Metaverse (2023) Volume 4 Issue 2
doi: 10.54517/m.v4i2.2295

ORIGINAL RESEARCH ARTICLE

A single-image human body reconstruction method combining normal recovery and frequency domain completion

Yanyan Li[1,2,3], Zhihao Yang[1,2], Weilong Peng[1,2,*], Meie Fang[1,2]

[1] *School of Computer Science and Cyber Engineering & Metaverse Research Institute, Guangzhou University, Guangzhou 510006, Guangdong Province, China*

[2] *Key Laboratory of Philosophy and Social Sciences in Guangdong Province of Maritime Silk Road, Guangzhou University, Guangzhou 510006, Guangdong Province, China*

[3] *School of Computer Engineering, Guangzhou City University of Technology, Guangzhou 510850, Guangdong Province, China*

*** Corresponding author:** Weilong Peng, wlpeng@gzhu.edu.cn

ABSTRACT

Reconstructing a 3D human body from a single image is convenient and efficient, but it faces challenges when the face is heavily occluded or the person is wearing loose clothing. In this paper, we propose a frequency domain-based method for completing missing parts of the human body using manifold harmonics and frequency domain analysis. Our approach involves linear interpolation of the incomplete 3D human body obtained through normal integration. The interpolated points are then projected onto the appropriate dimension of the frequency domain space using manifold harmonic bases. Through Laplacian mesh editing, the interpolated points are replaced, resulting in a refined and complete 3D human body. Our method surpasses the limitations of template human bodies and marching cubes algorithms, enabling more detailed feature reconstruction. By locally completing the body in a low-dimensional frequency domain space, our method avoids over smoothing and bulging issues, effectively filling the missing regions while maintaining mesh smoothness consistency. Experimental results demonstrate the effectiveness and superiority of our frequency domain-based completion method for accurate and detailed 3D human body reconstruction from single images.

Keywords: human body reconstruction; normal integration; manifold harmonics; Laplacian mesh editing

1. Introduction

With the rapid evolution of virtual reality, human-computer interaction, and digital entertainment, realistic 3D human body models have emerged as a pivotal resource for creating immersive experiences in various applications. Compared to the 3D human body reconstruction method based on multiple images[1–4], although a single image[5–7] provides limited depth information, it is more practical in application due to its ease of acquisition and its ability to avoid inconsistencies and matching problems between multiple images. Considering the high degree of freedom in human body movements and the wide variety of clothing, reconstructing a 3D human body from a single image remains an extremely challenging task.

ARTICLE INFO
Received: 8 September 2023 | Accepted: 20 October 2023 | Available online: 3 November 2023

CITATION
Li Y, Yang Z, Peng W, Fang M. A single-image human body reconstruction method combining normal recovery and frequency domain completion. *Metaverse* 2023; 4(2): 2295. doi: 10.54517/m.v4i2.2295

Currently, there are two main methods for reconstructing a 3D human body from a single image: explicit representation and implicit representation. The explicit representation method[8–11] usually uses the statistical template model SMPL[12] to reconstruct the human body. In order to represent clothing, some Methods[10,13] add displacement to the vertices of the SMPL model to reconstruct clothing. However, due to the limitations of the template model, it can only reconstruct tight clothing but cannot accurately reconstruct loose clothing, such as skirts. In contrast to the explicit representation method, the implicit representation methods[14–16] are able to represent detailed 3D shapes with arbitrary topologies, thereby efficiently reconstructing loose clothing. PIFu[17] and PIFuHD[18] use pixel-aligned implicit functions to reconstruct detailed clothed 3D humans. However, the drawback of this method is that it lacks explicit prior knowledge about the structure of the human body, which often results in issues such as limb artifacts for unseen poses.

Our goal is to combine the advantages of explicit and implicit representations to achieve more robust 3D human body reconstruction. We observed that neural networks are better at predicting 2D information than directly predicting 3D information, such as signed distance fields. A depth map is a 2D image with abundant 3D information that can provide distance information of objects in the scene. Through simple uniform sampling, we can obtain a single-sided uniform point cloud of the objects and then triangulate it to obtain a body mesh. Compared with the method using the statistical template model, the depth map method does not require model parameters and can provide more accurate 3D point cloud information. In addition, the depth map method does not appear to have a situation of missing geometry like the method based on the implicit function.

However, the use of front and back depth maps for 3D human body reconstruction will have the issue of missing 3D side information. In order to complement the missing 3D point cloud, the usual solution is to regress an implicit function[19–23] for surface reconstruction, but this method often reduces the geometric accuracy. Inspired by Vallet and Lévy[24], we noticed that the information stored in the frequency domain space of various dimensions is different for 3D objects, so it is possible that the missing side information is stored in the frequency domain space of a certain dimension. Therefore, we transform the reconstruction task from the spatial domain to the completion task in the frequency domain.

In this paper, we present a novel approach for 3D human body reconstruction by combining normal recovery and frequency domain completion. Our method starts by reconstructing the "hole" in the human body using the normal integration technique. This initial reconstruction is then refined through interpolation, resulting in a coarse human body mesh. Next, we project the mesh into the frequency domain space using manifold harmonic basis and further refine it using the Laplacian grid editing method. Unlike traditional methods that rely on statistical template models, our approach excels in reconstructing loose clothing, yielding superior results. The effectiveness of our method has been demonstrated through rigorous experimentation.

The main contributions of this work can be summarized in two points:

1) We introduce a method that utilizes normal integration to reconstruct the 3D human body, overcoming limitations in existing statistical template models, implicit function reconstruction, and the marching cube algorithm. This enables more detailed and robust 3D human body reconstruction.
2) We propose a novel human body completion method using manifold harmonic basis and Laplacian grid editing. This approach effectively completes missing regions, enhancing the fidelity of the reconstruction.

