

基于二维图像的参数化人体建模

魏雪霞, 徐增波*, 王巧丽

(上海工程技术大学 纺织服装学院, 上海 201620)

摘要:为解决现有基于二维图像的三维人体建模容易出现误差、模型参数估计不准确等问题,结合深度学习和优化后的人体建模算法,并利用深度学习算法提供的结果作为先验,缓解优化算法对参数初始值敏感的问题,以实现针对单帧 RGB 图像和多帧 RGB 图像的三维人体建模。同时基于 HMR 深度学习模型预测 SMPL-X 模型的初始参数,通过添加人体轮廓、二维关键点等对模型参数进一步优化求解,并利用视频序列的帧间连贯性对视频帧中人物的三维姿态进行约束。结果表明:添加人体分割轮廓约束可使重建模型更加贴合人体,能够提升模型拟合的精度;基于视频序列帧间连贯性重建的人体模型,可减少身体旋转、四肢弯曲等误差出现的可能性,与目标对象姿态更加相近。

关键词: 二维图像;三维人体建模;SMPL-X 模型;人体轮廓分割;能量函数

中图分类号: TS 941.2 **文献标志码:** A **文章编号:** 2096-1928(2023)01-0024-07

Parametric Human Modeling Based on Two-Dimensional Image

WEI Xuexia, XU Zengbo*, WANG Qiaoli

(School of Textiles and Fashion, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: The existing 3D human modeling based on two-dimensional images is prone to errors and inaccurate estimation of model parameters. In order to alleviate the problem that the optimization algorithm is sensitive to the initial value of parameters, the results provided by the deep learning algorithm are used as a priori, and the human modeling algorithm based on deep learning and optimization is combined. This enables 3D human modeling for single frame RGB images and multi frame RGB images. The initial parameters of SMPL-X model are predicted based on HMR deep learning model, and the model parameters are further optimized by constraints such as human contour and two-dimensional key points. The inter frame coherence of the video sequence is used to constrain the 3D pose of the characters in the video frame. The results showed that adding human segmentation contour constraints could make the reconstructed model more fit the human body and improve the accuracy of model fitting. The pose of the human model reconstructed based on the inter frame coherence of the video sequence was more similar to that of the target object, which reduced the possibility of errors such as body rotation and limb bending.

Key words: two-dimensional image, 3D human modeling, SMPL-X model, human contour segmentation, energy function

三维人体建模技术旨在获得一个与人体体型相近的三维模型,能够在虚拟场景中更加形象地展示用户的体型信息,其研究成果广泛应用于虚拟试衣、人体测量、虚拟现实、服装样板自动生成、影视

游戏制作、医疗医学等方面^[1]。然而,目前三维人体建模主要依赖三维扫描或者由专业建模设计师通过软件进行手工设计,这些方法的成本高,效率和精度低,而且无法实现在线试衣等功能的实时应

收稿日期:2022-09-22; 修订日期:2023-01-09。

作者简介:魏雪霞(1997—),女,硕士研究生。

* 通信作者:徐增波(1972—),男,教授,硕士生导师。主要研究方向为计算机视觉及三维虚拟建模。

Email: xuzengbo@aliyun.com

用,难以在实际生活中进行推广。因此,需要一种相对简单的方法,在保证成本较低的同时能够实现高精度三维人体建模。目前,三维建模方法有软件建模、三维扫描设备建模和基于二维图像建模 3 种。采用三维建模软件创建的人体模型真实性和精度不高,对人体的还原度也差强人意,为了获得高真实感的三维人体模型,其制作过程比较漫长,需要消耗建模人员大量的时间和精力^[2]。通过三维扫描设备建模虽然具有速度快、稳定性强和准确度高的优点,但是处理数据的算法繁琐、耗时长,且过度依赖于复杂而昂贵的硬件设备。基于二维图像建模的方式是近年来备受学者关注的三维人体重建方法,其操作更加简易方便。

