CVPR SEATTLE, WA JUNE 17-21, 2024

VS: Reconstructing Clothed 3D Human from Single Image via Vertex Shift

Leyuan Liu[a], Yuhan Li[a], Yunqi Gao[a], Changxin Gao[b], Yuanyuan Liu[c], Jingying Chen[a]

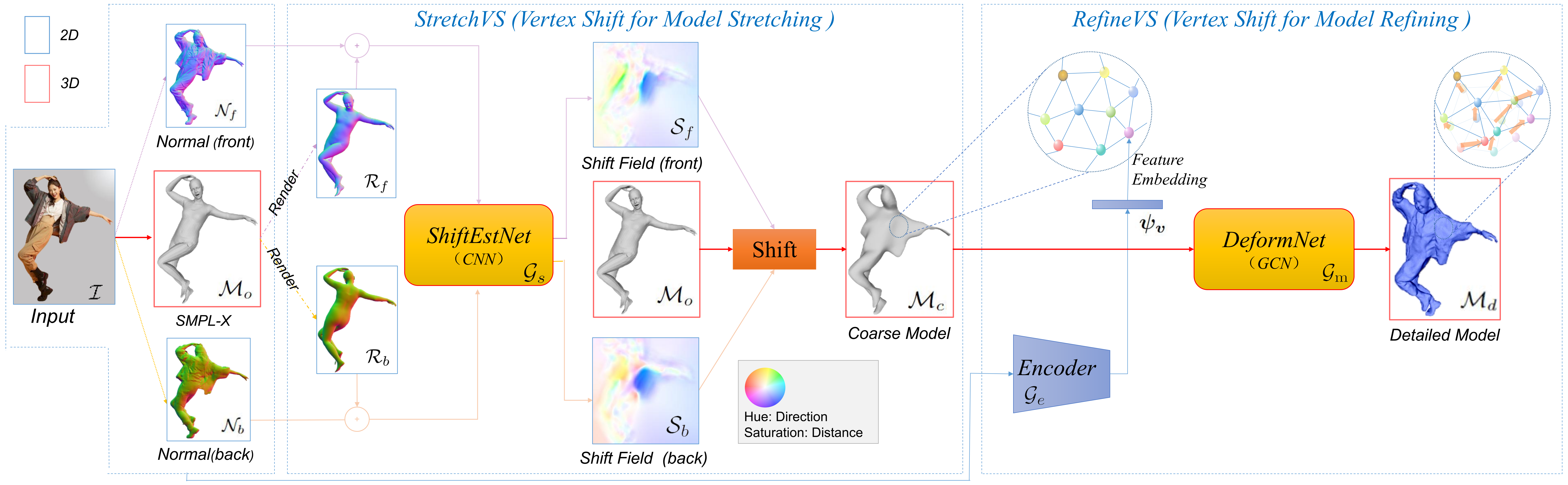
[a] Central China Normal University , [b] Huazhong University of Science and Technology, [c] China University of Geosciences (Wuhan)

Introduction

In this paper, we propose a two-stage **deformation method** named **Vertex Shift** (**VS**) to reconstruct *high-fidelity* and artifact-less clothed 3D humans from single images.

- ◆ We propose a two-stage deformation method that uses a "stretch-refine" strategy for clothed 3D human reconstruction, contributing to reconciling the contradiction between large deformations for reconstructing loose clothing and delicate formations for recovering surface details.
- ◆ We introduce shift fields inferred from normal maps for stretching the coarse model to align well with the input image, allowing our deformation method to handle loose clothing and correct inaccurate pose estimates.
- ◆ We combine implicit-function-learned features with a graph convolutional network, making VS not only recover surface details but also suppress artifacts.

