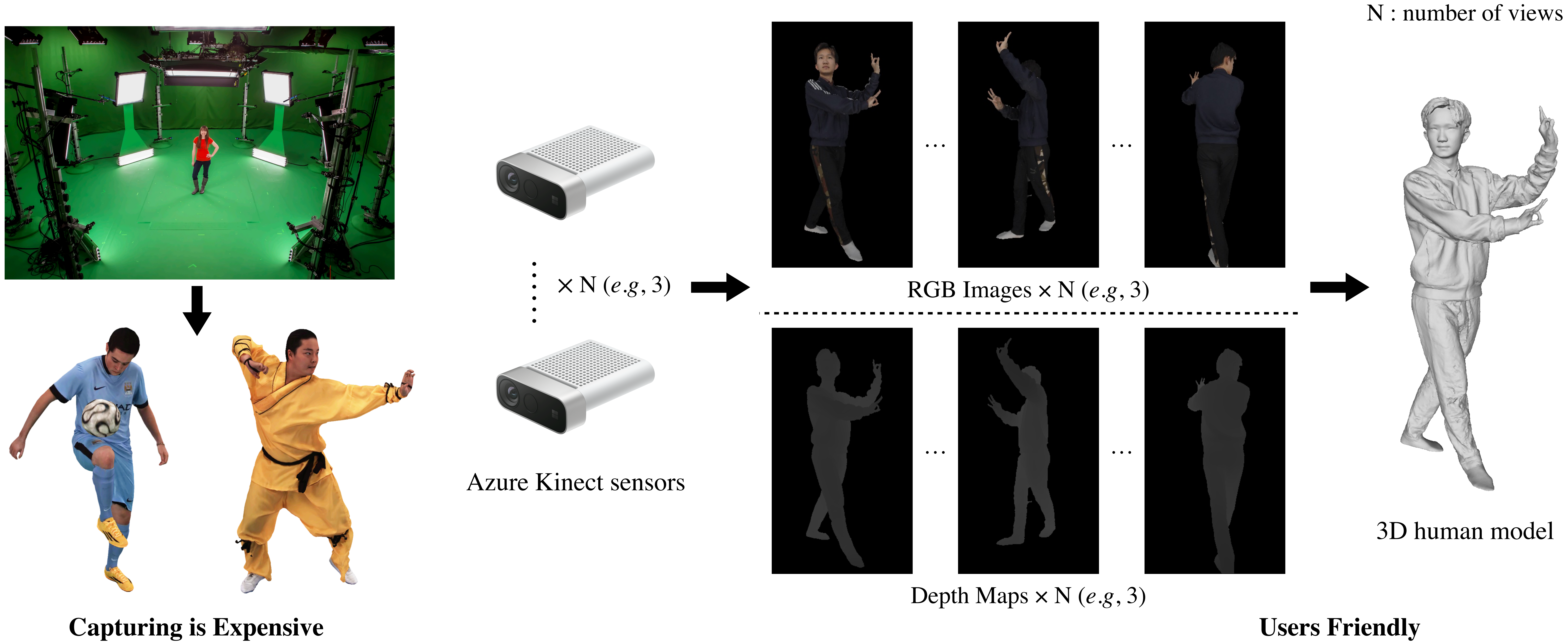
Geometry-aware Two-scale PIFu Representation for Human Reconstruction

Zheng Dong[1], Ke Xu[2], Ziheng Duan[1], Hujun Bao[1], Weiwei Xu[*1], Rynson W.H. Lau[2]

[1]State Key Lab of CAD&CG, Zhejiang University [2]City University of Hong Kong

Our Goal

Reconstruct high-quality human models in a sparse (*e.g.*, 3 RGBD sensors) capture setting.



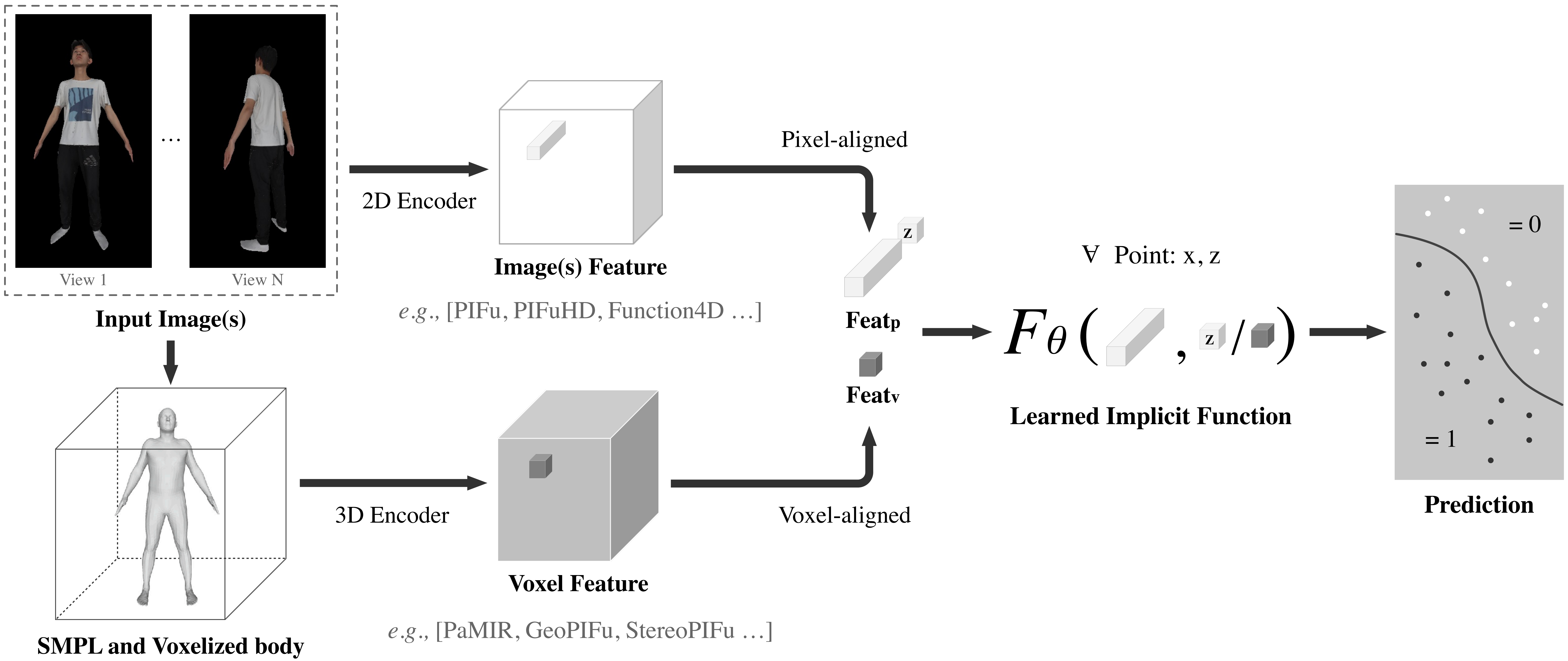
Collet et al. High-quality streamable free-viewpoint video (ACM TOG 2015)

THuman2.0-Dataset : https://github.com/ytrock/THuman2.0-Dataset

Human Reconstruction from Very Sparse Views

Due to ill-posed properties (*e.g.*, severe occlusion, input noise) in the very sparse capture settings, deep learning-based methods are applied to human reconstruction.

