# Learning to Implicitly Represent 3D Human Body From Multi-scale Features and Multi-view Images

ZHONGGUO LI, MAGNUS OSKARSSON, ANDERS HEYDEN

## Introduction

Capturing and reconstructing detailed 3D human body models from monocular images is a quite challenging task in computer vision. We propose a multi-scale features based method to learn an implicit function for 3D human body reconstruction from multi-view images in this paper. Our method has two main contributions:

- It is a model-free implicit function based method.
- The novel multi-scale features.

## Method

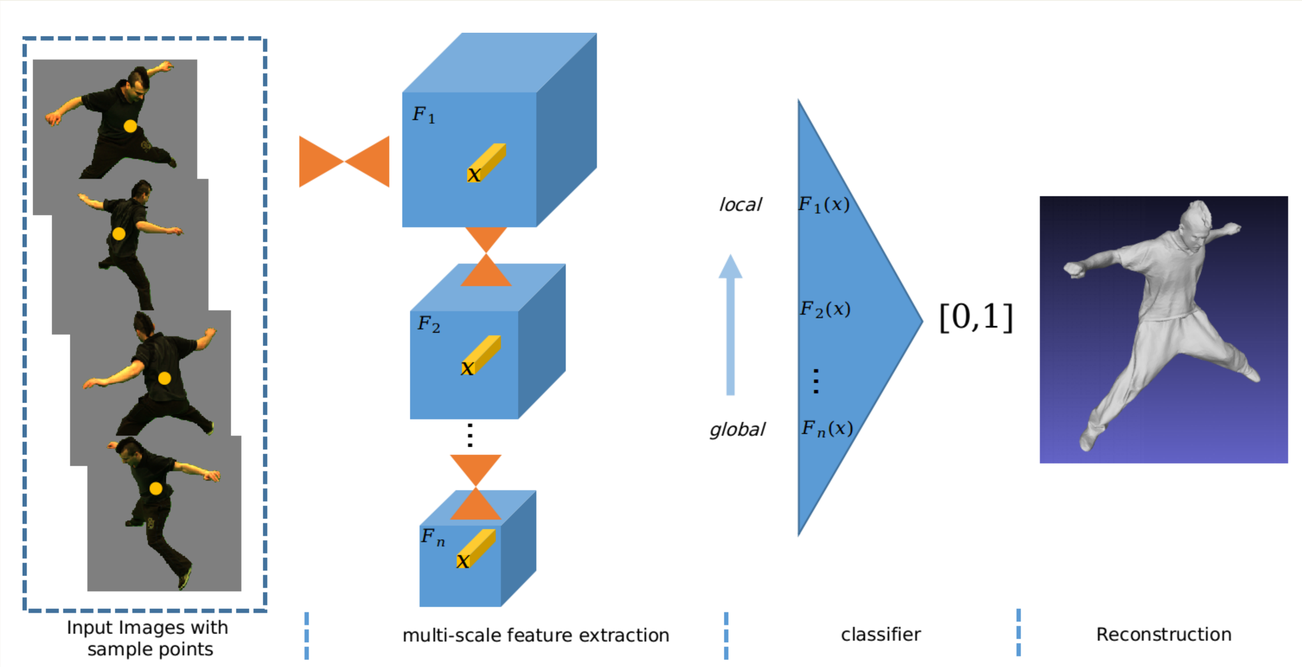



Fig. 1. The overview of our method. Multi-view images are fed into our model. Fi is the feature grids extracted by the hourglass network shown as the orange "►◄". For the point in the images (yellow "•"), the corresponding features can be extracted from multi-scale features. The features are passed to a classifier to decide the value of the implicit function representation. After training the model, we can reconstruct the 3D mesh from the implicit function.

