(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115951784 A
(43)申请公布日 2023.04.11

(21)申请号 202310215413.0

(22)申请日 2023.03.08

(71)申请人 南京理工大学
地址 210094 江苏省南京市玄武区孝陵卫街200号

(72)发明人 王康侃 丛素旭 李绍园

(74)专利代理机构 青岛锦佳专利代理事务所(普通合伙) 37283
专利代理师 朱玉建

(51)Int.Cl.
G06F 3/01(2006.01)
G06T 17/20(2006.01)
G06T 15/00(2011.01)
G06T 19/20(2011.01)
G06N 3/08(2023.01)

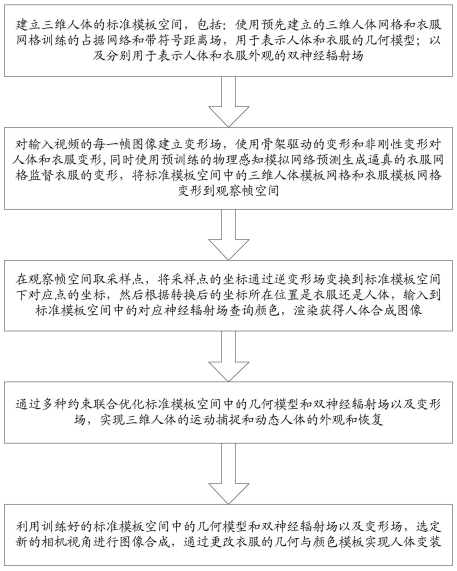
权利要求书5页 说明书8页 附图4页

(54)发明名称

一种基于双神经辐射场的穿衣人体运动捕捉和生成方法

(57)摘要

本发明公开了一种基于双神经辐射场的穿衣人体运动捕捉和生成方法,该方法包括建立三维人体的标准模板空间,包括表示人体和衣服的几何模型及表示人体和衣服外观的双神经辐射场;对输入视频的每一帧图像建立变形场,将标准模板空间中的三维人体模板网格变形到观察帧空间;在观察帧空间取采样点,通过逆变形场变换到标准模板空间的对应点的坐标,根据采样点所在方位属于人体还是衣服,输入不同神经辐射场查询颜色,渲染获得人体合成图像;通过多种约束联合优化标准模板空间中的几何模型和双神经辐射场以及变形场,实现三维人体的运动捕捉和动态人体的外观恢复。本发明不仅能实现三维人体的运动捕捉和外观恢复,还能实现新视角图像合成和衣服编辑。



CN 115951784 A

1.一种基于双神经辐射场的穿衣人体运动捕捉和生成方法,其特征在于,

包括如下步骤:

步骤1. 建立三维人体的标准模板空间,包括使用预先建立的三维人体模板网格和衣服模板网格训练的占据网络和带符号距离场,用于表示人体和衣服的几何模型;以及分别用于表示人体和衣服外观的双神经辐射场;

步骤2. 对输入视频的每一帧图像建立变形场,使用骨架驱动的变形以及非刚性变形对人体和衣服变形,同时使用预训练的物理感知模拟网络预测生成逼真的衣服模板网格监督衣服的变形,将标准模板空间中的三维人体模板网格和衣服模板网格变形到观察帧空间;

步骤3. 在观察帧空间取采样点,将采样点的坐标通过逆变形场变换到标准模板空间下对应点的坐标,然后根据转换后的采样点坐标所在位置是衣服还是人体,将该转换后的采样点坐标输入到标准模板空间中的对应神经辐射场查询颜色,渲染获得人体合成图像;

步骤4. 通过多种约束联合优化标准模板空间中的几何模型和双神经辐射场以及变形场,实现三维人体的运动捕捉以及动态人体的外观和恢复;

步骤5. 利用训练好的标准模板空间中的几何模型和双神经辐射场以及变形场,选定新的相机视角进行图像合成,通过更改衣服的几何与颜色模板实现人体变装。

2.根据权利要求1所述的穿衣人体运动捕捉和生成方法,其特征在于,

所述步骤1中,标准模板空间包括几何模型和颜色模型,且人体和衣服的表示相独立;人体和衣服的几何模型使用占据网络和带符号距离场表示;

标准模板空间的几何模型在时序变化的人体图像合成时保持不变;

人体和衣服的颜色模型使用双神经辐射场表示;在定义标准模板空间的颜色模型时,为人体和衣服分别定义一组隐式外观编码,对应了输入视频的每一帧图像;

在模板空间神经辐射场的颜色模型中融入隐式外观编码,表达并恢复时序变化的外观。

3.根据权利要求2所述的穿衣人体运动捕捉和生成方法,其特征在于,

所述步骤1具体为:

标准模板空间的几何模型使用占据网络和带符号距离场表示,标准模板空间的占据网络模型由函数Fo定义:$o(x)= Fo(\gamma_x(x))$;

其中,$o(x)\in\{0, 1\}$表示占据网络模型输出的在x坐标处的空间是否被占据;

$\gamma_x(x)$表示空间坐标的位置编码,其定义如下:

$\gamma_x(x)=[\sin(x),\cos(x),\sin(2x),\cos(2x),\cdots, \sin(2^{m-1}x),\cos(2^{m-1}x)]^T$,m为自然数;

带符号距离场将三维空间坐标映射到带符号的距离s,即$S:p\in R^3\rightarrow s\in R$;

其中,S表示带符号距离场模型,p表示三维点坐标;s为带符号的距离,表示三维点与最近物体表面的距离,符号代表所处位置是物体的内外,在内即为负,在外即为正;

标准模板空间的颜色模型使用神经辐射场表示,对输入视频的每一帧图像定义一个隐式外观编码,则颜色模型由函数F_c定义:$c_i(x)= F_c(\gamma_x(x), r_d(d),\psi_i)$;

其中,$c_i(x)$表示颜色模型输出的在x坐标处的颜色,d表示观察x坐标的视角方向,即x坐标所在射线的方向;ψ_i表示每一帧的隐式外观编码;

使用两个占据网络,两个颜色网络分别表示人体和衣服的几何和颜色;

具体为:利用一个占据网络 F_o^b、一个颜色网络(F_c^b,φ^b)表示人体的几何和颜色,使用另一个占据网络F_o^g、以及另一个颜色网络(F_c^g,φ^g)表示衣服的几何和颜色;

其中,F_c^b、φ^b分别为表示人体颜色的颜色模型以及隐式外观编码;

F_c^g、φ^g分别表示表示衣服颜色的颜色模型以及隐式外观编码。

4.根据权利要求3所述的穿衣人体运动捕捉和生成方法,其特征在于,

所述步骤2具体为:

步骤2.1. 非刚性变形;

首先通过嵌入变形对非刚性变形建模,该嵌入变形基于变形图计算一个弯曲场;一个嵌入变形图G包含K个节点,并且在三维人体模板网格中能够被自动建立;

