

From prediction to measurement, an efficient method for digital human model obtainment

Moyu Wang* and Qingping Yang

College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge, UK

Received: 31 July 2023 / Accepted: 1 November 2023

Abstract. Digital human has been increasingly used in industry, for example in Metaverse which has been a popular topic in recent years. The existing method of obtaining digital human models are either expensive or lack of accuracy. In this paper, we discuss a novel method to reconstruct a 3D human model from 2D images captured by a monocular camera. The input of our method only requires a set of rotated human body images that can accept slight movement. First, we apply a deep learning method to predict an initial 3D human body model from multi-view human body images. Then the total detailed digital human model will be computed and optimized. The typical method requires the human body and cameras fixed to obtain a visual hull from a significant number of camera images. This could be extremely expensive and inconvenient when such an application is developed for online users. Compared to the structural lighting measurement system, our predict-optimized framework only requires several input images captured by personal equipment to provide enough accuracy and online use resolution results.

Keywords: Digital human / deep learning / computer vision / data analysis

1 Introduction

Digital humans, which are computer-generated 3D representations of real people, are crucial in creating immersive experiences in a number of recent technological developments, e.g., virtual and augmented reality, and Metaverse which has been gaining a lot of attentions in recent years. They provide a sense of realism and interactivity that is difficult to achieve with traditional computer-generated graphics. As a result, there has been a significant increase in the development of new applications of digital humans in various industries, including manufacturing, gaming, entertainment, education, and healthcare. However, the existing measurement methods of obtaining digital human models are either too expensive or lack accuracy, which presents a challenge for developers looking to create quality and realistic digital humans. The cost of creating a digital human model by using active scanners can be prohibitively expensive, especially for smaller companies and indie developers. This is because the process involves a lot of time and resources, including specialized equipment, software, and skilled personnel, and has special requirements for target people such as standing still for a long time [1–3].

Additionally, the accuracy of the model can be compromised if the data used to create it is incomplete or of poor quality.

To address these challenges, researchers and developers are exploring new ways to create high-quality and affordable digital human models. A promising approach is to use deep learning algorithms to generate realistic human models from a small amount of data. Recently, there are a lot of learning-based work produced. Considering object or scene representation in 3D learning, those works can be simply categorized as explicit representation-based and implicit representation-based.

1.1 Explicit representation-based models

Polygon mesh statistical human body models [4–8] have been widely used in 3D human reconstruction as an explicit representation model. A polygon mesh is a data structure that represents a polyhedron by defining its surface as a collection of vertices and faces. This representation is useful for conveying topological information about the object's surface and provides a high-quality description of 3D geometric structures. Additionally, polygon meshes are memory-efficient and can be easily textured, making them a versatile tool for various applications in computer graphics and visualization. In [9–15], those single image-based work estimates a naked human body model from a

* Corresponding author: moyu.wang@brunel.ac.uk

monocular camera picture. Although those works produced some fine results, they still need further process to dress clothes up. To solve this problem, some other work [16–23] directly learns a mesh human body model with clothes offset from images. The resulted clothed 3D human models inherit the skeleton and surface covering weights of the based body model, facilitating their animation. However, a significant challenge lies in modelling clothing articles such as skirts and dresses, which exhibit substantial deviations from the body surface. The conventional approach of using body-to-cloth offsets is inadequate in such cases.

1.2 Implicit representation-based models

In contrast to meshes, deep implicit functions [24–26] could represent highly detailed 3D shapes with arbitrary topology and are not subject to resolution limitations. Recent research by Saito et al. [27,28] has employed deep implicit functions to reconstruct 3D human shapes from RGB images, achieving high levels of geometric detail and accurate alignment with image pixels. However, this approach suffers from a lack of regularization, resulting in various artifacts such as broken or missing limbs, incomplete details, and geometric noises. To address this issue, some researchers [29–31] have incorporated additional features, such as coarse-occupancy prediction and depth information from RGB-D cameras, to enhance the accuracy and robustness of the shape estimation. In addition, some [32,33] have proposed efficient volumetric sampling schemes to speed up the inference process. Nevertheless, a major limitation of all these methods is that the resulting 3D human shapes cannot be reposed, as implicit shapes do not possess a consistent mesh topology, a skeleton, or skinning weights that are typically found in statistical models.