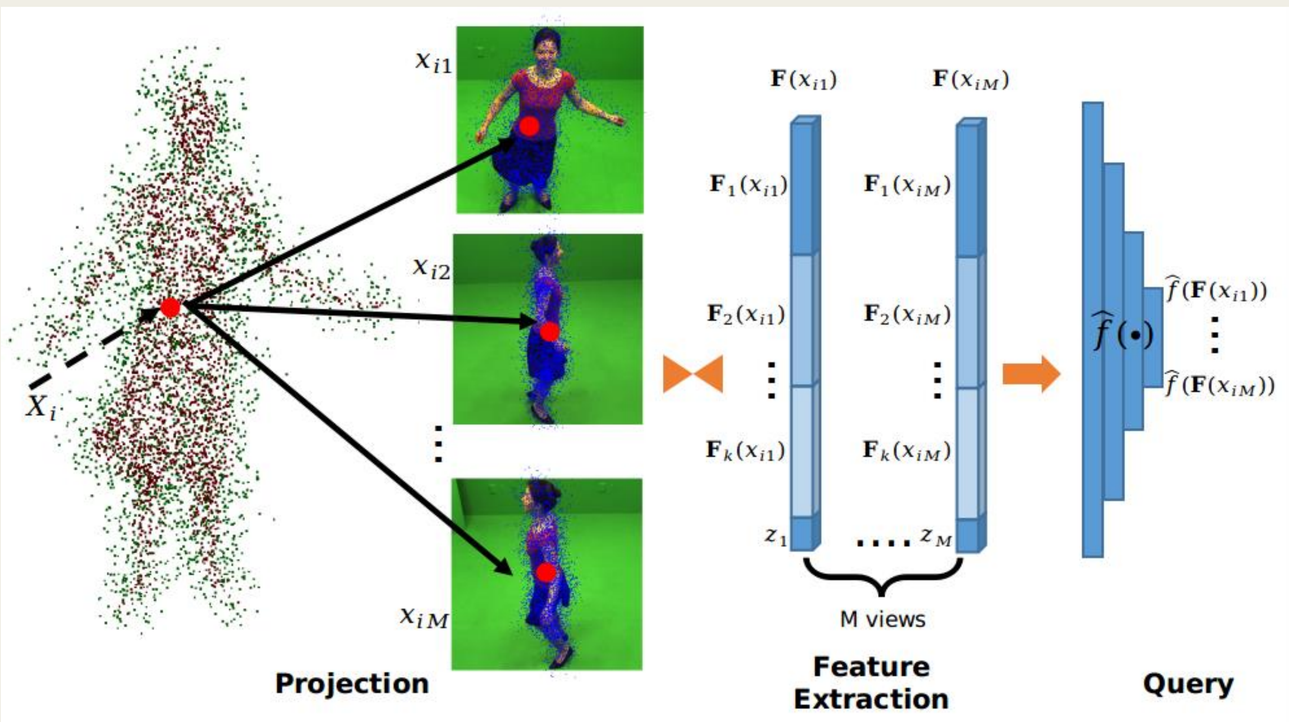


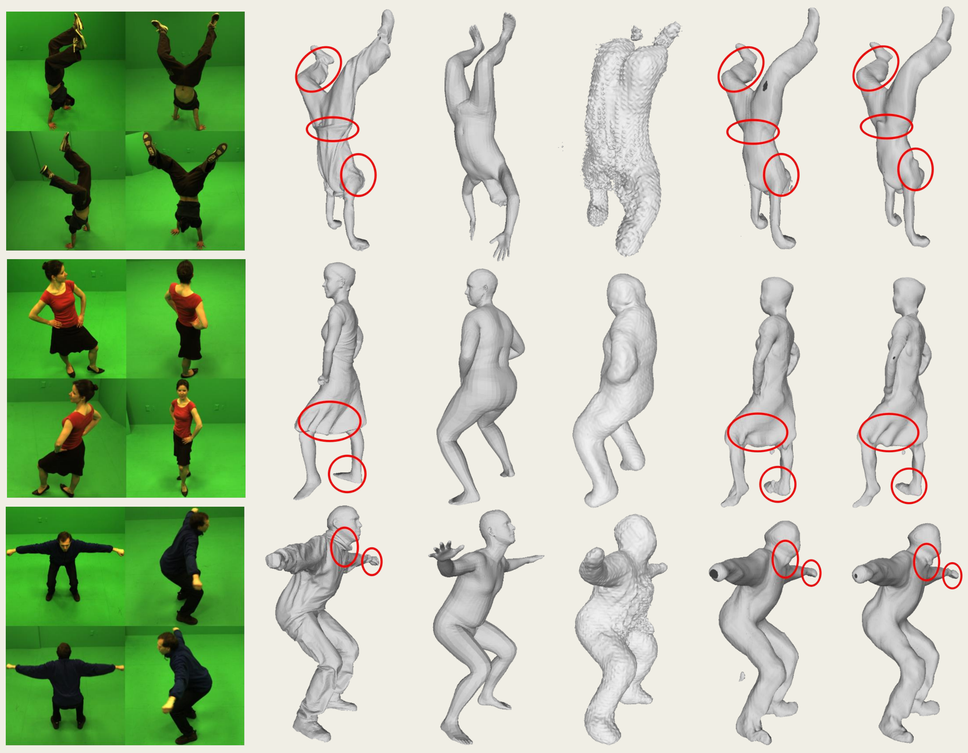

Fig. 2. The example of projection from 3D points to multi-view images, the multi-scale features extraction and query the multi-scale features

The loss function for learning the implicit function to represent the 3D human body model is defined as:

$$\mathcal{L}_f = \sum_{i=1}^{N}\sum_{j=1}^{M} L\left(\hat{f}\left(F^{(j)}(x_{ij})\right), o(\hat{X}_i)\right)$$