2. Related works

2.1. Explicit and implicit reconstruction methods

Reconstruction methods based on explicit and implicit combinations usually integrate the parametric human body model as prior knowledge into the implicit surface to achieve more accurate 3D human body reconstruction. PaMIR[16] uses the corresponding SMPL model to extract the 3D features of the human body and combines the 2D features to determine the location of the sampling point. However, this method learns global information, making the network extremely sensitive to the position of the human body. In contrast, ICON[25] uses the corresponding SMPL human body to obtain the basic normal map and reconstruct the 3D human body from basic feature information without global picture information. Bhatnagar et al.[26] and others employed a contrary approach, using implicit network process the input the sparse 3D point cloud of the human body surface to obtain a parameterized human body model, and then the details of clothes, face, and hair are captured by adding offsets to the vertices of SMPL. However, the explicit-implicit combination method[27–29] mentioned above has a significant drawback. It often misinterprets loose clothing, like skirts, as pants due to limitations in the model. Therefore, a more comprehensive approach is required to effectively combine the advantages of explicit and implicit reconstruction methods and achieve accurate and robust 3D human model reconstruction.

2.2. Depth map-based methods

Recently, a 3D human body reconstruction method based on positive and negative depth maps and geometric completion was proposed. As a data representation closely related to 3D geometry, depth maps can preserve distance and geometry information, which helps reconstruct the shape and pose of the 3D human body more accurately. Compared with directly using 3D information for reconstruction, the method based on the depth map[30–32] has the advantages of lower dimensionality and higher resolution. Gabeur et al.[33] proposed a non-parametric method using dual depth maps to reconstruct 3D human bodies. They estimated the front and back depth maps of the human body and used the Poisson reconstruction algorithm to reconstruct the corresponding front and back point clouds into a complete 3D human body. However, Poisson reconstruction can easily lead to loss of accuracy and bulge. ECON[34] proposed two new human body completion methods. One method uses the Poisson surface reconstruction algorithm to fuse the SMPL-X model with the point cloud data of the human body with holes to complement the area with holes. However, since the SMPL-X model only contains surface information of naked people, it cannot generate a coherent surface for the originally missing clothing and hair. Another method uses IF-Nets[35] to extract multi-scale features in the encoding stage to capture the global and local information of the input shape. The 3D human body is reconstructed by predicting an occupancy field in the decoding stage.

2.3. Frequency domain processing on grids

Frequency domain processing methods on grids mainly use techniques such as discrete Fourier transform and discrete cosine transform to convert geometric shapes from Euclidean space to spectral space for tasks such as shape reconstruction. Taubin proposed a Laplacian operator for grid smoothing in 1995[36], which laid the foundation for subsequent grid frequency domain processing. Vallet and Lévy[24] proposed a manifold-based frequency domain method of harmonic basis to process triangular grids. This method constructs the harmonic function space by calculating the eigenvectors of the discrete Laplacian operator. This frequency domain processing method on the grid has also been applied to the human body reconstruction field. Kim et al.[37] used Laplacian coordinates to represent the local structure of the input grid. This method uses the SMPL model to reconstruct a controllable base grid and learns a surface function using a neural network. The function can predict the Laplacian coordinates of surface details on the basic mesh, thereby achieving fine human body

reconstruction. Inspired by the work of Vallet and Lévy[24], we found that 3D objects retain various information in frequency domain spaces of different dimensions. Therefore, we consider converting the reconstruction task from the spatial domain to the frequency domain and completing the completion task in the frequency domain.

3. Method

3.1. Process of reconstructing 3D human body from a single image

Given an RGB image, we first estimate the front and back normal maps and then use the method of normal integral[38] to obtain the front and back depth maps, so as to reconstruct a 3D human body with a hole on the side. For the hole filling, we use the projection method of the manifold harmonic basis to transform the interpolated 3D human body model into the frequency domain space of appropriate dimensions. By applying the Laplacian grid transformation, we replace the points in the interpolated region and then reconstruct a smooth and complete human body model, as seen in **Figure 1**. The method of normal integration has the advantages of not being limited by the template human body and avoiding the accuracy limitation of the marching cube algorithm so that a high-fidelity 3D human body can be reconstructed. In addition, the manifold harmonic-based method can preserve the human body shape obtained by normal integral reconstruction as much as possible and effectively fill the holes in the human body so as to obtain a complete, high-fidelity 3D human body model.

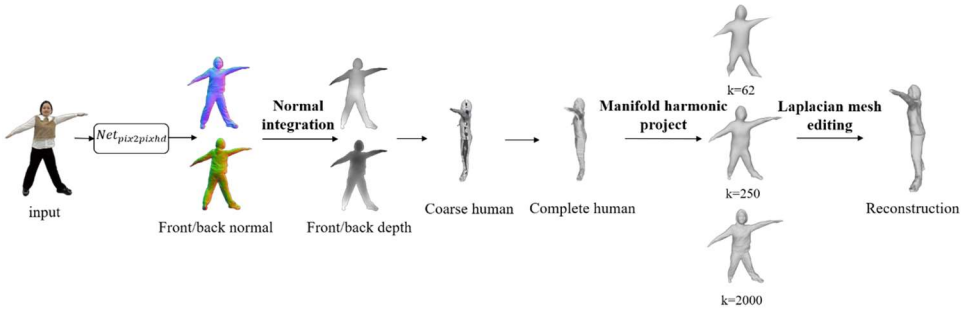

Figure 1. Reconstruct 3D human body from a single image.

