根据优化后的三维人体模型确定当前帧的人体姿态;
 所述三维人体模型为:

$$A(x) = \sum_{j=1}^{63} A_j(x) = \sum_{j=1}^{63} \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right)$$

其中, $A(x)$ 为三维人体模型, $A_j(x)$ 为第 j 个人体部位的三维人体模型, μ_j 表示第 j 个人体部位的坐标, σ_j 表示第 j 个人体部位的半径, x 表示人体的任意一点的位置;
 所述二维像素块模型为:

$$B_i(x) = c_i \cdot \exp\left(-\frac{\|x_i - \mu_i\|^2}{2\delta_i^2}\right)$$

其中, $B_i(x)$ 表示第 i 个像素块的二维像素块模型, c_i 表示第 i 个像素块中心的颜色, μ_i 表示第 i 个像素块中心的坐标, δ_i 表示第 i 个像素块的半边长, x_i 表示人体上任意一点的位置 x 在第 i 个像素块上的投影。

2. 根据权利要求1所述的人体姿态识别方法,其特征在于,所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型,之前还包括:
 根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标,预测当前帧时每个人体部位的位置坐标;
 根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整,获得预调整后的三维人体模型。

3. 根据权利要求1所述的人体姿态识别方法,其特征在于,所述根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型,具体包括:
 采用hourglass网络结构分别从每张所述人体图像中获得 K 个关键点;
 分别令 $k=1, 2, \dots, K$,基于相机的参数,计算出第 k 个关键点在每个视角下的人体图像中的反投影射线,获得每个关键点对应的多条反投影射线;
 确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标,作为所述关键点在三维空间下的坐标,得到每个关键点在三维空间下的坐标;
 根据每个关键点在三维空间下的坐标进行线性插值运算,得到包含人体的每个人体部位的三维坐标的三维人体模型。

4. 根据权利要求1所述的人体姿态识别方法,其特征在于,所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型,具体包括:
 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的

和,作为目标函数值;

判断所述目标函数值是否大于预设阈值,得到判断结果;

若所述判断结果表示否,则采用梯度下降法对三维人体模型中的第j个人体部位的模型进行优化,返回步骤“计算三维人体模型的第j个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值”;

若所述判断结果表示是,则令j的数值增加1,返回步骤“计算三维人体模型的第j个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值”,对三维人体模型的下一个人体部位的模型进行优化,直到优化完成三维人体模型中的每个人体部位的模型。

5. 根据权利要求4所述的人体姿态识别方法,其特征在于,所述计算三维人体模型的第j个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值,具体包括:

采用如下公式,计算三维人体模型的第j个人体部位的模型与每个所述二维像素块模型的相似度:

$$E_{ij} = d(c_i, c_j) \int B_i(x) A_j(x) dx$$

$$= d(c_i, c_j) 2\pi \frac{\sigma_i^2 (\sigma_j^{iD})^2}{\delta_i^2 + (\sigma_j^{iD})^2} \exp\left(-\frac{\|\mu_i - \mu_j^{iD}\|^2}{\delta_i^2 + (\sigma_j^{iD})^2}\right)$$

其中, E_{ij} 表示三维人体模型的第j个人体部位的模型与第i个像素块模型的相似度, $d(c_i, c_j)$ 表示第j个人体部位的模型的颜色 c_j 与第i个像素块的颜色 c_i 的相近程度, $B_i(x)$ 表示第i个像素块的二维像素块模型, $A_j(x)$ 表示三维人体模型的第j个人体部位的模型, x 表示人体的任意一点的位置, μ_i 表示第i个像素块中心的坐标, δ_i 表示第i个像素块的半边长, μ_j^{iD} 表示第j个人体部位的模型的坐标在第i个像素块上的投影坐标, σ_j^{iD} 表示第j个人体部位的半径在第i个像素块上的投影长度;

$$d(c_i, c_j) = \begin{cases} -w & \|c_i - c_j\| > \varepsilon \\ \|c_i - c_j\| & \|c_i - c_j\| < \varepsilon \end{cases};$$

$$\mu_j^{iD} = \begin{pmatrix} \mu_{jx} \\ \mu_{jz} \\ \mu_{jy} \\ \mu_{jz} \end{pmatrix}, \quad \sigma_j^{iD} = \sigma_j * f_{i1} / \mu_{jz};$$

其中, w 为颜色的惩罚值, ε 为颜色阈值, μ_{jx} 、 μ_{jy} 、 μ_{jz} 分别表示第j个人体部位的x轴、y轴和z轴坐标, σ_j 表示第j个人体部位的半径, f_{i1} 表示获得第i个像素块的相机的参数;

利用求和公式,计算三维人体模型的第j个人体部位的模型与每个所述二维像素块模型的相似度的和。

6. 一种人体姿态识别系统,其特征在于,所述识别系统包括:

图像获取模块,用于获取当前帧的不同视角下的多张人体图像;

三维人体模型建立模块,用于根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型;

像素聚类模块,用于分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块;