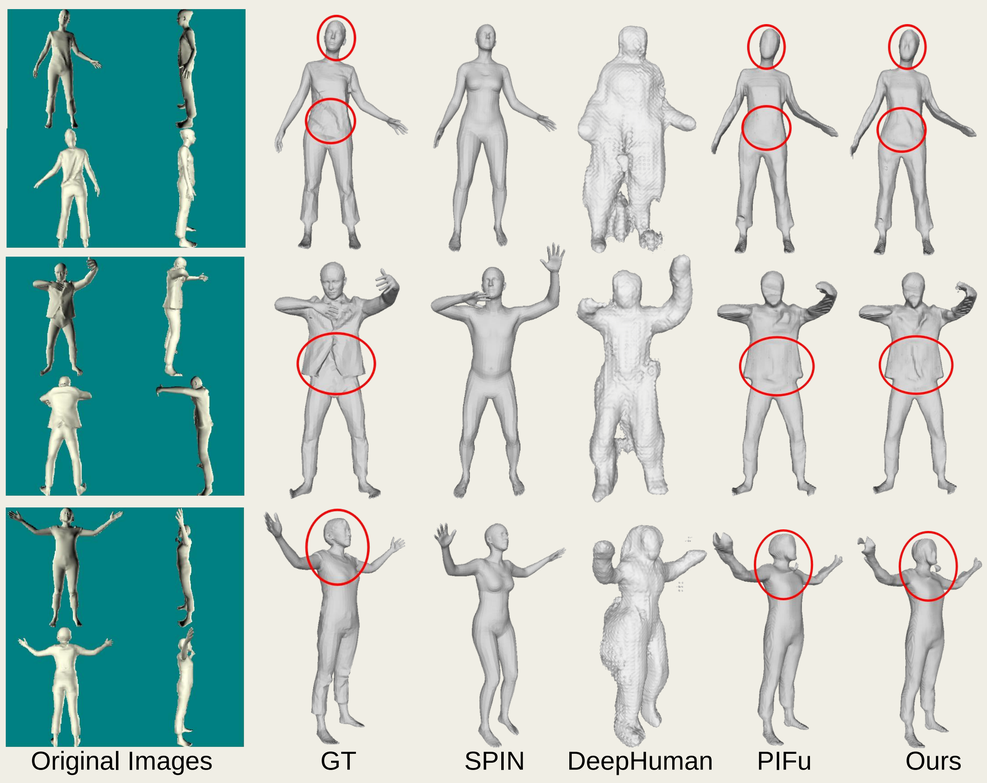
where $x_{ij}$ is the 2D projection on the j-th image for point $\hat{X}_i$. $o(\hat{X}_i)$ is the ground truth of occupancy value for $\hat{X}_i$. $L(\cdot)$ is the standard mean square error loss between $\hat{f}\left(F^{(j)}(x_{ij})\right)$ and $o(\hat{X}_i)$. Through minimizing $\mathcal{L}_f$, the multi-stage hourglass networks and the classifier fully connected network are trained.

## Experiments




| Methods | P2S ↓ | Chamfer-$L_2$ ↓ | IoU ↑ |
|---|---|---|---|
| SPIN | 3.5206 | 0.2679 | 0.3506 |
| DeepHuman | 3.9448 | 0.2675 | 0.3742 |
| PIFu | 0.8194 | 0.0210 | 0.8255 |
| Ours | **0.7332** | **0.0194** | **0.8484** |

Fig. 3: The qualitative results of previous methods and our method on the Articulated dataset




| Methods | P2S ↓ | Chamfer-$L_2$ ↓ | IoU ↑ |
|---|---|---|---|
| SPIN | 2.2134 | 0.1271 | 0.4044 |
| DeepHuman | 3.4028 | 0.1850 | 0.3861 |
| PIFu | 1.0330 | 0.0212 | 0.7571 |
| Ours | **0.9482** | **0.0196** | **0.7829** |

Fig. 4. The qualitative results of previous methods and our method on the CAPE dataset

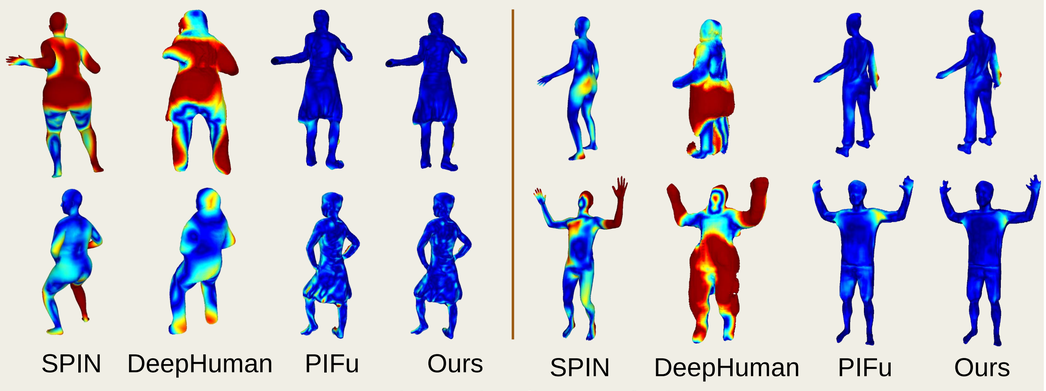



Fig. 5. Visualization of P2S

## Contact

Email: Zhongguo.li@math.lth.se