3.2. Reconstruction of "hole" human body by normal integral

Since a single picture cannot provide human body information on the back, in order to obtain the normal map of the human back, a neural network is required to infer from the known front human body information. PIFuHD[18] uses a large number of RGB images and their corresponding front normal maps to train the neural network to infer the back normal map. However, the back normal map obtained by this method is too smooth and lacks individual characteristics. To solve this problem, we use SMPL model corresponding to each image in the THuman 2.0 data set[39] to provide prior knowledge to assist normal generation network $Net_{pix2pixhd}$ to predict a more accurate back normal map and prevent the prediction from being too general. Specifically, we input the input image, and its front and back normal maps N_f^{Smpl} and N_f^{Smpl} rendered by corresponding SMPL mesh model M^{Smpl} into $Net_{pix2pixhd}$ to predict the complete front and back normal maps N_f^i, N_b^i. The training process is defined as:

$$N_f^i,\ N_b^i = Net_{pix2pixhd}\left(I_{rgb}, Render(M^{Smpl})\right) \tag{1}$$

where $Render\left(M^{Smpl}\right)$ is the rendering operations by using SMPL mesh model M^{Smpl} to obtain the coarse normal maps.

To train the network, we minimize the following loss function:

$$L_{total} = L_n(N^g, N^i) + L_{per}(N^g, N^i) + L_s(N^{Smpl}, M^i) \quad (2)$$

where L_n is the normal map loss, which is the $L1$ loss between the ground truth normal maps N_f^g, N_b^g and the predicted normal map N_f^i, N_b^i. L_{per} is perceptual loss between N_f^g, N_b^g and N_f^i, N_b^i and it can help to reconstruct details. L_s is $L1$ loss between the normal map N_f^{Smpl}, N_b^{Smpl} and masks M_f^i, M_b^i. This loss ensure that the normal map of the SMPL model is accurately aligned with the input image, which is a basic requirement for generating a correct normal map.

By using the above loss function to train the normal map network $Net_{pix2pixhd}$, and input the test picture, we can get the result shown in **Figure 2**. Although it is feasible to infer the depth map directly through the neural network, there will always be a lot of noise and wrong depth, resulting in the final reconstructed mesh, which is usually not ideal[40]. Therefore, we use the normal integral method to obtain the depth map. The traditional normal integration method is only suitable for recovering smooth surfaces because it assumes that the target surface is continuous and differentiable. However, when dealing with disconnected areas caused by occlusion and other reasons, the normal vector calculation of the traditional method will result in an error, thereby affecting the shape of the entire object surface in the normal integration. In this case, the method of smooth surface assumption will lead to distortion of the surface shape, as shown in **Figure 3**.

Figure 2. The of the front/back normal map.

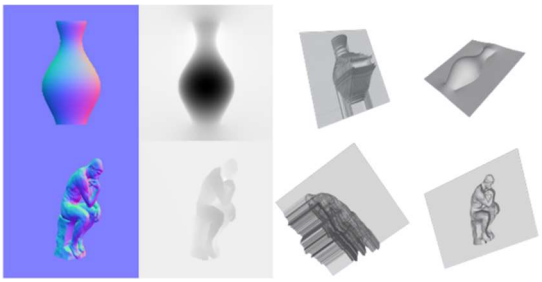

Figure 3. From left to right are the normal map, the depth map, the recovery map of the traditional normal integral, and the recovery map after optimization of the quadratic regular term.

Therefore, in order to deal with discontinuous normal integrals, they are usually transformed into an optimization problem and solved using quadratic regularization. Inspired by the semi-smooth surface assumption of Cao et al.[41], when the surface at a certain point is discontinuous, it will only be discontinuous on one side of the point and will not be discontinuous on both sides. Based on this assumption, when discontinuity occurs, we will only select the continuous side to participate in the normal integral computation, thus ensuring the depth map can be accurately reconstructed.

We use the following weight function and minimize it to recover the depth map z:

$$\min_{z} \iint w_u(n_z\partial_u^+ z + n_x)^2 + a(n_z\partial_u^- z + n_x)^2 + w_v\left(n_z\partial_v^+ z + n_y\right)^2 + b\left(n_z\partial_v^- z + n_y\right)^2 du\, dv \tag{3}$$

where $a = 1 - w_u$, $b = 1 - w_v$, $\partial_u^+ z$, $\partial_u^- z$ is the horizontal side derivative of the point, $\partial_v^+ z$, $\partial_v^- z$ is the vertical side derivative. w_u, w_v is the unilateral differentiability on both sides of each point. Differentiable on one side:

$$w_u = \begin{cases} 0(\text{left differentiable}) \\ 0.5(\text{both left and right differentiable}) \\ 1(\text{right differentiable}) \end{cases}$$

$$w_v = \begin{cases} 0(\text{bottom differentiable}) \\ 0.5(\text{both top and bottom differentiable}) \\ 1(\text{top differentiable}) \end{cases}$$

When the depth map is left differentiable at a certain point, but not right differentiable, the integral term remains left side $\partial_u^- z$, and the right side is ignored. The other two cases are similar, so this relative weight covers all possible cases at each point on the semi-smooth surface.

By discretizing Equation (3), we obtain the human body depth map $\{z_f,\ z_b\}$. By triangulating each pixel along with its three neighboring pixels, we reconstruct a complete human mesh, as is shown in **Figure 4**.

Figure 4. From left to right: Normal map, depth map, reconstructed mesh.

Due to the inconsistency in the coordinate system of the reconstructed depth map on the front and back, it cannot be directly aligned. Our solution is to choose the highest point of the depth map on the front and back as the reference point and set its depth value to zero. Such a simple operation can align the grid on the front and back of the human body, thereby obtaining a 3D human body mesh with holes on the side, as shown in **Figure 5**.

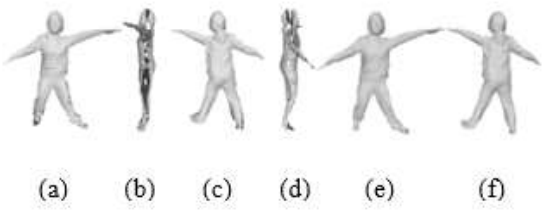

Figure 5. The "hole" human body after front and back stitching and preprocessing.

