E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2024)09-2610-15

论文引用格式:Huang Q P, Liu L, Fu X D, Liu L J and Peng W. 2024. Clothed feature learning for single-view 3D human reconstruction. Journal of Image and Graphics, 29(09):2610-2624(黄千芃, 刘骊, 付晓东, 刘利军, 彭玮. 2024. 单视角三维人体重建的着装特征学习. 中国图象图形学报, 29(09):2610-2624)[DOI:10.11834/jig.230623]

单视角三维人体重建的着装特征学习

黄千芃[1],刘骊[1,2*],付晓东[1,2],刘利军[1,2],彭玮[1,2]

1. 昆明理工大学信息工程与自动化学院,昆明　650500; 2. 云南省计算机技术应用重点实验室,昆明　650500

摘　要:**目的**　由于单视角着装人体重建中存在肢体遮挡、着装姿态复杂,且现有方法仅能精确提取和表示着装人体图像中的视觉特征,未考虑复杂的着装姿态引起的动态细节表达,较难生成具有动态褶皱的着装人体模型。因此,提出一种单视角三维人体重建的着装特征学习方法。**方法**　首先对着装人体图像集中的单视角图像进行肢体特征表示,通过二维关节点预测与姿态特征深度回归,提取人体的着装姿态特征;再基于着装姿态特征,定义以柔性变形关节点为中心的着装褶皱采样空间和柔性变形损失函数,通过引入服装模板对输入的着装人体真值模型学习着装柔性变形,获得着装褶皱特征;然后,结合姿态参数回归、特征图采样特征和编解码器,构建联合着装人体像素和体素的人体形状特征学习模块,得到着装人体形状特征;最后结合褶皱特征、着装人体形状特征和计算的三维采样空间,通过定义有向距离场进行三维人体重建,输出最终的着装人体模型。**结果**　为了验证方法的有效性,在公开的THuman2.0数据集进行对比实验。结果显示,构建姿态特征学习模块有助于重建完整的肢体与正确的姿态,以褶皱特征学习对形状特征进行优化,可以获得高精度的重建结果。与当前先进的单视角三维人体重建方法比较,相比于性能第2的模型,本文方法重建结果的点到面距离与倒角距离分别降低了4.4%和2.6%。**结论**　本文提出的单视角三维人体重建的着装特征学习方法,能有效学习单视角三维人体重建的着装特征,生成具有复杂姿态和动态褶皱的着装人体模型。

关键词:单视角三维人体重建;着装特征学习;采样空间;柔性变形;有向距离场

Clothed feature learning for single-view 3D human reconstruction

Huang Qianpeng[1], Liu Li[1,2*], Fu Xiaodong[1,2], Liu Lijun[1,2], Peng Wei[1,2]

1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming　650500, China;
2. Computer Technology Application Key Laboratory of Yunnan Province, Kunming　650500, China

Abstract: **Objective** Clothed human reconstruction is an important problem in the field of computer vision and computer graphics. This process aims to generate three-dimensional (3D) human body models, including clothes and accessories, through computer technology, and is widely used in virtual reality, digital human body, 3D clothing assistant design, film and television special effects production, and other scenes. Compared with a large number of single-view images available on the internet, multiview images are more difficult to obtain. Considering that single-view images are easier to obtain on the internet, which can greatly reduce the use conditions and hardware cost of reconstruction, we consider single-view

收稿日期:2023-10-11;**修回日期**:2024-02-23;**预印本日期**:2024-03-01

*通信作者:刘骊　ieall@kust.edu.cn

基金项目:国家自然科学基金项目(62262036, 62362043, 61962030);云南省中青年学术和技术带头人后备人才培养计划项目(202005AC160036)

Supported by: National Natural Science Foundation of China(62262036, 62362043, 61962030); Yunnan Provincial Foundation for Leaders of Disciplines in Science and Technology (202005AC160036)

images as input to establish a complete mapping between single-view human image and human shape and restore the 3D shape and geometric details of the human body. Most methods based on parametric models can only predict the shape and posture of the human body with a smooth surface, whereas nonparametric model methods lack a fixed grid topology when generating fine geometric shapes. High-precision 3D human model extraction can be realized through the combination of parametric human model and implicit function. Given that clothing can produce dynamic flexible deformation with the change in human posture, most methods focus on obtaining the fold details of a clothed human model from 3D mesh deformation. The clothing can be separated from the human body with the assistance of a clothed template, and flexible deformation of the clothing caused by human body posture can be directly obtained via the learning-based method. Given the overlapping of limbs, occlusion, and complex clothed posture of clothed human body in single-view 3D human body reconstruction, obtaining geometric shape representation under various clothed postures and angles is difficult. Moreover, existing methods can only accurately extract and represent visual features from clothed human body images without considering the dynamic detail expression caused by complex clothed posture. Difficulties are encountered regarding the representation and learning of the clothed features related to the posture of single-view clothed human and generation of a clothed mesh with complex posture and dynamic folds. In this paper, we propose a single-perspective 3D human reconstruction clothed feature learning method. **Method** We propose a feature learning approach to reconstruct clothed human with a single-view image. The experimental hardware platform in this paper used two NVIDIA GeForce RTX 1080Ti GPU. We utilized the clothing co-parsing fashion street photography dataset, which includes 2 098 human images, to analyze the physical features of clothing. Human3. 6M dataset was used to learn posture features of the human body, with the test set at 3DPW capture from the field environment. For fold feature learning, we used objects 00096 and 00159 in the CAPE dataset. For better training effects of the clothed mesh, we selected 150 meshes close to the dressed posture from the THuman2. 0 dataset as the training set for use in shape feature learning. First, we represented the limb features of the single-view image and extracted the clothed human pose features through 2D node prediction and deep regression of pose features. Then, based on the pose features of the clothing, the sampling space of clothing fold centered on the flexible deformation joint and flexible deformation loss function were defined. In addition, flexible clothing deformation was learned through the introduction of a clothing template to the input ground truth model of the clothing body to obtain fold features. We only focused on crucial details inside the space to acquire the fold features. Afterward, the human shape features learning module was constructed via the combination of posture parameter regression, feature map sampling feature, and codec. The pixel and voxel alignment features were learned from the corresponding image and grid in the 3D human mesh dataset, and the shape features of the human body were decoded. Finally, through the combination of fold features, shape features of a clothed human, and calculated 3D sampling space, the 3D human mesh was reconstructed by defining the signed distance field, and the final clothed human model was outputted. **Result** Aiming at the results of posture feature and single-view 3D clothed human reconstruction, we used 3DPW and THuman2. 0 datasets. Our experimental results findings compared with those of the three methods on the 3DPW dataset. The mean per joint position error (MPJPE) and MPJPE of Platts alignment (PA-MPJPE) were used to evaluate the differences between the predicted 3D joints and ground truth. The mean per vertex position error (MPVPE) was used to evaluate the predicted SMPL 3D human shape and the ground truth grid. Compared with that of the second-best model, the error was decreased by 1. 28 on MPJPE, 1. 66 on PA-MPJPE and 2. 38 on MPVPE, and the average error was reduced by 2. 4%. For the 3D reconstruction of the clothed human body, we conducted experiments to compare the four methods on the THuman2. 0 dataset. We used the Chamfer distance (CD/Chamfer) and point-to-surface distance (P2S) of the 3D space to evaluate the gap between the two 3D mesh groups. Notably, the P2S of the reconstructed result can be reduced by 4. 4% compared with the second-best model, and CD can be reduced by 2. 6%. Experimental results reveal that the posture feature learning module contributed to the reconstruction of the complete limb and correct posture, and fold feature learning for optimized learned shape features can be used to obtain high-precision reconstruction results. **Conclusion** In this paper, the clothed feature learning method for single-view 3D human-body reconstruction enables the effective learning of the clothed feature of single-view 3D human reconstruction and generates clothed human reconstruction results with complex posture and dynamic folds.

Key words: single-view 3D human reconstruction; clothed feature learning; sampling space; flexible deformation; signed distance field

0 引 言

着装人体重建旨在通过计算机技术生成包括着装与配饰的三维人体模型(Zins等,2021),目前广泛应用于虚拟现实、数字人体、三维服装辅助设计、影视特效制作等场景。着装人体重建方法按实现技术的不同可归纳为4类:1)基于深度采集设备连续获取的数据,重建最终的三维着装人体模型(Yu等,2017),但该类方法设备复杂且成本较高;2)使用多边形网格的显式表示方法(Alldieck等,2019)进行着装人体重建,此类方法计算量较大;3)通过占用场提取等值面的隐式方法(Saito等,2019)表示和重建着装人体模型,但仍需显式几何结构;4)结合显式和隐式表示的方法(Zheng等,2022),最终生成着装人体三维模型,但局限于表达人体着装的动态特征。

考虑到单视角图像在网络环境下更易于获取,可大大降低重建的使用条件与硬件成本(杨航 等,2023),本文旨在建立单视角着装人体图像与人体形状间的完整映射,恢复着装人体的三维形状与几何细节。然而,由于单视角图像具有深度歧义性,且缺少训练数据,如何基于一系列模糊、不完整的信息重建着装人体模型成为难点(Natsume等,2019)。为了恢复准确的人体结构及表面几何形状,Jinka等人(2023)将SMPL(skinned multi-person linear)模板先验与稀疏、非参数编码三维模型的深度预测相结合,实现了参数化人体先验与非参数深度表示的高效融合,而参数模型较难描述服装的几何形状和纹理,因此,Aggarwal等人(2022)在人体表面上生成由隐式函数场定义的多个无交叉着装层,以避免人体与着装之间以及不同着装表面之间的自相交。