节点变换由欧拉角$A\in R^{k\times 3}$和平移向量$T\in R^{k\times 3}$参数化;

对于三维人体模板网格的每个顶点v,经过非刚性变形后的新坐标y由下述公式得到:

$y=\sum_{k\in N(v)} w(v,g_k)[R(A_k)(v-g_k)+g_k+T_k]$;

其中,N(v)表示影响到顶点v的邻域顶点集合,$k\in N(v)$;

g_k表示第k个顶点的坐标;

A_k和T_k分别表示第k个邻域顶点变形所需的欧拉角和平移向量;

$R(\bullet):R^3\rightarrow SO(3)$ 将欧拉角转换为旋转矩阵;

$w(v,g_k)$是顶点v的第k个邻域顶点的变形权重,$w(v,g_k)$的值由如下公式计算:

$w(v,g_k)=(1-||v-g_k||/d_{max})^2$;

其中,d_{max}表示顶点v到k个最近顶点的距离;

非刚性形变中的欧拉角A和平移向量T都使用一个多层感知机模型训练获得,同时多层感知机还反向传播优化一个隐式变形编码w;

对于第i帧非刚性变形的欧拉角A_i,平移向量T_i和隐式变形编码w_i,由函数$F_{A,T}$定义:

$F_{A,T}:w_i\rightarrow(A_i,\ T_i)$;

步骤2.2. 估计骨架驱动的变形;

首先对输入视频的每一帧图像估计一个SMPL模型,并从估计的SMPL模型中计算蒙皮权重$w(y)_j$,$w(y)_j$表示顶点y的第j个部分的蒙皮权重;

人体的参数SMPL模型使用85维向量表示$\Theta=(\theta,\beta)$;

其中,$\beta\in R^{10}$,$\theta\in R^{75}$分别表示人体的形状参数和各个关节的相对角度;

基于非刚性变形获得的三维人体模板网格,进一步应用线性蒙皮变形来进行变形,对于三维人体模板网格的每个顶点y,其变形后的观察帧空间的顶点$\hat{v}$的计算公式如下:

$\hat{v}=[\sum_{j=1}^{J} w(y)_j\ G_j]y$;

其中,J是人体关节的数量;

$w(v)_j$表示顶点v的第j个部分的蒙皮权重,$G_j\in SE(3)$ 表示刚性变换矩阵;

步骤2.3. 使用预训练的物理感知模拟网络监督衣服的变形;

使用物理感知模拟网络学习衣服跟随人体动作所产生的变形,物理感知模拟网络由一个多层感知机模型D_ϕ定义;首先在Marvelous Designer软件中模拟各种衣服的变形;

对于每种衣服类别,使用25种衣服风格以及8种材质,将Marvelous Designer软件模拟

出的衣服形状作为多层感知机模型D_ϕ的监督，则模拟的衣服模板网格由如下公式定义；

$G_S=D_\phi(\gamma,\beta,\tau,\theta)$；

其中，G_S为模拟的衣服模板网格，$\gamma\in R^4$表示衣服风格，τ表示衣服的材质。

5.根据权利要求4所述的穿衣人体运动捕捉和生成方法，其特征在于，

所述步骤3具体为：

为了获得在输入视频第i帧图像的动态神经辐射场，即人体在不同时刻的颜色和几何，首先使用$x^{can}=T_i(x)$将观察帧空间的采样点x转换到标准模板空间的点x^{can}；

其中，T_i是三维人体模板网格变形场的逆变换；

根据相机位置和拍摄视角，由相机向观察帧空间发射多条射线，每条射线r对应最终人体合成图像上的一个像素，然后在每条射线上采样；

一条射线上的采样点表示为：r(t) = o + td；

其中，$o\in R^3$为射线起点，$d\in R^3$为射线方向，t为采样间隔；

将观察帧空间中的采样点坐标经过逆变形场变换到标准模板空间对应的坐标，然后将坐标输入标准模板空间中分别保存衣服和人体的模型查询几何与颜色；

对于来自像素p的射线r，找到这条射线r与衣服或人体网格相交的三角面；

如果相交的三角面来自三维人体模板网格，则将这条射线r上的所有采样点的掩码m_r^b设为1；否则，将这条射线r上的所有采样点的掩码m_r^b设为0；

同理，如果相交的三角面来自衣服模板网格，则将这条射线r上的所有采样点的掩码m_r^g设为1，否则，将这条射线r上的所有采样点的掩码m_r^g设为0；

分别使用以下公式来渲染完整的人体和衣服：

$$T_r^i=\prod_{j=1}^{i-1}\left(1-o_g^j\left(1-m_r^b\right)\right)\left(1-o_b^j m_r^b\right);$$

$$\hat{C}_r=\sum_{i=1}^{n}T_r^i\left(o_g^i\left(1-m_r^b\right)c_g^i+o_b^i m_r^b c_b^i\right);$$

其中，n为光线r上的采样点个数，T_r^i表示光线上各采样点的颜色权重；$\hat{C}_r$表示光线的颜色；c_g^i、c_b^i表示光线上第i个采样点的颜色；o_g^j、o_b^j分别表示衣服和人体神经辐射场输出的光线r上第j个采样点的密度；o_g^i、o_b^i分别表示衣服和人体神经辐射场输出的光线r上第i个采样点的密度。

6.根据权利要求5所述的穿衣人体运动捕捉和生成方法，其特征在于，

所述观察帧空间中的采样点逆变形过程为：

对于观察帧空间的一个采样点x，首先搜索与该采样点x距离最近的SMPL模型的顶点v，然后使用逆线性蒙皮变换采样点x的坐标，具体公式如下：

$\hat{x}=[\sum_{j=1}^{J}w(v)_j\ G_j]^{-1}x$；

其中，$\hat{x}$表示变形后的采样点坐标，$w(v)_j$表示顶点v的第j个部分的蒙皮权重，$G_j\in SE(3)$表示刚性变换矩阵；使用逆变形图将$\hat{x}$变换到标准模板空间下的坐标x^{can}，公式如下：

$x^{can}=[\sum_{k\in N(v)}w(v,g_k)A_k]^{-1}\cdot[\sum_{k\in N(v)}w(v,g_k)(\hat{x}-g_k-T_k+A_kg_k)]$。

7.根据权利要求6所述的穿衣人体运动捕捉和生成方法，其特征在于，

将观察帧空间下的采样点坐标通过逆变形转换到标准模板空间下的采样点坐标后，将变换后的采样点坐标输入标准模板空间神经辐射场查询该点的颜色和密度；

使用多种约束联合训练标准模板空间中的几何模型和神经辐射场以及变形场；具体为：

用于监督衣服和人体颜色的损失函数L_{rgb}：

$$L_{rgb}=\frac{1}{N_t}\sum_{r\in R}\left\|\hat{C}(r)-C(r)\right\|_2+\frac{1}{N_b}\sum_{r\in R}\|(\hat{C}(r)^b-C(r)^b)(1-m_r^g)\|_2+\frac{1}{N_g}\sum_{r\in R}\|(\hat{C}(r)^g-C(r)^g)(1-m_r^b)\|_2;$$