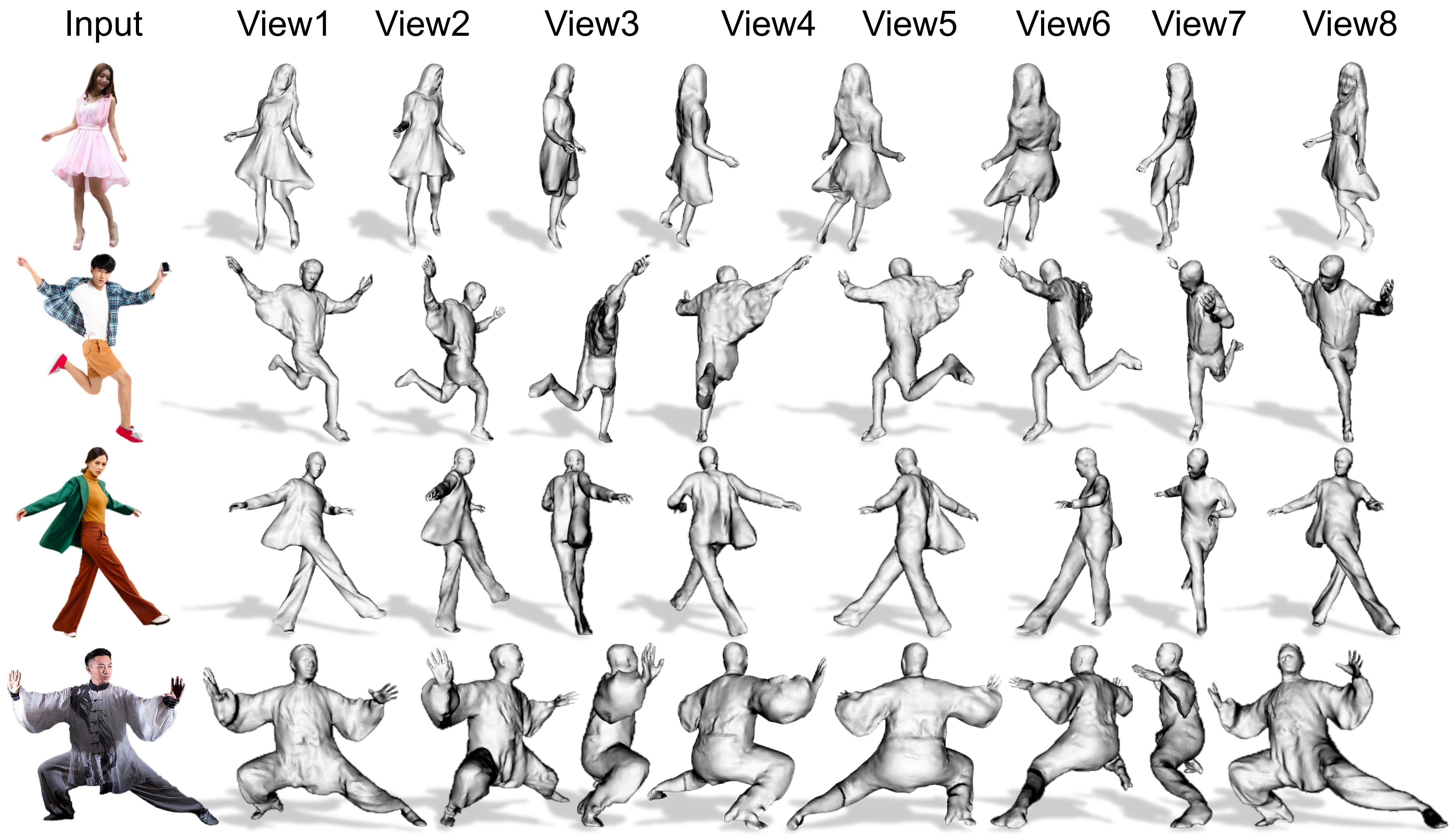
Method



- ◆ VS employs a "stretch-refine" strategy to stepwise deform the SMPL-X into a coarse human model and a detailed human model using the StretchVS and RefineVS modules, respectively.
- ◆ Two shift fields are inferred by warping the body normals into clothing normal maps via the ShiftEstNet (a CNN). Then, StretchVS shifts vertices of the SMPL-X to from the coarse model using the shift fields.
- ◆ Taking the coarse model as input, RefineVS employs the DeformNet (a GCN) embedded with implicit-function-learned features to infer vertex locations of the detailed human model.

Results

● Visual Results



● Quantitative Comparison

Methods	Publications	THuman 2.0				CAPE				RenderPeople			
		$\varepsilon_{cd}\downarrow$	$\varepsilon_{p2s}\downarrow$	$\varepsilon_{cos}\downarrow$	$\varepsilon_{l2}\downarrow$	$\varepsilon_{cd}\downarrow$	$\varepsilon_{p2s}\downarrow$	$\varepsilon_{cos}\downarrow$	$\varepsilon_{l2}\downarrow$	$\varepsilon_{cd}\downarrow$	$\varepsilon_{p2s}\downarrow$	$\varepsilon_{cos}\downarrow$	$\varepsilon_{l2}\downarrow$
PIFu	ICCV'19	1.760	1.904	0.0500	0.2408	2.967	2.738	0.0449	0.2409	2.781	2.857	0.0590	0.2773
PIFuHD	CVPR'20	3.088	3.113	0.0891	0.3663	4.714	3.823	0.0555	0.2796	3.311	3.3118	0.0846	0.3541
PaMIR	TPAMI'22	1.064	1.185	0.0438	0.1927	1.772	1.404	0.0337	0.1676	1.580	1.659	0.0486	0.2088
ICON	CVPR'22	0.947	0.925	0.0422	0.1761	1.133	1.096	0.0311	0.1431	1.265	1.251	0.0431	0.1871
ECON	CVPR'23	0.906	0.845	0.0379	0.1891	0.937	0.921	0.0335	0.1644	1.285	1.079	**0.0417**	0.1973
VS	Ours	**0.628**	**0.555**	**0.0373**	**0.1555**	**0.621**	**0.615**	**0.0262**	**0.1138**	**0.976**	**0.788**	0.0419	**0.1618**

Conclusion

- ◆ We propose VS to reconstruct high-fidelity and artifact-less clothed 3D humans from single images. Extensive experiments on five datasets demonstrate that VS can reconstruct high-fidelity and artifact-less clothed 3D humans and achieves SOTA performance.
- ◆ VS confirms that deformation methods can reconstruct high-quality clothed 3D humans with complex poses and loose clothing, and even have advantages over IF-based methods in eliminating artifacts.