二维像素块模型建立模块,用于建立每个所述像素块的二维像素块模型;

三维人体模型优化模块,用于根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型;

人体姿态识别模块,用于根据优化后的三维人体模型确定当前帧的人体姿态;

所述三维人体模型为:

$$A(x) = \sum_{j=1}^{63} A_j(x) = \sum_{j=1}^{63} \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right)$$

其中, $A(x)$ 为三维人体模型, $A_j(x)$ 为第 j 个人体部位的三维人体模型, μ_j 表示第 j 个人体部位的坐标, σ_j 表示第 j 个人体部位的半径, x 表示人体的任意一点的位置;

所述二维像素块模型为:

$$B_i(x) = c_i \cdot \exp\left(-\frac{\|x_i - \mu_i\|^2}{2\delta_i^2}\right)$$

其中, $B_i(x)$ 表示第 i 个像素块的二维像素块模型, c_i 表示第 i 个像素块中心的颜色, μ_i 表示第 i 个像素块中心的坐标, δ_i 表示第 i 个像素块的半边长, x_i 表示人体上任意一点的位置 x 在第 i 个像素块上的投影。

7.根据权利要求6所述的人体姿态识别系统,其特征在于,所述识别系统还包括:

位置坐标预测模块,用于根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标,预测当前帧时每个人体部位的位置坐标;

三维人体模型预调整模块,用于根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整,获得预调整后的三维人体模型。

8.根据权利要求6所述的人体姿态识别系统,其特征在于,所述三维人体模型建立模块,具体包括:

关键点获取子模块,用于采用hourglass网络结构分别从每张所述人体图像中获得 K 个关键点;

反投影运算符模块,用于分别令 $k=1,2,\dots,K$,基于相机的参数,计算出第 k 个关键点在每个视角下的人体图像中的反投影射线,获得每个关键点对应的多条反投影射线;

关键点坐标确定子模块,用于确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标,作为所述关键点在三维空间下的坐标,得到每个关键点在三维空间下的坐标;

三维人体模型建立子模块,用于根据每个关键点在三维空间下的坐标进行线性插值运算,得到包含人体的每个人体部位的三维坐标的三维人体模型。

一种人体姿态识别方法及系统

技术领域

[0001] 本发明涉及图像处理技术领域,特别是涉及一种人体姿态识别方法及系统。

背景技术

[0002] 3D人体姿态估计是指从图像、视频或点云中估计人物目标的姿态,是围绕人体3D研究中的一项基本任务。3D人体姿态估计是3D人体重建的重要前提,也可以是人体动作驱动中动作的重要来源。目前,获取人体姿态主要有两种实现方式。一、使用神经网络,通过对某一特定数据集训练,实现对该场景下人体姿态估计的目标。这种方法需要大量的人工标记数据训练神经网络,同时,该方法的精度较低。二、建立一个人体模型,使模型拟合图片上的人体。这种方法依赖于人体模型的建立,人体模型建立过程比较复杂,该方法的速度较慢。因此,如何在保证人体姿态估计精度的同时,提高人体姿态估计的速度成为一个亟待解决的技术问题。

发明内容

[0003] 本发明的目的是提供一种人体姿态识别方法及系统,以实现在保证人体姿态估计精度的同时,提高人体姿态估计的速度。

[0004] 为实现上述目的,本发明提供了如下方案:

[0005] 本发明提供一种人体姿态识别方法,所述识别方法包括如下步骤:

[0006] 获取当前帧的不同视角下的多张人体图像;

[0007] 根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型;

[0008] 分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块;

[0009] 建立每个所述像素块的二维像素块模型;

[0010] 根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型;

[0011] 根据优化后的三维人体模型确定当前帧的人体姿态。

[0012] 可选的,所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型,之前还包括:

[0013] 根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标,预测当前帧时每个人体部位的位置坐标,

[0014] 根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整,获得预调整后的三维人体模型。

[0015] 可选的,所述根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型,具体包括:

[0016] 采用hourglass网络结构分别从每张所述人体图像中获得K个关键点;

[0017] 分别令 $k=1,2,\dots,K$,基于相机的参数,计算出第k个关键点在每个视角下的人体图像中的反投影射线,获得每个关键点对应的多条反投影射线;

[0018] 确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标,作为所述关键点在三维空间下的坐标,得到每个关键点在三维空间下的坐标;

[0019] 根据每个关键点在三维空间下的坐标进行线性插值运算,得到包含人体的每个人体部位的三维坐标的三维人体模型。

[0020] 可选的,所述三维人体模型为:

$$[0021] \quad A(x) = \sum_{j=1}^{63} A_j(x) = \sum_{j=1}^{63} \exp \left(-\frac{\|x - \mu_j\|^2}{2 \sigma_j^2} \right)$$

[0022] 其中, $A(x)$ 为三维人体模型, $A_j(x)$ 为第 j 个人体部位的三维人体模型, μ_j 表示第 j 个人体部位的坐标, σ_j 表示第 j 个人体部位的半径, x 表示人体的任意一点的位置。