其中，R为投射的光线集合，N_t为一次训练中选取的像素的数量，N_b为像素中属于人体的像素数量，N_g为像素中属于衣服的像素数量；$\hat{C}(r)$为神经辐射场预测的颜色，$C(r)$为颜色真值，公式中的上角标b和g分别表示颜色属于人体还是衣服；

使用物理感知模拟网络约束变形的损失，即用于监督衣服变形的损失函数L_{sim}为：

$$L_{sim}=\frac{1}{|G|}\sum_{t\in G}\rho\left(\left\|\tilde{G}^t-G_s^t\right\|\right);$$

其中，t为网格G的顶点，|G|表示网格G的顶点数；

ρ表示Geman-McClure鲁棒性损失函数，$\tilde{G}^t$表示经过骨架驱动的变形和非刚性变形之后的网格顶点，G_s^t表示物理感知模拟网络输出的网格顶点坐标；

用于加强变形表面的局部光滑性的尽可能刚性的损失函数L_{arap}为：

$$L_{arap}=\sum_{g_i}\sum_{g_j\in N(g_i)}\omega(g_i,g_j)\left\|d_{i,j}(A,T)\right\|_2;$$

其中，g_i表示三维人体模板网格的顶点，g_j为g_i的邻域网格顶点，$N(g_i)$为g_i的邻域顶点集合，$g_j\in N(g_i)$，$w(g_i,g_j)$表示变形权重；

$d_{i,j}(A,T)=A_j(g_i-g_j)+g_j+T_j-(g_i+T_i)$；

其中，A_j、T_j分别表示第j个邻域顶点变形的欧拉角和平移向量；T_i表示顶点i变形的平移向量；

用于让变形后的网格重投影贴合真值mask的mask损失$L_{IoU}(T)$为：

$$L_{IoU}(T)=1-\frac{\left\|R(T)\otimes\bar{R}\right\|_1}{\left\|R(T)\oplus\bar{R}\right\|_1-\left\|R(T)\otimes\bar{R}\right\|_1};$$

其中，⊗和⊕表示矩阵按元素求积和求和；$\bar{R}$表示输入的人体mask真值；T=(M,G)表示三维人体模板网格与衣服模板网格的合集，R(•)表示对网格做投影；

在变形后的网格和模拟的网格上都使用mask损失L_{Iou}为：

$$L_{IoU}=L_{IoU}(\tilde{M},\tilde{G})+L_{IoU}(\tilde{M},G_s);$$

其中，$\tilde{M}$和$\tilde{G}$表示人体和衣服变形后的网格；

用于贴合衣服与三维人体模板网格的损失L_{attach}为：

$$L_{attach}=\frac{1}{|A|}\sum_{t\in A}\rho\left(\left\|\left(\tilde{G}^{t}-\tilde{M}^{t}\right)\right\|_{2}\right);$$

其中，A为衣服模板网格上与三维人体模板网格相接触的顶点集合，|A|为顶点集合A中的顶点个数，$\tilde{G}^{t}$和$\tilde{M}^{t}$分别表示变形后的衣服和三维人体模板网格顶点；

用于防止衣服模板网格和三维人体模板网格相交叉的损失$L_{interp}(M_1,M_2)$为：

$$L_{interp}(M_1,M_2)=\frac{1}{N_c}\sum_{i,j\in C}ReLU\left(\left(M_1^{i}-M_1^{j}\right)\cdot N_1^{i}\right);$$

其中，C表示网格M_1、M_2易发生交叉部位的顶点集合，N_1为网格M_1的法向量，M_1^{i}表示网格M_1的第i个顶点，M_1^{j}表示网格M_1的第j个顶点，N_1^{i}表示网格M_1第i个顶点的法向量，N_c表示C中顶点的个数；为了保证标准模板空间与观察帧空间的人体与衣服的合理性，同时约束两个空间的三维人体与衣服模板网格，给出如下损失L_{interp}；

$$L_{interp}=L_{interp}\left(\hat{M},\hat{G}\right)+L_{interp}\left(\tilde{M},\tilde{G}\right);$$

其中，$\hat{M}$和$\hat{G}$分别表示人体和衣服变形前的网格；

综上，总体损失函数L为：$L=\lambda_1L_{rgb}+\lambda_2L_{arap}+\lambda_3L_{sim}+\lambda_4L_{IoU}+\lambda_5L_{attch}+\lambda_6L_{interp}$；

其中，$\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6\}$为正则化权重。

一种基于双神经辐射场的穿衣人体运动捕捉和生成方法

技术领域

[0001] 本发明属于三维重建领域,涉及一种基于双神经辐射场的穿衣人体运动捕捉和生成方法。

背景技术

[0002] 穿衣服人体的动作捕捉(Clothed human performance capture and synthesis)在计算机视觉和图形学领域是一个重要的问题,不仅需要捕捉内部人体的动作,也要恢复外部衣服运动,这一工作可以用于很多有前景的应用,如虚拟试穿、视频编辑以及远程呈现等。基于人体的多视角或单目视频,三维人体运动捕捉和生成的目标是重建几何和外观时空一致的动态人体三维模型序列,并从新的视角渲染逼真的人体运动视频。人体存在随机的运动并且伴随着衣服的非刚性运动,而且光照变化、自阴影等因素会导致时序外观的变化。因此,三维人体运动捕捉和生成是一个具有很大挑战性的问题。

[0003] 先前的系统使用深度传感器或者将个性化的人体适应到观察帧图像来重建穿衣服的人体,只能恢复一体式的几何,其人体和衣服是一个整体,这些系统不能单独追踪衣服和编辑三维人体的衣服,而这是很多VR/AR应用如虚拟试穿的先决条件。相反的,因为这些方法需要从深度扫描中提取衣服和追踪,如果三维信息缺失,此应用将受限。现有的从彩色图像衣服估计方法需要人对着相机并且保持静态姿势;当人体处于运动中并且衣服在变形时,这些方法将不能真实地恢复三维衣服。最近的方法尝试从视频中模拟追踪人体和衣服的运动,但是此类方法需要为每一个表演者重建衣服模板,或者运行效率极低,因为需要在线模拟衣服或者需要算力耗费巨大的优化,这些使这些方法不能发展成广泛使用的日常应用。