In summary of these related work, the learning-based human body reconstruction method provided a significant result with only a few inputs. Although training neural networks may require large, labelled 3D digital human datasets and cost large computation resources and time, it is very convenient and efficient for end users. Consumers may only need to upload a small amount of data and wait for the returned result from the cloud service. But most learning-based works focus on recovering full human body from one image with the powerful prediction ability of neural network. This data-driven prediction method may achieve a great result in pose estimation tasks [34–36], but also lead to an ambiguity problem caused by a lack of unseen body information from only one image. It is hard to guess detailed back information from front body image, despite a strong pre-trained network. Hence, we address our problem of finding a balance and a connection between typical measurement method and the popular learning-based method to generate a digital human from inputs.

In this paper, we present our prediction-measurement pipeline to reconstruct a detailed human body model from a set of self-rotated target human images captured by a single monocular camera. We estimate an initial human body model from image sequences by a trained neural network and further vertex alignment to optimize it from image to image. Our research focuses on creating a human body model that is easily modifiable. To achieve this, we have chosen to

utilize a parametric representation of an explicit body model known as SMPL (Skinned Multi-Person Linear) [37]. Recent work [9–23] has shown that the SMPL model possesses excellent expansibility with high-quality open-source resources, which can assist in achieving good results for 3D reconstruction projects. This model allows us to generate body shapes that can be easily modified and adapted to different needs. We begin by collecting data on the SMPL pose and shape parameters, as well as the intrinsic camera parameters from input images. This information is then used to prepare for further optimization.

To create the initial body model, we will use the average pose and shape parameters from the SMPL model. This initial model will serve as a baseline for further modifications and adjustments. This involves projecting the initial SMPL model with the shape and pose of the target image and then minimizing the distance between the projected points and the silhouette of the target image. By doing so, we are able to obtain shape and pose information for every image. This method enables us to create a human body model that is easily adaptable to different needs and requirements. We can modify and adjust the model based on new input data, allowing us to create more accurate and realistic representations of the human body. Overall, our research aims to create a model that can be used in a wide range of applications, from computer graphics to medical simulations.

2 Methods

In order to make sure our predicted initial human body model is allowed to modify, we used a parametric representation of the explicit body model SMPL [37], which will be introduced in Section 2.1. Similar to [35], we collect the estimate results of SMPL pose and shape parameters and intrinsic camera parameters from input images for the preparation of further vertex aligned optimization. And we will build the initial body model with an average pose and shape parameters in the SMPL model, which will be discussed in Section 2.2. Section 2.3 will detail our optimization method. Since we obtain shape and pose information of every image, we project the initial SMPL model with shape and pose of target image and minimize the distance between projected points to silhouette of target image.

2.1 SMPL parameterized human body model

The SMPL model [37] is a powerful method for characterizing the human body in terms of both body shape and motion posture. It achieves this through the use of two sets of statistical parameters: body shape parameters and pose parameters.

The body shape parameters, denoted as β , are used to describe an individual's physique. This 10-dimensional vector allows for the quantification of a person's body shape along various dimensions such as height, weight, and overall body proportions. Each dimension of β can be thought of as a specific indicator of a person's physical characteristics, which collectively describe their overall body shape.

On the other hand, the pose parameters, denoted as θ , are used to describe the motion posture of the human body. This set of parameters comprises 24×3 dimensions, with 24 representing the number of joints and 3 representing the axis-angle representation used to describe rotations. This allows for a detailed and comprehensive description of the human body's motion posture.

To characterize the human body using these parameters, the SMPL model utilizes a base template or mean template T_m , which serves as a reference shape. The shape parameters are then linearly superimposed on this base template to produce the final 3D mesh, with the bias for each shape parameter being calculated using the $B_s(\beta)$ function learned from data. This allows for the generation of meshes that accurately reflect the desired body shape.

$$B_s(\beta) = \sum_{n=1}^{|\beta|} \beta_n S_n, \quad (1)$$

where S is learned through data and has dimensions of (6890, 3, 10).

