PERSONALIZED AVATARS CREATION BASED ON REALISTIC 3D FACE RECONSTRUCTION

Yueming Ding and Pik Yin Mok*
School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

Personalized 3D human avatars hold great potential in the digital era, and this also enables interesting fashion applications such as virtual try-on, virtual fashion shows, and immersive online shopping experiences. Nevertheless, personalized avatars are seldom used in the fashion industry due to the requirements of specialized equipment and the tedious and computationally demanding nature of traditional personalized face modeling processes. In this paper, a novel method is presented for creating personalized fashion avatars by reconstructing a realistic 3D face from a single image. The proposed method involves transferring customized 3D facial shape and appearance, reconstructed from a single face image, onto full-body avatars with varying topologies suitable for different fashion applications. Through a comprehensive comparison with state-of-the-art face reconstruction methods, the proposed method has demonstrated high realism and richer facial details. This technology has already been applied in an augmented reality (AR) mobile application, enabling users to engage in virtual try-on experience of fashion.

KEYWORDS

3D Reconstruction, Human Avatar, Virtual Try-on, 3DMM

1. INTRODUCTION

With the advancement of Internet and smart mobile devices, virtual reality technology soars in the fashion industry, with personalized fashion avatars playing a pivotal role in engaging applications, such as virtual trying-on (Meng et al., 2010), virtual fashion recommendation (Liu et al., 2014; Zhou et al., 2019), and immersive online shopping (Waltemate et al., 2018). According to Waltemate et al. (2018), personalization of avatar significantly increases self-identification and body ownership to the users (see Figure **1**). Nevertheless, most virtual try-on systems do not have personalized avatars. One of the reasons is that personalized avatars require professional equipment, such as camera arrays, and large computing resources, resulting in high costs of customization. In recent years, 3D morphable model (3DMM) provides an economic solution to personalize human face and body shape by reconstruction methods. 3DMM is a type of statistical model of shape and texture learned from 3D scans by principal components analysis (PCA). It utilizes a set of parameters to generate a 3D shape or texture by linearly weighting the principal components. A number of studies use 3DMM to reconstruct personalized human body (Bogo et al., 2016; Loper et al., 2015; Pavlakos et al., 2019).

Another key reason impeding the widespread use of personalized avatars is the diverse model topology. The topology of face models varies in different applications or datasets, making it difficult to implement 3DMM reconstruction across different fashion applications. In this regard, some methods were proposed to combine morphable face model with human body model. For example, Fang et al. (2021) stitched Basel face model (BFM) (Blanz & Vetter, 1999) to their human avatars to realize personalized shape and texture in virtual try-on application. TF-Flame (Li et al., 2017) transferred the parametric texture model of BFM (Blanz & Vetter, 1999) to the texture uv-map of its full head model. However, these models are far from realistic, for example, the BFM model was learned from only 200 identities. The scarce of 3D dataset and the Gaussian nature of statistic model limited the presentation ability of the statistic texture model, which cannot satisfy requirements for personalized appearance and makeup (Liu et al., 2014). In this paper, a new method is proposed to transfer a customized facial shape and appearance reconstructed from single face images to full-body avatars of different topologies for fashion applications.

*Corresponding author: tracy.mok@polyu.edu.hk

Figure 1. The try-on effect on (a) an avatar without texture, (b) an avatar without personalized texture, and (c) a personalized avatar

2. RELATED WORK

3D human body reconstruction: The research on 3D human reconstruction has mainly concentrated on reconstructing human poses and body shapes, with the face region often being neglected. For instance, in 3D human body reconstruction, a skinned multi-person linear model (SMPL) (Bogo et al., 2016; Loper et al., 2015) was put forward as a statistical full body animatable model to enable fast pose and shape reconstruction for computer games and animation applications. However, SMPL model only has coarse facial and body shape representation. Subsequently, Romero et al. (2022) and Pavlakos et al. (2019) augmented the hand and face model with additional joints and animation for SMPL. Joo et al. (2018) introduced a fusion model that merges the body from SMPL and head from the FaceWarehouse model (Cao et al., 2013) to acquire a comprehensive body model with more facial expression and animation. These models increased their amination ability, yet they still have limited shape and appearance.