[0004] 神经辐射场(Neural radiance fields,简称NeRF)是一种对三维静态场景连续、隐式的表达方式,其灵活地表示了三维场景的几何和外观,实现了逼真的新视角二维图像合成。近两年NeRF被成功推广到动态场景的图像合成,通过定义一个变形场,变形场通常表示为刚体变形场或位移向量场,将观察帧空间的三维点变换到标准空间,联合优化标准空间NeRF和变形场,实现动态场景NeRF。在不使用运动先验的情况下,同时优化标准空间下的NeRF和变形场是一个欠约束问题,这些方法不适用于运动人体。最近,NerfCap和HumanNeRF分别采用SMPL模型和基于骨架驱动的变形表达动态人体,有效约束了人体变形场的学习,生成了高质量的新视角动态人体视频,但是他们使用一个单独的NeRF表示人体,而没有对衣服建模,因此衣服的运动不能被提取,这限制了其在虚拟显示、增强现实等下游任务上的应用。

发明内容

[0005] 本发明的目的在于提出一种基于双神经辐射场的穿衣人体运动捕捉和生成方法,该方法通过对运动中的人体和衣服分别重建,不仅能实现三维人体的运动捕捉和外观恢复,还能实现新视角图像合成以及衣服编辑。

[0006] 本发明为了实现上述目的，采用如下技术方案：

一种基于双神经辐射场的穿衣人体运动捕捉和生成方法，包括如下步骤：

步骤1．建立三维人体的标准模板空间，包括使用预先建立的三维人体模板网格和衣服模板网格训练的占据网络和带符号距离场，用于表示人体和衣服的几何模型；以及分别用于表示人体和衣服外观的双神经辐射场；

步骤2．对输入视频的每一帧图像建立变形场，使用骨架驱动的变形以及非刚性变形对人体和衣服变形，同时使用预训练的物理感知模拟网络预测生成逼真的衣服模板网格监督衣服的变形，将标准模板空间中的三维人体模板网格和衣服模板网格变形到观察帧空间；

步骤3．在观察帧空间取采样点，将采样点的坐标通过逆变形场变换到标准模板空间下对应点的坐标，然后根据转换后的采样点坐标所在位置是衣服还是人体，将该转换后的采样点坐标输入到标准模板空间中的对应神经辐射场查询颜色，渲染获得人体合成图像；

步骤4．通过多种约束联合优化标准模板空间中的几何模型和双神经辐射场以及变形场，实现三维人体的运动捕捉以及动态人体的外观和恢复；

步骤5．利用训练好的标准模板空间中的几何模型和双神经辐射场以及变形场，选定新的相机视角进行图像合成，通过更改衣服的几何与颜色模板实现人体变装。

[0007] 本发明具有如下优点：

如上所述，本发明述及了一种双神经辐射场的穿衣人体运动捕捉和生成方法，该方法通过对运动中的人体和衣服分别重建，不仅能实现三维人体的运动捕捉和外观恢复，还能实现新视角图像合成以及衣服编辑。本发明方法重建的几何精度高、能合成任意视角的逼真图像，且人体与衣服相分离，能够实现对人体的衣服编辑操作，应用场景广泛。

附图说明

[0008] 图1为本发明实施例中基于双神经辐射场的穿衣人体运动捕捉和生成方法的流程框图。

[0009] 图2为本发明实施例中基于双神经辐射场的穿衣人体运动捕捉和生成方法的流程示意图。

[0010] 图3为本发明使用单目相机视频恢复的人体几何模型和新视角生成示意图。

[0011] 图4为本发明提出的方法的结果和其他方法的比较示意图。

[0012] 图5为本发明衣服与人体分别渲染的结果与其他方法的比较示意图。

[0013] 图6为本发明人体换衣结果的示意图。

具体实施方式

[0014] 下面结合附图以及具体实施方式对本发明作进一步详细说明：

如图1所示，一种基于双神经辐射场的穿衣人体运动捕捉和生成方法，包括如下步骤：

步骤1．建立三维人体的标准模板空间，包括：

使用预先建立的三维人体模板网格和衣服模板网格训练的占据网络和带符号距

离场,用于表示人体和衣服的几何模型;以及分别用于表示人体和衣服外观的双神经辐射场。

[0015] 其中,双神经辐射场(NeRF)在后续步骤中逐渐优化。

[0016] 标准模板空间包括几何模型和颜色模型,且人体和衣服的表示相独立;人体和衣服的几何模型使用占据网络和带符号距离场表示。

[0017] 标准模板空间的几何模型在时序变化的人体图像合成时保持不变。

[0018] 人体和衣服的颜色模型使用双神经辐射场表示;在定义标准模板空间的颜色模型时,为人体和衣服分别定义一组隐式外观编码,对应了输入视频的每一帧图像。

[0019] 在模板空间神经辐射场的颜色模型中融入隐式外观编码,表达并恢复时序变化的外观。

[0020] 标准模板空间的几何模型在准备阶段使用人体和衣服网格预先训练好,使用占据网络和带符号距离场表示,标准模板空间的占据网络模型由函数Fo定义:$o(x)=Fo(\gamma_x(x))$。

[0021] 其中,$o(x)\in\{0,1\}$表示占据网络模型输出的在x坐标处的空间是否被占据;x为采样点坐标,$\gamma_x(x)$表示空间坐标x的位置编码,其定义如下:

$\gamma_x(x)=[\sin(x),\cos(x),\sin(2x),\cos(2x),\cdots,\sin(2^{m-1}x),\cos(2^{m-1}x)]^T$,m为自然数。

[0022] 为了处理变化的衣服和人体形状,本发明还使用带符号距离场(Signed Distance Field,SDF)表示人体和衣服的几何,使用多层感知机神经网络S表示。

[0023] 带符号距离场将三维空间坐标映射到带符号的距离s,即$S:p\in R^3\rightarrow s\in R$;

S表示带符号距离场模型,p表示三维点坐标;s为带符号的距离,表示三维点与最近物体表面的距离,符号代表所处位置是物体的内外,在内即为负,在外即为正。

[0024] 使用如下公式对带符号距离场做约束L_{sdf}:

$$L_{sdf}=\sum_{p\in\Phi}\left|S(p)\right|+\sum_{p\in S}\left(1-\nabla S(p),\bar{n}\right)+\sum_{p\in\Omega}\left|\nabla S(p)_2-1\right|+\sum_{p\in\Omega\setminus\Phi}\exp\left(-\delta\cdot\nabla S(p)\right)。$$

[0025] 其中,$\bar{n}$为表面法向,$\nabla S(p)$为三维空间的梯度,Ω和Φ表示三维空间和物体的表面,δ表示一个远大于1的常数,S(p)表示带符号距离场输出距离值s。

[0026] 占据场网络F_o由带符号的距离场网络S监督L_{occ}:

$$L_{occ}=\sum_{x\in\Omega}L_C\left(F_o(\gamma_x(x),o\right)。$$

[0027] 其中,o为由带符号距离场确定的占据值;如果$S(x)<=0$,o=1;否则o=0。Lc表示交叉熵损失。因此,几何模型网络的损失L_{geo}可以表示为:$L_{geo}=\mu_1L_{sdf}+\mu_2L_{occ}$。