[0023] 可选的,所述二维像素块模型为:

$$[0024] \quad B_i(x) = c_i \cdot \exp \left(-\frac{\|x_i - \mu_i\|^2}{2 \delta_i^2} \right)$$

[0025] $B_i(x)$ 表示第 i 个像素块的二维像素块模型, c_i 表示第 i 个像素块中心的颜色, μ_i 表示第 i 个像素块中心的坐标, δ_i 表示第 i 个像素块的半边长, x_i 表示人体上任意一点的位置 x 在第 i 个像素块上的投影。

[0026] 可选的,所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型,具体包括:

[0027] 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值;

[0028] 判断所述目标函数值是否大于预设阈值,得到判断结果;

[0029] 若所述判断结果表示否,则采用梯度下降法对三维人体模型中的第 j 个人体部位的模型进行优化,返回步骤“计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值”;

[0030] 若所述判断结果表示是,则令 j 的数值增加1,返回步骤“计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值”,对三维人体模型的下一个人体部位的模型进行优化,直到优化完成三维人体模型中的每个人体部位的模型。

[0031] 可选的,所述计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和,作为目标函数值,具体包括:

[0032] 采用如下公式,计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度:

$$[0033] \quad E_{ij} = d(c_i, c_j) \int B_i(x) A_j(x) dx$$

$$= d(c_i, c_j) 2 \pi \frac{\sigma_i^2 (\sigma_j^{iD})^2}{\delta_i^2 + (\sigma_j^{iD})^2} \exp \left(-\frac{\|\mu_i - \mu_j^{iD}\|^2}{\delta_i^2 + (\sigma_j^{iD})^2} \right)$$

[0034] 其中, E_{ij} 表示三维人体模型的第 j 个人体部位的模型与第 i 个像素块模型的相似

度, $d(c_i, c_j)$ 表示第 j 个人体部位的模型的颜色 c_j 与第 i 个像素块的颜色 c_i 的相近程度, $B_i(x)$ 表示第 i 个像素块的二维像素块模型, $A_j(x)$ 表示三维人体模型的第 j 个人体部位的模型, x 表示人体的任意一点的位置, μ_i 表示第 i 个像素块中心的坐标, δ_i 表示第 i 个像素块的半边长, μ_j^{iD} 表示第 j 个人体部位的模型的坐标在第 i 个像素块上的投影坐标, σ_j^{iD} 表示第 j 个人体部位的半径在第 i 个像素块上的投影长度;

$$[0035] \quad d(c_i, c_j) = \begin{cases} -w & \|c_i - c_j\| > \varepsilon; \\ \|c_i - c_j\| & \|c_i - c_j\| < \varepsilon; \end{cases}$$

$$[0036] \quad \mu_j^{iD} = \begin{pmatrix} \frac{\mu_{jx}}{\mu_{jz}} \\ \frac{\mu_{jz}}{\mu_{jz}} \\ \frac{\mu_{jy}}{\mu_{jz}} \end{pmatrix}, \quad \sigma_j^{iD} = \sigma_j * f_{i1} / \mu_{jz};$$

[0037] w 为颜色的惩罚值, ε 为颜色阈值, μ_{jx} 、 μ_{jy} 、 μ_{jz} 分别表示第 j 个人体部位的 x 轴、 y 轴和 z 轴坐标, σ_j 表示第 j 个人体部位的半径, f_{i1} 表示获得第 i 个像素块的相机的参数。

[0038] 利用求和公式, 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和。

[0039] 一种人体姿态识别系统, 所述识别系统包括:

[0040] 图像获取模块, 用于获取当前帧的不同视角下的多张人体图像;

[0041] 三维人体模型建立模块, 用于根据多张所述人体图像, 采用卷积神经网络算法建立三维人体模型;

[0042] 像素聚类模块, 用于分别将每张所述人体图像上的相邻像素进行聚类, 获得多个像素块;

[0043] 二维像素块模型建立模块, 用于建立每个所述像素块的二维像素块模型;

[0044] 三维人体模型优化模块, 用于根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化, 获得优化后的三维人体模型;

[0045] 人体姿态识别模块, 用于根据优化后的三维人体模型确定当前帧的人体姿态。

[0046] 可选的, 所述识别系统还包括:

[0047] 位置坐标预测模块, 用于根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标, 预测当前帧时每个人体部位的位置坐标;

[0048] 三维人体模型预调整模块, 用于根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整, 获得预调整后的三维人体模型。

[0049] 可选的, 所述三维人体模型建立模块, 具体包括:

[0050] 关键点获取子模块, 用于采用 hourglass 网络结构分别从每张所述人体图像中获得 K 个关键点;

[0051] 反投影运算子模块, 用于分别令 $k=1, 2, \dots, K$, 基于相机的参数, 计算出第 k 个关键点在每个视角下的人体图像中的反投影射线, 获得每个关键点对应的多条反投影射线;

[0052] 关键点坐标确定子模块, 用于确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标, 作为所述关键点在三维空间下的坐标, 得到每个关键点在三维空