- Explicit shape regression : SiCloPe [CVPR 2019], DeepHuman [ICCV 2019], NormalGAN [ECCV 2020] …
- Learning **Implicit fields** : **PIFu** [ICCV 2019], StereoPIFu [CVPR 2021], Function4D [CVPR 2021] …

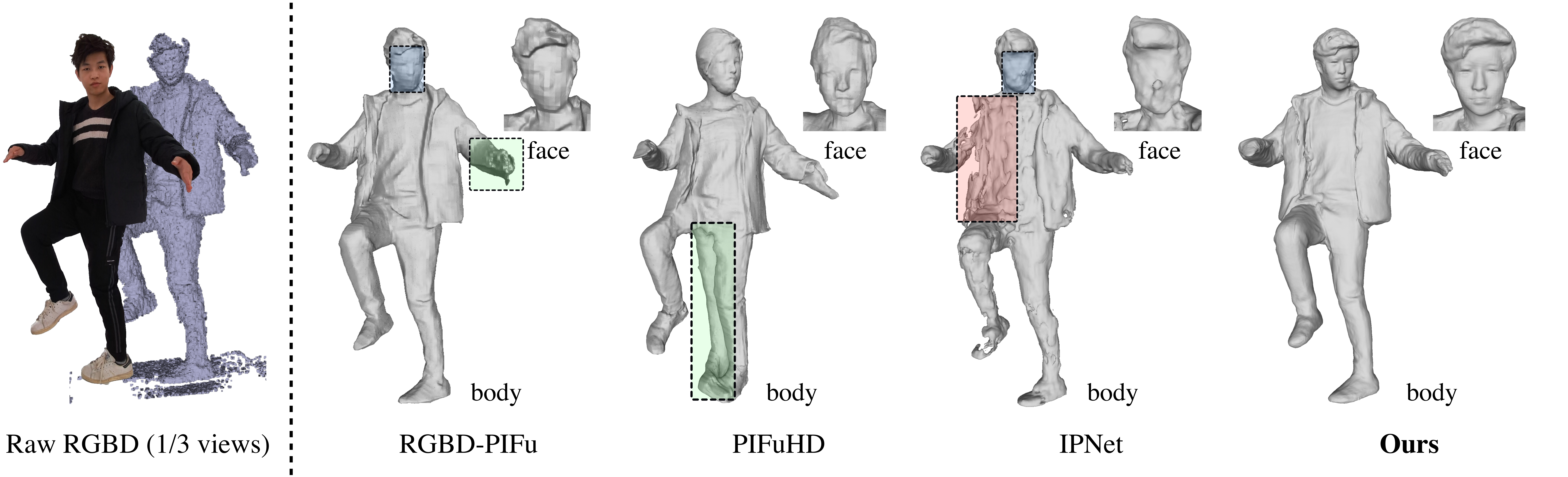


Human reconstruction from pixels (RGB, Depth) in an implicit manner

Our Observations and Contributions

The reconstructed quality of the existing implicit methods is still unsatisfactory under the sparse capture settings.

- Body geometries with topology errors (depth noise is amplified under sparse views)
- Lack of high-frequency details (*e.g.*, Flat or incorrect facial surfaces, hair geometries)



Saito et al. PIFu : Pixel-aligned implicit function for high-resolution clothed human digitization (ICCV 2019)

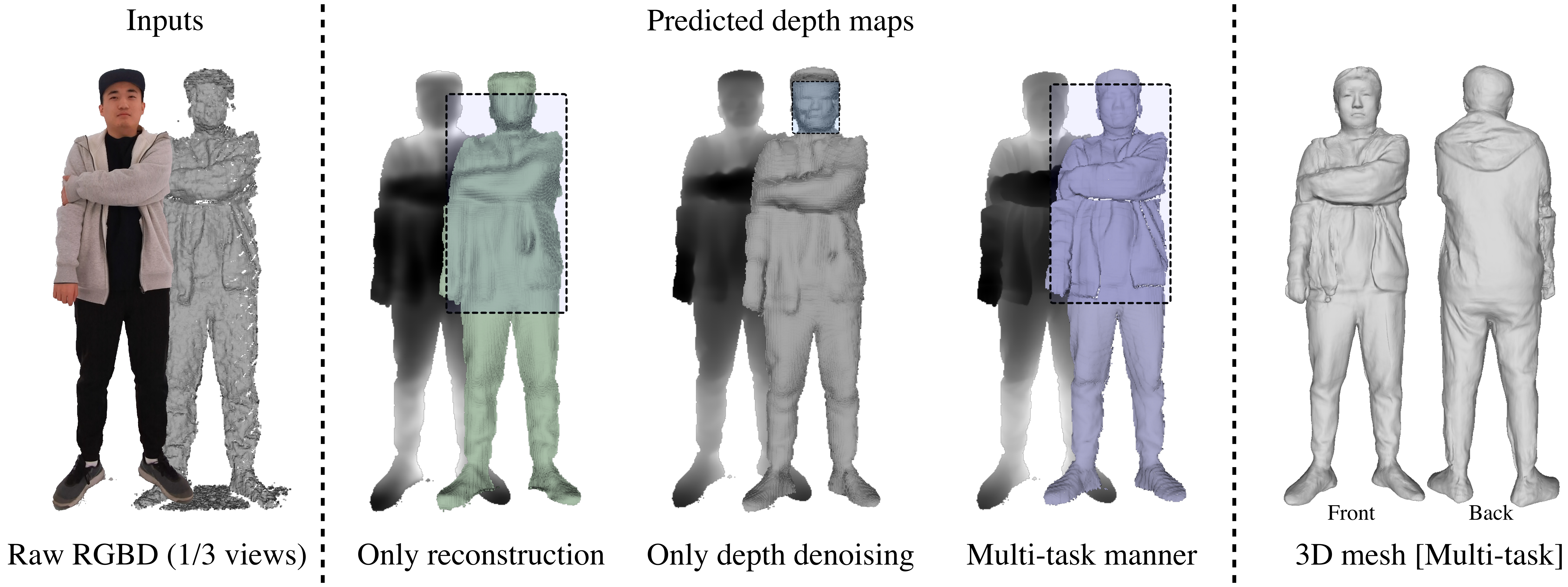
Saito et al. PIFuHD : Multi-level pixel-aligned implicit function for high-resolution 3d human digitization (CVPR 2020)

Bhatnagar et al. Combining implicit function learning and parametric models for 3d human reconstruction (ECCV 2020)

Multi-task Formulation

The two tasks, **depth denoising** and **3D reconstruction** are complementary to each other.

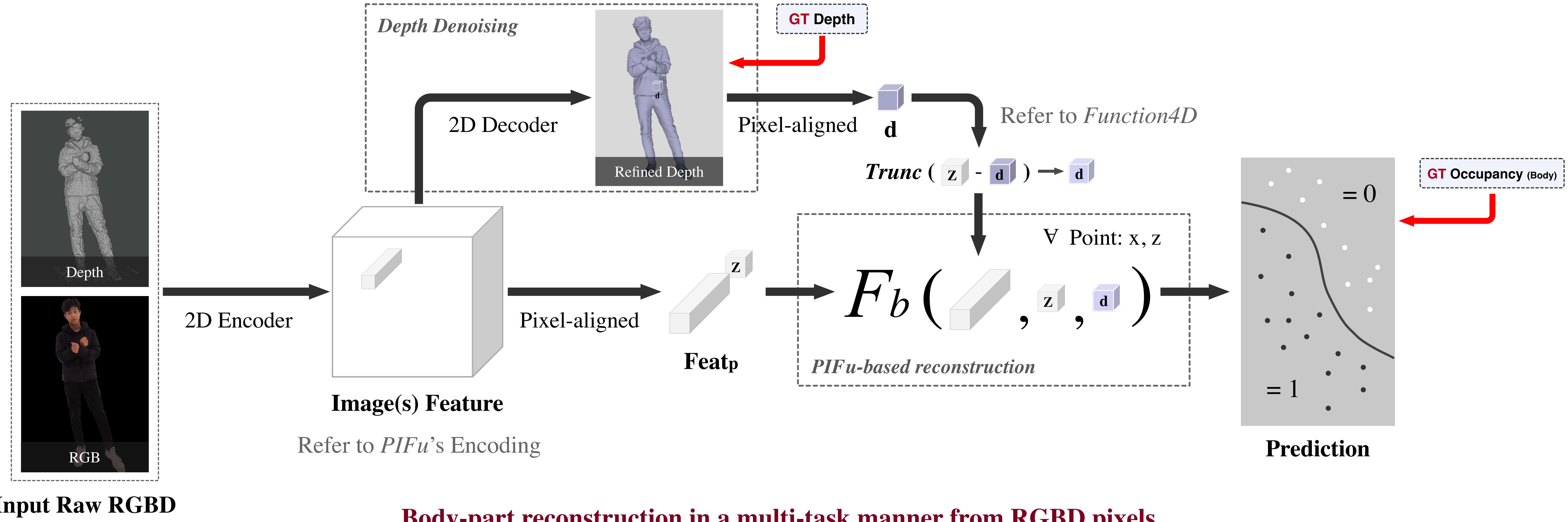
- Image-to-image **depth denoising** : preserves local geometric fidelity, prone to introducing incorrect details
- PIFu-based **3D reconstruction** : provides global topology guidance, lacking high-frequency local details



Multi-task Formulation

Proposed : formulate **depth denoising** and **body reconstruction** processes in a multi-task learning manner to exploit their complementary properties.