[0028] 其中,μ_1和μ_2为正则化权重,实际操作中取值均为1.0。

[0029] 标准模板空间的颜色模型使用神经辐射场表示,对输入视频的每一帧图像定义一个隐式外观编码,则颜色模型由函数F_c定义:$c_i(x)=F_c(\gamma_x(x),r_d(d),\psi_i)$。

[0030] 其中,$c_i(x)$表示颜色模型输出的在x坐标处的颜色,d表示观察x坐标的视角方向,即x坐标所在射线的方向;ψ_i表示每一帧的隐式外观编码。

[0031] 使用两个占据网络,两个颜色网络分别表示人体和衣服的几何和颜色。

[0032] 具体为:利用一个占据网络F_o^b、一个颜色网络(F_c^b,φ^b)表示人体的几何和颜色,使用

另一个占据网络F_o^g、以及另一个颜色网络(F_c^g,φ^g)表示衣服的几何和颜色。

[0033] 其中,F_c^b、φ^b分别为表示人体颜色的颜色模型以及隐式外观编码。

[0034] F_c^g、φ^g分别表示表示衣服颜色的颜色模型以及隐式外观编码。

[0035] 步骤2. 对输入视频的每一帧图像建立变形场,使用骨架驱动的变形以及非刚性变形对人体和衣服变形,为保证衣服变形的准确性,使用预训练的物理感知模拟网络预测生成逼真的衣服模板网格监督衣服的变形,将标准模板空间中的三维人体模板网格和衣服模板网格变形到观察帧空间。该步骤2具体为:

步骤2.1. 非刚性变形。

[0036] 首先通过嵌入变形对非刚性变形建模,该嵌入变形基于变形图计算一个弯曲场;一个嵌入变形图G包含K个节点,并且在三维人体模板网格中能够被自动建立。

[0037] 节点变换由欧拉角$A\in R^{k\times 3}$和平移向量$T\in R^{k\times 3}$参数化。

[0038] 对于三维人体模板网格的每个顶点v,经过非刚性变形后的新坐标y由下述公式得到:

$$y=\sum_{k\in N(v)} w(v,g_k)\left[R(A_k)(v-g_k)+g_k+T_k\right]。$$

[0039] 其中,N(v)表示影响到顶点v的邻域顶点集合,$k\in N(v)$;g_k表示第k个顶点的坐标;A_k和T_k分别表示第k个邻域顶点变形所需的欧拉角和平移向量;$R(\bullet):R^3\rightarrow SO(3)$ 将欧拉角转换为旋转矩阵;$w(v,g_k)$是顶点v的第k个邻域顶点的变形权重,$w(v,g_k)$的值由如下公式计算:$w(v,g_k)=(1-||v-g_k||/d_{max})^2$;其中,$d_{max}$表示顶点v到k个最近顶点的距离。

[0040] 非刚性形变中的欧拉角A和平移向量T都使用一个多层感知机模型训练获得,同时多层感知机还反向传播优化一个隐式变形编码w。

[0041] 对于第i帧非刚性变形的欧拉角A_i,平移向量T_i和隐式变形编码w_i,由函数$F_{A,T}$定义:

$$F_{A,T}:w_i\rightarrow(A_i,\ T_i)。$$

[0042] 步骤2.2. 估计骨架驱动的变形。

[0043] 首先对输入视频的每一帧图像估计一个SMPL模型,并从估计的SMPL模型中计算蒙皮权重$w(y)_j$,$w(y)_j$表示顶点y的第j个部分的蒙皮权重。

[0044] 人体的参数SMPL模型使用85维向量表示$\Theta=(\theta,\beta)$。其中,$\beta\in R^{10}$,$\theta\in R^{75}$分别表示人体的形状参数和各个关节的相对角度。

[0045] 基于非刚性变形获得的三维人体模板网格,进一步应用线性蒙皮变形来进行变形,对于三维人体模板网格的每个顶点y,其变形后的观察帧空间的顶点$\hat{v}$的计算公式如下:

$\hat{v}=[\sum^{J}_{j=1}w(y)_jG_j]y$。其中,J是人体关节的数量;$w(v)_j$表示顶点v的第j个部分的蒙皮权重,$G_j\in SE(3)$ 表示刚性变换矩阵。

[0046] 步骤2.3. 使用预训练的物理感知模拟网络监督衣服的变形。

[0047] 使用物理感知模拟网络学习衣服跟随人体动作所产生的变形,物理感知模拟网络由一个多层感知机模型D_ϕ定义;首先在Marvelous Designer软件中模拟各种衣服的变形。

[0048] 对于每种衣服类别,使用25种衣服风格以及8种材质,将Marvelous Designer软件模拟出的衣服形状作为多层感知机模型D_ϕ的监督,则模拟的衣服模板网格由如下公式定

义。

[0049] $G_s=D_\phi(\gamma,\beta,\tau,\theta)$。

[0050] 其中，G_s为模拟的衣服模板网格，$\gamma\in R^4$表示衣服风格，τ表示衣服的材质。

[0051] 步骤3．在观察帧空间取采样点，将采样点的坐标通过逆变形场变换到标准模板空间下对应点的坐标，然后根据转换后的采样点坐标所在位置是衣服还是人体，将转换后的采样点坐标输入到标准模板空间中的对应神经辐射场(ReNF)查询颜色，渲染获得人体合成图像。

[0052] 该步骤3具体为：

为了获得在输入视频第i帧图像的动态神经辐射场，即人体在不同时刻的颜色和几何，首先使用$x^{can}=T_i(x)$将观察帧空间的采样点x转换到标准模板空间的点x^{can}。

[0053] 其中，T_i是三维人体模板网格变形场的逆变换。

[0054] 根据相机位置和拍摄视角，由相机向观察帧空间发射多条射线，每条射线r对应最终人体合成图像上的一个像素，然后在每条射线上采样。

[0055] 一条射线上的采样点表示为：r(t) = o + td。

[0056] 其中，$o\in R^3$为射线起点，$d\in R^3$为射线方向，t为采样间隔。

[0057] 将观察帧空间中的采样点坐标经过逆变形场变换到标准模板空间对应的坐标，然后将坐标输入标准模板空间中分别保存衣服和人体的模型查询几何与颜色。

[0058] 对于来自像素p的射线r，找到这条射线r与衣服或人体网格相交的三角面。

[0059] 如果相交的三角面来自三维人体模板网格，则将这条射线r上的所有采样点的掩码m_r^b设为1；否则，将这条射线r上的所有采样点的掩码m_r^b设为0。

[0060] 同理，如果相交的三角面来自衣服模板网格，则将这条射线r上的所有采样点的掩码m_r^g设为1，否则，将这条射线r上的所有采样点的掩码m_r^g设为0。