Similarly, the effect of different pose parameters is determined using the $B_p(\theta)$ function, which is calculated relative to the T-pose state to account for changes in posture. This enables the creation of meshes that accurately reflect the desired motion posture.

$$B_p(\theta) = \sum_{n=1}^{9K} (R_n(\theta) - R_n\theta^*) P_n. \quad (2)$$

Each pose parameter is represented by a rotation matrix R , so there are $9K$ dimensions. P (i.e., the weight matrix) is learned through data and has dimensions of (6890, 3, 207), where 207 is obtained from 23×9 .

Finally, the SMPL model accounts for skin deformation caused by joint motion through a skinning process. This involves a weighted linear combination of skin nodes that change with the joint, with the weights determined based on the distance of the endpoint from the joint. Closer endpoints are more strongly influenced by joint rotation or translation, resulting in a more realistic and accurate representation of the human body's motion. Here the template is defined as:

$$T(\beta, \theta) = T_m + B_s(\beta) + B_p(\theta). \quad (3)$$

Since SMPL body template is a representation of a naked human body, we add an offset S as a detailed cloth supplement:

$$T(\beta, \theta, S) = T_m + B_s(\beta) + B_p(\theta) + S. \quad (4)$$

A pose and shape driven detailed SMPL model is further defined as:

$$M(\beta, \theta, S) = W(T(\beta, \theta, S) + J(\beta), \theta, W), \quad (5)$$

where W is the Linear Blend Skinning (LBS) function, $J(\beta)$ is the locations of 24 skeleton joints; W is the learned blend weights.

2.2 Images information extraction

In this part, we extract information from input images with several deep learning technologies. We collect SMPL model shape and pose parameters with a network of PARE [35] whose main method is to propose a novel deep learning-based approach for estimating 3D human body shape and pose from a single 2D image. The method is centred around a part attention regressor, which divides the human body into various parts and focuses on each one independently to generate accurate 3D body estimations.

- The key components of PARE's methodology include:
- Part Attention: The network utilizes an attention mechanism to focus on specific body parts, enabling it to handle occlusions and varying poses. This mechanism helps the network learn and emphasize individual part features, leading to more precise 3D shape and pose estimations.
 - Multi-stage Estimation: PARE employs a multi-stage estimation process, using an initial coarse estimation followed by multiple refinement stages. This hierarchical approach allows the network to progressively refine its predictions, leading to higher accuracy.
 - Joint 2D-3D Representation Learning: PARE learns a joint embedding space of 2D and 3D features, enabling it to leverage both 2D and 3D information during the estimation process. This joint learning process allows the model to handle a wide range of poses and improve overall accuracy.
 - Part-based Loss Function: The model uses a part-based loss function, which encourages the network to focus on each body part individually. This loss function helps the network to handle complex poses and occlusions, as well as achieve better generalization across various body shapes.

In summary, the PARE method leverages a part attention mechanism, multi-stage estimation, joint 2D-3D representation learning, and a part-based loss function to achieve the accurate 3D human body shape and pose estimations from a single 2D image.

We simply initialize an SMPL body model with average estimated shape and pose parameters of input images, and further detailed offset optimization will be discussed in the next section.

2.3 Full detailed body model optimization

Given that we have acquired the human body pose and camera position data for all input images, we can obtain the projection results of the initialized model concerning angles and poses. By comparing the derived contour images with those of the input images, we can optimize the vertex parameters of the SMPL model. For the i_{th} input image, the associated contour of the human body model is denoted as S_i , while the contour of the human body in the input image is represented as S'_i . In accordance with a differentiable renderer approach [38], we employ an Intersection-over-Union error metric for the optimization process.

$$L_{sil} = \frac{1}{f} \sum_1^f \left(1 - \frac{\|S_i \boxtimes S'_i\|_1}{\|S_i \boxtimes S'_i - S_i S'_i\|_1} \right), \quad (6)$$

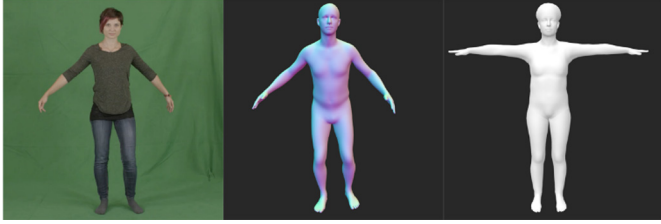


Fig. 1. From left to right, input images, initial SMPL model, optimized SMPL-offset model.

where \otimes is an element wise product and \otimes is a sum operator.