- **Depth denoising** and **Occupancy estimation** tasks share the image(s) features

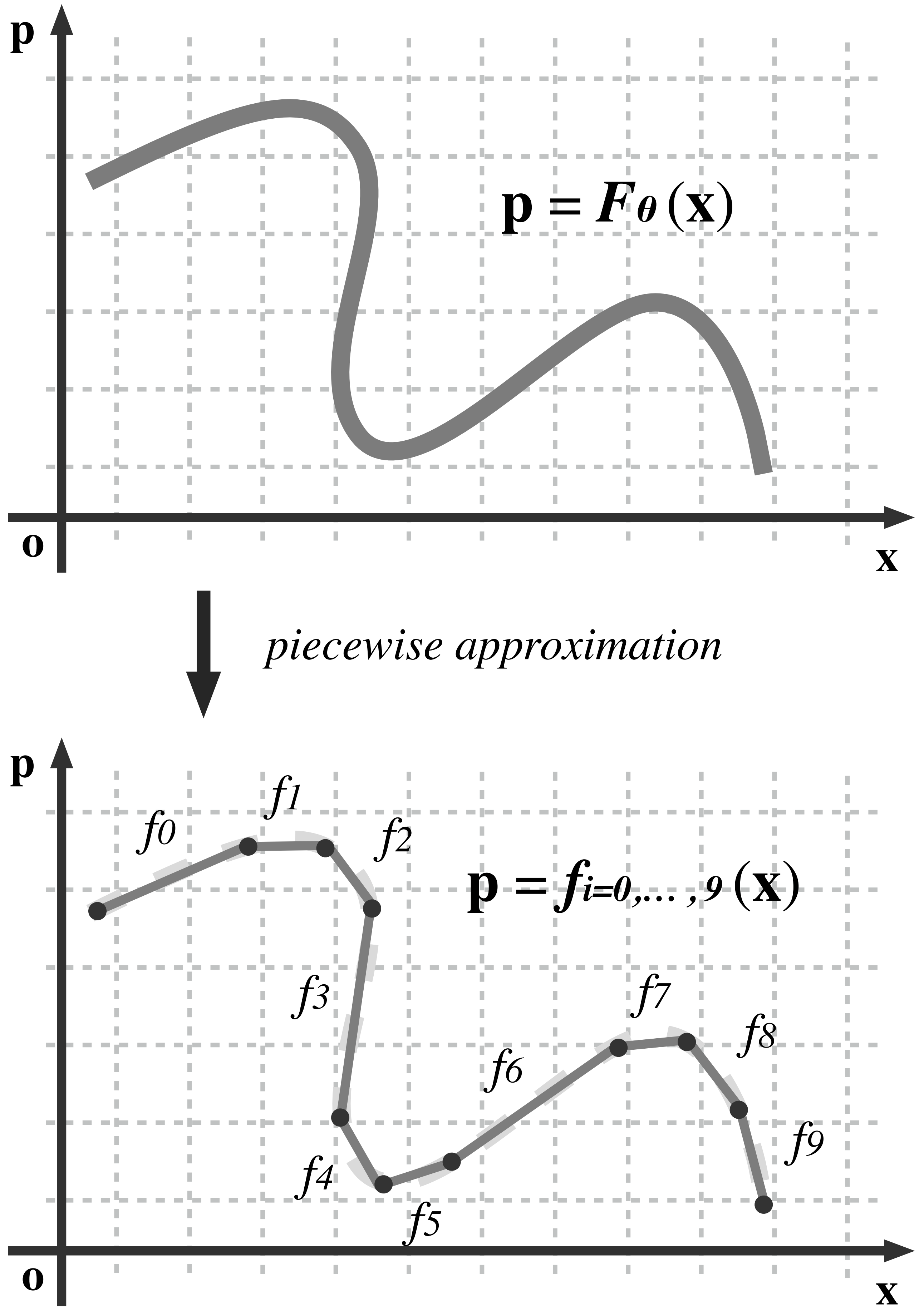


Body-part reconstruction in a multi-task manner from RGBD pixels

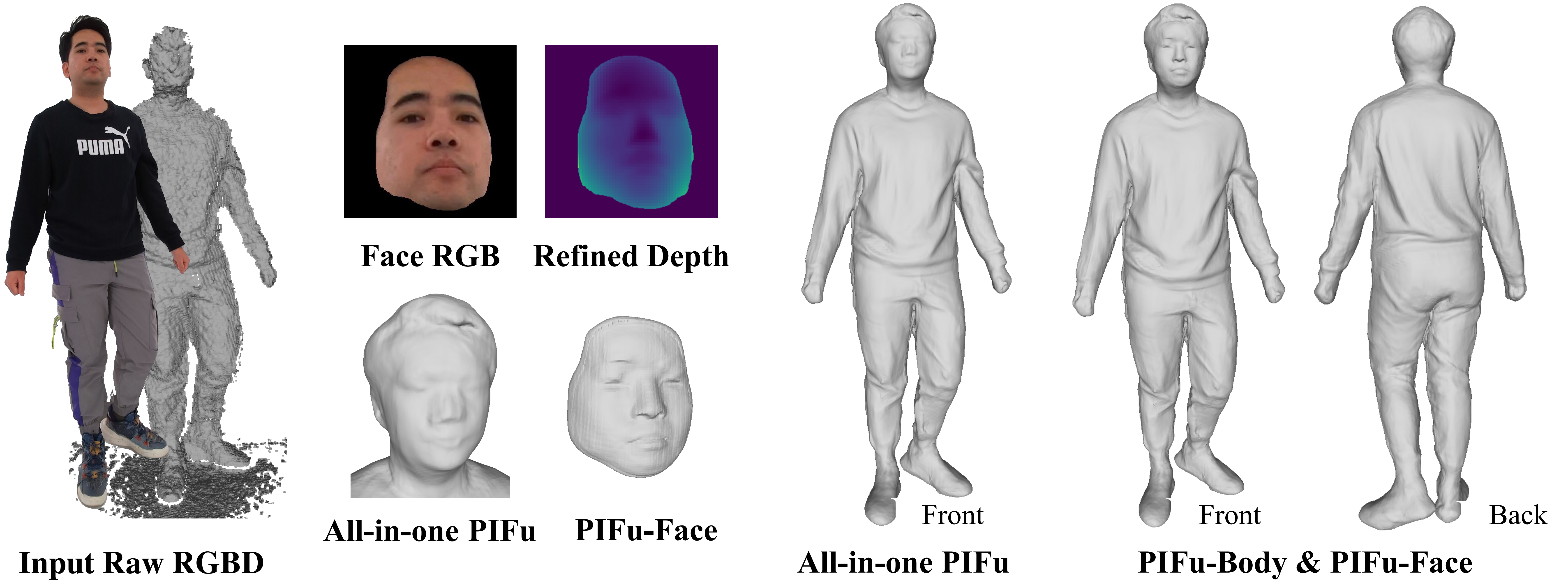
Yu et al. Function4D: Real-time Human Volumetric Capture from Very Sparse Consumer RGBD Sensors (CVPR 2021)

PIFu-Body and PIFu-Face

A function F_θ of high complexity may not be easy to express, but it is much easier to approximate F_θ piecewise (*e.g.*, in two parts) for high-frequency information.