目前,3 种建模方法都有相关研究成果,在三维建模软件方面,劳秀霞等^[3]针对被建模物体的结构特征将建模方法分为两种,外形比较规则的物体选择堆砌建模法,外形不规则的物体选择细分建模法。吕珍等^[4]对现有技术进行改进,结合 OpenGL 构建三维人体模型及服装试穿来读取 3DS 文件。在三维扫描设备建模方面,刘雁等^[5]通过三维扫描仪获得大量的人体模型数据,并运用小平面网格建立人体模型,从而提高了建模的效率和模型的可控性。韩凯等^[6]利用多组扫描仪分别获得高辨识度的人体无关点云数据、人体上身和全身轮廓点云数据,将采集到的 3 组点云数据进行对齐、替换、拼接,从而构建出高辨识度的三维人体模型。在基于二维图像建模方面,GUAN P 等^[7]从单幅 RGB 图像中估计人体的形状和姿势,然后生成一个人体的二维边界并对图像中的人体轮廓进行分割,最后将分割出的人体轮廓与 SCAPE 模型拟合生成三维人体模型。由于充分利用了图像中的轮廓、阴影和边缘信息,所以可以解决遮挡问题,但在人体姿态估计时仍需手动对人体关节点进行标注,无法自动完成单视图三维人体建模。BOGO F 等^[8]提出了从单幅图像重建三维人体的自动化方法 SMPL-X,利用基于卷积神经网络的 Deepcut 方法预测图像中的 2D 身体关节的位置,然后将 SMPL-X 模型与关节点进行拟合完成人体的姿态估计和体形重建。但由于没有对姿态施加约束,导致生成的三维人体有不合理的姿态。张广翩^[9]提出一种新型方法,只需输入单张二维点云图就可以得到最终的三维人体模型。这种方法考虑到人体表面形状复杂度的同时处理破损的二维点云,但该方法对人体姿势的要求十分严格,而且单张图片获取到的轮廓信息有限,重建出来的三维人体模型侧面信息会有一定误差。

基于二维图像的三维人体建模容易出现误差、模型参数估计不准确等问题,文中提出一种结合基于深度学习和优化人体建模算法的方法,以相对简单一般化的方式完成三维人体模型的构建,生成精确度更高的人体模型,且在实际应用中有更大的适用性。

1 基于单幅 RGB 图像的三维人体建模

1.1 图像采集

选取 3 名女大学生作为实验对象,采用常规标准站姿,通过分辨率为 1 920 × 1 080 的普通智能手机采集样本。3 名目标对象的正面图像如图 1 所示。

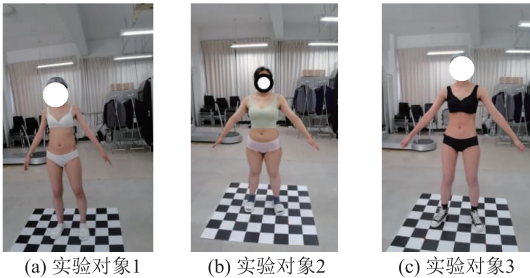
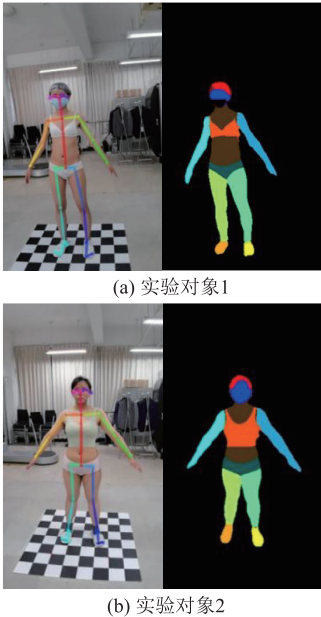


图 1 原始图像采集

Fig. 1 Original image acquisition

1.2 数据预处理

利用 OpenPose 姿态估计算法^[10]和Graphonomy^[11]深度学习模型对图像进行人体关节点估计与轮廓分割,得到 3 个目标人体对应的二维关节点与轮廓分割结果。OpenPose 关节点估计和人体轮廓分割结果具体如图 2 所示。