Besides, due to the errors in reconstructing the depth map, aligning through the highest point may lead to a situation where the depth of the back is smaller than the depth of the front, as shown by a foot in **Figure 5(a)** and **Figure 5(c)** black shadow. To solve this issue, we can simply discard those points with wrong depth values. From **Figure 5(e)** and **Figure 5(f)**, it can be seen that the "dark shadow" of the legs has disappeared.

3.3. Human completion based on manifold harmonic basis

For the completion of the 3D human body mesh, we first use the linear interpolation method to complement it into a manifold human body and then transform it into a frequency domain space of appropriate dimensions with the manifold harmonic basis, using the Laplacian mesh editing method replaces the interpolation points to obtain a complete 3D human body mesh.

Linear interpolation for areas with "holes". We use the linear interpolation method for preliminary grid completion. However, since the distance between the corresponding points on the front and back are too large, randomly inserting points may result in distorted girds. These grids can visually appear as a very flat surface even after the subsequent smoothing process. Therefore, we consider finding the maximum distance between the corresponding points on the front and back, and, according to a certain ratio, determining the number of points that need to be inserted. Insert the same number of points evenly between the corresponding points of each front and back to form a regular grid structure, as shown in **Figure 6(a)**. By using a regular grid-building strategy, we were able to speed up the process and generate more regular grids.

Although this interpolation method can avoid some bad grids, it is still difficult for the newly added grids to be consistent with the smoothness of the original mesh, and there may be slice-like defects, as shown in **Figure 6(b)**. Therefore, we need a smooth operation. From **Figure 6(c)** and **Figure 6(d)**, it can be seen that if Poisson reconstruction not only produces bulging situations but also leads to the loss of high-frequency information such as face details. Therefore, we need to find an appropriate smoothing method to ensure consistent smoothness of the reconstructed 3D human body.

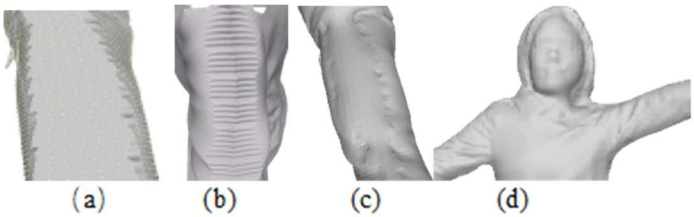

Figure 6. Linear interpolation grid completion and Poisson surface reconstruction.

Calculation of the harmonic basis of the manifold. In order to compute the manifold harmonic basis, we first discretize the Laplacian operator using standard finite element cotangent weights. For a mesh with n vertices, we calculate the cotangent weight of each vertex V_i and its adjacent vertices, and assign it to the corresponding edge. Assuming that the set of adjacent vertices of vertex V_i is n_i, each cotangent weight W_{ij} of the adjacent side j is:

$$w_{ij} = \frac{cot(\alpha_{ij}) + cot(\beta_{ij})}{2A_{ij}} \tag{4}$$

where α_{ij} and β_{ij} are the two angles between the vertex V_i and the adjacent vertex V_j respectively, and A_{ij} is the triangle area corresponding to the shared side ij between the vertex V_i and the vertex V_j. Therefore, the $n \times n$ dimensional cotangent weight matrix W can be obtained by Equation (5):

$$W_{ij} = \begin{cases} w_{ij}, & if(i,j) \in E \\ 0, & otherwise \end{cases} \tag{5}$$

where $(i,j) \in E$ means that there is an edge between the vertices V_i and V_j.

After obtaining the cotangent weight matrix W, we perform an eigendecomposition on W. The eigenvectors $\lambda_1, \lambda_2, \ldots, \lambda_n$ corresponding to a series of eigenvalues $v_1, v_2, \ldots, v_n$ are obtained, and the

pairwise linear independence is known by definition.

Subsequently, we sort the obtained eigenvectors according to the eigenvalues from small to large, and standardize, so as to obtain the arranged eigenvector set V. V is also defined the manifold harmonic basis, and they are orthogonal to each other, and the length is $\|V_1\| = \|V_2\| = \|V_3\| = 1$. Since the information of the grid itself can be divided into n dimensions, the relevant information of the grid can be projected to the frequency domain space by means of the manifold harmonic basis, such as coordinates or normals.

Projection and reconstruction. In manifold harmonic analysis, the calculation of eigenvectors and eigenvalues is achieved through solving the eigendecomposition of the Laplacian matrix. Eigenvalues represent the importance of each eigenvector, while eigenvectors indicate the shape variations under that particular feature. For a 3D mesh, eigenvectors associated with smaller eigenvalues correspond to minor shape variations, resulting in a smoother and simpler model. Conversely, eigenvectors associated with larger eigenvalues correspond to significant shape variations, resulting in a more complex model. By sorting the eigenvectors in ascending order of eigenvalues, a reconstruction can be achieved from coarse to fine.

We can project the surface grid from the geometric space to the *k*-dimensional frequency domain space through *k* manifold harmonic bases. Through the projection, we can obtain three *k*-dimensional coefficients, representing the original The *x*, *y*, and *z* components of the vertex set in the frequency domain space. The coefficients can be reconstructed based on the *k*-dimensional manifold harmonic basis through Equation (6).

$$\sum_{i=0}^{k}\sum_{j=1}^{3} X_{ji} * A_i \tag{6}$$

where X_{ji} means the *i-th* coefficient of the *j-th* *k*-dimensional coefficient, and A_i is the *i-th*-dimensional manifold harmonic basis.

Figure 7 shows the 3D human body mesh reconstructed in different dimensions. It can be observed that different frequency domain dimensions preserve the 3D human body with different levels of detail. According to the point coordinates of the 3D human body reconstructed in the frequency domain space of appropriate dimensions, we can replace the interpolation coordinates of the point so as to achieve the smoothing effect of the model.