3D face reconstruction: For 3D face reconstruction, 3DMM-based methods are widely employed using RGB-D data (Bao et al., 2021), single or multiple images (Bai et al., 2020). Blanz and Vetter (1999) proposed an analysis-by-synthesis reconstruction method using PCA-based face model. Nevertheless, PCA facial texture is often of low resolution and with limited expressions because of its Gaussian nature and small sample size. Feng et al. (2021) presented a self-supervised DECA method to reconstruct detailed head shape based on TF-FLAME model (Li et al., 2017) using in-the-wild images.

For virtual try-on of fashion products, avatars with personalized appearances are definitely beneficial. Considering the difficulty of reconstructing human avatars with highly individualized identities, Yuan et al. (2013) suggested a mixed reality virtual clothes try-on system, which combined the rendered reconstructed body model with the real face image of the user. Nonetheless, the result is not lifelike and presents inconsistent between the face and body. Some scholars proposed a pixel-aligned reconstruction of body with clothes from 2D images through deep neural networks in voxel space, such as PIFu (Saito et al., 2019) and PIFuHD (Saito et al., 2020). However, these voxel-based methods only can output low resolution models due to the high demands on storage and computing resources. In addition, these models do not have uniform topology, models reconstructed by PIFu are not animable and thus not suitable for fashion applications. Recently, for improving the personalization of human avatars in virtual try-on system, Fang et al. (2021) proposed a weak registration method to align a customized face model to a body model, similar to wearing a `mask'. This method can partly provide customized fashion avatar for virtual try-on system which only possess a generic human avatar. However, this method still has a few drawbacks, including distortion of the face model with noisy reference landmarks, only low resolution and less individual identities. In summary, a new method is needed for personalized fashion avatar creation that is ready for customized facial shape and appearance and adaptable to a full-body avatar with various topologies for plug-and-play fashion applications.

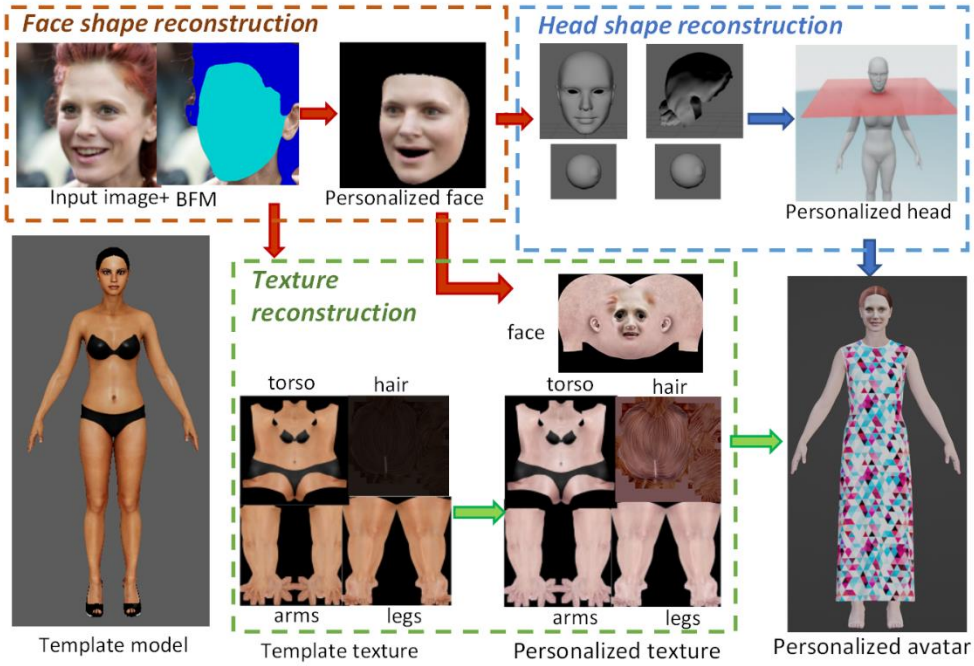

Figure 2. Overview of the personalized avatar reconstruction

3. METHOD

Figure **2** shows the outline of the proposed method for the creation of personalized fashion avatars based on realistic face reconstruction. The creation process of personalized fashion avatars includes four parts: (1) Reconstructing face shape based on input face image; (2) Transferring face shape to a full head model of any fashion avatar which can have varied topology; (3) Reconstructing face texture and aligning skin tone and color from face to full-body texture; (4) aligning and stitching the customized head model to body model.