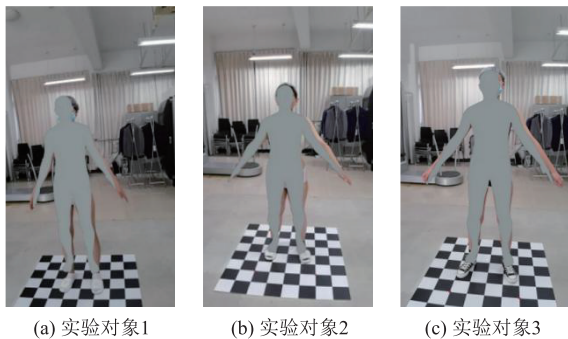
(c) 实验对象3

图2 OpenPose 关节点估计和人体轮廓分割结果

Fig.2 Results of OpenPose joint point estimation and human body contour segmentation

1.3 初始模型获取

利用人体网格重构 (human mesh recovery, HMR) 算法^[12] 进行三维人体姿态建模, 从而得到 SMPL-X 模型的体型和姿态参数以及相机的内参, 并将这些参数作为后续步骤的初始值。基于 HMR 算法得到的模型估计结果如图3所示。



(a) 实验对象1

(b) 实验对象2

(c) 实验对象3

图3 HMR 算法模型估计结果

Fig.3 Estimation results of the HMR algorithm model

1.4 能量函数构造

对2D关节点进行匹配(将OpenPose所得的二维关节点与SMPL-X模型的三维关节点进行对应), 发现姿态拟合后的形体与实际形体仍然存在一定差异。由于SMPL-X模型是根据欧美人群生成的参数化模型, 其形体与亚洲人体形具有适配性问题, 导致模型拟合误差偏大。因此, 可以通过添加轮廓约束项降低拟合误差。考虑到文中模型求解时主要是求取形体, 且采用的是标准站姿, 故在具体应用时对SMPL-X算法^[13]进行相应修改, 改进后的能量函数为

$$E(\beta, \theta, \psi) = E_J + \lambda_\alpha E_\alpha + \lambda_\beta E_\beta + \lambda_c E_c(\beta, \theta; K, U_c) \quad (1)$$

式中: β, θ, ψ 分别为 SMPL-X 模型的体型参数、姿态参数和面部表情参数; E_J 为重投影损失; λ 为权重参数; E_α 为用于膝盖和肘部弯曲的简单先验; E_β 为身体形状先验; $E_c(\beta, \theta; K, U_c)$ 为添加的轮廓约束项, 表示从 SMPL-X 模型渲染得到的轮廓与基于

Graphonomy 分割得到的轮廓之间的拟合误差, 其定义为

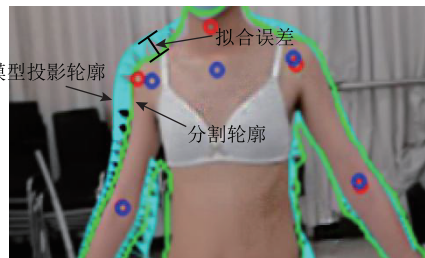
$$E_c(\beta, \theta; K, U_c) = \sum_{x \in S_c(\beta, \theta)} \min(\|x - U_c\|^2) \quad (2)$$

式中: $S_c(\beta, \theta)$ 为模型投影后的轮廓边界点集合; U_c 为分割图像的轮廓边界点集合; $\sum_{x \in S_c(\beta, \theta)} \min(\|x - U_c\|^2)$ 为模型投影轮廓边界点至分割轮廓边界点之间的最近距离之和。

文中主要是对形体进行估计, 因此在计算轮廓拟合误差时屏蔽头和四肢的轮廓, 只保留躯干的轮廓。轮廓分割计算如图4所示。图中红色和蓝色圆圈分别为求解模型的投影关节点和 OpenPose 估计的关节点。



(a) 轮廓分割效果



(b) 细节放大效果

图4 轮廓约束求解示意

Fig.4 Schematic diagram of solving contour constraints

1.5 主要实现过程

基于单幅 RGB 图像的三维人体建模实验步骤如下:

1) 导入 RGB 图像、OpenPose 参数、人体轮廓分割图以及 SMPL-X 模型;

2) 基于 HMR 深度学习模型, 导入 OpenPose 参数和 RGB 图像, 获取相机内参 K 和 SMPL-X 模型参数的预测值;

3) 对模型参数进行优化, 先固定相机参数, 再使用 Vposer 模型进行姿态估计, 最后调整 SMPL-X 模型的全局姿态参数和体型参数, 使能量函数值最小化。

1.6 实验结果与分析

在实验过程中, 改进的能量函数参数定义为:

$\lambda_{\alpha}=4.78, \lambda_{\beta}=5.0, \lambda_{\gamma}=5.0$, 发现当 λ_{γ} 值太小时, E_c 对 E 几乎没有影响, 太大会出现轮廓过拟合效果。为证明提出添加轮廓约束方法的有效性, 将添加人体分割轮廓约束项和无人体分割轮廓约束项

的模型拟合结果进行对比, 如图 5 所示。由图 5 可知, 每张图中左侧为添加人体分割轮廓约束项的模型拟合结果, 右侧为无人体分割轮廓约束项的模型拟合结果。

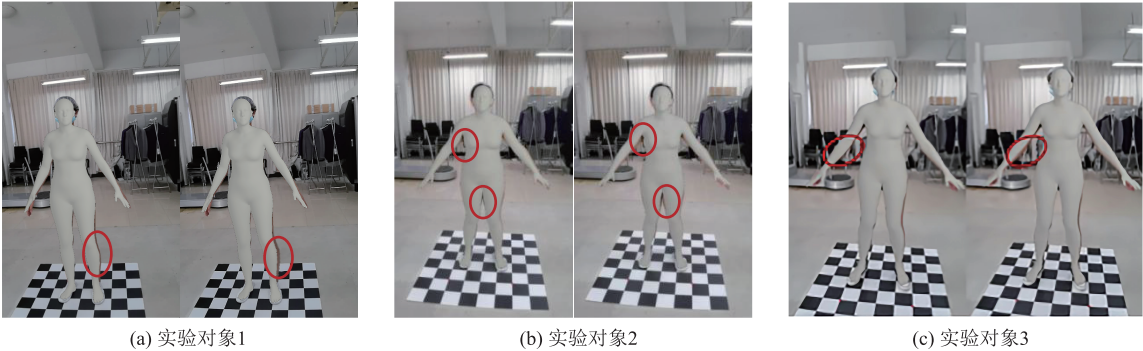


图 5 有无轮廓约束项的模型拟合结果对比

Fig. 5 Comparisons of model fitting results between with and without contour constraints

选取重建模型的细节差异处进行对比, 如图 5 红圈部分: (a) 中左小腿处, (b) 中腋下位置、裆部位置, (c) 中右胳膊肘处, 由此明显看出添加人体轮廓约束项的模型重建结果更加贴合人体。将有添加人体分割轮廓约束项和无人体分割轮廓约束项的模型拟合结果中关节平均误差和顶点平均误差数据进行对比, 具体见表 1。

表 1 误差对比

Tab. 1 Error comparison table

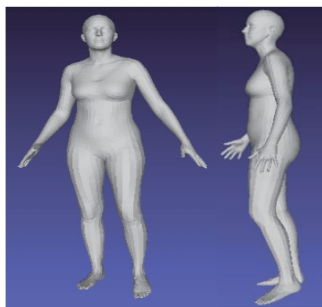
实验对象	算法	关节平均 误差/cm	顶点平均 误差/cm
1	添加人体分割轮廓约束项	0.61	1.15
	无人体分割轮廓约束项	0.76	1.38
2	添加人体分割轮廓约束项	0.62	1.18
	无人体分割轮廓约束项	0.74	1.37
3	添加人体分割轮廓约束项	0.60	1.20
	无人体分割轮廓约束项	0.72	1.35