- **Face surface** typically contains more high-frequencies (*e.g.*, vivid expression) than other parts, and plays a vital role when assessing the reconstruction quality

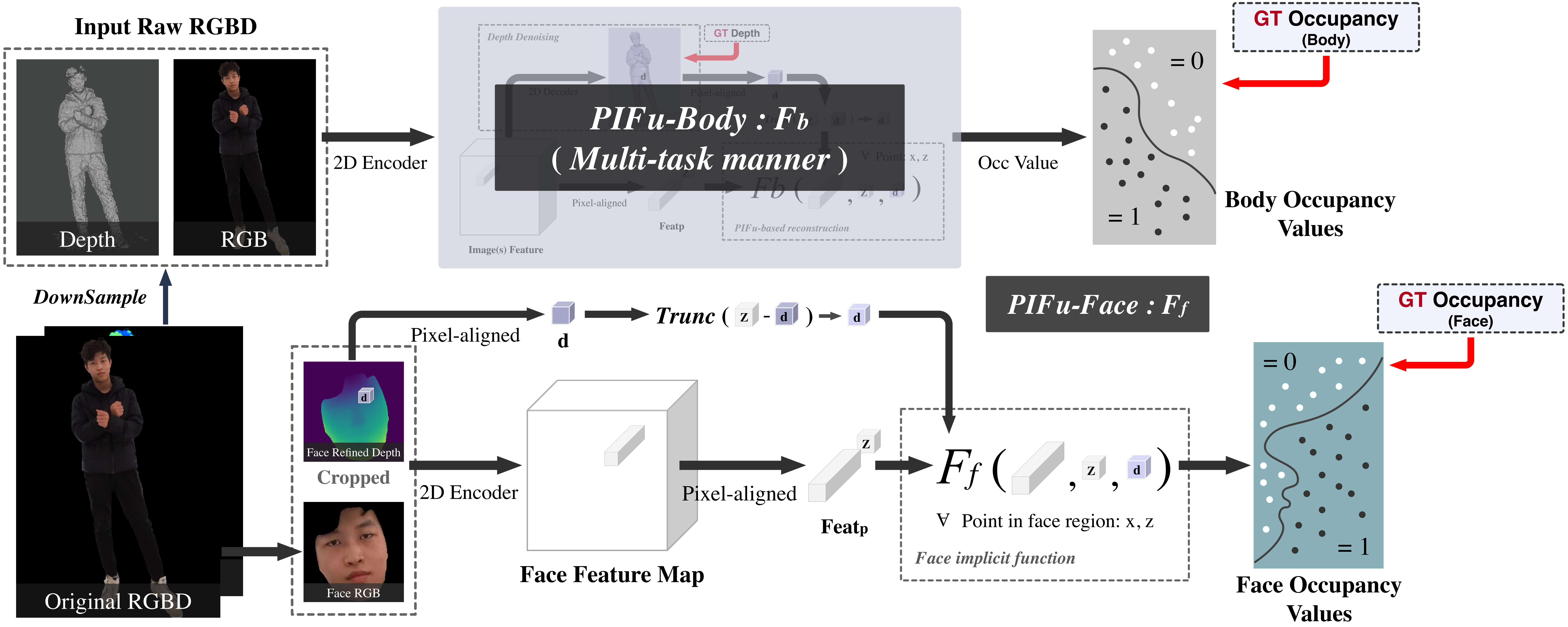


Face and body reconstruction based on piecewise (two parts) approximation

PIFu-Body and PIFu-Face

Proposed : express the implicit function F_θ in a **piecewise manner** (*i.e.*, PIFu-Body: F_b and PIFu-Face: F_f) to reduce the complexity of joint occupancy estimation while producing vivid facial and body details.

- The PIFu-Face F_f is conditioned on the high-resolution face image and the denoised facial depth map
- F_f can be pretrained on the existed 3d Face Dataset (*e.g.*, ***FaceScape***) to introduce stronger **priors**

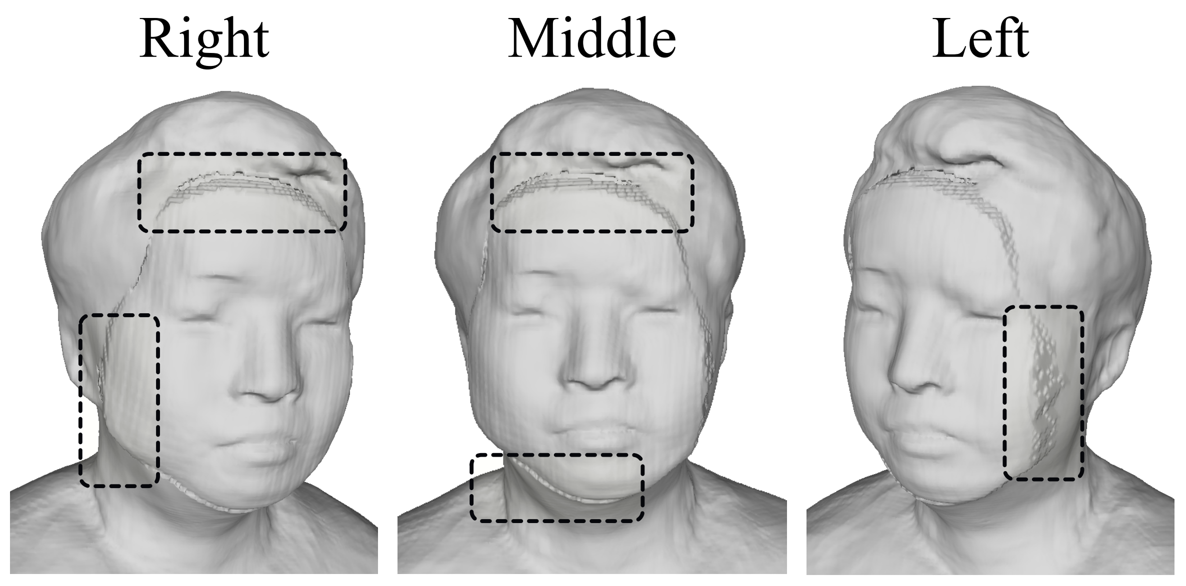


Face-to-body Occupancy Fields Fusion

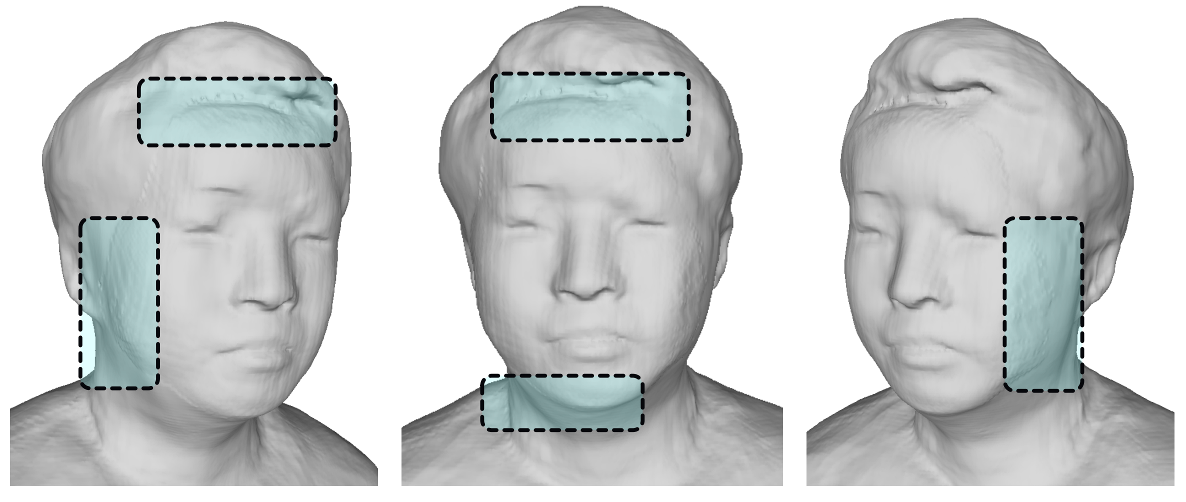
Simply merging the reconstructed face and body (*i.e.*, replacing body occupancy value : o_b with face occupancy value : o_f for the facial points) would result in the discontinuity artifacts at the stitching.