[0061] 分别使用以下公式来渲染完整的人体和衣服：

$$T_r^i=\prod_{j=1}^{i-1}\left(1-o_g^j\left(1-m_r^b\right)\right)\left(1-o_b^j m_r^b\right);$$

$$\hat{C}_r=\sum_{i=1}^{n}T_r^i\left(o_g^i\left(1-m_r^b\right)c_g^i+o_b^i m_r^b c_b^i\right);$$

其中，n为光线r上的采样点个数，T_r^i表示光线上各采样点的颜色权重；$\hat{C}_r$表示光线的颜色；c_g^i、c_b^i表示光线上第i个采样点的颜色；

o_g^j、o_b^j分别表示衣服和人体神经辐射场输出的光线r上第j个采样点的密度；o_g^i、o_b^i分别表示衣服和人体神经辐射场输出的光线r上第i个采样点的密度。

[0062] 观察帧空间中的采样点逆变形过程为：

对于观察帧空间的一个采样点x，首先搜索与该采样点x距离最近的SMPL模型的顶点v，然后使用逆线性蒙皮变换采样点x的坐标，具体公式如下：

$$\hat{x}=[\Sigma^{J}_{j=1}w(v)_j G_j]^{-1}x。$$

[0063] 其中，$\hat{x}$表示变形后的采样点坐标，$w(v)_j$表示顶点v的第j个部分的蒙皮权重，G_j

∈SE(3) 表示刚性变换矩阵;使用逆变形图将$\hat{x}$变换到标准模板空间下的坐标x^{can},公式如下:

$$x^{can}=[\sum_{k\in N(v)}w(v,g_k)A_k]^{-1}\cdot[\sum_{k\in N(v)}w(v,g_k)(\hat{x}-g_k-T_k+A_kg_k)]。$$

[0064] 步骤4. 通过多种约束联合优化标准模板空间中的几何模型和双神经辐射场以及变形场,实现三维人体的运动捕捉以及动态人体的外观和恢复。

[0065] 将观察帧空间下的采样点坐标通过逆变形转换到标准模板空间下的采样点坐标后,将变换后的采样点坐标输入标准模板空间神经辐射场查询该点的颜色和密度。

[0066] 使用多种约束联合训练标准模板空间中的几何模型和神经辐射场以及变形场。具体为:

用于监督衣服和人体颜色的损失函数L_{rgb}为:

$$L_{rgb}=\frac{1}{N_t}\sum_{r\in R}\left\|\hat{C}(r)-C(r)\right\|_2+\frac{1}{N_b}\sum_{r\in R}\|(\hat{C}(r)^b-C(r)^b)(1-m_r^g)\|_2+\frac{1}{N_g}\sum_{r\in R}\|(\hat{C}(r)^g-C(r)^g)(1-m_r^b)\|_2。$$

[0067] 其中,R为投射的光线集合,N_t为一次训练中选取的像素的数量,N_b为像素中属于人体的像素数量,N_g为像素中属于衣服的像素数量;$\hat{C}(r)$为神经辐射场预测的颜色,$C(r)$为颜色真值,公式中的上角标b和g分别表示颜色属于人体还是衣服。

[0068] 使用物理感知模拟网络约束变形的损失,即用于监督衣服变形的损失函数L_{sim}为:

$$L_{sim}=\frac{1}{|G|}\sum_{t\in G}\rho\left(\left\|\tilde{G}^t-G_s^t\right\|\right)。$$

[0069] 其中,t为网格G的顶点,|G|表示网格G的顶点数。

[0070] ρ表示Geman-McClure鲁棒性损失函数,$\tilde{G}^t$表示经过骨架驱动的变形和非刚性变形之后的网格顶点,G_s^t表示物理感知模拟网络输出的网格顶点坐标。

[0071] 用于加强变形表面的局部光滑性的尽可能刚性的损失函数L_{arap}为:

$$L_{arap}=\sum_{g_i}\sum_{g_j\in N(g_i)}\omega(g_i,g_j)\left\|d_{i,j}(A,T)\right\|_2。$$

[0072] 其中,g_i表示三维人体模板网格的顶点,g_j为g_i的邻域网格顶点,$N(g_i)$为g_i的邻域顶点集合,$g_j\in N(g_i)$,$w(g_i,g_j)$表示变形权重。

[0073] $d_{i,j}(A,T)=A_j(g_i-g_j)+g_j+T_j-(g_i+T_i)$。

[0074] 其中,A_j、T_j分别表示第j个邻域顶点变形的欧拉角和平移向量。T_i表示顶点i变形的平移向量。用于让变形后的网格重投影贴合真值mask的mask损失$L_{IoU}(T)$为:

$$L_{IoU}(T)=1-\frac{\left\|R(T)\otimes\bar{R}\right\|_1}{\left\|R(T)\oplus\bar{R}\right\|_1-\left\|R(T)\otimes\bar{R}\right\|_1}。$$

[0075] 其中,⊗和⊕表示矩阵按元素求积和求和;$\bar{R}$表示输入的人体mask真值;T=(M,G)表示三维人体模板网格与衣服模板网格的合集,R(•)表示对网格做投影。

[0076] 此处,M,G表示函数Liou的自变量,只笼统表示人体和衣服网格,不具体指哪个网格。

[0077] 在变形后的网格和模拟的网格上都使用mask损失L_{Iou}为:

$$L_{IoU}=L_{IoU}\left(\tilde{M},\tilde{G}\right)+L_{IoU}\left(\tilde{M},G_s\right)。$$

[0078] 其中，$\tilde{M}$和$\tilde{G}$表示人体和衣服变形后的网格。用于贴合衣服与三维人体模板网格的损失L_{attach}为：

$$L_{attach}=\frac{1}{|A|}\sum_{t\in A}\rho\left(\left\|\left(\tilde{G}^t-\tilde{M}^t\right)\right\|_2\right)。$$

[0079] 其中，A为衣服模板网格上与三维人体模板网格相接触的顶点集合，|A|为顶点集合A中的顶点个数，$\tilde{G}^t$和$\tilde{M}^t$分别表示变形后的衣服和三维人体模板网格顶点。

[0080] 用于防止衣服模板网格和三维人体模板网格相交叉的损失$L_{interp}(M_1,M_2)$为：

$$L_{interp}\left(M_1,M_2\right)=\frac{1}{N_c}\sum_{i,j\in C}ReLU\left(\left(M_1^i-M_1^j\right)\cdot N_1^i\right)。$$

[0081] 其中，C表示网格M_1、M_2易发生交叉部位的顶点集合，N_1为网格M_1的法向量，M_1^i表示网格M_1的第i个顶点，M_1^j表示网格M_1的第j个顶点，N_1^i表示网格M_1第i个顶点的法向量，N_c表示C中顶点的个数；为了保证标准模板空间与观察帧空间的人体与衣服的合理性，同时约束两个空间的三维人体与衣服模板网格，给定如下损失函数L_{interp}。