从表 1 可以看出, 添加人体轮廓约束项的模型拟合结果关节误差和顶点误差均有所降低, 实验对象 1 的关节平均误差降低了 19.7%, 顶点平均误差

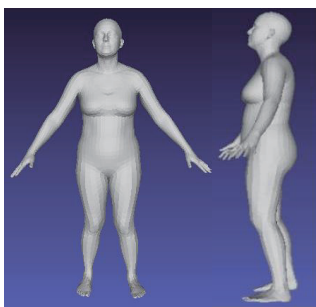
降低了 16.7%; 实验对象 2 的关节平均误差降低了 16.2%, 顶点平均误差降低了 13.9%; 实验对象 3 的关节平均误差降低了 16.7%, 顶点平均误差降低了 11.1%。整体而言, 添加人体分割轮廓约束项能够明显提升模型拟合的精度。

人体模型重建结果如图 6 所示, 其中, (a - c) 分别为 3 个实验对象的原始标准姿态图, (d - f) 分别为 (a - c) 对应的三维人体模型正侧面网格图。将模型与原始图像进行对比: 从正面三维网格图看, 重建出的第 1 个人体模型的身体方向发生一定的旋转; 3 个人体模型的头部均发生不同程度的上抬; 从侧面三维网格图看, 与原始侧面图像中 3 人标准站姿相比, 重建后的人体模型腿部均发生不同程度的弯曲, 而且实验对象 1 的左脚出现踮起现象, 与原始图像姿态有一定误差; 再对比手部姿态, 重建后的人体模型手部出现不同程度的上抬弯曲现象, 而且五指均呈张开状, 与原始图像姿态有误差。同时, 由于投影遮挡原因, 基于单张正面的人体图像进行三维人体重建设没有正确拟合出形体围度。

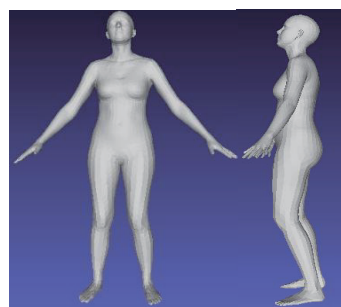




(d) 实验对象1的三维人体模型



(e) 实验对象2的三维人体模型



(f) 实验对象3的三维人体模型

图 6 人体模型重建结果

Fig. 6 Human model reconstruction results

2 基于 RGB 视频序列的三维人体建模

针对单幅图像建模中出现的问题,利用视频序列的帧间连贯性进行纠正。

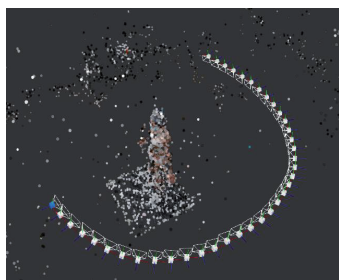
2.1 视频采集与视频解析

用手机环绕目标对象 90° 进行视频采集,采集时手机距离人体 1.5 m,拍摄时长约为 3 s。拍摄过

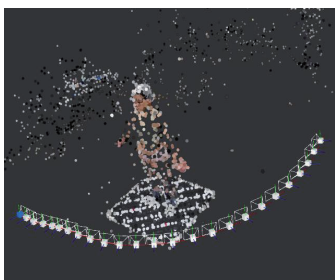
程中应保持手机处于直立状态,减少手机晃动引起的误差。然后通过 FFmpeg 进行视频解析,每个目标对象视频解析出 25 张图像,后续用于相机空间位置稀疏点阵的重建。

2.2 稀疏点阵重建

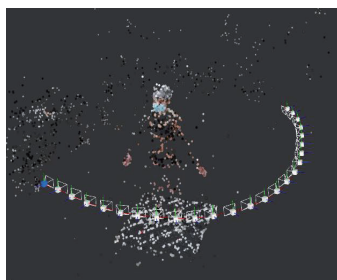
采用 OpenMVG^[14] 中增量式 SFM 完成相机空间位置稀疏点阵重建,具体步骤参考文献[14]。对 3 个目标图像稀疏点阵重建得到的结果如图 7 所示。