Figure 7. The reconstruction results of projecting the human body from the geometric space to the *k* (from left to right: *k* = 62, 250, 1000, 2000) dimensional frequency domain space based on the manifold harmonic basis.

Laplacian grid editing to replace interpolation points. If directly replacing the coordinates of the spatial interpolation points in the frequency domain to the spatial domain, we obtain the inconsistency of grid smoothness as shown in **Figure 8(a)** and **Figure 8(b)**. Laplacian grid editing is a manifold processing method, it obtains delta coordinates by applying Laplacian matrix to vertex coordinates. By using delta coordinates to edit the grid, it can maintain the local geometric characteristics of the grid and achieve a smooth editing effect. Therefore, we can use this method to replace the interpolated point coordinates.

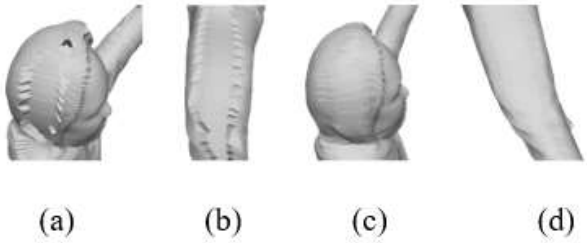

Figure 8. **(a)(b)** is the reconstruction result after directly replacing the interpolation point coordinates; **(c)(d)** is the reconstruction result edited with Laplacian grid.

Specifically, based on the idea of local replacement, we record these interpolation points as the point set to be replaced p. We respectively calculate the manifold grid M_L after linear interpolation and the Laplac S-Beltrami operator, which is an $n \times n$ matrix, where n is the number of vertices of the grid. The off-diagonal element L_{ij} of the matrix represents the weight between vertex i and vertex j, and the diagonal element L_{ii} is the negative value of the sum of the area weights and sparse cotangent weights of all adjacent vertices around vertex i. We can calculate the delta coordinates of each point in p through Equation (7):

$$\Delta p_i = \sum_{j \in N_i} w_{ij} p_i \tag{7}$$

where N_i is the neighborhood of vertex p_i, w_{ij} is the weight coefficient between p_i and p_j.

Delta coordinates are differential space coordinates, which represent the distance between each point and all 1-neighbor points, and all 1-neighbor points of this point affect each other. We directly set the delta coordinates of the point set p to be replaced in the M_L grid as the M_f grid delta coordinates of corresponding points in the grid to smooth the grid at the interpolated points.

In order to reconstruct the original 3D grid from the delta coordinates, we need to convert the edited grid from the differential space to the Euclidean space, which means solving for the edited vertex coordinate set $\hat{p}$. To avoid affecting the coordinates of non-interpolated points in the subsequent solving process, we set the points obtained by taking the inverse of point set p as anchor points.

However, since the number of anchor points is greater than 1, resulting in over-constraint, the Laplacian-Beltrami operator is no longer full rank, so it cannot be solved directly. Our solution is using the least squares method to find the closest to the original vertex coordinate solution to obtain the 3D coordinates of the interpolation point in the original grid. For each point in p, we calculate its point-to-plane distance, and if the distance exceeds a certain threshold, the grid formed by the point will visually appear as bulging. Therefore, we set the coordinates of these points to the average value of their 1-neighbors to smooth these grids. **Figure 9** shows the results of our 3D human body reconstruction based on the frequency domain completion method.

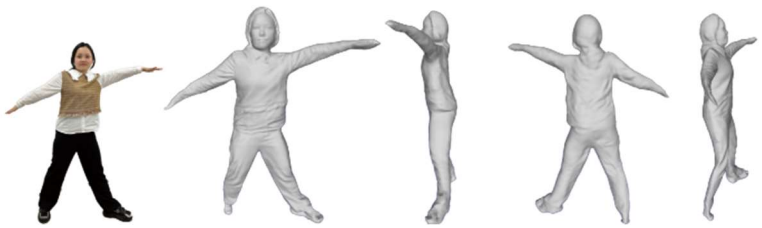

Figure 9. 3D human body reconstruction based on frequency domain completion method.

4. Experiment results

4.1. Datasets and evaluation metrics

In this work, we use the THuman 2.0 dataset as our training dataset, which contains 525 high-quality human scans captured by dense DSLR equipment. Each human scans includes a 3D model with corresponding texture maps and their corresponding SMPL-X fitting parameters and corresponding grid.

Because the CAPE dataset[42] was not used as a training dataset among the compared methods, we used the dataset CAPE to evaluate the generality of various methods. The CAPE dataset is high-precision human scan data. The acquisition frame rate is 60 frames/second, which provides an accurate shape of the human body under clothes for each frame, including details such as body contours, muscles, and fat. The data set includes 10 males and 5 females, covering more than 600 people. Action sequences and 140K+ frames. It covers a wide range of human body pose variations, including various postures and movements, which can be used to test the robustness and stability of algorithms.

4.2. Qualitative assessment

Comparing our method with PIFu[17], ICON[25] on the data set SHHQ[43], the results are shown in **Figure 10**, and the detailed four sides comparison of human body reconstruction is shown in **Figure 11**.

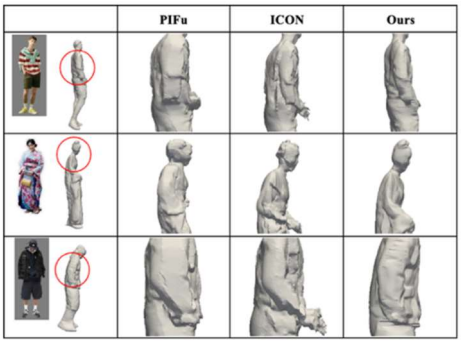

Figure 10. Comparison of the qualitative evaluation of our method with PIFu and ICON on the SSHQ data set.