[0082] $$L_{interp}=L_{interp}\left(\hat{M},\hat{G}\right)+L_{interp}\left(\tilde{M},\tilde{G}\right)。$$

[0083] 其中，$\hat{M}$和$\hat{G}$分别表示人体和衣服变形前的网格。

[0084] 综上，总体损失函数L为：$L=\lambda_1L_{rgb}+\lambda_2L_{arap}+\lambda_3L_{sim}+\lambda_4L_{IoU}+\lambda_5L_{attch}+\lambda_6L_{interp}$，其中，$\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6\}$为正则化权重，实际操作中取值为{1.0,0.1,0.02,30,0.1,100.0}。

[0085] 步骤5. 利用训练好的标准模板空间中的几何模型和双神经辐射场以及变形场，选定新的相机视角进行图像合成，通过更改衣服的几何与颜色模板实现人体变装。

[0086] 具体为：选定新的相机位置和相机视角，由相机向观察帧空间发射射线并在射线上取采样点，并通过逆变形场变换到标准模板空间的对应点的坐标。

[0087] 将经过变换后的对应点的坐标，根据射线触碰到的标准模板空间中的网格是人体还是衣服，输入不同的NeRF查询密度和颜色，即可合成三维人体的新视角逼真图像。

[0088] 分别对两个个体训练两个基于双神经辐射场隐式表示，将其中一个个体的衣服的几何与颜色模型替换为另一个个体的，即可实现人体的变装。

[0089] 通过以上步骤，使得本发明方法能够同时灵活地对人体的几何和外观建模，因此，重建的几何精度更高，且能合成任意视角的逼真图像，具有更广泛的应用场景。

[0090] 由于本发明方法能够能准确地恢复时序变化地人体几何和外观，因而能够从多视角或单目视频中准确地捕捉三维人体运动，并生成逼真的任意视角人体运动视频。

[0091] 同时，由于本发明方法能够将人体和衣服分开建模，因此能够实现人体的变装。

[0092] 图3中给出了本发明方法效果的3组例子，每组例子给出了4个视角，每个视角从左向右分别是真值图像、经过本发明方法恢复的三维人体几何图像以及三维人体合成图像。

[0093] 图4给出了本发明方法的结果和其他方法的比较，总共两组对比数据。每组数据从左到右为：真值、DeepCap、NerfCap、ICON、BCNet、TailorNet和本发明方法。

[0094] 在每一个例子中，均展示了通过以上几种方法重建的几何的两个视角图像。

[0095] 由图4中各幅图对比发现:与本发明方法相比,传统方法在恢复人体的衣服,特别是宽松的裙子时准确度不高,其中DeepCap、NerfCap和ICON不能将人体和衣服分开。

[0096] 而本发明方法重建的人体表面则能够很好地将人体和衣服分开,对于穿一般衣服(包括宽松衣服)的人体也能重建较大的运动和几何细节,因而,本发明方法具有很强的人体运动表达能力,并且由于本发明方法对于衣服和人体分开建模,能够实现人体的变装。

[0097] 图5给出了衣服和人体分开渲染的结果和其他方法的对比,在图5中从左到右依次为真值,本发明方法渲染的单独衣服、单独人体和整个人体,Dynamic view synthesis from dynamic monocular video方法渲染的单独衣服、单独人体和整个人体。

[0098] 由图5中两组图对比不难发现:

本发明方法在分离衣服和人体时表现更好,人体上残留的衣服细节更少。

[0099] 图6为人体换装之后的几何渲染结果的两个例子,每个例子给出了三个不同视角不同姿势的图像。其中,两个人体的衣服相互交换了。

[0100] 由图6能够看出,本发明方法能够实现人体的变装,且几何精度高。

[0101] 当然,以上说明仅仅为本发明的较佳实施例,本发明并不限于列举上述实施例,应当说明的是,任何熟悉本领域的技术人员在本说明书的教导下,所做出的所有等同替代、明显变形形式,均落在本说明书的实质范围之内,理应受到本发明的保护。