(a) 实验对象1



(b) 实验对象2



(c) 实验对象3

图 7 序列图像的相机空间位置稀疏点阵

Fig. 7 Sparse lattice of camera space position of sequence image

2.3 关键帧提取及轮廓分割

为了提高计算效率,对每个视频解析出的序列图像抽取 4 幅关键帧原始图像,抽取原则是相机主轴与人体正面的夹角分别大致对应 0°, 30°, 60°, 90°。

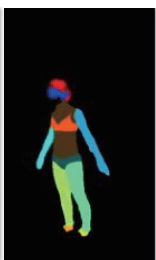
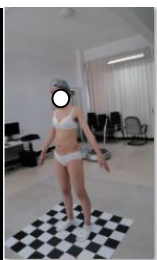
90°。然后采用基于 Graphonomy 深度学习的方法对原始图像进行人体轮廓分割。以目标对象 1 为例,关键帧抽取及对应的轮廓分割结果如图 8 所示。



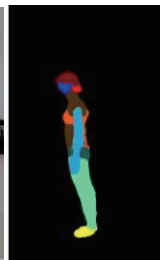
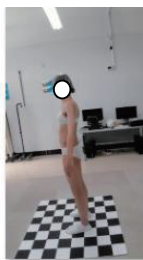
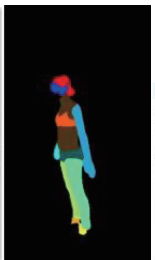
(a) 0°对应结果



(b) 30°对应结果



(c) 60°对应结果



(d) 90°对应结果

图 8 原始视频帧抽取及轮廓分割结果

Fig. 8 Results of original video frame extraction and contour segmentation

2.4 能量函数构造

在式(1)的基础上得到基于多视图三维重建的能量函数,即

$$E(\beta, \theta, \psi) = \sum_{i=0}^M E_{j_i}(\beta, \theta; K_i, J_{\text{est}_i}) + \lambda_{\alpha} E_{\alpha} +$$

$$\lambda_{\beta} E_{\beta} + \lambda_h E_h + \lambda_c \sum_{i=0}^M E_c(\beta, \theta; K_i, U_{c_i}) \quad (3)$$

式中: E_{j_i} 为 SMPL-X 模型的三维关节点坐标在第 i 个图像上的投影; J_{est_i} 为 OpenPose 预测的二维关节点; K_i 为第 i 个图像帧处相机的内外参; U_{c_i} 为第 i 个

图像帧的轮廓分割图像; M 为关键帧图像个数, 这里 $M=4$ 。

通过该算法进行若干次优化迭代最终重建出与目标对象姿态和体形相似的三维人体模型。

2.5 实验结果与分析

以目标对象 1 为例, SMPL-X 模型估计轮廓拟

合中间结果如图 9 所示。图 9 中绿色线条表示模型投影后轮廓与实际轮廓的误差欧式距离。SMPL-X 模型估计的相机图像反投影标记结果如图 10 所示。图中红色点为 OpenPose 估计的关节点, 绿色点为 SMPL-X 拟合模型的关节投影点, 白色网格为 SMPL-X 拟合模型的网格投影。