Figure 11. Comparison of the reconstruction results of our method with PIFu and ICON on the SSHQ data set.

From the first row of **Figure 10**, we can see that when the input image has a human hand in a pocket posture, PIFu and ICON will incorrectly estimate the side hand posture, and our method is significantly better. For human body reconstruction in loose clothes, PIFu based on the implicit method for human body reconstruction has seriously lost the details of the clothes, while ICON, which combines explicit and implicit methods, cannot reconstruct reasonable clothes. Our method can recover both the head and clothes. better than those two methods.

In addition, the reconstruction results using PIFu often exhibit an unrealistic bulging phenomenon on the sides and ICON is unable to reconstruct reasonable geometric shapes when the human body is wearing loose-fitting clothing. Our normal integration and frequency domain completion algorithms are able to reconstruct clothing geometry details and achieve higher consistency with the original input images.

4.3. Quantitative assessment

In the quantitative evaluation we use three evaluation metrics: chamfer distance, P2S, and normals difference. Chamfer distance and P2S distance is used to measure the difference between the real mesh generated by the scanned data and the mesh generated by our method. To eliminate the effect of the resolution difference, we uniformly sample the same points on the scanned data and the reconstructed mesh, and calculate the two-way (chamfer distance)/one-way (P2S) average distance from the point to the surface. The chamfer distance and P2S distance can capture larger geometric differences, but will ignore smaller geometric details. For high-frequency geometric details difference, we use the normal map difference to measure. It means the L2 distance between the front normal map I_1 of the reconstructed mesh rendered from a fixed perspective and the ground-truth front normal map I_2.

Table 1 shows the differences in chamfer distance, P2S distance and normal between our method and PIFu[17], PIFuHD[18], PaMIR[16] and ICON[25] under the CAPE data set. From **Table 1**, it can be found that, Our method uses normal integration to reconstruct the front and back mesh, and does not operate the points of normal integration in the Laplacian mesh editing to maintain the fine 3D human body obtained by normal integration. Our method outperforms other algorithms in the normal difference.

Table 1. Quantitative comparison with different methods under the CAPE dataset.

	PIFu	PIFuHD	PaMIR	ICON	Ours
Chamfer	3.627	3.237	2.122	1.142	2.946
P2S	3.729	3.123	1.495	1.065	3.005
Normal	0.116	0.112	0.088	0.066	0.048

For the chamfer distance and P2S distance, our algorithm outperforms PIFu and PIFuHD but falls behind PaMIR and ICON. After analyzing the normal map, depth map and other intermediate products of the reconstruction process, we observed that for loose-fitting clothing, our algorithm exhibits significant errors in depth prediction near the feet. Specifically, our algorithm predicts that the depth near the feet is in the same plane as the skirt, which is clearly incorrect. Additionally, due to our front-back alignment strategy starting from the head, errors accumulate towards the feet, resulting in the 'swollen' appearance of the feet as shown in **Figure 12**. Because the disparity at the feet is substantial, our algorithm's performance is not optimal when calculating chamfer distance and P2S distance. This issue can be addressed in the future through optimization of foot depth prediction or refining the front-back alignment strategy. However, due to our normal integration producing fine front and back human body reconstructions, our algorithm still outperforms PIFu and PIFuHD, which exhibit similar bulky leg issues. Furthermore, our recovery of the skirt is noticeably better than ICON, which performs better than us in terms of chamfer distance and P2S distance.

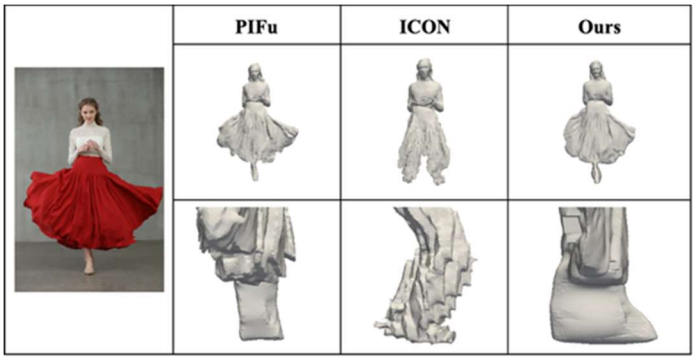

Figure 12. ICON based on the SMPL prior model reconstruct the skirt into legs (middle), PIFu based on the implicit function (left) and our method (right) results in a swollen foot.

4.4. Ablation experiment

We designed an ablation experiment to verify the effectiveness of the proposed frequency domain completion method. We compared it with the method of directly using linear interpolation method and directly replacing the frequency domain spatial coordinates with the interpolation point coordinate completion method. **Figure 13** shows results of various completion methods. It can be seen from the results that the completion result using the linear interpolation method presents a flat surface at the seam, while directly replacing the coordinates of the frequency domain space with the coordinates of the interpolation points results in the collapse of triangles in the hole area. In contrast, our frequency domain completion method successfully stitches the body and maintains the smoothness of the mesh. The results of the ablation experiment verify the effectiveness of our completion method using frequency domain projection reconstruction.

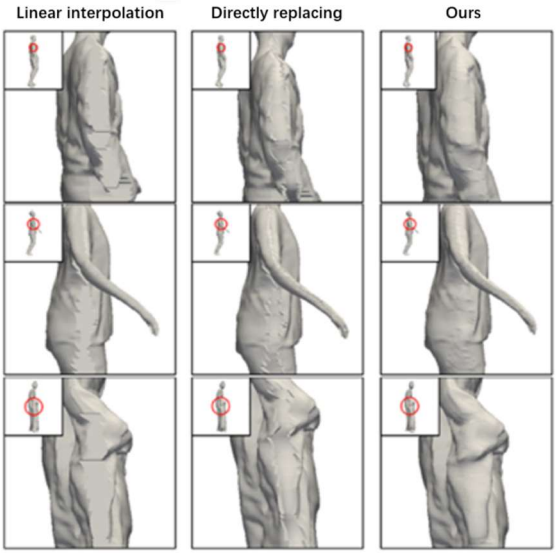

Figure 13. Comparison with each part of the design in the frequency domain completion method.