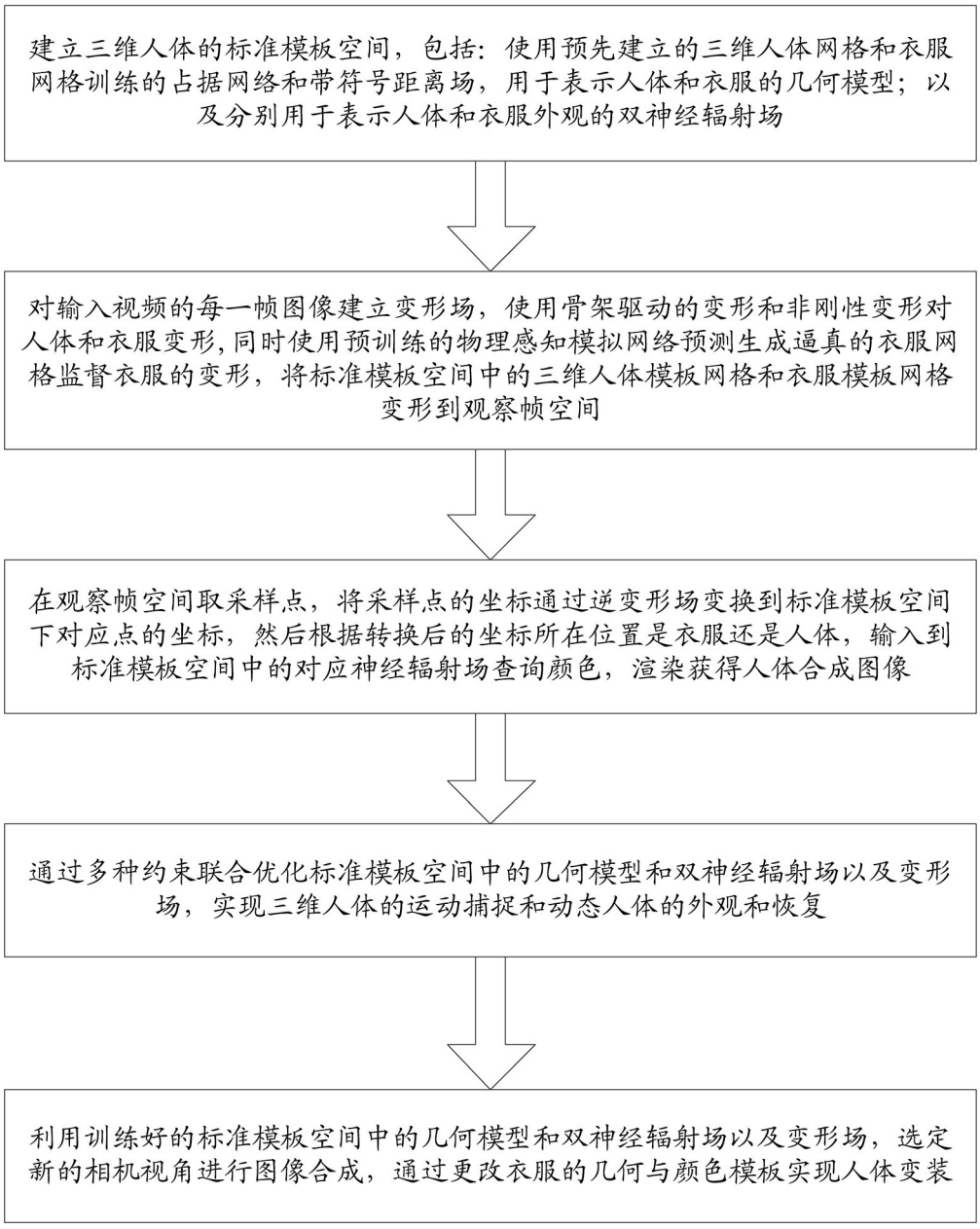

建立三维人体的标准模板空间，包括：使用预先建立的三维人体网格和衣服网格训练的占据网络和带符号距离场，用于表示人体和衣服的几何模型；以及分别用于表示人体和衣服外观的双神经辐射场
对输入视频的每一帧图像建立变形场，使用骨架驱动的变形和非刚性变形对人体和衣服变形，同时使用预训练的物理感知模拟网络预测生成逼真的衣服网格监督衣服的变形，将标准模板空间中的三维人体模板网格和衣服模板网格变形到观察帧空间
在观察帧空间取采样点，将采样点的坐标通过逆变形场变换到标准模板空间下对应点的坐标，然后根据转换后的坐标所在位置是衣服还是人体，输入到标准模板空间中的对应神经辐射场查询颜色，渲染获得人体合成图像
通过多种约束联合优化标准模板空间中的几何模型和双神经辐射场以及变形场，实现三维人体的运动捕捉和动态人体的外观和恢复
利用训练好的标准模板空间中的几何模型和双神经辐射场以及变形场，选定新的相机视角进行图像合成，通过更改衣服的几何与颜色模板实现人体变装

图 1

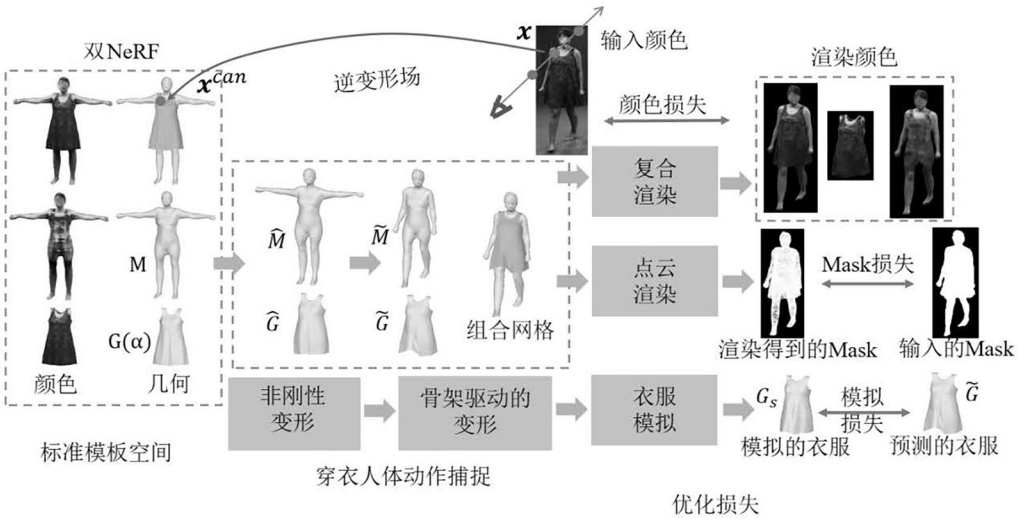

双NeRF
x^{can}
逆变形场
x
输入颜色
渲染颜色
颜色损失
复合渲染
M
G(α)
颜色
几何
$\widehat{M}$
$\widetilde{M}$
$\widehat{G}$
$\widetilde{G}$
组合网格
点云渲染
Mask损失
渲染得到的Mask
输入的Mask
非刚性变形
骨架驱动的变形
衣服模拟
G_s
模拟损失
$\widetilde{G}$
模拟的衣服
预测的衣服
标准模板空间
穿衣人体动作捕捉
优化损失

图 2

图 3

图 4

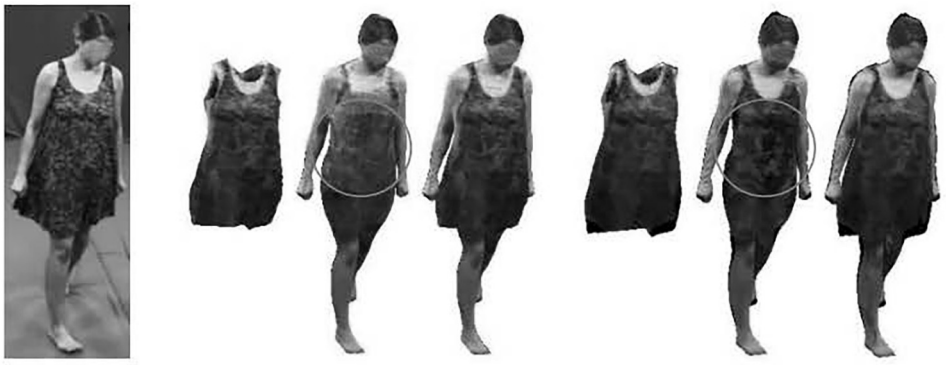


图 5

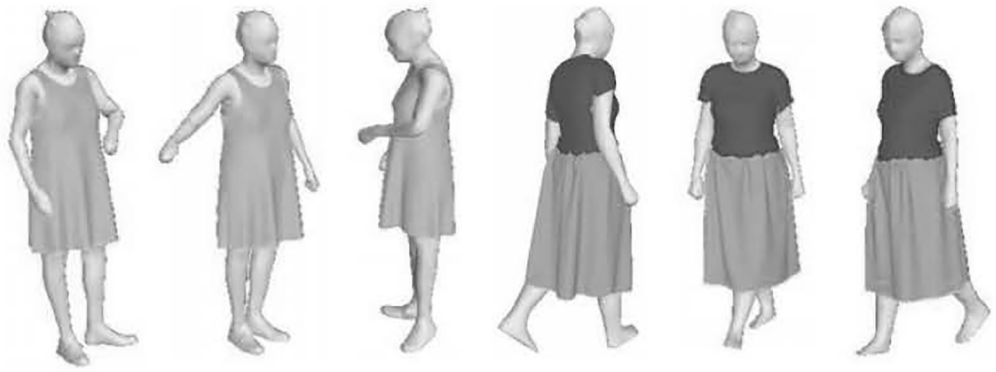


图 6