随着数字人体、虚拟试衣等技术不断发展(徐爱婧和周捷,2020),单视角三维人体重建中的着装特征表示和优化成为热点。然而,现有的着装人体三维重建方法通常侧重于表示服装款式特征(普骏程等,2021),由于单视角着装人体图像存在遮挡、着装姿态复杂等问题,上述方法未考虑复杂着装姿态,仅能提取和表示着装人体图像中的静态视觉特征,较难生成具有复杂姿态和动态褶皱的着装人体模型。综上所述,单视角着装人体重建还存在以下难点:1)着装人体存在肢体交叠、遮挡,着装姿态复杂等问题,现有方法未将人体关节与着装姿态进行有效关联,较难表示和学习人体在三维空间下的复杂着装姿态特征;2)现有方法仅能精确提取和表示着装人体图像中的视觉特征,未考虑由于复杂着装姿态引起的动态细节表达,较难生成具有动态褶皱的着装人体模型;3)真实环境下,人体自身动作会导致服装的柔性变形,由于单视角图像存在不可见视角以及缺少相应的深度信息,较难获得不同着装姿态和角度下的几何形状表达。

因此,本文以单幅着装人体图像作为输入,提出三维人体重建的着装特征学习方法,实现对着装人体中着装姿态特征、着装褶皱特征、着装人体形状特征的学习与表示。本文方法流程如图1所示,主要贡献如下:1)结合二维肢体特征表示和人体关节点深度值预测,对输入的单视角图像进行着装姿态特征学习,得到准确的着装人体三维姿态,以解决单视角着装人体图像不易获得复杂的三维姿态特征的问题;2)基于着装姿态特征,定义柔性变形损失函数,通过构建以柔性变形关节点为中心的着装褶皱采样空间,对着装人体重建真值进行网格表面顶点的蒙皮权重学习,预测着装姿态下产生的褶皱细节信息,以增强由着装姿态引起的各角度的动态细节表达;3)结合着装褶皱特征、计算的三维采样空间,以及通过构建的人体形状特征学习模块得到的着装人体形状特征,基于定义的有向距离场进行三维人体重建,提升最终生成的三维着装人体模型的精度。

1 相关工作

基于参数化模板(Loper等,2015)可显式表示人体的网格,精确地重建人体自然姿态下的不同身体形状。大多数三维人体姿态网格估计方法是从输入图像中回归得到人体网格模型的参数(Zhang等,2021),由于其高度非线性映射,会破坏输入图像中像素之间的空间关系。为此,Moon和Lee(2020)预测每个网格顶点坐标在热图上的像素似然性,以保留输入图像像素间的空间关系。为解决三维坐标无

法直接转换为参数化人体模板可用的相应旋转角的问题，Yu 等人（2021b）构建姿态细化与形状细化模块，以减少生成人体网格的错误与模糊性。近几年的方法引入多边形网格的线性混合蒙皮（linear blend skinning，LBS）以帮助实现网格形状的变形，但网格受限于自身的固定拓扑，结合神经隐式表达与线性混合蒙皮（Chen 等，2021），能使离散网格形成平滑、连续的表面，生成更复杂姿态下的网格。Cho 等人（2022）解决了由 Transformer 编解码器架构内指令交互产生的内存开销大、推理速度慢等问题，利用人体形态学关系的先验知识，加快了模型的收敛速度，提高了人体重建的精度。上述方法采用特征提取以及编解码器直接预测单视角图像中人体的三维姿态，但由于单视角着装人体图像中存在复杂肢体交叠、遮挡，导致二维姿态预测三维关节点的准确性降低。Choi 等人（2022）利用二维人体姿态辅助三维人体的估计，基于关节的回归器从深度卷积特征图中对图像特征进行采样，以保留目标人体的空间激活。受到该方法的启发，本文构建二维着装肢体特征表示，以提高图像二维姿态预测的准确性，与其不同，本文侧重基于二维着装肢体特征学习和表示单视角人体图像的三维着装姿态特征。

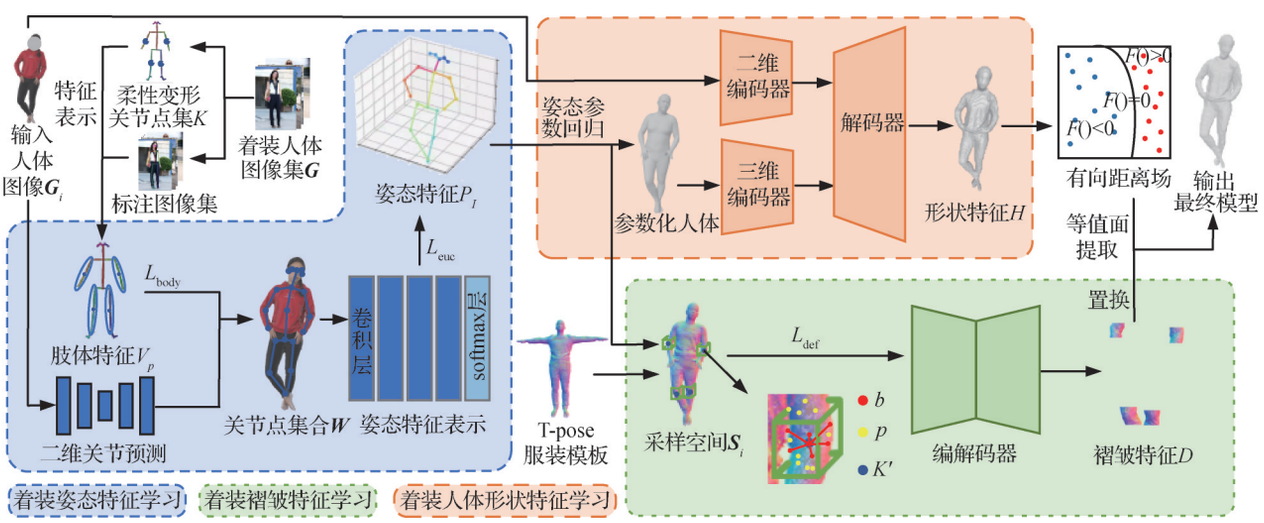


图 1　单视角三维人体重建的着装特征学习流程图

Fig. 1　The framework of clothed feature learning for single-view 3D human reconstruction

基于服装会随人体姿态的变化产生动态的柔性变形，许多研究工作侧重从三维网格变形获得着装人体的褶皱细节。Hu 等人（2020）通过拟合度量着装与身体之间相互作用，生成着装人体的褶皱细节。借助服装模板可以将着装与人体分离，并使用基于学习的方法分别单独处理，直接获得由人体姿态引起的着装柔性变形，Ma 等人（2021）使用单个模型对多个不同身体形状与不同服装类型进行建模，用于生成具有姿态相关几何形状的高保真着装人体。Su 等人（2022）将着装人体分解为多个几何层，即身体网格层和着装层，使用多层表示进行拓扑重建与纹理渲染。Lin 等人（2022）结合扩散蒙皮，学习表示粗糙服装拓扑的隐式表面模板，以获得姿态相关的服装变形。受该方法启发，本文基于服装模板学习因人体动态引起的着装柔性变形。不同的是，本文定义了着装人体关节点采样空间的柔性变形损失函数，以提取和表示单视角图像中与姿态相关的动态细节特征，增强着装人体动态细节的几何表达。

近年来，很多研究工作利用结合隐式函数和参数化人体模板的混合方法，重建高精度的三维着装人体模型。Huang 等人（2020）结合参数化三维人体估计和具有空间局部特征的隐式函数表示，减少由姿态变化引起的几何模糊，提升着装人体的表面细节。Liu 等人（2021）基于混合的显式与隐式模型进行二维与三维特征对齐，最终生成着装人体模型。为了解决由着装人体在规范姿态下的拓扑变化引起的占用场模糊问题，He 等人（2021）联合估计姿态空间与规范空间中的占用率，以提高任意视角下的真实性。Xiu 等人（2022）构建由 SMPL 引导的法线预测推理详细的着装人体法线，隐式表面回归器获得占用场等值面。在像素对齐的隐式模型中引入面向

深度的采样(Chan等,2022),预测和利用深度和人体解析信息,减少像素对齐的隐式模型在重建中产生的噪声与伪影。受上述方法启发,本文结合特征编解码器、特征图采样和姿态参数回归,构建联合着装人体像素和体素的人体形状特征学习模块,确保采用参数化模板约束人体形状的同时,引入隐式函数提取着装人体表面几何细节,获得着装人体形状特征。