We also add a Laplacian mesh regularizer [39] to ensure the deformation process smoothly. The regularizer is defined as:

$$L_{lp} = \sum_1^N \|L(v_i) - L(v_i(\beta, 0))\|^2, \quad (7)$$

where L is a Laplace operator, v is the vertices set.

Similar to [18], we penalize the difference between the optimized detailed body model vertices and the standard SMPL template body model vertices to avoid large differential error.

$$L_{dif} = \sum_{i=1}^N \|v_i(\beta, S) - v_i(\beta, 0)\|^2. \quad (8)$$

Our joint optimized formula is defined as:

$$L = L_{sil} + w_{lp}L_{lp} + w_{dif}L_{dif}, \quad (9)$$

where w_{lp} and w_{dif} are the balance weights.

By minimizing the loss function L , we modify the vertices of SMPL model and finally collect a detailed body model with cloth information offset. Since SMPL is a pose and shape parametric driven model, the result model can be further animated, which is suitable for more applications.

3 Results

3.1 People-snapshot dataset test

We test our method in People-snapshot dataset [40], Figure 1 shows the results of every step. The input images are captured by a stable camera, and photographed person is self-rotated with a fixed pose. We do not need photographed person keep this pose strictly, a slight change is acceptable. In our method, we extract some frames from the video of dataset, our test used $f=100$ frames to reconstruct body model.

The mid image in Figure 1 shows the initial SMPL model reconstructed from information extracted from input images in step one. We take an estimated average pose and shape parameter of images applying to the SMPL template. The main computing cost here is information extraction with deep neural network, also the accuracy is determined by the efficiency of the state-of-the-art network. However, we have also found that the SMPL



Fig. 2. Detailed optimized SMPL-offset model.

parameter prediction network incorrectly computed the gender of the target individual. This error in prediction, in the subsequent detailed reconstruction steps, will be corrected to adjust the model's vertex offsets.

Our approach takes about 100s for optimizing every frame. We remove the pose parameter in the result, and a standard T-pose SMPL model is showing in the right panel of Figure 1. And we provide some rotated results in Figure 2. Our result can be further modified and rendered. Compared with the initial model, we successfully recovered some hair, face and cloth details in the SMPL model with offset.

3.2 Detailed normal map refinement

As the results generated by our method still have shortcomings in terms of detail representation. We tried a normal map aligned method to refine more details in our result. Traditionally, more refined details have been captured using Shape from Shading (SfS) [18]. However, for monocular clothing capture in unconstrained environments, we have empirically found it challenging to reliably extract such refined details using SfS due to the complexity of garment albedo, wide variations in lighting conditions, and self-shadowing effects. Recently, the success of learning-based approaches [27,28] in estimating accurate surface normal for human appearance using neural networks has been observed. These estimated surface normal provide robust and direct indications for incorporating wrinkles into our clothing capture results to achieve better alignment with the original images. Our results shown in Figure 3 and a more generalized test of a daily indoor work environment images shown in Figure 4.

4 Discussions

In this paper, we proposed a vertices-pixels aligned method and jointly use deep learning method and key idea of traditional computer 3D graphics to achieve a fine level digital human geometry reconstruction from images. Our method relies on several deep learning-based methods such as pose and shape estimates from single images. Although significant progress has been made in deep learning-based methods for 3D human body reconstruction from 2D images, several challenges and limitations still need to be addressed.

4.1 Handling of complex clothing and occlusions

Most current methods rely on the SMPL model, which primarily represents the human body with minimal clothing. Incorporating complex clothing, accessories,



Fig. 3. Detail-refined with normal map result. Compared with [Figures 1 and 2](#), our reconstructed details such as hair, face and clothes have been significantly improved by normal map refinement.

and occlusions remains a significant challenge. Future research could explore the integration of garment-specific models, leveraging semantic information, or employing unsupervised learning techniques to improve the reconstruction of clothed human bodies.