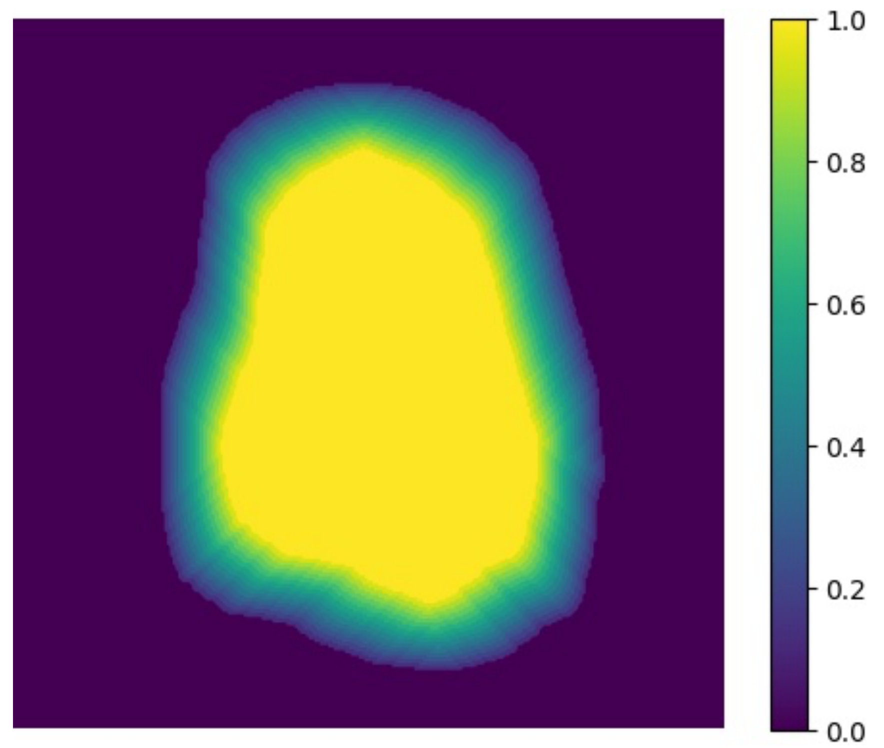
Proposed : fuse the face and body occupancy fields (O_b, O_f) via adaptive weights ω calculated in 3D space.

I. In x-y plane, compute a 2D fusion weight map via eroding edges of the facial mask
II. Along z axis, compute the weights through a Gaussian distribution model of the PSDF values

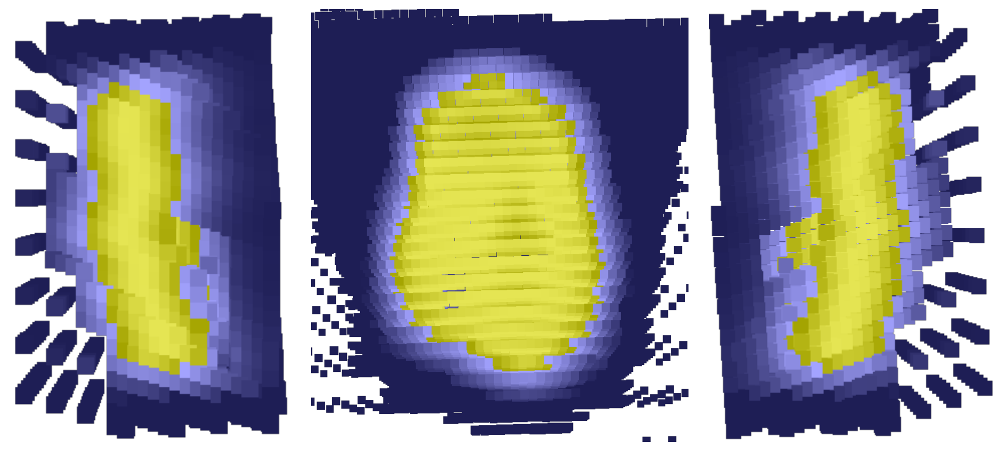


Simply merging face and body

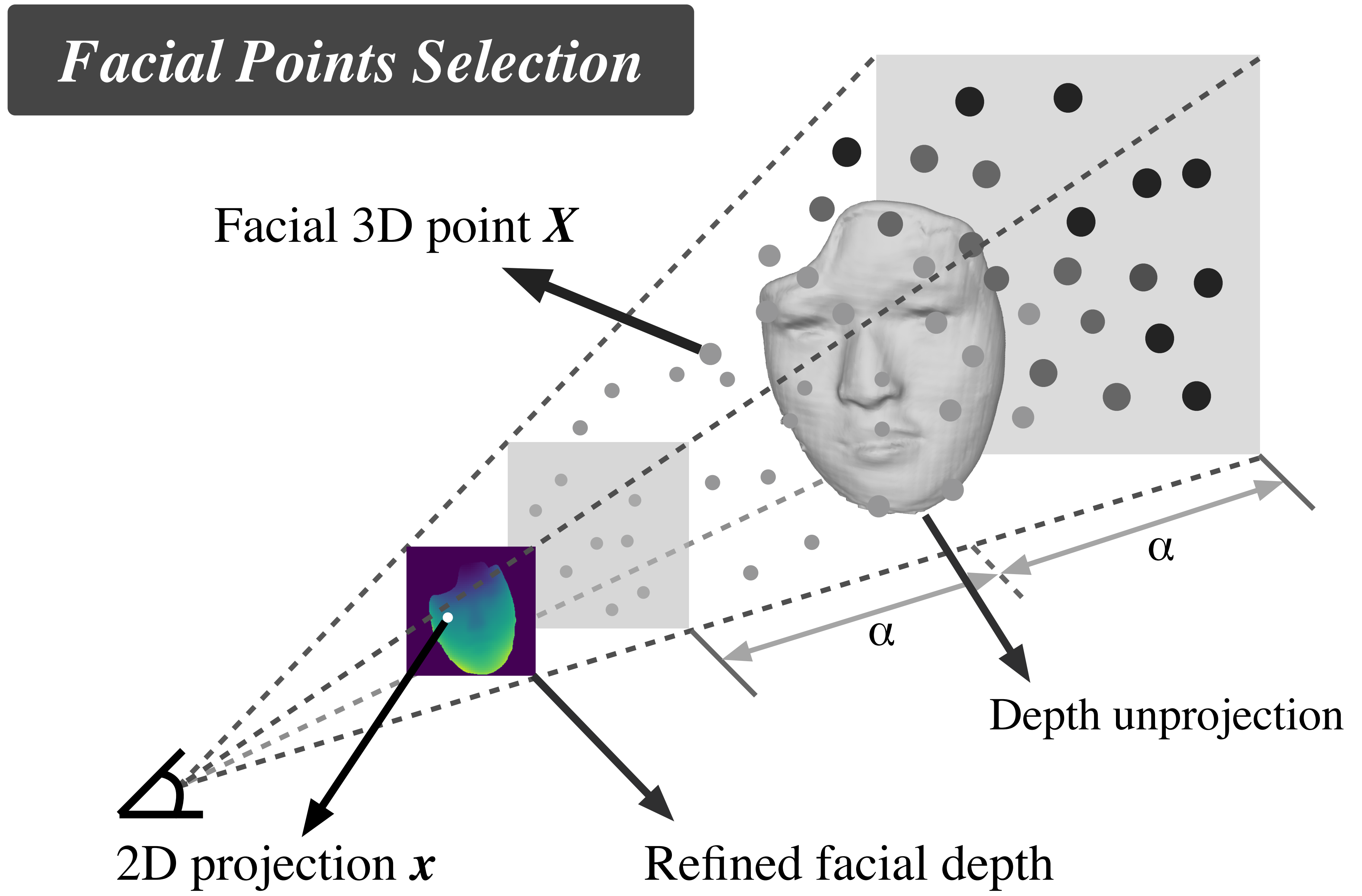

Adaptive face-to-body Fusion (**Ours**)