由于着装动态特征体现了动态角度下人体与着装的变形,难以从静态图像中恢复。为此,Santesteban等人(2022)基于物理的变形模型,以自监督的方式训练动态三维服装变形。Yoon等人(2022)基于运动外观仿真,从三维人体表面的空间和时间生成可推广至任意姿态的表示。Chen等人(2022)提出多主题前向蒙皮技术,从每个人体的非刚性扫描中学习生成具有相应皮肤权重的精细三维人体形状。毛爱华等人(2022)采用蒙皮多人线性模型的多级拓扑构建的图卷积神经网络,将SMPL模型顶点投影到特征图中,以获取具体位置的局部特征。为了捕捉服装中高频细节,Zhang等人(2023)分解显式模板与服装模板,模拟不同姿态下的服装变形,能够更好地学习任意姿态相关的褶皱细节。Ho等人(2023)基于混合表示和端到端训练的网络架构,在可变形身体模型的网格顶点上存储局部几何特征。然而,以上方法仅适用于给定的着装模板,不能较好地适应着装图像场景下的复杂着装姿态。为获得动态下的着装变形,本文引入蒙皮权重,通过定义柔性变形损失函数,以获得人体骨骼运动与着装变形间的关系,此外,通过定义的采样空间,学习不同着装姿态下的褶皱特征,相比于学习整体网格可减少训练时间。

2 着装特征学习

本文提出的着装特征主要包括姿态特征、褶皱特征和形状特征,学习过程如图1所示。首先,结合自定义的柔性变形关节点对着装人体图像集进行标注,通过对标注的图像集进行二维着装肢体特征表示和着装人体关节点深度值预测,学习三维着装姿态特征。然后,基于三维着装姿态特征,定义着装人体关节点采样空间(包含变形骨骼点与网格表面顶点)的柔性变形损失函数,对输入的着装人体网格学习着装褶皱特征。最后,结合姿态参数回归和特征编解码器,构建联合着装人体像素和体素的人体形状特征学习模块,得到着装人体的形状特征。通过对着装特征的分步学习实现细节丰富的着装人体网格模型。

2.1 姿态特征学习

2.1.1 着装肢体特征表示

着装人体的肢体容易被服装遮挡,易导致姿态预测结果出现关节点位置错误、肢体长度(关节点间距)不合理。本文通过构建着装肢体特征表示,以提高二维肢体预测的准确性,并通过损失函数进行着装人体比例修正,以降低三维空间下人体肢体比例的不合理性。

针对着装场景下,人体骨骼较易被着装姿态引起活动的关节点,定义柔性变形关节点集 $\boldsymbol{K}=\{K_{\text{l\_elbow}},K_{\text{r\_elbow}},K_{\text{l\_knee}},K_{\text{r\_knee}}\}$,点集中各关节点 $K_{\text{l\_elbow}}$,$K_{\text{r\_elbow}}$,$K_{\text{l\_knee}}$,$K_{\text{r\_knee}}$ 分别表示着装人体的左肘、右肘、左膝、右膝,由二维空间坐标位置 (x,y) 表示。基于定义的柔性变形关节点集 $\boldsymbol{K}$,利用openpose关节点模型(Cao等,2017)的标注格式对输入的单视角着装人体图像集 $\boldsymbol{G}$ 进行柔性变形关节点标注。

首先,根据柔性变形关节点集 $\boldsymbol{K}$ 中每个关节点所在的着装肢体(包括左臂、右臂、左腿、右腿),对单幅单视角着装人体图像 $\boldsymbol{G}_i\in\boldsymbol{G}$,进行二维着装姿态特征聚类。其中,左肘 $K_{\text{l\_elbow}}$ 和右肘 $K_{\text{r\_elbow}}$ 分别与左臂和右臂进行联合聚类,左膝 $K_{\text{l\_knee}}$ 和右膝 $K_{\text{r\_knee}}$ 分别与左腿和右腿进行联合聚类,得到二维着装肢体特征表示 $\boldsymbol{V}_p=\{V_{\text{larm}},V_{\text{rarm}},V_{\text{lleg}},V_{\text{rleg}}\}$,各个特征对应的关节点如表1所示。

再基于着装肢体特征表示,定义二维着装肢体

表1 着装肢体特征与对应关节点

Table 1 Clothing limb feature and corresponding joint points

着装肢体特征	所含关节点
V_{larm}(左臂)	$K_{\text{l\_shoulder}}$、$K_{\text{l\_elbow}}$、$K_{\text{l\_wrist}}$ (左肩、左肘、左腕)
V_{rarm}(右臂)	$K_{\text{r\_shoulder}}$、$K_{\text{r\_elbow}}$、$K_{\text{r\_wrist}}$ (右肩、右肘、右腕)
V_{lleg}(左腿)	$K_{\text{l\_hip}}$、$K_{\text{l\_knee}}$、$K_{\text{l\_ankle}}$ (左髋、左膝盖、左脚踝)
V_{rleg}(右腿)	$K_{\text{r\_hip}}$、$K_{\text{r\_knee}}$、$K_{\text{r\_ankle}}$ (右髋、右膝盖、右脚踝)

特征损失函数，具体为

$$L_{\text{body}} = -\sum_{k=1}^{n} \boldsymbol{B}_k \ln p_k \tag{1}$$

式中，$\boldsymbol{B}_k$ 为一个 $1 \times n$ 的向量，n 为 $\boldsymbol{V}_p$ 中各二维肢体特征中包含的特征个数，p_k 为输入图像二维姿态中着装肢体属于 $\boldsymbol{V}_p$ 的第 k 个特征的概率。

最后，定义二维肢体特征仿射变换矩阵 $\boldsymbol{M}$，其中包含旋转角度、比例因子、x 轴平移、y 轴平移、向左翻转与向右翻转 6 个变量。仿射变换矩阵 $\boldsymbol{M}$ 使二维姿态预测结果经过变换能够接近 $\boldsymbol{V}_p$ 中各个特征的坐标。通过仿射变换矩阵计算 $\boldsymbol{V}_p$ 中每个二维肢体特征的得分，具体形式定义为

$$\boldsymbol{M}^* = \arg\min_{\boldsymbol{M}} \left\| \boldsymbol{M} \times w - \boldsymbol{V}_p \right\| \tag{2}$$

式中，w 为每次的二维姿态预测结果，由得分可计算概率 p_k，定义为

$$p_k = \frac{score_k}{\sum_{m=1}^{n} score_m} \tag{3}$$

式中，$\sum_{m=1}^{n} score_m$ 为 n 个特征得分的总和，得分的计算式为 $score = \exp(-\left\| \boldsymbol{M}^* \times w - \boldsymbol{V}_p \right\|)$，$score_k$ 为第 k 个特征的得分。

2.1.2　姿态关节点预测

结合二维着装肢体特征 $\boldsymbol{V}_p$ 与定义的损失函数 L_{body}，对 $\boldsymbol{G}_i$ 进行二维着装姿态预测，得到表示 $\boldsymbol{G}_i$ 二维着装姿态的关节点坐标集合 $\boldsymbol{W}$，其每个二维坐标点由二维空间坐标位置 (x, y) 表示。将二维着装姿态坐标集合 $\boldsymbol{W}$ 表示为热图 $\boldsymbol{T}^{2\text{D}} \in \mathbf{R}^{J \times 64 \times 64}$，其中 $J = 15$ 表示坐标集合 $\boldsymbol{W}$ 中关节点的数量，并通过特征提取器，得到图像 $\boldsymbol{G}_i$ 的早期特征 $\boldsymbol{F} \in \mathbf{R}^{c \times 64 \times 64}$，其中 $c = 64$ 为通道尺寸。再采用残差网络，通过提取图像 $\boldsymbol{G}_i$ 的鲁棒特征 $\boldsymbol{I}_f \in \mathbf{R}^{c' \times 8 \times 8}$ 进行着装人体关节点深度值预测，其中 $c' = 2\,048$ 为通道尺寸，然后从 $\boldsymbol{I}_f$ 中回归得到坐标点的深度 z 值，将 $\boldsymbol{W}$ 中的每个二维坐标点抬升为三维坐标点，并用三维空间坐标位置 (x, y, z) 进行表示。

由于二维关节点预测结果可能会出现人体比例不正确的情况，如肢体间长度不符合人体结构，且三维关节点的深度坐标值是在已知二维关节点的基础上进行回归，往往导致三维关节点预测结果出现人体比例错误。为了保证关节点的准确性，本文对 $\boldsymbol{W}$ 与 $\boldsymbol{K}$ 对应的关节点，定义损失函数为

$$L_{\text{euc}} = \min_{\boldsymbol{K}_i \in \boldsymbol{K}_p} \left| \left\| \boldsymbol{K}_i - \boldsymbol{K}_i' \right\| - \left\| \boldsymbol{K}_i - \boldsymbol{K}_i'' \right\| \right| \tag{4}$$

式中，$\boldsymbol{K}_i$ 为 $\boldsymbol{K}$ 中的第 i 个关节点，$\boldsymbol{K}_i'$、$\boldsymbol{K}_i''$ 为与 $\boldsymbol{K}_i$ 相邻的关节点，其与着装肢体特征中的相应关节点对应，每个关节点都有两个相邻的关节点。以左臂着装肢体特征 V_{larm} 为例，当 $\boldsymbol{K}_i$ 为 $\boldsymbol{K}_{\text{l\_elbow}}$ 时，$\boldsymbol{K}_i'$、$\boldsymbol{K}_i''$ 分别为 $\boldsymbol{K}_{\text{l\_shoulder}}$、$\boldsymbol{K}_{\text{l\_wrist}}$，$\| \cdot \|$ 为向量范数。通过 L_{euc} 对三维坐标点进行着装人体比例修正，以减小预测结果中不正确的人体比例与现实人体比例的误差，最终得到图像 $\boldsymbol{G}_i$ 的着装姿态特征 $\boldsymbol{P}_I$。