3.1 Face Shape Reconstruction

3DMM In this work, we employ the Basel Face model (BFM), one of 3DMM face models in face reconstruction stage, to reconstruct the initial face shape and appearance. BFM generates a new 3D face shape $S_f \in R^{3N}$ or face texture $T_f \in R^{3N}$ of N mesh vertices by adding to a mean face shape S_{mean_f} or a mean face texture T_{mean_f} a linear combination of weight parameters α_{S_f}, α_{T_f} and principle parameters P_{S_f}, P_{T_f} respectively, which are formulated as following,

$$S_f = S_{mean_f} + \alpha_{S_f} \cdot P_{S_f}$$
$$T_f = T_{mean_f} + \alpha_{T_f} \cdot P_{T_f}. \quad (1)$$

Preprocessing: Prior to the personalized face reconstruction, face landmark detection and region segmentation are processed. Afterwards, the landmarks of the 3D head model must be estimated. This is because in most fashion avatars, facial landmarks are absent. Thus, the landmarks of the 3D head model must be estimated for following processes.

Landmarks detection. For aligning the face model to face image and initializing face reconstruction, it is necessary to detect facial landmarks from input face images. Because our work requires high accuracy of landmarks localization, we employ Dlib (Kazemi & Sullivan, 2014) to detect 68 points landmarks for this work.

Face segmentation. A face parsing network MaskGAN (Lee et al., 2020) pretrained on CelebAHQ dataset is then used to segment a face image into face skin region and hair region.

Landmarks estimation for 3D head mesh. Most 3D fashion avatars lack annotated facial, however landmarks are necessary for our subsequent processing, such as shape transfer and alignment. Therefore, we have to estimate 3D facial landmarks for any input fashion avatars. Compared to detect landmarks from 3D data directly, detecting landmarks from 2D images has more thoroughly explored. We take advantage of 2D landmark estimation to assist 3D landmark detection. To do so, we render the 3D avatar into an image and use 2D landmarks detector to back-project these landmarks to 3D mesh.

Reconstruction of face model: Reconstructing 3D face from face images through 3DMM can be decomposed into facial shape, texture and pose estimation. As discussed, BFM is employed to reconstruct 3D face from input face image, and orthogonal projection $P(\cdot)$ is assumed as our camera model. In face reconstruction, a 3D face mesh S_f is generated from an input image I first. Next, a face image I' is rendered from the generated face mesh S_f. The reconstruction of face can be considered as optimizing the μ parameters of 3DMM face model including shape α_{S_f} and texture parameters α_{T_f}, as well as the pose parameters $[R, t, s]$ consisting of rotation, translation and scaling factors.

$$\mu = \left[\alpha_{S_f}, \alpha_{T_f}, R, t, s\right] \tag{2}$$

The face reconstruction is optimized by minimizing the loss functions defined below.

Landmark loss measures the difference between the landmarks v^{lmk} of an input image I and the landmarks v'^{lmk} extracted from the rendered image I'

$$L_{lmk}(\mu) = \frac{1}{N_{lmk}} \sum_{k=1}^{N_{lmk}} \left\| v_i^{lmk} - P\left(v'^{lmk}_i\right) \right\|^2, \tag{3}$$

where $P(\cdot)$ represents the orthographic projection operation, v'^{lmk}_i is a landmark point i detected from the input image I, v_i^{lmk} is the corresponding landmark on the 3d face S_f, and N_{lmk} is the total number of detected landmark points.

Contour edge loss is used to fine-tune face shape reconstruction, since landmarks can only provide coarse shape information. Contour loss measures the distance between the points on the detected facial edges of the 2D image and the points on the projected occluding contours from the 3D face S_f, and is defined as follows,

$$L_{edge}(\mu) = \frac{1}{N_C} \sum_{j \in C} \left\| v_i^{cont} - P\left(v'^{edge}_i\right) \right\|^2, \tag{4}$$

where v_i^{cont} is vertex j on the occluding contour of the predicted 3D face model, $j \in C$ and C is the number of vertices on the contour. v'^{edge}_i denotes the pixel location on the detected occluding contours of the input image that is the closest to vertex j. The contour loss is mainly used to fine-tune the reconstructed 3D shape (Bas et al., 2016).