Our Fusion Weight Map: $\mathbf{M}_f^e$



Weights of 3D Points (Face Region)



For the facial point X, the fused occupancy value :

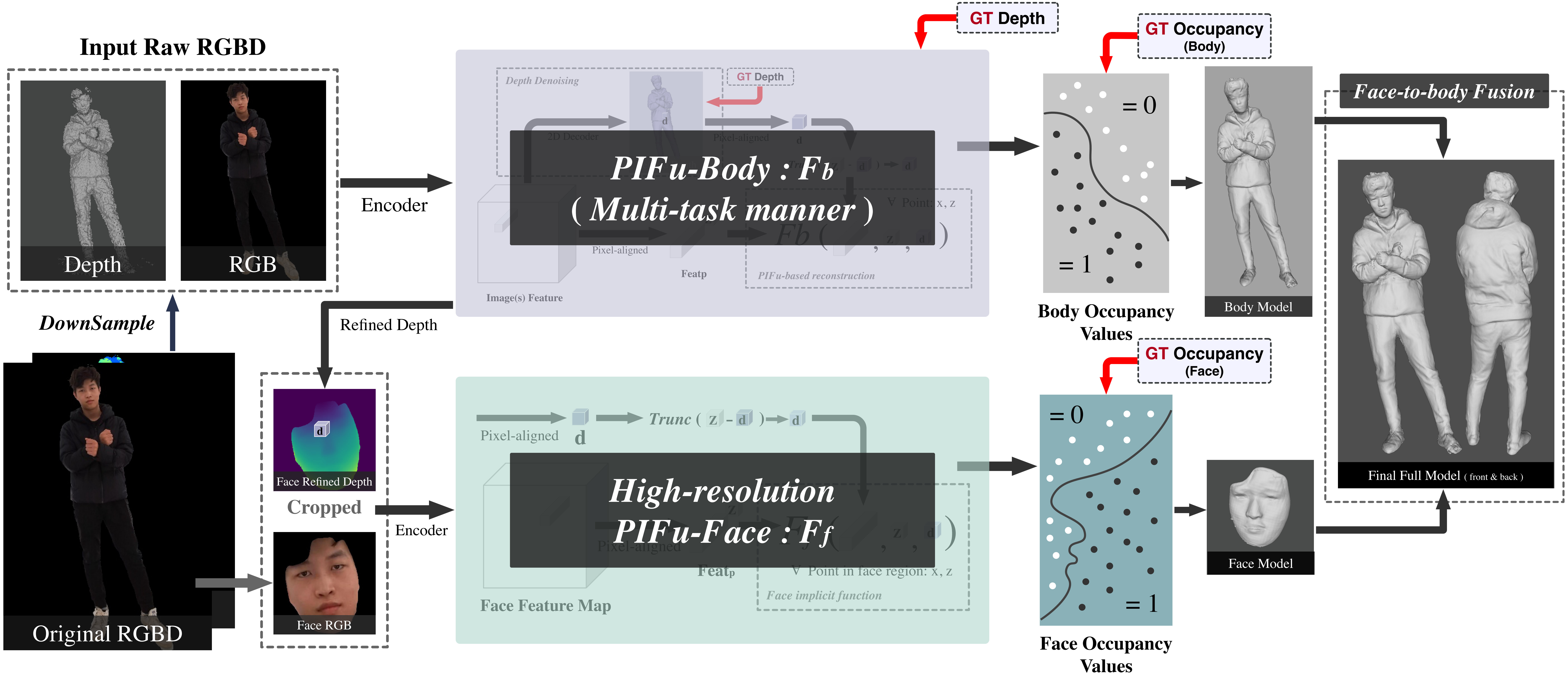
$$o(\mathbf{X}) = \omega \cdot o_f(\mathbf{X}) + (1-\omega) \cdot o_b(\mathbf{X})$$

Pipeline Overview

I. PIFu-Body : F_b predicts the body field O_b and the refined depth maps from sparse and noisy RGBDs

II. PIFu-Face : F_f obtains the fine-grained face field O_f, using the refined depth from F_b and the high-resolution face RGB

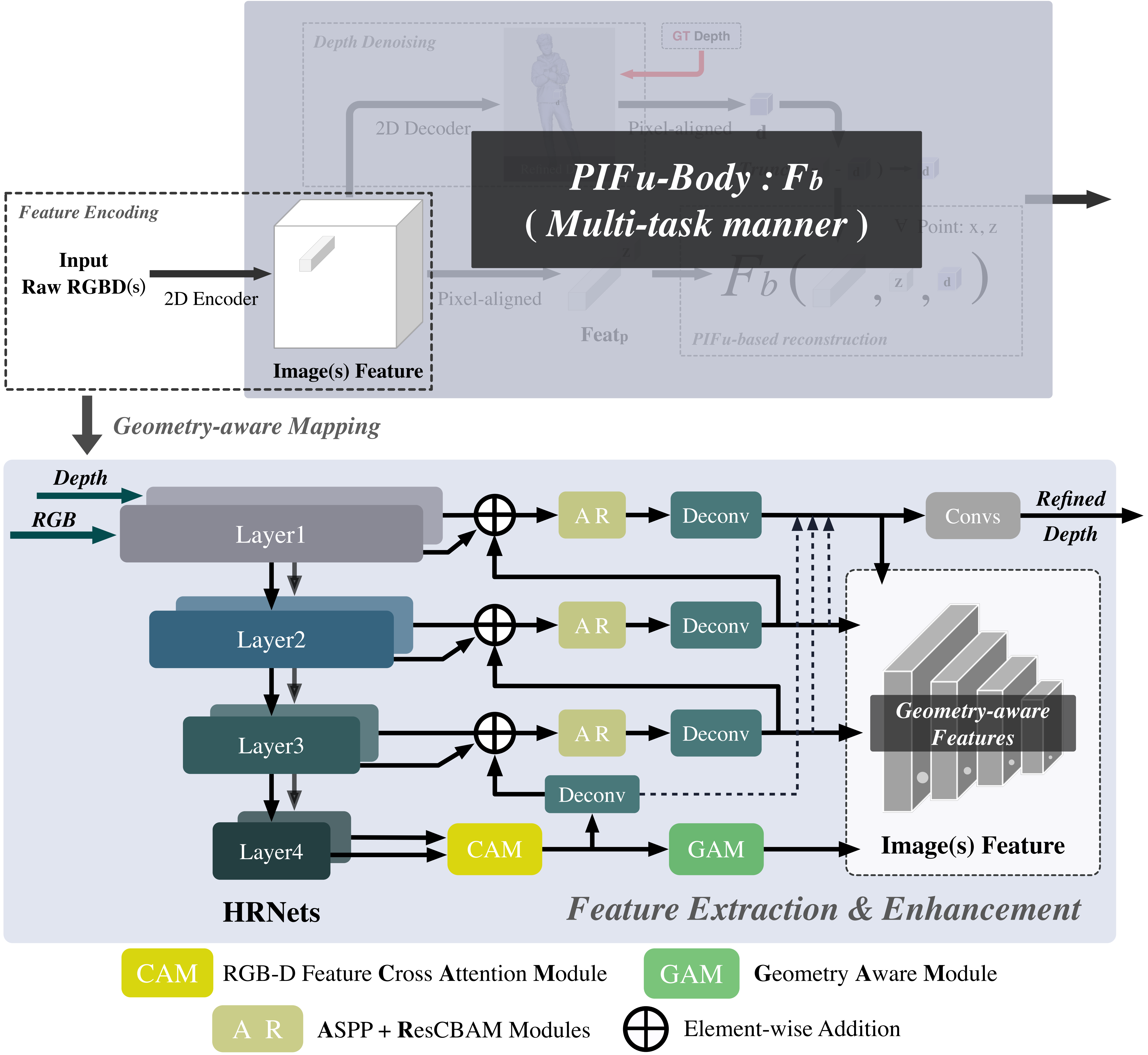
III. Face-to-body Fusion : W reconstructs full human model by fusing O_b and O_f