间下的坐标;

[0053] 三维人体模型建立子模块,用于根据每个关键点在三维空间下的坐标进行线性插值运算,得到包含人体的每个人体部位的三维坐标的三维人体模型。

[0054] 根据本发明提供的具体实施例,本发明公开了以下技术效果:

[0055] 本发明公开了一种人体姿态识别方法,所述识别方法包括如下步骤:获取当前帧的不同视角下的多张人体图像;根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型;分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块;建立每个所述像素块的二维像素块模型;根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型;根据优化后的三维人体模型确定当前帧的人体姿态。本发明首先采用卷积神经网络算法,简单快速的建立一个三维人体模型,然后利用图像中人体姿态对所述三维人体模型进行优化,利用优化后的三维人体模型进行姿态识别,实现了在保证人体姿态估计精度的同时,提高人体姿态估计的速度。

[0056] 而且本发明利用人体动作的连续性,更具前几帧的结果对当前帧的结果进行预测,减小优化过程中的迭代次数,进一步的提高了人体姿态估计的速度。

附图说明

[0057] 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性的前提下,还可以根据这些附图获得其他的附图。

[0058] 图1为本发明提供的一种人体姿态识别方法的流程图;

[0059] 图2为本发明提供的一种人体姿态识别方法的原理图;

[0060] 图3为本发明提供的用于获取不同视角下的多张人体图像的相机的排布方式图;

[0061] 图4为本发明提供的三维人体模型图;

[0062] 图5为本发明提供的多个二维像素块模型组成的人体模型图;

[0063] 图6为本发明提供的优化后的三维人体模型图。

具体实施方式

[0064] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0065] 本发明的目的是提供一种人体姿态识别方法及系统,以实现在保证人体姿态估计精度的同时,提高人体姿态估计的速度。

[0066] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0067] 本发明是一种无需人工标注数据或传感器数据的人体姿态估计方法。通过多个彩色相机采集目标人物在固定场景中不同角度的彩色图像,使用已经预训练的卷积神经网络得到粗精度的人体关键点坐标,然后使用反投影技术,得到人体三维坐标;使用该三维人体

坐标生成通用三维人体模型,计算三维人体模型与图像中人体的相似度,同时结合多个视角在三维世界中的一致性作为约束,优化此人体模型,得到最终的人体姿态。该计算过程在CUDA上进行,约束人体模型的参数,可以使得在常数时间内完成计算,约每秒处理25帧图像,可以实时处理视频。本发明只需要彩色图像作为输入,无需人工手动操作,也无需额外的传感器设备,在需要采集人体姿态的领域可以得到广泛的应用。

[0068] 如图1和2所示,本发明提供本发明提供一种人体姿态识别方法,所述识别方法包括如下步骤:

[0069] 步骤101,获取当前帧的不同视角下的多张人体图像。

[0070] 使用多个彩色相机,位于室内不同位置,从多个视角获得人体的连续、同步的彩色图像序列。所有彩色相机受同步盒控制。

[0071] 多个彩色相机的呈圆周分布,获得不同角度的人体信息,一个典型的八相机阵列如图3所示。同步盒发出方波信号,当相机收到信号后,同时拍下,以保证采集到的是同一时刻的人体信息。尽可能的使人体衣服的色彩信息丰富,以区分人体与背景物体,同时更容易识别人体的不同部位。背景信息应该尽可能的简单,得到的人体姿态更具有鲁棒性。

[0072] 步骤102,根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型。

[0073] 根据第一帧神经网络输出的关节点坐标,反投影得到三维人体坐标,进而生成符合图像中人体的粗精度姿态。身体模型应尽可能的描述人体的胖瘦,高矮,特征颜色等信息。在三维高斯函数集合的身体模型中,用每个高斯函数的均值与方差描述。为了避免高自由度导致的过拟合,使用L描述人体躯干的长度,R描述人体躯干的宽度,进而通过计算获得每一个高斯函数的均值和方差,这样可以使得自由度大幅降低,以获得更好的泛化效果。本发明调整这个模型来生成一个行为者特定的身体模型,大致代表了形状和每个高斯函数的颜色。

[0074] 该步骤针对输入不同视角下的8张照片,使用神经网络hourglass从每张照片提取16个人体关键点 (x_i, y_i) 。对同一关键点的8个视角,基于相机参数,计算出反投影的8条射线。使用最小二乘法,得到距离8条射线总距离最短的公共点 (x_i, y_i, z_i) ,此公共点为该关键点在三维空间下的坐标。使用16个关键点的三维坐标 (x_i, y_i, z_i) ,通过线性插值,得到具有63个人体部位的三维坐标的三维人体模型:

$$[0075] \quad A(x) = \sum_{j=1}^{63} A_j(x) = \sum_{j=1}^{63} \exp \left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2} \right)$$

[0076] 其中, $A(x)$ 为三维人体模型, $A_j(x)$ 为第j个人体部位的三维人体模型, μ_j 表示第j个人体部位的坐标, σ_j 表示第j个人体部位的半径, x 表示人体的任意一点的位置。