Photometric loss measures the pixel-wise RGB color distance between the input image I and the rendered image I' as follows,

$$L_{photo}(\mu) = \frac{1}{N_R} \sum_{(x,y) \in R} \left\| I_{(x,y)} - I'_{(x,y)} \right\|^2, \tag{5}$$

where the (x, y) represents a pixel in the face region R of an image, and N_R represents the number of pixels in face region, $I_{(x,y)}$ and $I'_{(x,y)}$ present the RGB value of a pixel on input image and the rendered image. This loss is mainly used in shape fine-tune and texture reconstruction.

The total loss function is defined as weighted landmark loss, edge loss and photometric loss as follows

$$L(\mu) = \omega_{lmk} \cdot L_{lmk}(\mu) + \omega_{edge} \cdot L_{edge}(\mu) + \omega_{photo} \cdot L_{photo}(\mu). \tag{6}$$

3.2 Head Shape Reconstruction Based on Face Model

Transfer face shape to head model: We use PCA parameters and linear least squares method to transfer the face shape of S_f to an arbitrary head model S_h. First of all, we random the face shape parameters of S_f and use nonrigid ICP (NICP) method (Amberg et al., 2007) to generate paired synthetic head shape data. Then, the PCA method is used to generate principal components of the head model. The head shape model is defined as:

$$S_h = S_{mean_h} + \alpha_{S_h} \cdot P_{S_h} \tag{7}$$

where S_{mean_h} is the mean head shape and α_{S_h} denotes the parameters of the head principal components P_{S_h}. We mainly consider shape and texture reconstructions by setting the expression parameters of the face model to zero. Transferring face shape S_f of eq. (1) to head model S_h of

$$\begin{matrix} S_f = S_{mean_f} + \alpha_{S_f} \cdot P_{S_f} \\ T_f = T_{mean_f} + \alpha_{T_f} \cdot P_{T_f} \end{matrix}.$$

eq. ($S_h = S_{mean_h} + \alpha_{S_h} \cdot P_{S_h}$ (7) can be considered as a regression problem. After shape registration of face and head models, the head and face shape parameters are represented as matrices $C_h \in R^{n_h \times r}$ and $C_f \in R^{n_f \times r}$, respectively. The face-to-head transformation is solved as a regression matrix (Ploumpis et al., 2020),

$$T_{h,f} = C_h C_f^T C_f C_f^{T-1}, \tag{8}$$

where the $C_f^T C_f C_f^{T-1}$ is the pseudo-inverse of C_f. By solving the regression matrix, the head shape parameters can be calculated by

$$P_{S_h} = T_{h,f} P_{S_f}. \tag{9}$$

Reconstruction of complete 3D full head model: In most 3D full-head models, eyeballs and hairs are usually isolated models in addition to the head shape model (Figure **2**). These models are transformed according to the customized head shape model. To do so, the joint of the head is considered as the point of origin, and the center coordinates $v_{c\{.\}}$ of each eyeball and hair model in the template are transferred to the head local coordinates. When adjusting the size of the isolated models, the scale factors are calculated in three directions by proportion, considering the reference points $v_{i,j} \in \varphi$ estimated head model and $v_{i_0,j_0} \in \varphi_{ori}$ template. Landmarks of eyes $\varphi_{\{eyes\}}$ and vertices on bounding box of cranium $\varphi_{\{hair\}}$ are used as reference points, the isolated scale factors for eyes and hair are formulated as,

$$H_{x,y,z\{eyes,hair\}} = \frac{\left\| \max_{i,j \in \varphi_{\{eyes,hair\}}} v_{i(x,y,z)} - v_{j(x,y,z)} \right\|}{\left\| \max_{i_0,j_0 \in \varphi_{\{eyes,hair\}}} v_{i0(x,y,z)} - v_{j0(x,y,z)} \right\|}. \tag{10}$$