Geometry-aware Features

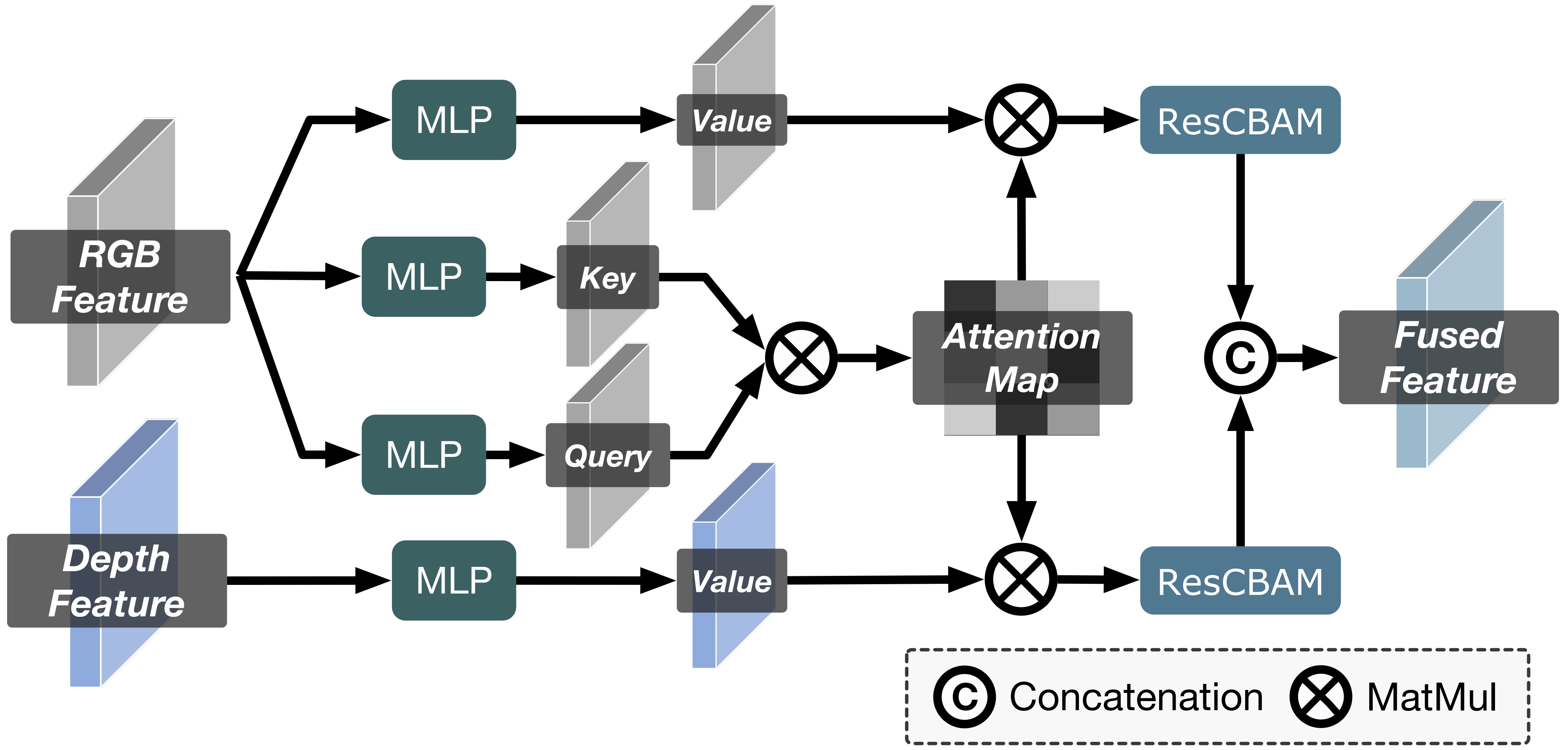
The encoded image(s) features aim to exploit the complementary properties of depth denoising and occupancy field estimation.

- Two ***HRNets*** are used to encode the RGB and Depth data respectively for handling modal discrepancy
- A novel **Cross Attention Module** (CAM) is proposed to fuse RGB and Depth top-level backbone features
- A novel **Geometry Aware Module** (GAM) is proposed to enrich the CAM output features with high-frequency information
- The enhanced features and the fused low-level features form the **Geometry-aware Features**



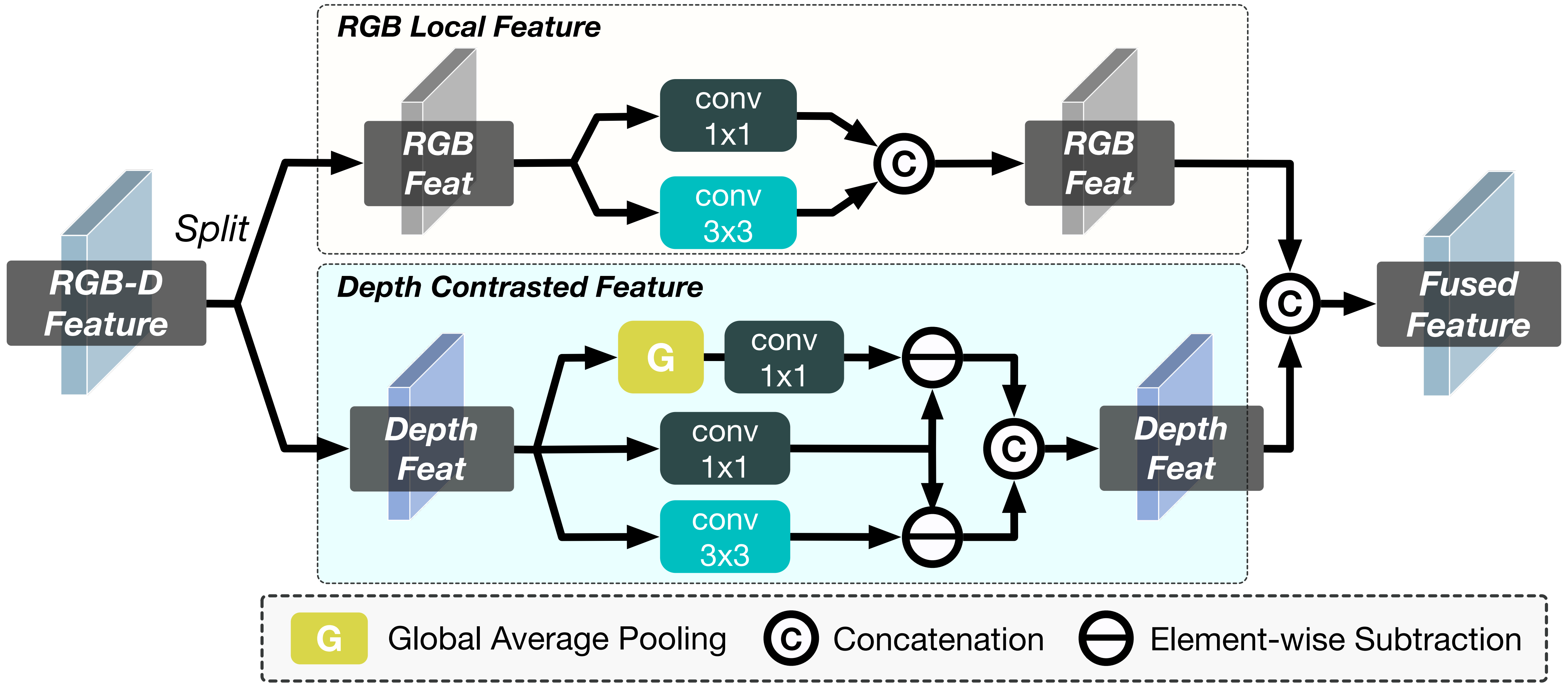
CAM and GAM

Cross Attention Module (CAM)