2.2　褶皱特征学习

由于局部服装几何形状会随着人体的运动而改变，导致着装人体包含多样的着装类型与拓扑结构，使得正确处理由人体姿态变化带来的动态褶皱的表示在重建带有柔性变形的着装人体模型中十分重要。因此，不仅需要保证服装拓扑的完整性，还需重建由复杂姿态引起的动态褶皱细节。本文通过构建以柔性变形关节点为中心的着装褶皱采样空间，对着装人体重建真值进行空间中网格表面顶点的蒙皮权重学习，预测着装姿态下产生的褶皱细节信息，提高人体模型的真实性，增强由着装姿态引起的着装人体模型各角度的动态细节表达。

2.2.1　着装褶皱采样

如图 2 所示，着装褶皱通常会集中在着装的局部区域，为了模拟服装的姿态相关变形，通常利用 SMPL 人体模板来解析人体关节运动，利用服装模板表示服装产生的变化。Moon 等人（2022）将服装模板引入三维人体重建，为训练数据集之外的野外人体图像生成稳健的重建结果。受此方法启发，本文引入服装模板，学习因人体动态引起的着装柔性变形。然而，区别于上述方法以姿态参数作为输入，通过已有模板生成姿态拟合的模型，本文则是以图像作为输入，通过以柔性变形关节点集构建的着装褶皱采样空间和定义的隐式函数，学习集中于空间内部的几何细节表达，生成不同目标图像内的着装。

如图 3 所示，以经过二维坐标点抬升的柔性变形关节点集 $\boldsymbol{K}'$ 中每个着装人体关节点 K_i' 为中心，构建包含人体网格内部变形骨骼点 b 和网格表面顶点 p 的着装褶皱采样空间 S_i，空间 S_i 的三维边界框为立方体，其中点的三维坐标取值范围由 K_i' 的三维坐标值 (x, y, z) 确定。对于 $K_{\text{l\_elbow}}$，$K_{\text{r\_elbow}}$，空间范围为 $(x \pm 0.25, y \pm 0.25, z \pm 0.25)$，对于 $K_{\text{l\_knee}}$，$K_{\text{r\_knee}}$，空间范

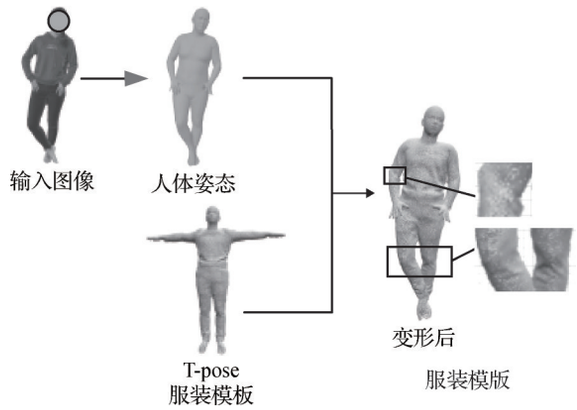


图2　人体姿态变形实例

Fig. 2　Example of human posture deformation

围为($x \pm 0.2, y \pm 0.2, z \pm 0.3$),其中$\{b, p\} \in S_i$。上述所选取的空间范围均能获得较好的实验结果。点$b$的坐标位置决定于着装姿态特征$\boldsymbol{P}_I$,而当点$b$相对于标准姿态发生坐标改变时,将引起一定数量的点$p$的位移,每一个点$b$可影响的点$p$数量及位置是固定的,不同位置的点$b$影响不同数量的点$p$,并适应不同位置。对空间$S_i$中的每个网格表面顶点$p$进行姿态蒙皮加权,定义为

$$\boldsymbol{w}_p = w_0 \boldsymbol{B} \tag{5}$$

式中,$\boldsymbol{w}_p = \{w_1, w_2, \cdots, w_{bn}\}$表示权重,$bn$为该顶点关联的骨骼数量,$w_0$为初始化的$w_p$,$\boldsymbol{B}$为$K_i'$处的骨骼变换矩阵。

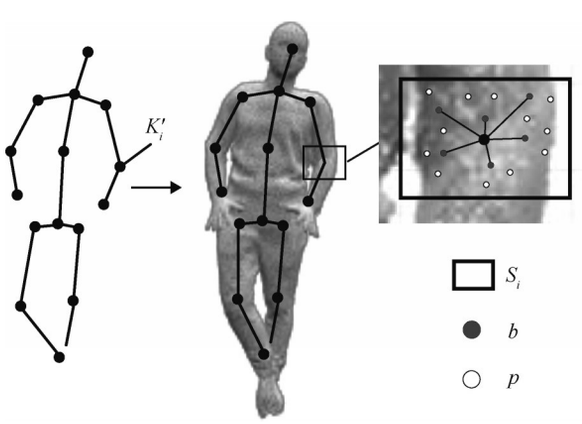


图3　采样空间及采样点定义

Fig. 3　Definition of sampling space and sampling point

为使服装模板能够在姿态作用下产生相应变形,本文引入的服装模板与SMPL模型(Loper等,2015)一致,着装人体网格$M_c(\cdot)$表示为

$$M_c(\beta, \theta) = W(T_c(\beta, \theta), J(\beta), \theta) \tag{6}$$

式中,$T_c(\beta, \theta) = \boldsymbol{T} + r_p, \theta \in \mathbf{R}^{21 \times 3}$为人体的姿态参数,$\beta \in \mathbf{R}^{10}$为人体的形状参数,$W(\cdot)$为蒙皮函数,$T_c(\beta, \theta)$为线性函数,用于向标准姿态下的T-pose网格顶点$\boldsymbol{T} \in \mathbf{R}^{n \times 3}$添加姿态相关的变形位移$r_p$,$J(\beta)$代表人体骨骼。通过褶皱特征表示获得位移$r_p$,学习人体由于姿态产生的动态变形。

2.2.2　柔性变形损失

为在蒙皮权重迭代过程中保留使网格表面顶点距离与真值更加接近的权重值,以避免模拟结果与真值相差过大,因此对空间S_i中的每个网格表面顶点p,定义柔性变形损失函数为

$$L_{\text{def}} = \min(g(w_p)p - t(p), w_p p - t(p)) \tag{7}$$

式中,$g(w_p) = w_0 f(\left\| \boldsymbol{K}_i - \boldsymbol{p}^* \right\|, \theta_p) + w_p$为姿态相关的蒙皮权重更新函数,$\boldsymbol{p}^*$为采样空间$S_i$中的剩余顶点,$\theta_p$为该着装姿态特征$P_I$下与$K_i'$相邻两处骨骼所形成的夹角;$t(p)$为顶点$p$的地面真值。

假设正向蒙皮权重场$w: \mathbf{R}^3 \rightarrow \mathbf{R}^{24}$满足以下约束条件:1)$w_p$应与位于标准姿态下SMPL网格$\boldsymbol{T}$上的$p$的蒙皮权重$w_p^s$相同;2)权重场$w$应从SMPL网格表面自然地扩散,即权重沿法线方向的变化率应该为0,在每个网格表面顶点p定义为

$$\begin{cases} w_p = w_p^s \\ \nabla w_p \cdot n^s(p) = 0 \end{cases} \tag{8}$$

式中,w_p^s表示p处的SMPL蒙皮权重,$n^s(p)$表示在$\boldsymbol{T}$中p处的法线方向。

2.2.3　褶皱特征表示

本文基于得到的着装姿态特征$\boldsymbol{P}_I$,结合定义的柔性变形损失函数L_{def},通过对变形后服装模板下着装褶皱采样空间S_i中的每个网格顶点p进行着装褶皱特征表示,以学习着装人体网格表面的褶皱特征。

首先,在着装褶皱采样空间S_i中,通过采样得到网格顶点p的点集$\boldsymbol{P}^u = \{p_n\}_{n=1}^N$,查询并得到每个顶点$p$在扩散蒙皮场$w(p_n)$中的蒙皮权重$w_p$。$\boldsymbol{P}^u = \{p_n\}_{n=1}^N$保存了采样空间$S_i$中所有网格顶点$p$的位置,其中$N$代表空间中网格顶点数量。

由于变形后的服装模板与T-pose模板中相应的顶点$\boldsymbol{P}^t = \{p_n^t\}_{n=1}^N$一一对应,采用姿态编码器生成姿态相关的几何特征$\{\phi_p(p_n^t)\}_{n=1}^N = E_p(\boldsymbol{P}^u, \boldsymbol{P}^t)$,再将几何特征$\phi_p(p_n^t)$输入到姿态解码器$D_p$中。通过几何特征张量$\boldsymbol{F} \in \mathbf{R}^{h \times w \times 64}$对解码器进行显式约束,其中,$h, w$为特征张量的空间维度,将着装人体内在的、与姿态无关的几何特征从解码器中解耦,只关注

与姿态相关、经过姿态产生的变形。

最后通过姿态解码器D_p得到每个网格顶点p处姿态相关褶皱的位移$r^p = D_p(\phi_p(p_n^t), p_n^t)$后，通过定义的损失函数$L_{\text{def}}$更新位移$r^p$，得到位移后的各顶点的集合，生成着装褶皱特征$D$。