To adjust model location, landmarks as well as the center of the bounding box of hair model are used to compute the translation and rotation matrix,

$$t_{\{eyes,hair\}} = \left[v_{C\{eyes,hair\}} - v_{C0\{eyes,hair\}} \right]^T. \tag{11}$$

The transformation of isolated models is calculated as,

$$M_{eyes,hair} = \begin{bmatrix} H_x & 0 & 0 \\ 0 & H_y & 0 \\ 0 & 0 & H_z \end{bmatrix} R \left(\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \right) + t_{eyes,hair}. \tag{12}$$

3.3 Texture reconstruction

The third step of the fashion avatar customization pipeline is texture reconstruction, which covers face texture and body model texture. Parametric face models usually have low resolution texture information. To improve this, we combine the texture sampled from input face images with the reconstructed BFM albedo map as supplement in face texture reconstruction (see Figure 2). The reconstructed head mesh can be unwrapped into UV-space using the cylinder unwarp, then the texture of BFM can be mapped to UV-space through barycentric coordinates. We obtain a face segmentation result from the input face image using MaskGAN, a pretrained face parsing model (Lee et al., 2020). The combination of sampled texture and BFM texture is processed through a segmentation mask with a gaussian kernel with 30-pixel radius.

The paper also presents a method to customize the texture of body and hair by using the color features extracted from face skin and hair as reference. For hair reconstruction, because hair reconstruction is complicated and time-consuming, we mainly focus on hair color reconstruction. In terms of body texture, the skin color is the main attribute for human body texture. Thus, in this process, we first extract color features from the skin and hair segmentation of face image, and color features from template body textures and hair textures. Then, we transfer these features on CIELAB color space. These textures transformation allows a more realistic representation of the hair and full-body skin color of the user in practice (see Figure **2**).

3.4 Align and Stitch Head to Fashion Avatar

Previous research on reconstructing personalized fashion avatars proposed to replace the face part of a full body avatar with a 3DMM face model through cutting and stitching. For example, Fang et al. (2021) used the Bezier curve to stitch the BFM face model with the head of fashion avatar. Unfortunately, this will lead to the compression and distortion of the face model along the depth direction due to the degradation of depth

information and inaccurate landmarks. Moreover, the texture of the mesh model in the stitching regions is not smooth and consistent.

In this work, the head is separated from the fashion avatar by cutting along the plane on the neck. After the personalized head is reconstructed, the rigid transformation matrix is calculated using the estimated landmarks, and the head is aligned to the fashion body using rigid transformation. The personalized head is then stitched to the fashion body along the boundary point set.

4. EXPERIMENTS AND RESULT DISCUSSION

Experiments were conducted on a PC with Intel i7 CPU and NVIDIA Titan GPU, and the programming environment is python. We use the differentiable rendered of Pytorch3D to render the reconstructed model.

Dataset: The CelebA dataset (Liu et al., 2015) was used as a face image dataset for simulating application scenarios in our experiment. This dataset contains 10,177 identities and 202,599 face images. The BFM (Blanz & Vetter, 1999) with 35709 vertices was used to reconstruct 3D faces from images. For full-body avatar customization, we used a template body model with 40763 vertices, in which the head region contains 8187 vertices and has different topology than the BFM. In fact, an arbitrary body model can be used in the proposed method.

Face only comparison: For face only customization, the analysis-by-synthesis method from BFM was used as baseline for comparison with the proposed method, and we also compare to other state-of-art methods, including DECA method (Feng et al., 2021) and weak-registration algorithm (Fang et al., 2021).

Quantitative analysis: We use perceptual similarity (Kim et al., 2021) to evaluate the effectiveness of different face reconstruction methods. Perceptual similarity is a metric using the cosine similarity of embedding features extracted from VGGface (Parkhi et al., 2015), a pretrained face recognition model. We conducted experiment on a subset of CelebA dataset (Liu et al., 2015). The results in Table 1 indicate that our method has the lowest mean cosine distance between the reconstructed personal face images with the inputs, compared to other state-of-the-art methods.