[0077] 具体步骤为:

[0078] 采用hourglass网络结构分别从每张所述人体图像中获得K个关键点。

[0079] 分别令 $k=1, 2, \dots, K$,基于相机的参数,计算出第k个关键点在每个视角下的人体图像中的反投影射线,获得每个关键点对应的多条反投影射线。

[0080] 确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标,作为所述关键点在三维空间下的坐标,得到每个关键点在三维空间下的坐标。

[0081] 根据每个关键点在三维空间下的坐标进行线性插值运算,得到包含人体的每个人

体部位的三维坐标的三维人体模型,如图4所示。

[0082] 步骤103,分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块。

[0083] 对于每一视角的图片,首先根据上一帧的优化后的三维人体模型,投影得到感兴趣的人体区域。只提取出人体区域,可以去掉无关的背景。对相邻相似的像素聚类,生成颜色块(像素块),结果如图5所示。每一个颜色块使用一个二维高斯函数来近似。聚类的过程中,本发明使用某个特定阈值来确定哪些像素聚集在一起。这种将图片以二维高斯函数集合的图片模型表示,比直接使用图片的像素进行相似度匹配,可以节约大量的计算能力,获得整体实验速度的大幅提升。

[0084] 步骤102,建立每个所述像素块的二维像素块模型。

[0085] 每个像素块的二维像素块模型为:

$$[0086] \quad B_i(x) = c_i \cdot \exp \left(-\frac{\|x_i - \mu_i\|^2}{2 \delta_i^2} \right)$$

[0087] $B_i(x)$ 表示第*i*个像素块的二维像素块模型, c_i 表示第*i*个像素块中心的颜色, μ_i 表示第*i*个像素块中心的坐标, δ_i 表示第*i*个像素块的半边长, x_i 表示人体上任意一点的位置*x*在第*i*个像素块上的投影。

[0088] 整张图像被分割成多个像素块,因此,整张图像可以表示为:

$$[0089] \quad I_m(x) = \sum c_i \cdot B_i(x)$$

[0090] 步骤105,根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型。

[0091] 将人体模型投影到各视角:

$$[0092] \quad \mu_j^{iD} = \begin{pmatrix} \frac{\mu_{jx}}{f_{i1}} \\ \frac{\mu_{jz}}{f_{i1}} \\ \frac{\mu_{jy}}{f_{i1}} \\ \frac{\mu_{jz}}{f_{i1}} \end{pmatrix}, \quad \sigma_j^{iD} = \sigma_j * f_{i1} / \mu_{jz}$$

[0093] 其中, μ_{jx} 、 μ_{jy} 、 μ_{jz} 分别表示第*j*个人体部位的*x*轴、*y*轴和*z*轴坐标, σ_j 表示第*j*个人体部位的半径, f_{i1} 表示获得第*i*个像素块的相机的参数。通过此公式,将三维高斯函数集合的三维人体模型投影到二维,即可与二维高斯函数集合的二维像素块模型计算相似度。计算的是真实投影的近似值,真实值是一个椭球体,但是这个近似值引入的误差可以忽略不计。三维人体模型与图像模型的相似度,可以表示为:

$$[0094] \quad E_{ij} = d(c_i, c_j) \int B_i(x) A_j(x) dx \\ = d(c_i, c_j) 2 \pi \frac{\sigma_i^2 (\sigma_j^{iD})^2}{\delta_i^2 + (\sigma_j^{iD})^2} \exp \left(\frac{\|\mu_i - \mu_j^{iD}\|^2}{\delta_i^2 + (\sigma_j^{iD})^2} \right)$$

[0095] 其中, $d(c_i, c_j)$ 为两个颜色的相近程度:

$$[0096] \quad d(c_i, c_j) = \begin{cases} -w & \|c_i - c_j\| > \varepsilon \\ \|c_i - c_j\| & \|c_i - c_j\| < \varepsilon \end{cases}$$

[0097] w 为对差异过大的颜色的惩罚, ε 为判断颜色是否接近的阈值。一般情况下, $\varepsilon = 0.1$, $w = 0.05$, 可以取得不错的结果。RGB颜色空间与光照强度密切相关, 使用其他颜色空间, 如Lab颜色空间可以得到更好的结果。对相似度 E_{ij} 求导, 沿梯度方向, 使 E_{ij} 变大:

$$[0098] \quad E_{ijk+1} = E_{ijk} + \rho_k s^{(k)}$$

[0099] 其中, s^k 表示梯度方向 ρ_k 表示梯度方向上的搜索步长。迭代一定次数后, 当 E_{ij} 趋近于常数后, 记录此时的人体模型参数作为当前的人体姿态。其中, 步长为动态的。根据多个视频的统计规律, 确定一个初始步长。在优化过程中, 当两次迭代计算的导数符号一致时, 意味着当前姿态仍未到达最优点, 我们将步长扩大到1.1倍。当两次迭代计算的导数符号不一致时, 意味着当前姿态跳过了最优点, 我们将步长减小到0.5 (并不需要修改为-0.5, 因为导数符号本身发生了变化, 优化方向已经变化过了)。考虑到人体动作是连续的, 根据前几帧的结果, 可以预测下一帧的结果。

$$[0100] \quad \text{pose}_{i+1} = t_1 * \text{pose}_i + t_2 * \text{pose}_{i-1} + t_3 * \text{pose}_{i-2}$$