5. Conclusions

In this study, we propose a method for 3D human body reconstruction from a single image that combines normal recovery and frequency domain completion. Our approach involves reconstructing an initial "holed" human body through normal integration and then locally completing it using manifold harmonic transform and Laplacian grid editing techniques. The key contribution of our method is the projection of the interpolated grid from the spatial domain to the frequency domain space, followed by smoothing of the interpolation points using Laplacian grid editing. Unlike methods that rely on statistical template models or implicit reconstruction

marching cube algorithms, our approach does not require prior knowledge of human body models and achieves improved accuracy in reconstructing human bodies wearing loose-fitting clothing. However, Our method of body reconstruction may cause leg curvature in certain baggy pants areas. In addition, the reconstruction effect of our method is relatively poor when the hand occlusion area is too large and when the feet are reconstructed under loose clothes. To solve the problem of inaccurate prediction of feet by depth maps, the effect of our algorithm should be better than the previous comparison method. Further optimization will be necessary to address these challenges in future research.

Author contributions

Conceptualization, WP and MF; methodology, YL and ZY; software, YL; validation, YL and ZY; formal analysis, YL; investigation, WP and YL; resources, YL; data curation, YL; writing—original draft preparation, YL, WP; writing—review and editing, WP and MF; visualization, ZY; supervision, WP and MF; project administration, YL; funding acquisition, WP and MF. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (62072126), the Guangdong Basic and Applied Basic Research Foundation (2022A1515010138), the Science and Technology Program of Guangzhou (202201020229), Key Laboratory of Philosophy and Social Sciences in Guangdong Province of Maritime Silk Road of Guangzhou University (GD22TWCXGC15).

Conflict of interest

The authors declare no conflict of interest.

References

1. Li Z, Oskarsson M, Heyden A. Detailed 3D human body reconstruction from multi-view images combining voxel super-resolution and learned implicit representation. *Applied Intelligence* 2022; 52(6): 6739–6759. doi: 10.1007/s10489-021-02783-8
2. Yu Z, Zhang L, Xu Y, et al. Multiview human body reconstruction from uncalibrated cameras. In: Proceedings of the 2022 Conference on Neural Information Processing Systems; 28 November–9 December 2022; New Orleans, LA, USA.
3. Li X. Multi-view canonical pose 3D human body reconstruction based on volumetric TSDF. In: Karlinsky L, Michaeli T, Nishino K (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2022 Workshops; 23–27 October 2022; Tel Aviv, Israel. Springer; 2023. Volume 13805, pp. 293–397. doi: 10.1007/978-3-031-25072-9_27
4. Choi H, Moon G, Park JK, Lee KM. Learning to estimate robust 3D human mesh from in-the-wild crowded scenes. In: Proceedings of the 2022 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21–24 June 2022; New Orleans, LA, USA. pp. 1475–1484.
5. Jiang W, Kolotouros N, Pavlakos G, et al. Coherent reconstruction of multiple humans from a single image. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 5579–5588.
6. Choi H, Moon G, Lee KM. Pose2mesh: Graph convolutional network for 3D human pose and mesh recovery from a 2D human pose. In: Vedaldi A, Bischof H, Brox T, Frahm JM (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2020; 23–28 August 2020; Glasgow, UK. Springer; 2020. Volume 12352, pp. 769–787. doi: 10.1007/978-3-030-58571-6_45
7. Kocabas M, Huang CHP, Hilliges O, Black MJ. PARE: Part attention regressor for 3D human body estimation. In: Proceedings of the 2020 IEEE/CVF International Conference on Computer Vision; 20–25 June 2020; Nashville, TN, USA. pp. 11127–11137.
8. Jiang B, Zhang J, Hong Y, et al. Bcnet: Learning body and cloth shape from a single image. In: Vedaldi A, Bischof H, Brox T, Frahm JM (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—