Metric. VGGface (Parkhi et al., 2015) is a deep learning facial recognition model developed by the Visual Geometry Group (VGG) at Oxford University. It uses a convolutional neural network (CNN) to extract features from a face image and then compares them to a reference image to measure the perceptual similarity. The cosine distance is used to measure the similarity among the VGGFace-features $F(\cdot)$ extracted from ground true image I and images I' synthesized from different methods, which is formulated as,

$$Cosine\ similarity(I, I') = \frac{F(I)\cdot F(I')}{\|F(I)\|\times\|F(I')\|}. \tag{13}$$

Table 1. Quantitative analysis by perceptual similarity

	BFM (Blanz & Vetter, 1999)	DECA (Feng et al., 2021)	Weak-rigid registration (Fang et al., 2021)	Our method
Subset of CelebA (Liu et al., 2015)	0.57±0.01	0.39±0.12	0.59±0.16	0.37±0.10

Qualitative analysis: The qualitative comparison is shown Figure **3**. As shown, the texture from the method based on BFM has less details, such as in the eyes, and has low resolution due to the nature of PCA-based texture models. This can lead to low self-identity of the facial representations. The DECA model (Feng et al., 2021) realized full head reconstruction based on TF-FLAME morphable face model but without complicated texture. The coarse version of this model can realize head shape reconstruction with sampled texture. The reconstructions resulting from the weak-registration algorithm proposed by Fang et al. (2021) are depicted in column (d). This method enables the fusion of the BFM results with a coarse head model. However, the face model is twisted in the z-axis during the registration process, resulting in an inconsistency between the face and head regions. In addition, the texture is based on BFM model, suffering from the same limitation of BFM. The reconstructions of the proposed method are shown in column (e) of Figure **3**. As shown, our reconstruction results possess high self-identity comparing to BFM, and the connection between face and head is smoother than that of .Fang et al. (2021). In sum, the proposed method has better face reconstruction shape and appearance and is suitable for fashion avatar personalization.

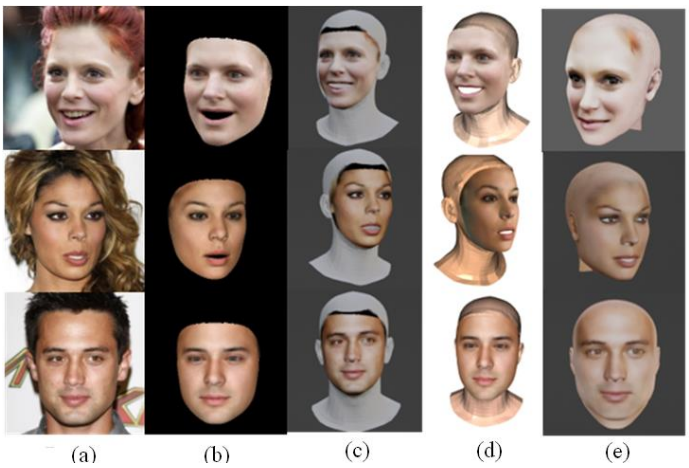

Figure 3. Comparison with different face reconstruction methods: (a) input images, (b)results by methods of BFM (Blanz & Vetter, 1999), (c) DECA (Feng et al., 2021), (d) Weak-rigid registration (Fang et al., 2021) and (e) ours

Personalized avatar comparison: For personalized full-body avatar, we only conducted qualitative evaluation in comparison to other methods including SMPL-X (Pavlakos et al., 2019) and weak-rigid registration method (Fang et al., 2021). As illustrated in Figure **4**, SMPL-X can reconstruct human models from single images with limited facial details and without full-body texture. The weak-registration method can reconstruct personalized human model; however, the resulting face is twisted and the texture of the face and body are inconsistent. In contrast, our method can reconstruct personalized human avatars with more details and consistent texture and shape. More examples are given in Figure **5**, demonstrating that the proposed method can create avatars with different skin colors and facial shapes. We have applied this personalized avatar reconstruction method to a mobile AR app, and a video is shown in the supplementary information.