[0101] 其中 pose 为不同帧的姿态结果, t 为预测时使用每一帧结果的权重。使用历史姿态预测当前帧的姿态的结果, 以替换直接使用上一帧结果作为当前帧的初始姿态, 可以更接近当前帧的二维高斯函数集合的图片模型, 以减少优化的迭代次数, 获得整体实验的加速。

[0102] 步骤105所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化, 获得优化后的三维人体模型, 具体包括: 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和, 作为目标函数值; 判断所述目标函数值是否大于预设阈值, 得到判断结果; 若所述判断结果表示否, 则采用梯度下降法对三维人体模型中的第 j 个人体部位的模型进行优化, 返回步骤“计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和, 作为目标函数值”; 若所述判断结果表示是, 则令 j 的数值增加1, 返回步骤“计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和, 作为目标函数值”, 对三维人体模型的下一个人体部位的模型进行优化, 直到优化完成三维人体模型中的每个人体部位的模型。

[0103] 其中, 所述计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和, 作为目标函数值, 具体包括: 采用如下公式, 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度:

$$[0104] \quad E_{ij} = d(c_i, c_j) \int B_i(x) A_j(x) dx = d(c_i, c_j) 2\pi \frac{\sigma_i^2 (\sigma_j^{iD})^2}{\delta_i^2 + (\sigma_j^{iD})^2} \exp \left(-\frac{\|\mu_i - \mu_j^{iD}\|^2}{\delta_i^2 + (\sigma_j^{iD})^2} \right)$$

[0105] 其中, $d(c_i, c_j)$ 表示第 j 个人体部位的模型的颜色 c_j 与第 i 个像素块的颜色 c_i 的相近程度, $B_i(x)$ 表示第 i 个像素块的二维像素块模型, $A_j(x)$ 表示三维人体模型的第 j 个人体部位的模型, x 表示人体的任意一点的位置, μ_i 表示第 i 个像素块中心的坐标, δ_i 表示第 i 个像素块的半边长, μ_j^{iD} 表示第 j 个人体部位的模型的坐标在第 i 个像素块上的投影坐标, σ_j^{iD} 表示第 j 个人体部位的半径在第 i 个像素块上的投影长度;

$$[0106] \quad d(c_i, c_j) = \begin{cases} -w & \|c_i - c_j\| > \varepsilon \\ \|c_i - c_j\| & \|c_i - c_j\| < \varepsilon \end{cases};$$

$$[0107] \quad \mu_j^{iD} = \begin{pmatrix} \mu_{jx} \\ \mu_{jz} \\ \mu_{jy} \\ \mu_{jz} \end{pmatrix}, \quad \sigma_j^{iD} = \sigma_j * f_{i1} / \mu_{jz};$$

[0108] w 为颜色的惩罚值, ε 为颜色阈值, μ_{jx} 、 μ_{jy} 、 μ_{jz} 分别表示第 j 个人体部位的 x 轴、 y 轴和 z 轴坐标, σ_j 表示第 j 个人体部位的半径, f_{i1} 表示获得第 i 个像素块的相机的参数。

[0109] 利用求和公式, 计算三维人体模型的第 j 个人体部位的模型与每个所述二维像素块模型的相似度的和。

[0110] 图5中每一个像素块计算颜色相似度。然后沿梯度方向, 优化三维人体模型的参数, 使相似度更大。当相似度稳定后, 将三维身体模型投影到图片上, 结果如图6所示。当前的三维人体模型的参数, 即为当前人体的姿态。

[0111] 步骤106, 根据优化后的三维人体模型确定当前帧的人体姿态。

[0112] 为了减少优化过程中的迭代次数, 在步骤105所述根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化, 获得优化后的三维人体模型, 之前还包括: 根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标, 预测当前帧时每个人体部位的位置坐标; 根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整, 获得预调整后的三维人体模型。

[0113] 本发明在CPU上进行图片预处理, 得到图片的数学表示。将图片与人体模型参数传入CUDA (compute unified device architecture, 并行计算框架) 并计算。对于每一张图片, 使用相同数量的像素块。每一个CUDA核心计算各像素块与人体一个部位的相似度, 人体模型的参数量也是固定的, 因此每一次计算是一个常数时间。

[0114] 本发明还提供一种人体姿态识别系统, 所述识别系统包括:

[0115] 图像获取模块, 用于获取当前帧的不同视角下的多张人体图像;

[0116] 三维人体模型建立模块, 用于根据多张所述人体图像, 采用卷积神经网络算法建立三维人体模型;