2.3 形状特征学习

大多数基于参数化模型的方法仅限于预测具有平滑表面的人体形状、姿态，而在生成精细的几何形状时，非参数化模型方法缺乏固定的网格拓扑，容易出现肢体形状不完整。本文受到Zheng等人(2022)的启发，融合参数化人体模型与隐式函数，生成高精度的三维人体形状。

首先将单视角着装人体图像输入深度图像编码器，得到图像$\boldsymbol{G}_i$的特征图$\boldsymbol{E}_i$，使用双线性插值对$\boldsymbol{E}_i$的值进行采样的采样函数定义为

$$S(\boldsymbol{E}_i, \pi(q)) \tag{9}$$

式中，$\pi(q)$为空间中三维点q在特征图$\boldsymbol{E}_i$中的二维投影像素点。基于给定像素的局部图像特征，使用采样函数$S(\cdot,\cdot)$对三维点q进行占据概率值估计，学习得到像素对齐特征$\boldsymbol{P}$。

然后，将着装姿态特征$\boldsymbol{P}_I$重塑为三维热图$\boldsymbol{T}^{3D} \in \mathbf{R}^{J\times 8\times 8\times 8}$，使用$\boldsymbol{P}_I$中的二维空间坐标$(x, y)$从图像$\boldsymbol{G}_i$的鲁棒特征$\boldsymbol{I}_f$中采样每个关节的图像特征，并从$\boldsymbol{T}^{3D}$中采样预测置信度，联合得到$\boldsymbol{F}^M$，将$\boldsymbol{F}^M$输入图卷积网络得到生成SMPL人体所需的姿态参数$\theta \in \mathbf{R}^{21\times 3}$与形状参数$\beta \in \mathbf{R}^{10}$，并由参数计算SMPL人体网格，其形式为

$$M(\beta,\theta) = W(T(\beta,\theta), J(\beta)) \tag{10}$$

式中，$W(\cdot)$为蒙皮函数，$T(\beta,\theta)$为姿态参数θ与形状参数β下的T-pose模板；$J(\beta)$代表人体骨骼。将SMPL人体网格$M(\beta,\theta)$经过网格体素化后转换为占用体积，输入三维特征编码器得到特征空间E_v，使用$S(E_v, q)$对空间中三维点q进行双线性插值，得到体素对齐特征$\boldsymbol{V}$。

最后，联合像素对齐特征$\boldsymbol{P}$与体素对齐特征$\boldsymbol{V}$并进行解码，得到将空间占用概率，通过提取等值面得到着装人体形状特征$\boldsymbol{H}$。

基于Saito等人(2019)将像素对齐的隐式函数定义为空间占用概率，结合参数化人体模型与像素对齐隐式函数，本文将着装人体形状空间占用率定义为

$$\begin{cases} F(C(q)) \\ C(q) = (S(E_i, \pi(q)), S(E_v, q))^{\mathrm{T}} \end{cases} \tag{11}$$

式中，$F(\cdot)$为空间占用概率值，当$F(C(q)) < 0$时，表示点q存在于网格内部；当$F(C(q)) > 0$时，则q存在于网格外部；当$F(C(q)) = 0$时，则q存在于网格表面。隐式函数使用像素级特征作为条件变量，重建与输入图像对齐的精细网格表面。

给定图像上的一个像素，沿z轴方向投射一条穿过该像素的射线，隐式函数根据给定像素的局部像素特征与体素特征来估计该射线上空间的占用概率值$F(\cdot)$，以显示网格在空间中的形状与位置，提取$F(C(q)) = 0$处的空间点集，从有向距离场中生成初步的着装人体网格，由于着装人体网格与SMPL拟合，通过查询与着装褶皱特征D对应的SMPL网格顶点，与初步网格中的相应顶点进行置换，得到最终的着装人体网格模型。

3 实验结果与分析

3.1 数据集与评估指标

本文实验硬件平台选用Intel Core i9-9900k CPU @3.60 GHz，NVIDIA GeForce RTX 2080Ti GPU 2，32 GB DDR4 2666 MHz RAM，集成式开发环境为Visual Studio Code等。

本文基于Clothing Co-Parsing时尚街拍数据集(Yang等，2014)进行着装肢体特征表示，数据集包含2 098幅着装人体图像，能较好体现着装姿态下的肢体特征；Human3.6M数据集(Ionescu等，2014)用于学习着装人体的姿态特征，本文只使用Human3.6M的训练集(对象S1，S5，S6，S7，S8)，测试集选用三维测试集3DPW(Von Marcard等，2018)，以评估本文方法的鲁棒性；对于褶皱特征学习，本文使用CAPE数据集中的对象00096(男性，20动作序列，共4 800帧)与00159(女性，12动作序列，共2 088帧)，每个对象都包含规范姿态下的T-pose模板；受限于使用设备等因素，为达到更好的着装人体模型训练效果，本文从THuman2.0数据集(Yu等，2021a)的共525个网格模型中，挑选了接近着装姿态下的150个网格作为训练集进行形状特征学习。数据集中各姿态分布如图4所示，其中前4个为手部姿态，高举代表手部举过头顶，除下垂、叉腰、高举等的其他手部姿态

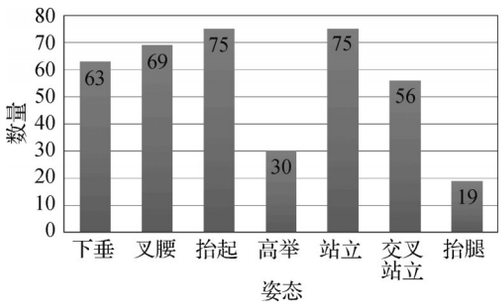


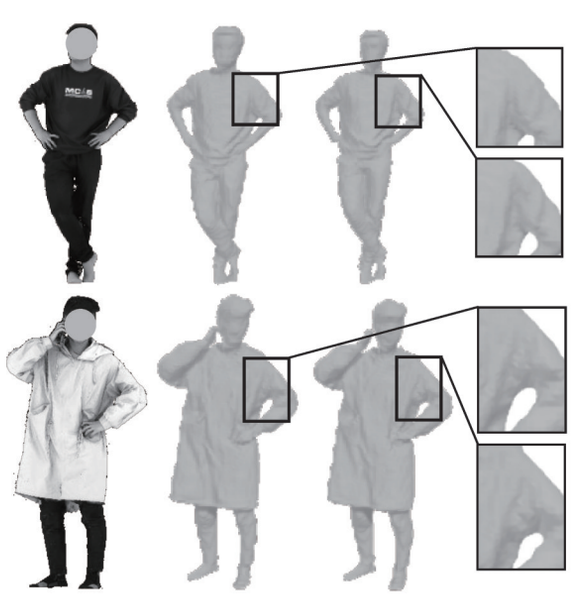
图4　选择的着装姿态统计

Fig. 4　The statistics of selected clothed postures

统称为抬起;后3个为腿部姿态。

实验采用的THuman2.0数据集中,除常见着装姿态外,还包括舞蹈、武术等特殊场景下的动作。图5展示了顺序选取150个人体模型与筛选150个人体模型训练后的对比,结果表明经过筛选后着装褶皱较为集中的区域能生成更丰富的细节。该数据集需要进行预处理以获得与网格拟合的SMPL模型用于训练。

本文使用平均每关节位置误差(mean per joint position error,MPJPE)、普氏对齐的平均每关节位置误差(mean per joint position error of platts alignment,PA-MPJPE),评价预测的三维关节点与真值的误差;使用平均每顶点位置误差(mean per vertex position error,MPVPE),评价预测的三维人体网格与真值网格间的误差。在着装人体三维重建中,使用三维空间的倒角距离(chamfer distance,CD/Chamfer)与点到面距离(point-to-surface distance,P2S)以及两点间欧氏距离(L2范数)进行评估。上述所有评估指标的值越低则代表方法的效果越好。



(a) 输入图像　(b) 未筛选数据集 (c) 筛选数据集　(d) 细节对比

图5　有无筛选数据集的结果对比

Fig. 5　Results comparison with or without filtering dataset ((a) input images; (b) unfiltered dataset; (c) filtered dataset; (d) details)

3.2　实验结果与对比分析

本文分别在3DPW数据集和THuman2.0数据集上与其他三维姿态预测方法和三维人体重建方法进行了姿态特征学习结果、三维重建精度的定量与定性结果对比与分析。

3.2.1　定量结果与对比分析

如表2所示,本文将姿态特征学习与GCMR(graph convolutional mesh regression)(Kolotouros等,2019)、Pose2Mesh(Choi等,2020)、3Dcrowd(Choi等,2022)3种三维人体姿态估计方法进行了对比,可看出本文方法能够提高关节点预测的准确性,相较3DCrowd方法在MPJPE上提高1.28、在PA-MPJPE上提高1.66、在MPVPE上提高2.38,误差平均降低2.4%,其中着装肢体特征表示对二维关节点预测具有明显作用,性能提升较大;深度回归提高了深度值预测的准确性,人体比例损失函数能够获得更好的三维关节点结果。

表2　姿态特征学习与其他方法的定量对比

Table 2　Quantitative comparison between posture feature learning and other methods