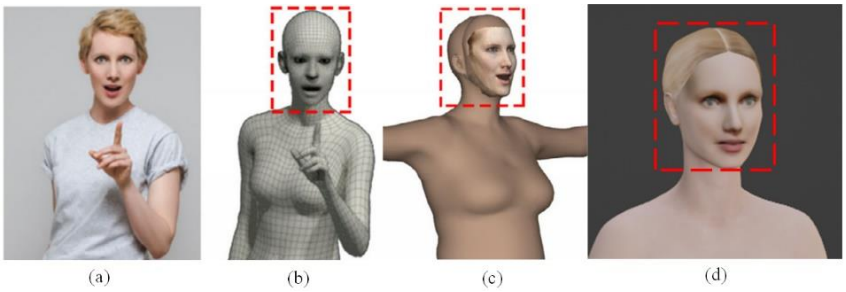

Figure 4. Comparison with different full-body model customization methods: (a) input image, results of methods by (b) SMPL-X (Pavlakos et al., 2019), (c) weak-registration(Pavlakos et al., 2019), and (d) ours

5. CONCLUSION

This paper develops a new reconstruction method for personalizing fashion avatars that can be used on different platforms, of which avatars have different topology and texture, for virtual try-on application. Compared with pure 3DMM-based human personalization methods, our personalization method for human model personalization can be used on models of arbitrary topology. In addition, this method can also reconstruct facial appearance with higher resolutions and self-identity. When comparing with the method which replace face mesh with rigid registration, our method not only presents high self-identity on face appearance, but also presents consistence on texture and shape. The current method has a few limitations, for instance it only reconstructs static model and the performance will be affected if the avatar has a head mesh with extreme sparse vertices. In the future, we will extend this method to avatars with animation and personalized hair shape. On the other hand, we will try to extend this method to avatars with poor mesh conditions.

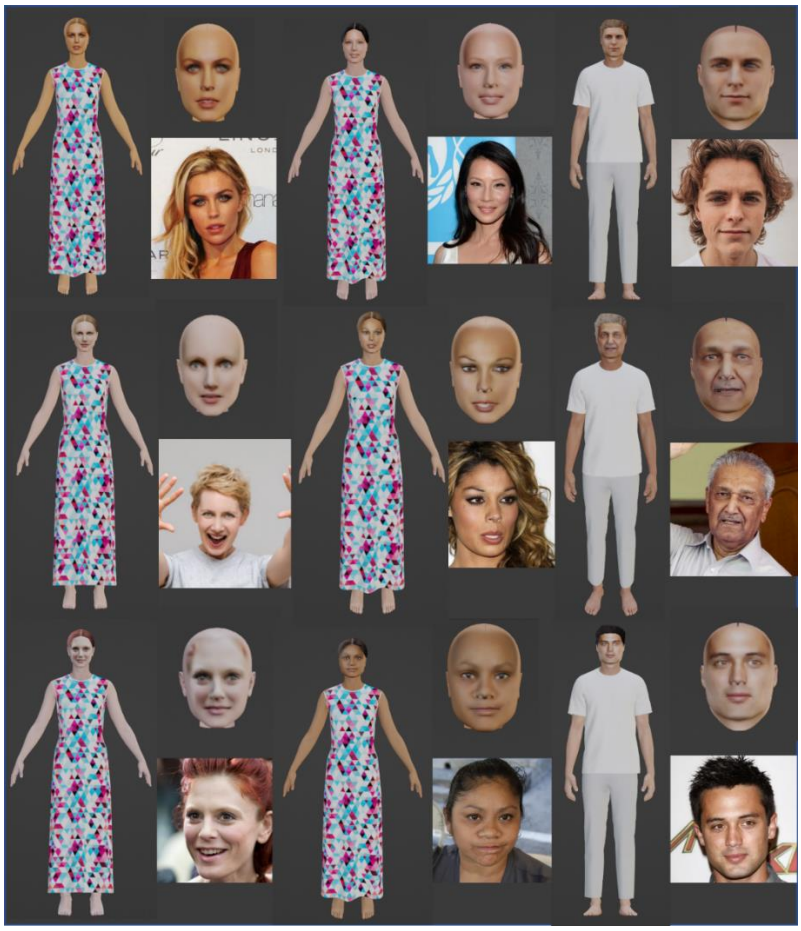

Figure 5. Personalized avatar in try-on application

ACKNOWLEDGEMENT

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant Number 152112/19E).