4.2 Robustness to lighting and shadows

Deep learning models may struggle to generalize to varying lighting conditions and shadows, which can significantly impact on the accuracy of 3D reconstruction. Developing methods that are more robust to these factors, such as incorporating illumination-invariant features, is an essential direction for future work.

4.3 Utilization of multi-view and temporal information

The majority of current methods focus on single-view images. Exploiting multi-view or temporal information from videos could potentially improve the accuracy and

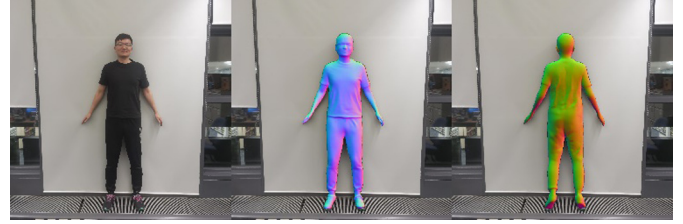


Fig. 4. Daily scene test result. In [Figure 3](#) we show our reconstructed result from a target person standing in front of a green screen. We also test our method in a simple and daily environment. And the result reveals our method is adaptable.

robustness of 3D human body reconstruction. This would require the development of novel network architectures and loss functions that can effectively leverage such additional data.

4.4 Evaluation metrics and benchmarks

Evaluating the performance of 3D human body reconstruction methods is non-trivial due to the lack of ground truth data and the subjectivity of visual quality. Developing standardized evaluation metrics and benchmarks, including datasets with accurate ground truth 3D annotations, is crucial for enabling a fair comparison of methods and guiding future research.

4.5 Real-time performance and computational efficiency

Many deep learning-based methods for 3D human body reconstruction require significant computational resources, limiting their applicability in real-time scenarios or on resource-constrained devices. Future research should focus on developing efficient algorithms and network architectures that can deliver high-quality reconstructions with minimal computational overhead.

In summary, while deep learning has shown tremendous potential in the domain of 3D human body reconstruction from images, there is still ample room for improvement and exploration. Addressing the challenges and limitations discussed in this section will pave the way for more accurate, robust, and efficient 3D human body reconstruction techniques, ultimately benefiting a wide range of applications, from entertainment and virtual reality to healthcare and sports analytics.

Our method is limited by the accuracy and precision of some of the deep learning techniques used. Although we have employed multi-angle image optimization to minimize the inherent ambiguity of the prior prediction model method as much as possible, we still need to spend a considerable amount of computational power and time to optimize our loss function. Therefore, in order to achieve faster and higher-precision human body model reconstruction, more work needs to be done to optimize the method. One approach is to train a deep learning network with multi-angle view priors, allowing the network to learn more 3D human body knowledge. Another approach is to improve the speed of the multi-view optimization process.

5 Conclusions

In this paper, we discussed a vertices-pixels aligned method jointly using deep learning method and the key idea of traditional computer 3D graphics to achieve a fine level of digital human geometry reconstruction from images. Our method relies on several deep learning-based methods such as pose and shape estimate from single images. Compared with related deep learning-based methods, our method eliminates the inherent ambiguity of predicting the complete body model from a single image. With the assistance of deep learning techniques such as pose estimation and human parameter model prediction, we have improved computational speed and reduced experimental conditions compared to traditional optical measurement techniques for obtaining human models. Despite some shortcomings in our work, we have successfully demonstrated the possibility and potential of combining deep learning with traditional techniques.