[0117] 像素聚类模块, 用于分别将每张所述人体图像上的相邻像素进行聚类, 获得多个像素块;

[0118] 二维像素块模型建立模块, 用于建立每个所述像素块的二维像素块模型。

[0119] 所述三维人体模型建立模块, 具体包括: 关键点获取子模块, 用于采用hourglass网络结构分别从每张所述人体图像中获得 K 个关键点; 反投影运算符模块, 用于分别令 $k=1, 2, \dots, K$, 基于相机的参数, 计算出第 k 个关键点在每个视角下的人体图像中的反投影射线, 获得每个关键点对应的多条反投影射线; 关键点坐标确定子模块, 用于确定距离每个关键点对应的多条反投影射线的总距离最短的公共点的坐标, 作为所述关键点在三维空间下的坐标, 得到每个关键点在三维空间下的坐标; 三维人体模型建立子模块, 用于根据每个关键点在三维空间下的坐标进行线性插值运算, 得到包含人体的每个人体部位的三维坐标的三维人体模型。

[0120] 三维人体模型优化模块,用于根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型。

[0121] 人体姿态识别模块,用于根据优化后的三维人体模型确定当前帧的人体姿态。

[0122] 所述识别系统还包括:位置坐标预测模块,用于根据当前帧之前的预设帧数内的每帧时每个人体部位的位置坐标,预测当前帧时每个人体部位的位置坐标;三维人体模型预调整模块,用于根据当前帧时每个人体部位的位置坐标对所述三维人体模型进行预调整,获得预调整后的三维人体模型。

[0123] 根据本发明提供的具体实施例,本发明公开了以下技术效果:

[0124] 本发明公开了一种人体姿态识别方法及系统,所述识别方法包括如下步骤:获取当前帧的不同视角下的多张人体图像;根据多张所述人体图像,采用卷积神经网络算法建立三维人体模型;分别将每张所述人体图像上的相邻像素进行聚类,获得多个像素块;建立每个所述像素块的二维像素块模型;根据每个所述像素块的二维像素块模型对所述三维人体模型进行优化,获得优化后的三维人体模型;根据优化后的三维人体模型确定当前帧的人体姿态。本发明首先采用卷积神经网络算法,简单快速的建立一个三维人体模型,然后利用图像中人体姿态对所述三维人体模型进行优化,利用优化后的三维人体模型进行姿态识别,实现了在保证人体姿态估计精度的同时,提高人体姿态估计的速度。