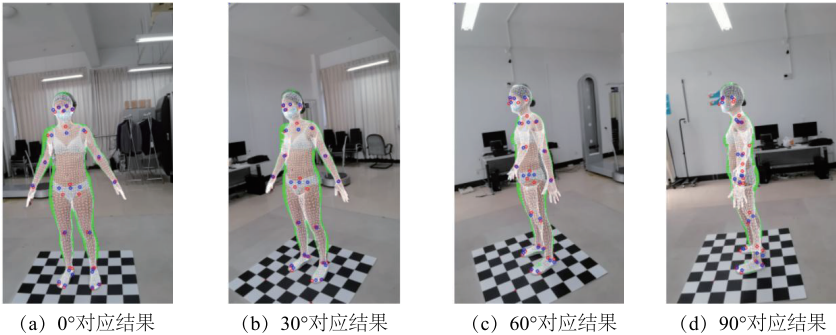


图 9 SMPL-X 模型估计的轮廓拟合中间结果

Fig. 9 Intermediate results of contour fitting estimated by SMPL-X model

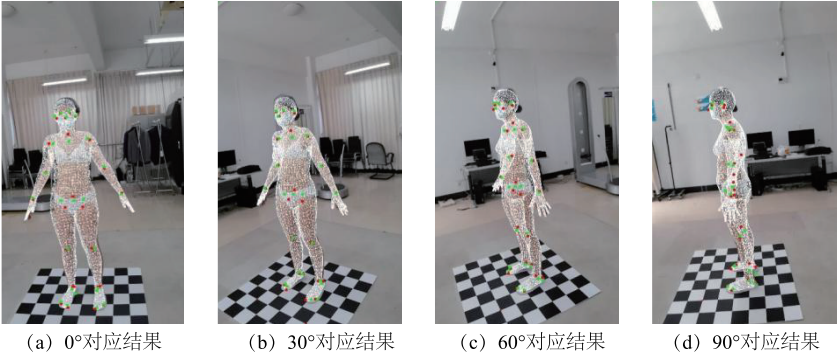


图 10 SMPL-X 模型估计的相机图像反投影结果

Fig. 10 Camera image back-projection results estimated by SMPL-X

多视图与单视图建模结果对比如图 11 所示。为了证明利用视频序列的帧间连贯性可以得到更好的三维建模结果, 将 3 个目标对象的最终人体模型正侧面结果与单视图的建模结果相比较。

从图 11(b) 红框部分可以看出单视图中重建的三维人体模型出现身体扭转、四肢弯曲、头部上抬

的现象, 而图 11(a) 多视图中重建模型人体的姿态与形体是一致的, 身体没有发生旋转, 四肢没有出现弯曲, 头部也没有出现上抬问题, 更好地刻画了人体体态形貌特征。结果证明利用视频序列的帧间连贯性来纠正单幅图像建模出现的问题是有效的。

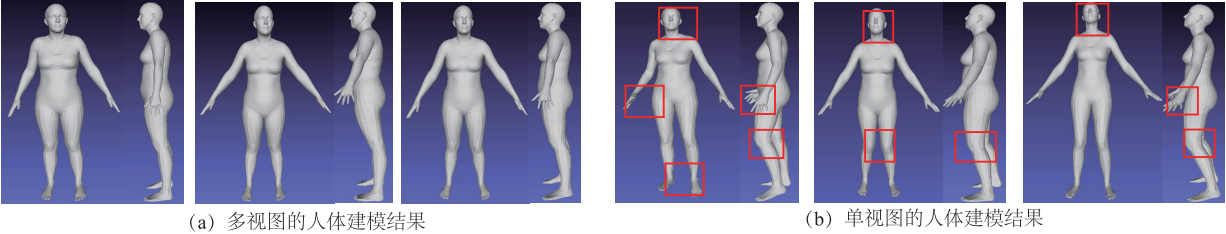


图 11 多视图与单视图建模结果对比

Fig. 11 Comparisons of modeling results between multi-view and single-view

3 结 语

文中将基于深度学习和优化的人体建模算法结合重建人体三维模型, 利用深度学习算法 HMR

提供的结果作为先验, 来缓解优化算法对 SMPL-X 模型参数初始值敏感的问题。并对 SMPL-X 算法的能量函数进行改进, 主要通过添加轮廓约束项来降低人体模型的拟合误差, 经过若干次优化迭代最

终获得与图像中人体姿态和体型相似的人体模型。与无添加轮廓约束的模型拟合结果相比,文中方法得到的结果更加贴合人体,能够明显提升模型拟合的精度。但由于单视图图像缺乏深度信息,所建的三维人体模型与原始图像中设定的标准站姿有一定误差,如出现身体扭转、腿部弯曲、头部上抬等现象。