方法	MPJPE	PA-MPJPE	MPVPE
GCMR	-	70.20	-
Pose2Mesh	89.50	56.30	105.30
3DCrowd	81.70	51.50	98.30
姿态特征学习	**80.42**	**49.84**	**95.92**

注:加粗字体表示各列最优结果,"-"表示缺少相应数据。

本文与PIFu(pixel-aligned implicit function)(Saito等,2019)、PIFuHD(pixel-aligned implicit function for high-resolution 3D human digitization)(Saito等,2020)、Pamir(parametric model-conditioned implicit representation)(Zheng等,2022)和ICON(implicit clothed humans obtained from normals)(Xiu等,2022)4种单视角三维人体重建方法进行了重建效果的定量对比。如表3所示,本文方法的CD值与P2S值结

果均有明显降低，相较其他方法中效果最佳的模型分别降低了2.6%与4.4%，说明本文方法能够降低重建误差，提高重建的鲁棒性；L2值结果达到最优，说明本文能有效重建人体姿态相关的局部褶皱信息。如表4所示，PIFu的隐式表示能够以较快的速度进行重建，本文方法所需时间达到次优，但能够具有更为丰富的细节。

表3　本文方法与其他方法的重建精度对比

Table 3　Reconstruction accuracy comparison between proposed method and other methods

方法	CD	P2S	L2
PIFu	2.49	2.35	0.19
PIFuHD	2.34	2.19	0.17
Pamir	2.01	1.92	0.16
ICON	1.86	1.81	0.15
本文	**1.81**	**1.73**	**0.15**

注：加粗字体表示各列最优结果。

表4　本文方法与其他方法的时间对比

Table 4　Time comparison between proposed method and other methods

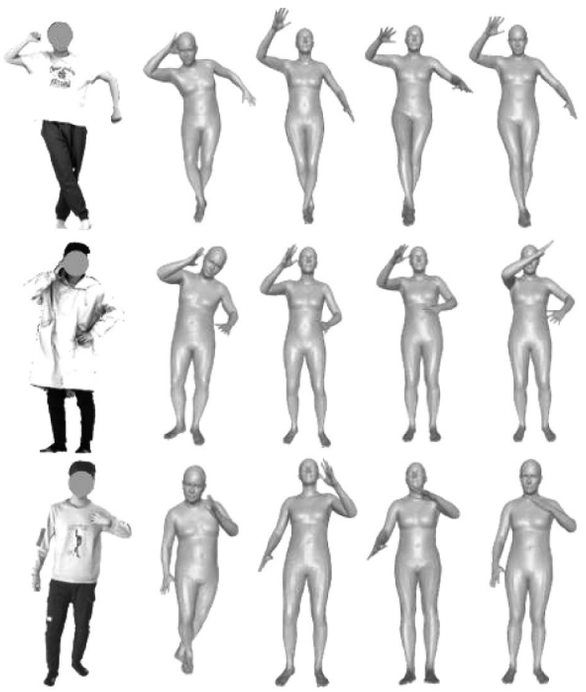
方法	时间/s
PIFu	**41.92**
PIFuHD	49.33
Pamir	48.25
ICON	47.70
本文	47.02

注：加粗字体表示最优结果。

3.2.2　定性结果与对比分析

为进一步验证本文方法对于着装姿态预测的准确性，将姿态特征学习与其他3种三维人体姿态估计方法进行定性对比分析。本文选取包含站立、交叉站立、举起、下垂、叉腰等着装姿态的图像进行测试，如图6所示。

从图6可以看出，GCMR（Kolotouros等，2019）方法的手部姿态较为准确，但身体角度与腿部姿态都存在明显误差；Pose2Mesh（Choi等，2020）在第1、2行示例中的手部姿态较为准确，但第1行示例的腿部交叠方向错误，第3行示例的手部有较大误差；3Dcrowd（Choi等，2022）方法的腿部姿态较为准确，但第1、3行示例中的手部姿态不够准确。



(a) 输入图像　(b) GCMR　(c) Pose2Mesh(d) 3Dcrowd (e) 本文方法

图6　姿态特征学习与其他方法的对比

Fig. 6　Comparison between posture feature learning and other methods ((a) input images; (b) GCMR; (c) Pose2Mesh; (d) 3Dcrowd; (e) ours)

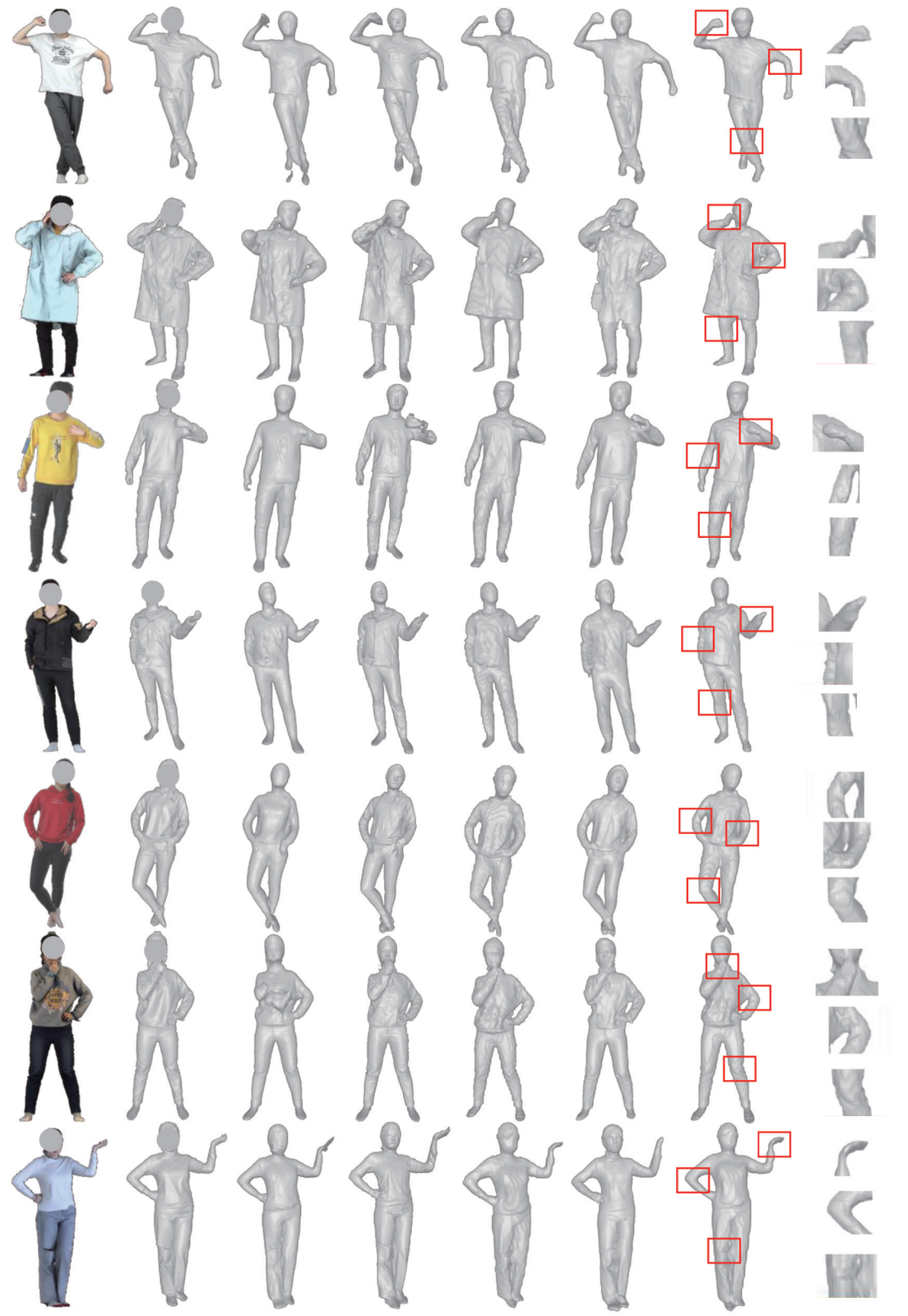
综上，本文方法在手部与腿部姿态上均能得到较准确的预测结果。

如图7所示，本文在重建效果上与PIFu（Saito等，2019）、PIFuHD（Saito等，2020）和Pamir（Zheng等，2022）相比，能够得到更加完整的人体结构，如手部、腿部末端等未出现缺失，且着装褶皱细节更丰富；与ICON（Xiu等，2022）相比，本文的部分重建结果在细节方面有所欠缺，如第2行示例中ICON方法的结果缺失部分着装的形状，第1、4行示例中右手的模型角度相比输入图像不够准确，第3、7示例中缺失手部末端。但相比于ICON，本文方法能获得更准确的着装人体姿态与形状。

3.3　消融实验结果分析

由于着装肢体特征数量的选取会对肢体预测的结果产生影响，本文选取手臂肢体特征为3、4，双腿肢体特征为4、5、6，通过排列组合进行对比。如表5所示，在左臂特征数量为4、右臂为4、双腿为5时，误差将达到最低。

为了验证肢体特征表示与肢体比例修正的有效性，本文将左臂、右臂、双腿肢体特征数量分别确定



(a) 输入图像 (b) 真值模型 (c) PIFu (d) PIFuHD (e) Pamir (f) ICON (g) 本文方法 (h) 细节展示

图7 本文方法与现有同任务方法结果对比

Fig. 7 The visual results of our method compared with existing related methods

((a) input images; (b) ground truth; (c) PIFu; (d) PIFuHD; (e) Pamir; (f) ICON; (g) ours; (h) details)