References

1. S. Fuhrmann, F. Langguth, M. Goesele, Mve-a multiview reconstruction environment, Eurograph. Workshops Graph. Cult. Herit. 11 – 18 (2014)
2. R.A. Newcombe, S.J. Lovegrove, A.J. Davison, Dtam: Dense tracking and mapping in real-time, IEEE Int. Conf. Comput. Vis. 2320–2327 (2011)
3. Y. Xu, X. Liu, L. Qin, S.-C. Zhu, Multi-view people tracking via hierarchical trajectory composition, AAAI Conf. Artif. Intell. 1, (2017)
4. H. Joo, T. Simon, Y. Sheikh, Total capture: a 3D deformation model for tracking faces, hands, and bodies, Comput. Vis. Pattern Recognit. (CVPR) 8320–8329 (2018)
5. M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M.J. Black, SMPL: a skinned multi-person linear model, Trans. Graph. (TOG). 34, 1–16 (2015)
6. G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A.A.A. Osman, D. Tzionas, M.J. Black, Expressive body capture: 3D hands, face, and body from a single image, Comput. Vis. Pattern Recognit. (CVPR). 10975–10985 (2019)
7. J. Romero, D. Tzionas, M.J. Black, Embodied hands: modeling and capturing hands and bodies together, Trans. Graph. (TOG). 36, 1–17 (2017)
8. H. Xu, E.G. Bazavan, A. Zafir, W.T. Freeman, R. Sukthankar, C. Sminchisescu, GHUM & GHUML: generative 3D human shape and articulated pose models, Comput. Vis. Pattern Recognit. (CVPR). 6183–6192 (2020)
9. V. Choutas, L. Muller, C.-H.P. Huang, S. Tang, D. Tzionas, M.J. Black, Accurate 3D body shape regression via linguistic attributes and anthropometric measurements, Comput. Vis. Pattern Recognit. (CVPR). (2022)
10. A. Kanazawa, M.J. Black, D.W. Jacobs, J. Malik, End-to-end recovery of human shape and pose, Comput. Vis. Pattern Recognit. (CVPR). 7122–7131 (2018)
11. M. Kocabas, N. Athanasiou, M.J. Black, VIBE: Video inference for human body pose and shape estimation, Comput. Vis. Pattern Recognit. (CVPR). 5252–5262 (2020)
12. N. Kolotouros, G. Pavlakos, M.J. Black, K. Daniilidis, Learning to reconstruct 3D human pose and shape via model-fitting in the loop, Int. Conf. Comput. Vis. (ICCV). 2252–2261 (2019)
13. D. Smith, M. Loper, X. Hu, P. Mavroidis, J. Romero, FACSIMILE: Fast and accurate scans from an image in less than a second, Int. Conf. Comput. Vis. (ICCV). 5330–5339 (2019)
14. Y. Sun, W. Liu, Q. Bao, Y. Fu, T. Mei, M.J. Black, Putting people in their place: monocular regression of 3D people in depth, Comput. Vis. Pattern Recognit. (CVPR). (2022)
15. H. Yi, C.-H.P. Huang, D. Tzionas, M. Kocabas, M. Hassan, S. Tang, J. Thies, M.J. Black, Human-aware object placement for visual environment reconstruction, Comput. Vis. Pattern Recognit. (CVPR). (2022)
16. T. Alldieck, M.A. Magnor, B.L. Bhatnagar, C. Theobalt, G. Pons-Moll, Learning to reconstruct people in clothing from a single RGB camera, Comput. Vis. Pattern Recognit. (CVPR). 1175–1186 (2019)
17. T. Alldieck, M.A. Magnor, W. Xu, C. Theobalt, G. Pons-Moll, Detailed human avatars from monocular video, Int. Conf. 3D Vis. (3DV). 98–109 (2018)
18. T. Alldieck, M.A. Magnor, W. Xu, C. Theobalt, G. Pons-Moll, Video based reconstruction of 3D people models, Comput. Vis. Pattern Recognit. (CVPR). 8387–8397 (2018)
19. T. Alldieck, G. Pons-Moll, C. Theobalt, M.A. Magnor, Tex2Shape: detailed full human body geometry from a single image, Int. Conf. Comput. Vis. (ICCV). 2293–2303 (2019)
20. V. Lazova, E. Insafutdinov, G. Pons-Moll, 360-degree textures of people in clothing from a single image, Int. Conf. 3D Vis. (3DV). 643–653 (2019)
21. G. Pons-Moll, S. Pujades, S. Hu, M.J. Black, ClothCap: seamless 4D clothing capture and retargeting, Trans. Graph. (TOG). 36, 1–15 (2017)
22. D. Xiang, F. Prada, C. Wu, J.K. Hodgins, MonoClothCap: towards temporally coherent clothing capture from monocular RGB video, Int. Conf. 3D Vis. (3DV). 322–332 (2020)
23. H. Zhu, X. Zuo, S. Wang, X. Cao, R. Yang, Detailed human shape estimation from a single image by hierarchical mesh deformation, Comput. Vis. Pattern Recognit. (CVPR). 4491–4500 (2019)
24. Z. Chen, H. Zhang, Learning implicit fields for generative shape modelling, Comput. Vis. Pattern Recognit. (CVPR). 5939–5948 (2019)
25. L.M. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger, Occupancy networks: learning 3D reconstruction in function space, Comput. Vis. Pattern Recognit. (CVPR). 4460–4470 (2019)
26. J.J. Park, P. Florence, J. Straub, R.A. Newcombe, S. Lovegrove, DeepSDF: learning continuous signed distance functions for shape representation, Comput. Vis. Pattern Recognit. (CVPR). 165–174 (2019)
27. S. Saito, Z. Huang, R. Natsume, S. Morishima, H. Li, A. Kanazawa, PIFu: pixel-aligned implicit function for high-resolution clothed human digitization, Int. Conf. Comput. Vis. (ICCV). 2304–2314 (2019)
28. S. Saito, T. Simon, J.M. Saragih, H. Joo, PIFuHD: multi-level pixel-aligned implicit function for high-resolution 3D human digitization, Comput. Vis. Pattern Recognit. (CVPR). 81–90 (2020)
29. T. He, J.P. Collomosse, H. Jin, S. Soatto, Geo-PIFu: geometry and pixel aligned implicit functions for single-view human reconstruction, Conf. Neural Inf. Process. Syst. (NeurIPS). (2020)
30. Z. Li, T. Yu, C. Pan, Z. Zheng, Y. Liu, Robust 3D self-portraits in Seconds, Comput. Vis. Pattern Recognit. (CVPR). 1341–1350 (2020)