ECCV 2020; 23–28 August 2020; Glasgow, UK. Springer; 2020. Volume 12365, pp. 18–35. doi: 10.1007/978-3-030-58565-5_2
9. Li Z, Yu T, Pan C, et al. Robust 3D self-portraits in seconds. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 1344–1353.
10. Alldieck T, Magnor M, Bhatnagar BL, et al. Learning to reconstruct people in clothing from a single RGB camera. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 15–20 June 2019; Long Beach, CA, USA. pp. 1174–1186.
11. Zhang H, Tian Y, Zhou X, et al. PyMAF: 3D human pose and shape regression with pyramidal mesh alignment feedback loop. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 11446–11456.
12. Loper M, Mahmood N, Romero J, et al. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics* 2015; 34(6): 248. doi: 10.1145/2816795.2818013
13. Alldieck T, Pons-Moll G, Theobalt C, Magnor M. Tex2Shape: Detailed full human body geometry from a single image. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 27 October–2 November 2019; Seoul, Korea. pp. 2293–2303.
14. Dong Z, Guo C, Song J, et al. PINA: Learning a personalized implicit neural avatar from a single RGB-D video sequence. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 18–24 June 2022; New Orleans, LA, USA. pp. 20470–20480.
15. Li R, Xiu Y, Saito S, et al. Monocular real-time volumetric performance capture. In: Vedaldi A, Bischof H, Brox T, Frahm JM (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2020; 23–28 August 2020; Glasgow, UK. Springer; 2020. Volume 12368, pp. 49–67. doi: 10.1007/978-3-030-58592-1_4
16. Zheng Z, Yu T, Liu Y, Dai Q. PaMIR: Parametric model-conditioned implicit representation for image-based human reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2022; 44(6): 3170–3184. doi: 10.1109/TPAMI.2021.3050505
17. Saito S, Huang Z, Natsume R, et al. PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 27 October–2 November 2019; Seoul, Korea. pp. 2304–2314.
18. Saito S, Simon T, Saragih J, Joo H. PIFuHD: Multi-level pixel-aligned implicit function for high-resolution 3D human digitization. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 84–93.
19. Park JJ, Florence P, Straub J, et al. Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 15–20 June 2019; Long Beach, CA, USA. pp. 165–174.
20. Erler P, Guerrero P, Ohrhallinger S, et al. Points2Surf learning implicit surfaces from point clouds. In: Vedaldi A, Bischof H, Brox T, Frahm JM (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2020; 23–28 August 2020; Glasgow, UK. Springer; 2020. Volume 12350, pp. 108–124. doi: 10.1007/978-3-030-58558-7_7
21. Ren S, Hou J, Chen X, et al. Geoudf: Surface reconstruction from 3D point clouds via geometry-guided distance representation. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2–3 October 2023; Paris, France. pp. 14214–14224.
22. Pavlakos G, Choutas V, Ghorbani N, et al. Expressive body capture: 3D hands, face, and body from a single image. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 15–20 June 2019; Long Beach, CA, USA. pp. 10975–10985.
23. Zanfir M, Zanfir A, Bazavan EG, et al. Thundr: Transformer-based 3D human reconstruction with markers. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 12971–12980.
24. Vallet B, Lévy B. Spectral geometry processing with manifold harmonics. *Computer Graphics Forum* 2008; 27(2): 251–260. doi: 10.1111/j.1467-8659.2008.01122.x
25. Xiu Y, Yang J, Tzionas D, et al. ICON: Implicit clothed humans obtained from normals. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 18–24 June 2022; New Orleans, LA, USA. pp. 13286–13296. doi: 10.1109/CVPR52688.2022.01294
26. Bhatnagar BL, Sminchisescu C, Theobalt C, et al. Combining implicit function learning and parametric models for 3D human reconstruction. In: Vedaldi A, Bischof H, Brox T, Frahm JM (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2020; 23–28 August 2020; Glasgow, UK. Springer; 2020. Volume 12347, pp. 311–329. doi: 10.1007/978-3-030-58536-5_19
27. Zheng Y, Shao R, Zhang Y, et al. DeepMultiCap: Performance capture of multiple characters using sparse multiview cameras. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 6239–6249.

28. Huang Z, Xu Y, Lassner C, et al. Arch: Animatable reconstruction of clothed humans. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 3093–3102.
29. He T, Xu Y, Saito S, et al. Arch++: Animation-ready clothed human reconstruction revisited. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 11046–11056.
30. Biggs B, Novotny D, Ehrhardt S, et al. 3D multibodies: Fitting sets of plausible 3D human models to ambiguous image data. In: Proceedings of 2020 Conference on Neural Information Processing Systems; 6–12 December 2020.
31. Kolotouros N, Pavlakos G, Jayaraman D, Daniilidis K. Probabilistic modeling for human mesh recovery. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 11605–11614.
32. Wehrbein T, Rudolph M, Rosenhahn B, Wandt B. Probabilistic monocular 3D human pose estimation with normalizing flows. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10–17 October 2021; Montreal, QC, Canada. pp. 11199–11208.
33. Gabeur V, Franco JS, Martin X, et al. Moulding humans: Non-parametric 3D human shape estimation from single images. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 27 October–2 November 2019; Seoul, Korea. pp. 2232–2241.
34. Xiu Y, Yang J, Cao X, et al. ECON: Explicit clothed humans optimized via normal integration. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 17–24 June 2023; Vancouver, BC, Canada. pp. 512–523.
35. Chibane J, Alldieck T, Pons-Moll G. Implicit functions in feature space for 3D shape reconstruction and completion. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 6970–6981.
36. Taubin G. A signal processing approach to fair surface design. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques; 6–11 August 1995; Los Angeles, California, USA. pp. 351–358.
37. Kim H, Nam H, Kim J, et al. LaplacianFusion: Detailed 3D clothed-human body reconstruction. *ACM Transactions on Graphics* 2022; 41(6): 1–14. doi: 10.1145/3550454.3555511
38. Quéau Y, Durou JD, Aujol JF. Normal integration: A survey. *Journal of Mathematical Imaging and Vision* 2018; 60: 576–593. doi: 10.1007/s10851-017-0773-x
39. Yu T, Zheng Z, Guo K, et al. Function4D: Real-time human volumetric capture from very sparse consumer RGBD sensors. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 19–25 June 2021; Online conference. pp. 5746–5756.
40. Smith D, Loper M, Hu X, et al. Facsimile: Fast and accurate scans from an image in less than a second. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 27 October–2 November 2019; Seoul, Korea. pp. 5330–5339.
41. Cao X, Santo H, Shi B, Okura F. Bilateral normal integration. In: Karlinsky L, Michaeli T, Nishino K (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2022 Workshops; 23–27 October 2022; Tel Aviv, Israel. Springer; 2023. pp. 552–567. doi: 10.1007/978-3-031-19769-7_32
42. Ma Q, Yang J, Ranjan A, et al. Learning to dress 3D people in generative clothing. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13–19 June 2020; Seattle, WA, USA. pp. 6469–6478.
43. Fu J, Li S, Jiang Y, et al. StyleGAN-human: A data-centric odyssey of human generation. In: Karlinsky L, Michaeli T, Nishino K (editors). *Lecture Notes in Computer Science*, Proceedings of Computer Vision—ECCV 2022 Workshops; 23–27 October 2022; Tel Aviv, Israel. Springer; 2023. pp. 1–19. doi: 10.1007/978-3-031-19787-1_1