REFERENCES

Amberg, B., Romdhani, S., & Vetter, T. (2007). Optimal step nonrigid ICP algorithms for surface registration. 2007 IEEE conference on computer vision and pattern recognition.

Bai, Z., Cui, Z., Rahim, J. A., Liu, X., & Tan, P. (2020). Deep facial non-rigid multi-view stereo. Proceedings of the ieee/cvf conference on computer vision and pattern recognition.

Bao, L., Lin, X., Chen, Y., Zhang, H., Wang, S., Zhe, X., Kang, D., Huang, H., Jiang, X., & Wang, J. (2021). High-Fidelity 3D Digital Human Head Creation from RGB-D Selfies. ACM Transactions on Graphics (TOG), 41(1), 1-21.

Bas, A., Smith, W. A., Bolkart, T., & Wuhrer, S. (2016). Fitting a 3D morphable model to edges: A comparison between hard and soft correspondences. Asian Conference on Computer Vision.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. Proceedings of the 26th annual conference on Computer graphics and interactive techniques.

Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., & Black, M. J. (2016). Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14.

Cao, C., Weng, Y., Zhou, S., Tong, Y., & Zhou, K. (2013). Facewarehouse: A 3d facial expression database for visual computing. IEEE transactions on visualization and computer graphics, 20(3), 413-425.

Fang, N., Qiu, L., Zhang, S., Wang, Z., Wang, Y., Gu, Y., & Tan, J. (2021). A Modeling Method for the Human Body Model with Facial Morphology. Computer-Aided Design, 141, 103106.

Feng, Y., Feng, H., Black, M. J., & Bolkart, T. (2021). Learning an animatable detailed 3D face model from in-the-wild images. ACM Transactions on Graphics (TOG), 40(4), 1-13.

Joo, H., Simon, T., & Sheikh, Y. (2018). Total capture: A 3d deformation model for tracking faces, hands, and bodies. Proceedings of the IEEE conference on computer vision and pattern recognition.

Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. Proceedings of the IEEE conference on computer vision and pattern recognition.

Kim, J., Yang, J., & Tong, X. (2021). Learning high-fidelity face texture completion without complete face texture. Proceedings of the IEEE/CVF International Conference on Computer Vision.

Lee, C.-H., Liu, Z., Wu, L., & Luo, P. (2020). Maskgan: Towards diverse and interactive facial image manipulation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Li, T., Bolkart, T., Black, M. J., Li, H., & Romero, J. (2017). Learning a model of facial shape and expression from 4D scans. ACM Trans. Graph., 36(6), 194:191-194:117.

Liu, L., Xing, J., Liu, S., Xu, H., Zhou, X., & Yan, S. (2014). Wow! you are so beautiful today! ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 1-22.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. Proceedings of the IEEE international conference on computer vision.

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. ACM Transactions on Graphics (TOG), 34(6), 1-16.

Meng, Y., Mok, P. Y., & Jin, X. (2010). Interactive virtual try-on clothing design systems. Computer-Aided Design, 42(4), 310-321.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Ploumpis, S., Ververas, E., O'Sullivan, E., Moschoglou, S., Wang, H., Pears, N., Smith, W. A., Gecer, B., & Zafeiriou, S. (2020). Towards a complete 3D morphable model of the human head. IEEE transactions on pattern analysis and machine intelligence, 43(11), 4142-4160.

Romero, J., Tzionas, D., & Black, M. J. (2022). Embodied hands: Modeling and capturing hands and bodies together. arXiv preprint arXiv:2201.02610.

Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. Proceedings of the IEEE/CVF international conference on computer vision.

Saito, S., Simon, T., Saragih, J., & Joo, H. (2020). Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. IEEE transactions on visualization and computer graphics, 24(4), 1643-1652.

Yuan, M., Khan, I. R., Farbiz, F., Yao, S., Niswar, A., & Foo, M.-H. (2013). A mixed reality virtual clothes try-on system. IEEE Transactions on Multimedia, 15(8), 1958-1968.

Zhou, W., Mok, P., Zhou, Y., Zhou, Y., Shen, J., Qu, Q., & Chau, K. (2019). Fashion recommendations through cross-media information retrieval. Journal of Visual Communication and Image Representation, 61, 112-120.