Distribution-aligned Diffusion for Human Mesh Recovery

Lin Geng Foo[1], Jia Gong[1], Hossein Rahmani[2], and Jun Liu[1]

[1] Singapore University of Technology and Design; [2] Lancaster University

ICCV23 PARIS

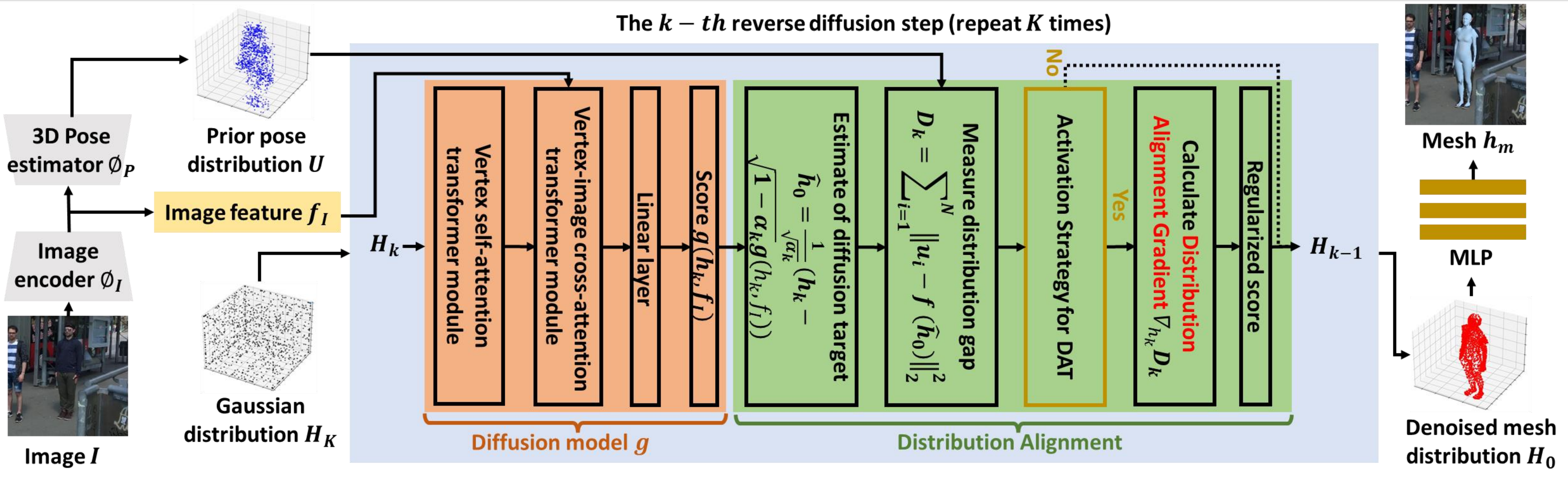
Motivation

- **Goal:** Eliminate the uncertainty in 3D mesh recovery caused by depth ambiguity and self-occlusion.
- **Motivation:** Diffusion models have become popular as an effective way to generate high-quality samples (e.g., image, video, text) starting from random noise via progressive denoising, displaying a strong ability to recover high-quality outputs from uncertain and noisy input data. Inspired by this capability, we seek to **recover a high-quality mesh prediction from uncertain and noisy input data via diffusion models** for human mesh recovery.

Key Design

- To tackle the challenging human mesh recovery via diffusion models, **we formulate human mesh reconstruction as a (reverse) human mesh diffusion process** to obtain denoised mesh distribution H_0 from a Gaussian noise distribution H_K via our *HMDiff framework*.
- To simplify the task of monocular HMR, **we propose to estimate the posterior to guide the initial stages of the diffusion process.** Specifically, we align the initial mesh distribution towards an extracted prior pose distribution via our proposed *Distribution Alignment Technique (DAT)*.

Our Proposed Method



HMDiff (Human Mesh Diffusion Process):

Forward Process: To generate intermediate distributions $\{H_1, H_2, \ldots, H_{K-1}\}$ as step-by-step supervisory signals for training, we apply noise to the vertex coordinates while keeping the topology between vertices fixed.

Reverse Process: We build our mesh reverse diffusion process based on DDIM formulation, which 1) reduces the number of diffusion steps required during inference; 2) offers a convenient way to jump over k steps to obtain an early prediction to facilitate our DAT. Using this mesh reverse diffusion process, we progressively denoise the input distribution H_K into the target H_0.

DAT (Distribution Alignment Technique):

Firstly, we extract and use a **prior distribution $\boldsymbol{U}$** that strongly correlates to H_0. Our prior distribution U is an extracted pose heatmap which contains rich semantic and uncertainty information.

At the k-th step, we compute a Distribution Alignment Gradient using U which guides us to align samples h_k**, such that after $\boldsymbol{k}$ diffusion steps, the prediction $\boldsymbol{\hat{h}_0}$ is pulled closer to $\boldsymbol{U}$** (and the target H_0), in a way that does not disrupt the diffusion process.

We also introduce an **Activation Strategy** for DAT, which deactivates DAT when H_k converges to a more compact and high-quality distribution.

Experimental Results

Results on 3DPW and Human3.6M.

Method	3DPW MPVE↓	3DPW MPJPE↓	3DPW PA-MPJPE↓	Human3.6M MPJPE↓	Human3.6M PA-MPJPE↓
Kanazawa et al. [20]	-	-	81.3	88.0	56.8
GraphCMR [27]	-	-	70.2	-	50.1
SPIN [26]	116.4	-	59.2	-	41.1
Pose2Mesh [7]	-	89.2	58.9	64.9	47.0
I2LMeshNet [39]	-	93.2	57.7	55.7	41.1
VIBE [24]	99.1	82.0	51.9	65.6	41.4
METRO [32]	88.2	77.1	47.9	54.0	36.7
Mesh Graphormer [33]	87.7	74.7	45.6	51.2	34.5
FastMETRO [6]	84.1	73.5	44.6	52.2	33.7
Ours	**82.4**	**72.7**	**44.5**	**49.3**	**32.4**

Results on FreiHAND.

Method	PA-MPVPE↓	PA-MPJPE↓	F@5 mm↑	F@15 mm↑
Hasson et al. [15] *	13.2	-	0.436	0.908
Boukhayma et al. [5] *	13.0	-	0.435	0.898
FreiHAND [63] *	10.7	-	0.529	0.935
Pose2Mesh [7]	7.8	7.7	0.674	0.969
I2LMeshNet [39]	7.6	7.4	0.681	0.973
METRO [32]	6.7	6.8	0.717	0.981
Mesh Graphormer [33]	5.9	6.0	0.764	0.986
FastMETRO [6]	-	6.5	-	0.982
Ours	**5.7**	**5.6**	**0.781**	**0.986**

Visualization