31. Z. Dong, C. Guo, J. Song, X. Chen, A. Geiger, O. Hilliges, PINA: learning a personalized implicit neural avatar from a single RGB-D video sequence, *Comput. Vis. Pattern Recognit. (CVPR)*. (2022)
32. R. Li, K. Olszewski, Y. Xiu, S. Saito, Z. Huang, H. Li, Volumetric human teleportation, *ACM SIGGRAPH 2020 Real-Time Live*. 1–1 (2020)
33. R. Li, Y. Xiu, S. Saito, Z. Huang, K. Olszewski, H. Li, Monocular real-time volumetric performance capture, *Eur. Conf. Comput. Vis. (ECCV)*. **12368**, 49–67 (2020)
34. F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, M.J. Black, Keep it SMPL: automatic estimation of 3D human pose and shape from a single image, *Eur. Conf. Comput. Vis. Springer International Publishing*, (2016)
35. M. Kocabas, C.-H.P. Huang, O. Hilliges, M.J. Black, PARE: part attention regressor for 3D human body estimation, *Int. Conf. Comput. Vis. (ICCV)*. 11127–11137 (2021)
36. Z. Cao, G.H. Martinez, T. Simon, S.-E. Wei, Y.A. Sheikh, Openpose: realtime multi-person 2d pose estimation using part affinity fields, *IEEE Trans. Pattern Anal. Mach. Intell.* (2019)
37. M.M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M.J. Black, SMPL: a skinned multi-person linear model, *ACM Trans. Graph.* **34**, 1–16 (2015)
38. S. Liu, T. Li, W. Chen, H. Li, Soft rasterizer: a differentiable renderer for image-based 3d reasoning, *Proc. IEEE Int. Conf. Comput. Vis.* 7708–7717 (2019)
39. O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rossil, H.P. Seidel, Laplacian surface editing, *Eurogr./ACM SIGGRAPH Symp. Geom. Process.* 175–184 (2004)
40. <https://graphics.tu-bs.de/people-snapshot>

Cite this article as: Moyu Wang, Qingping Yang, From prediction to measurement, an efficient method for digital human model obtainment, *Int. J. Metrol. Qual. Eng.* **15**, 1 (2024)