表 5　着装肢体特征学习消融实验结果分析

Table 5　Ablation experimental results analysis of clothed body feature learning

左臂	右臂	双腿	MPJPE	PA-MPJPE	MPVPE
3	3	4	81.16	50.01	96.76
3	3	5	81.63	50.37	97.51
3	3	6	82.41	50.82	99.26
4	4	4	81.38	50.01	97.22
4	4	5	**80.80**	**49.89**	**96.36**
4	4	6	81.31	50.15	96.85

注：加粗字体表示各列最优结果。

为 4、4、5，通过姿态特征学习进行对比。表 6 中的定量分析结果表明，相较于只添加着装肢体特征表示和人体比例修正而言，两者结合后的误差降低，说明姿态预测的准确性有所提高，其中肢体特征带来的提升更为明显。

本文以形状特征为基本网络，通过增加姿态特征模块与褶皱特征模块以评估其有效性。如表 7 所示，通过模块的增加，CD、P2S 有所降低，代表重建结果更加接近真实值，准确性提高；L2 的降低说明模型的精度有所提高，即褶皱细节重建的精度提高。

表 6　姿态特征学习消融实验结果分析

Table 6　Ablation experimental results analysis of posture feature learning

方法	MPJPE	PA-MPJPE	MPVPE
基本网络	80.73	50.24	96.45
基本网络 + 肢体特征	80.57	49.89	96.00
基本网络 + 比例修正	80.80	50.17	96.36
基本网络 + 肢体特征 + 比例修正	**80.42**	**49.84**	**95.92**

注：加粗字体表示各列最优结果。

表 7　着装特征学习消融实验结果分析

Table 7　Ablation experimental results analysis of clothed feature learning

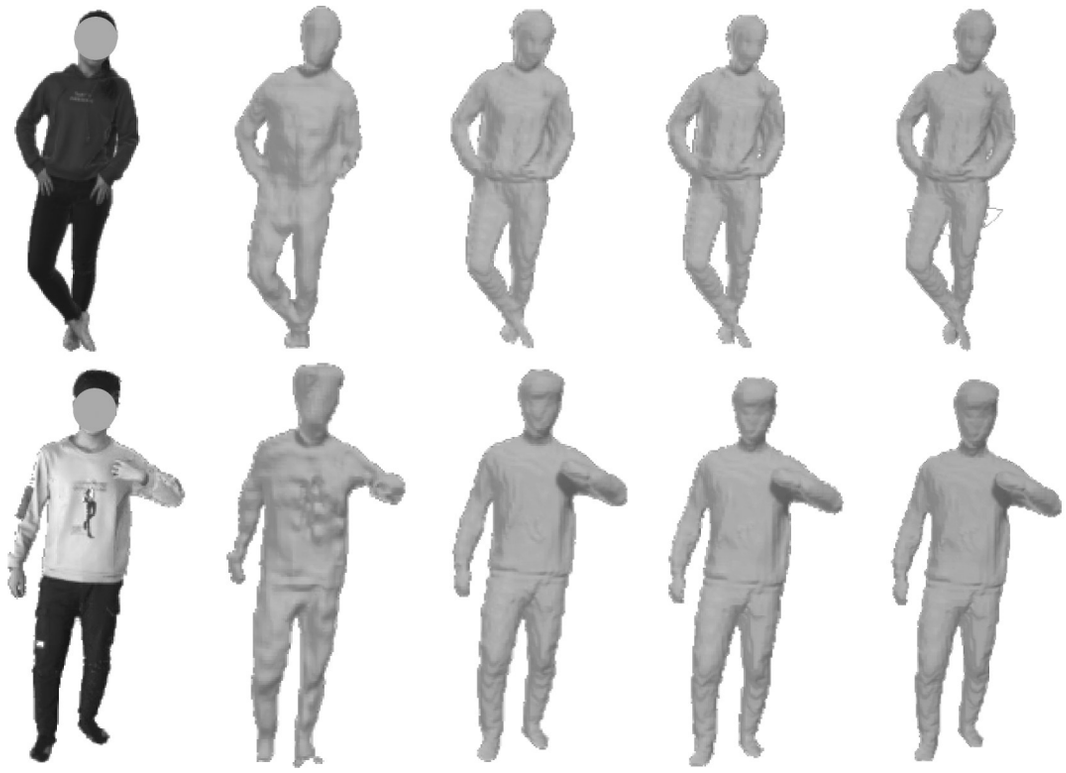
方法	CD	P2S	L2
形状特征	1.99	1.87	0.17
形状特征 + 姿态	1.91	1.79	0.16
形状特征 + 褶皱	1.88	1.78	**0.15**
形状特征 + 姿态 + 褶皱	**1.81**	**1.73**	**0.15**

注：加粗字体表示各列最优结果。

各模块添加后的重建结果如图 8 所示，证明本文构建的姿态特征学习模块有助于重建完整的肢体与正确的姿态以约束重建结果；以褶皱特征学习得到的褶皱细节特征对学习到的形状特征进行优化，可有效重建正确的人体姿态与动态变形。

4　结　论

针对单视角三维人体重建较难获得不同着装姿态和角度下的几何形状表达，未考虑复杂的着装姿态引起的动态细节表达，本文提出一种着装人体的着装特征学习方法。通过姿态特征学习，能够提高人体姿态预测的准确度；褶皱特征学习能够预测着装姿态下产生的褶皱细节信息，增强动态细节表达；形状特征学习能够实现高精度三维人体形状的提取。实验结果表明，本文方法在人体姿态数据集上的 MPJPE 误差降低 1.28，PA-MPJPE 误差降低 1.66，MPVPE 误差降低 2.38，平均降低 2.4%；重建结果的点到面距离可降低 4.4%，倒角距离可降低 2.6%。本文的姿态特征学习仅针对常见的着装姿态，未考虑运动、舞蹈等特殊场景下的复杂着装姿态。由于人体姿态和动作姿势较多，用有限的服装模板等方法可以获得的重建着装人体效果有限，方法上的通用性不足；此外，由于服装褶皱较多且较复杂，仅通



(a) 输入图像 (b) 形状特征 (c) 形状特征+姿态 (d) 形状特征+褶皱 (e) 形状特征+姿态+褶皱

图8 着装特征学习消融实验定性对比

Fig. 8 Qualitative comparison of clothed feature learning ablation experiments ((a) input images; (b) shape features; (c) shape features + pose; (d) shape features + fold; (e) shape features + pose + fold)

过服装模板不能获得不同种类、不同环境的褶皱。因此,本文的后续工作将着重围绕上述问题展开,以进一步提高单视角三维着装人体重建的精度。

参考文献(References)

Aggarwal A, Wang J K, Hogue S, Ni S F, Budagavi M and Guo X H. 2022. Layered-garment net: generating multiple implicit garment layers from a single image//Proceedings of the 16th Asian Conference on Computer Vision. Macao, China: Springer: 378-395 [DOI: 10.1007/978-3-031-26319-4_23]

Alldieck T, Magnor M, Bhatnagar B L, Theobalt C and Pons-Moll G. 2019. Learning to reconstruct people in clothing from a single RGB camera//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1175-1186 [DOI: 10.1109/CVPR.2019.00127]

Cao Z, Simon T, Wei S E and Sheikh Y. 2017. Realtime multi-person 2d pose estimation using part affinity fields//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1302-1310 [DOI: 10.1109/CVPR.2017.143]

Chan K Y, Lin G S, Zhao H Y and Lin W S. 2022. Integratedpifu: integrated pixel aligned implicit function for single-view human reconstruction//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 328-344 [DOI: 10.1007/978-3-031-20086-1_19]

Chen X, Jiang T J, Song J, Yang J L, Black M J, Geiger A and Hilliges O. 2022. gDNA: towards generative detailed neural avatars//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20395-20405 [DOI: 10.1109/CVPR52688.2022.01978]

Chen X, Zheng Y F, Black M J, Hilliges O and Geiger A. 2021. SNARF: differentiable forward skinning for animating non-rigid neural implicit shapes//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 11574-11584 [DOI: 10.1109/ICCV48922.2021.01139]

Cho J, Youwang K and Oh T H. 2022. Cross-attention of disentangled modalities for 3d human mesh recovery with transformers//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 342-359 [DOI: 10.1007/978-3-031-19769-7_20]

Choi H, Moon G and Lee K M. 2020. Pose2Mesh: graph convolutional network for 3D human pose and mesh recovery from a 2D human pose//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 769-787 [DOI: 10.1007/978-3-

030-58571-6_45］

Choi H，Moon G，Park J K and Lee K M. 2022. Learning to estimate robust 3D human mesh from in-the-wild crowded scenes//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA：IEEE：1465-1474［DOI：10.1109/CVPR52688.2022.00153］

He T，Xu YL，Saito S，Soatto S and Tung T. 2021. ARCH++：animation-ready clothed human reconstruction revisited//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：11026-11036［DOI：10.1109/ICCV48922.2021.01086］