基于此,文中利用视频序列的帧间连贯性对视频帧中人物的三维姿态进行约束,减小出现上述误差的可能性。由三维建模结果可知,基于视频序列的帧间连贯性得到的三维人体模型精确度更高,与原始图像中人体的姿态和体形基本一致。不足之处在于文中只是对简单姿势下的人体进行模型重建,未来可以考虑加入复杂的动作和有深度的动作(即Z轴方向的人体姿势有深度)来进行三维人体重建。

参考文献:

- [1] 顾佳雯,周晓琪,王军. 三维人体建模方法研究与应用[J]. 纺织科技进展, 2021(9): 15-18.
GU Jiawen, ZHOU Xiaopi, WANG Jun. Research and application of 3D human body modeling[J]. Progress in Textile Science and Technology, 2021(9): 15-18. (in Chinese)
- [2] 王巧丽,徐增波,田丙强. 虚拟试衣中个性化三维人体建模的研究进展[J]. 纺织导报, 2020(7): 79-82.
WANG Qiaoli, XU Zengbo, TIAN Bingqiang. Research progress of personalized 3D human body modeling in virtual fitting[J]. China Textile Leader, 2020(7): 79-82. (in Chinese)
- [3] 劳秀霞,柯略. 基于3Dmax的三维动画角色建模探析[J]. 科技传播, 2011, 3(4): 236-237.
LAO Xiuxia, KE Lue. Analysis of 3D animation character modeling based on 3Dmax[J]. Public Communication of Science and Technology, 2011, 3(4): 236-237. (in Chinese)
- [4] 吕珍,谢倩,祝双武,等. OpenGL导入3DS模型模拟服装试穿[J]. 西安工程大学学报, 2014, 28(4): 440-444.
LYU Zhen, XIE Qian, ZHU Shuangwu, et al. Importing 3DS models to simulate garment fitting by OpenGL[J]. Journal of Xi'an Polytechnic University, 2014, 28(4): 440-444. (in Chinese)
- [5] 刘雁,耿兆丰. 三维人体及服装建模方法分析与实现[J]. 计算机应用与软件, 2004, 21(10): 47-48, 95.
LIU Yan, GENG Zhaofeng. Analysis and realization of 3D mannequin and garment modeling[J]. Computer Applications and Software, 2004, 21(10): 47-48, 95. (in

Chinese)

- [6] 韩凯,庞宗强,王龙,等. 基于深度扫描仪的高辨识度三维人体模型重建方法[J]. 图学学报, 2015, 36(4): 503-510.
HAN Kai, PANG Zongqiang, WANG Long, et al. High identification 3D human body model reconstruction method based on the depth scanner[J]. Journal of Graphics, 2015, 36(4): 503-510. (in Chinese)
- [7] GUAN P, WEISS A, BÄLAN A O, et al. Estimating human shape and pose from a single image[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto:IEEE, 2010: 1381-1388.
- [8] BOGO F, KANAZAWA A, LASSNER C, et al. Keep it SMPL: automatic estimation of 3D human pose and shape from a single image[M]//Computer Vision – ECCV 2016. Berlin: Springer International Publishing, 2016: 561-578.
- [9] 张广翩,计忠平. 基于二维点云图的三维人体建模方法[J]. 计算机工程与应用, 2020, 56(19): 205-215.
ZHANG Guangpian, JI Zhongping. Method of 3D human body modeling based on 2D point cloud image[J]. Computer Engineering and Applications, 2020, 56(19): 205-215. (in Chinese)
- [10] CAO Z, SIMON T, WEI S H, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 1302-1310.
- [11] GONG K, GAO Y M, LIANG X D, et al. Graphonomy: universal human parsing via graph transfer learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 7442-7451.
- [12] KANAZAWA A, BLACK M J, JACOBS D W, et al. End-to-end recovery of human shape and pose[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7122-7131.
- [13] PAVLAKOS G, CHOUTAS V, GHORBANI N, et al. Expressive body capture: 3D hands, face, and body from a single image[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 10967-10977.
- [14] MOULON P, MONASSE P, PERROT R, et al. OpenMVG: open multiple view geometry[M]//Reproducible Research in Pattern Recognition. Berlin: Springer International Publishing, 2017: 60-74.

(责任编辑:张雪)