[0125] 而且本发明利用人体动作的连续性,更具前几帧的结果对当前帧的结果进行预测,减小优化过程中的迭代次数,进一步的提高了人体姿态估计的速度。

[0126] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。

[0127] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。

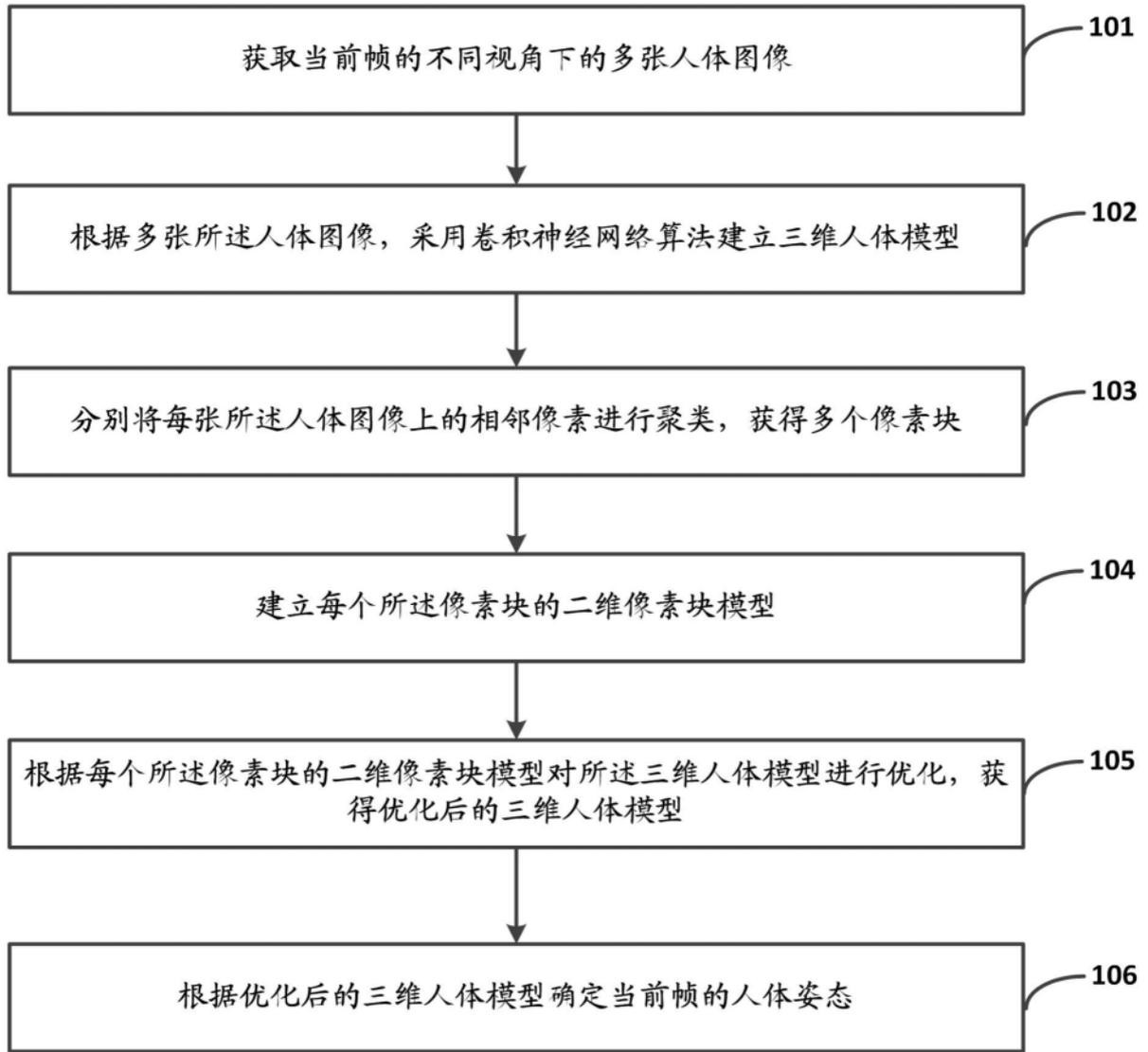


图1

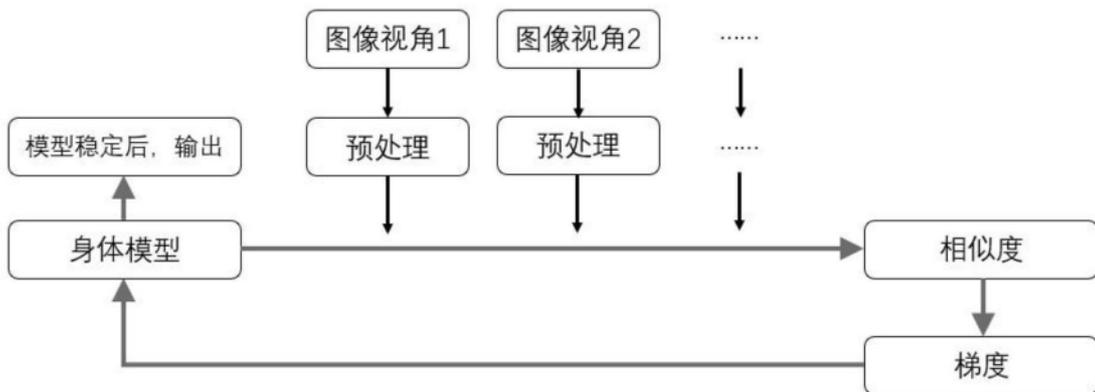


图2

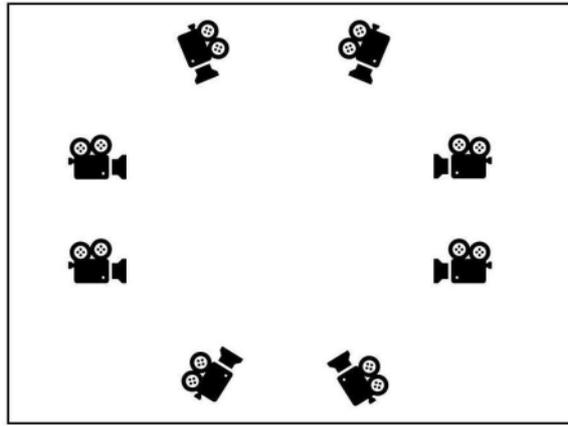


图3

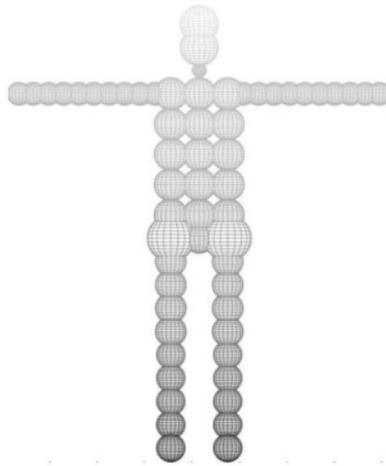


图4

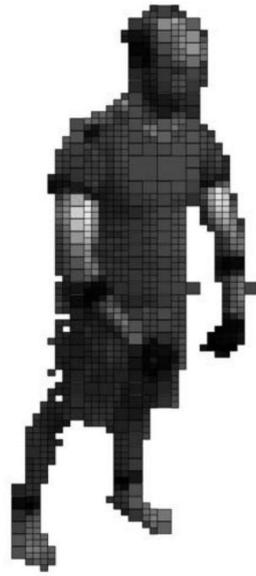


图5



图6