Ho H I，Xue L X，Song J and Hilliges O. 2023. Learning locally editable virtual humans//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver，Canada：IEEE：21024-21035［DOI：10.1109/CVPR52729.2023.02014］

Hu P P，Nourbakhsh N，Tian J，Sturges S，Dadarlat V and Munteanu A. 2020. A generic method of wearable items virtual try-on. Textile Research Journal，90（19/20）：2161-2174［DOI：10.1177/0040517520909995］

Huang Z，Xu Y L，Lassner C，Li H and Tung T. 2020. ARCH：animatable reconstruction of clothed humans//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：3090-3099［DOI：10.1109/CVPR42600.2020.00316］

Ionescu C，Papava D，Olaru V and Sminchisescu C. 2014. Human3.6M：large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence，36（7）：1325-1339［DOI：10.1109/TPAMI.2013.248］

Jinka S S，Srivastava A，Pokhariya C，Sharma A and Narayanan P J. 2023. Sharp：shape-aware reconstruction of people in loose clothing. International Journal of Computer Vision，131（4）：918-937［DOI：10.1007/s11263-022-01736-z］

Kolotouros N，Pavlakos G and Daniilidis K. 2019. Convolutional mesh regression for single-image human shape reconstruction//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA：IEEE：4496-4505［DOI：10.1109/CVPR.2019.00463］

Lin S Y，Zhang H W，Zheng Z R，Shao R Z and Liu Y B. 2022. Learning implicit templates for point-based clothed human modeling//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv，Israel：Springer：210-228［DOI：10.1007/978-3-031-20062-5_13］

Liu L Y，Sun J C，Gao Y Q and Chen J Y. 2021. HEI-Human：a hybrid explicit and implicit method for single-view 3D clothed human reconstruction//Proceedings of the 4th Chinese Conference on Pattern Recognition and Computer Vision. Beijing，China：Springer：251-262［DOI：10.1007/978-3-030-88007-1_21］

Loper M，Mahmood N，Romero J，Pons-Moll G and Black M J. 2015. SMPL：a skinned multi-person linear model. ACM Transactions on Graphics，34（6）：#248［DOI：10.1145/2816795.2818013］

Ma Q L，Yang J L，Tang S Y and Black M J. 2021. The power of points for modeling humans in clothing//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：10954-10964［DOI：10.1109/ICCV48922.2021.01079］

Mao A H，Zhuo G J and Xuan J. 2022. Deformable 3D clothed humans reconstruction by a multi-level topological graph convolutional network. Journal of Computer-Aided Design and Computer Graphics，34（12）：1899-1910（毛爱华，禚冠军，褶骏. 2022. 采用多级拓扑图卷积网络的可变形三维着装人体重建. 计算机辅助设计与图形学学报，34（12）：1899-1910）［DOI：10.3724/SP.J.1089.2022.19225］

Moon G and Lee K M. 2020. I2l-meshnet：image-to-lixel prediction network for accurate 3d human pose and mesh estimation from a single RGB image//Proceedings of the 16th European Conference on Computer Vision. Glasgow，UK：Springer：752-768［DOI：10.1007/978-3-030-58571-6_44］

Moon G，Nam H，Shiratori T and Lee K M. 2022. 3D clothed human reconstruction in the wild//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv，Israel：Springer：184-200［DOI：10.1007/978-3-031-20086-1_11］

Natsume R，Saito S，Huang Z，Chen W K，Ma C Y，Li H and Morishima S. 2019. SiCloPe：silhouette-based clothed people//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA：IEEE：4475-4485［DOI：10.1109/CVPR.2019.00461］

Pu J C，Liu L，Fu X D，Liu L J and Huang Q S. 2022. Clothing visual representation for 3d human reconstruction. Journal of Computer-Aided Design and Computer Graphics，34（3）：352-363（普骏程，刘骊，付晓东，刘利军，黄青松. 2022. 三维人体重建中的服装视觉信息表示. 计算机辅助设计与图形学学报，34（3）：352-363）［DOI：10.3724/SP.J.1089.2022.18921］

Saito S，Huang Z，Natsume R，Morishima S，Li H and Kanazawa A. 2019. PiFu：pixel-aligned implicit function for high-resolution clothed human digitization//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul，Korea（South）：IEEE：2304-2314［DOI：10.1109/ICCV.2019.00239］

Saito S，Simon T，Saragih J and Joo H. 2020. PiFuHD：multi-level pixel-aligned implicit function for high-resolution 3d human digitization//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：81-90［DOI：10.1109/CVPR42600.2020.00016］

Santesteban I，Otaduy M A and Casas D. 2022. SNUG：self-supervised neural dynamic garments//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA：IEEE：8130-8140［DOI：10.1109/CVPR52688.2022.00797］

Su Z Q，Wan W L，Yu T，Liu L J，Fang L，Wang W Pand Liu Y B.

2022. MulayCap: multi-layer human performance capture using a monocular video camera. IEEE Transactions on Visualization and Computer Graphics, 28(4): 1862-1879 [DOI: 10.1109/TVCG.2020.3027763]

Von Marcard T, Henschel R, Black M J, Rosenhahn B and Pons-Moll G. 2018. Recovering accurate 3D human pose in the wild using Imus and a moving camera//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 614-631 [DOI: 10.1007/978-3-030-01249-6_37]

Xiu Y L, Yang J L, Tzionas D and Black M J. 2022. ICON: implicit clothed humans obtained from normals//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13286-13296 [DOI: 10.1109/CVPR52688.2022.01294]

Xu A J and Zhou J. 2020. Research on key technologies of virtual fitting system. Progress in Textile Science and. Technology, (3): 28-32 (徐爱婧, 周捷. 2020. 虚拟试衣系统关键技术研究. 纺织科技进展, (3): 28-32) [DOI: 10.3969/j.issn.1673-0356.2020.03.009]

Yang H, Chen R, An S P, Wei H and Zhang H. 2023. The growth of image-related three dimensional reconstruction techniques in deep learning-driven era: a critical summary. Journal of Image and Graphics, 28(8): 2396-2409 (杨航, 陈瑞, 安仕鹏, 魏豪, 张衡. 2023. 深度学习背景下的图像三维重建技术进展综述. 中国图象图形学报, 28(8): 2396-2409) [DOI: 10.11834/jig.220376]

Yang W, Luo P and Lin L. 2014. Clothing co-parsing by joint image segmentation and labeling//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 3182-3189 [DOI: 10.1109/CVPR.2014.407]

Yoon J S, Ceylan D, Wang T Y, Lu J W, Yang J M, Shu Z X and Park H S. 2022. Learning motion-dependent appearance for high-fidelity rendering of dynamic humans from a single camera//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 3397-3407 [DOI: 10.1109/CVPR52688.2022.00340]

Yu T, Guo K W, Xu F, Dong Y, Su Z Q, Zhao J H, Li J G, Dai Q H and Liu Y B. 2017. BodyFusion: real-time capture of human motion and surface geometry using a single depth camera//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 910-919 [DOI: 10.1109/ICCV.2017.104]

Yu T, Zheng Z R, Guo K W, Liu P P, Dai Q H and Liu Y B. 2021a. Function4D: real-time human volumetric capture from very sparse consumer RGBD sensors//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 5742-5752 [DOI: 10.1109/CVPR46437.2021.00569]

Yu Z B, Wang J J, Xu J W, Ni B B, Zhao C L, Wang M S and Zhang W J. 2021b. Skeleton2Mesh: kinematics prior injected unsupervised human mesh recovery//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 8599-8609 [DOI: 10.1109/ICCV48922.2021.00850]

Zhang H W, Lin S Y, Shao R Z, Zhang Y X, Zheng Z R, Huang H, Guo Y D and Liu Y B. 2023. CloSET: modeling clothed humans on continuous surface with explicit template decomposition//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 501-511 [DOI: 10.1109/CVPR52729.2023.00056]

Zhang H W, Tian Y T, Zhou X C, Ouyang W L, Liu Y B, Wang L M and Sun Z N. 2021. PyMAF: 3D human pose and shape regression with pyramidal mesh alignment feedback loop//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 11426-11436 [DOI: 10.1109/ICCV48922.2021.01125]

Zheng Z R, Yu T, Liu Y B and Dai Q H. 2022. PaMIR: parametric model-conditioned implicit representation for image-based human reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 3170-3184 [DOI: 10.1109/TPAMI.2021.3050505]

Zins P, Xu Y L, Boyer E, Wuhrer S and Tung T. 2021. Data-driven 3D reconstruction of dressed humans from sparse views//Proceedings of 2021 International Conference on 3D Vision. London, UK: IEEE: 494-504 [DOI: 10.1109/3DV53792.2021.00059]

作者简介

黄千芃,女,硕士研究生,主要研究方向为图形图像处理。E-mail:20212204233@stu.kust.edu.cn

刘骊,通信作者,女,教授,博士生导师,主要研究方向为计算机图形学与计算机视觉、图像处理。E-mail:ieall@kust.edu.cn

付晓东,男,教授,博士生导师,主要研究方向为服务计算、决策理论与方法。E-mail:xdfu@kust.edu.cn

刘利军,男,副教授,主要研究方向为图像处理、云计算和信息检索。E-mail:cloneiq@kust.edu.cn

彭玮,女,教授,博士生导师,主要研究方向为机器学习和数据挖掘。E-mail:weipeng@kust.edu.cn