- Fuse the RGB and depth features by computing their non-local correlations

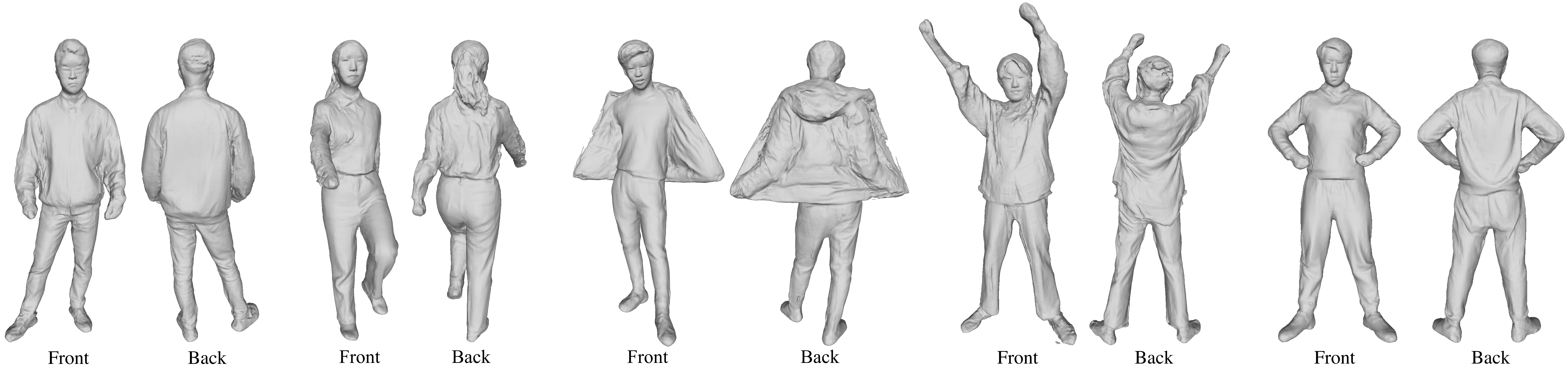
Geometry Aware Module (GAM)



- Calculate the depth contrasted features to enrich the fused features with high-frequency information

Results

Our reconstructed results under various poses of different human bodies :



Qualitative comparisons on our captured real data :

Refer to our paper for more experiments and results

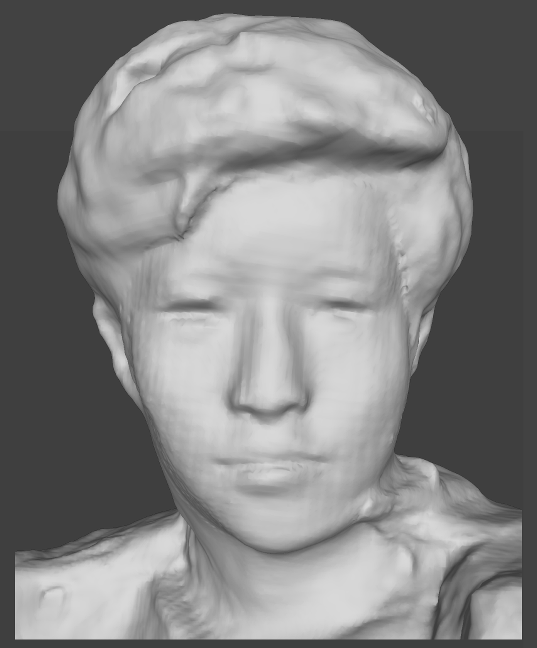
Demos

View 1

View 2

View 3

Our Reconstructed Model



Face

View 1

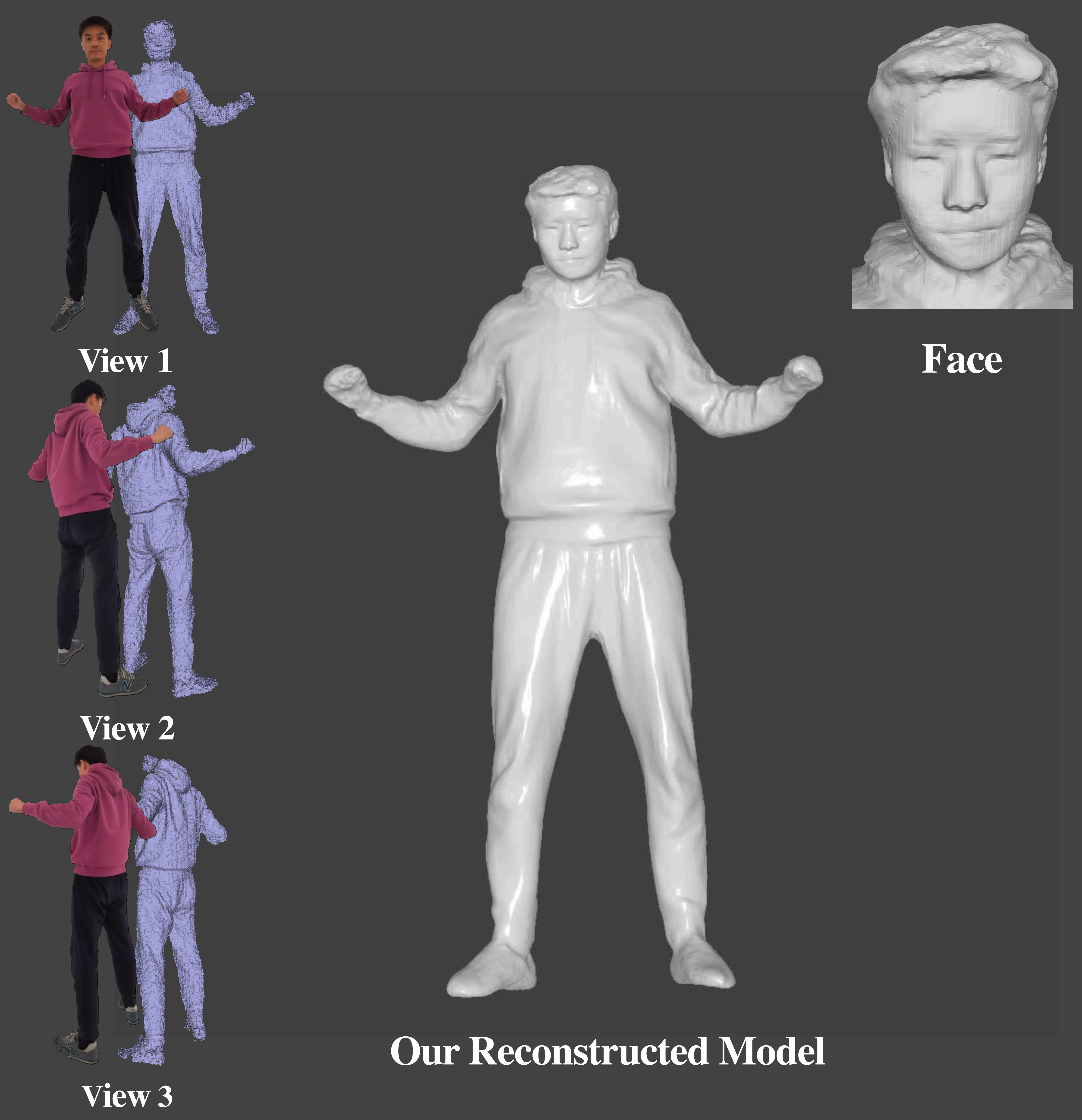
View 2

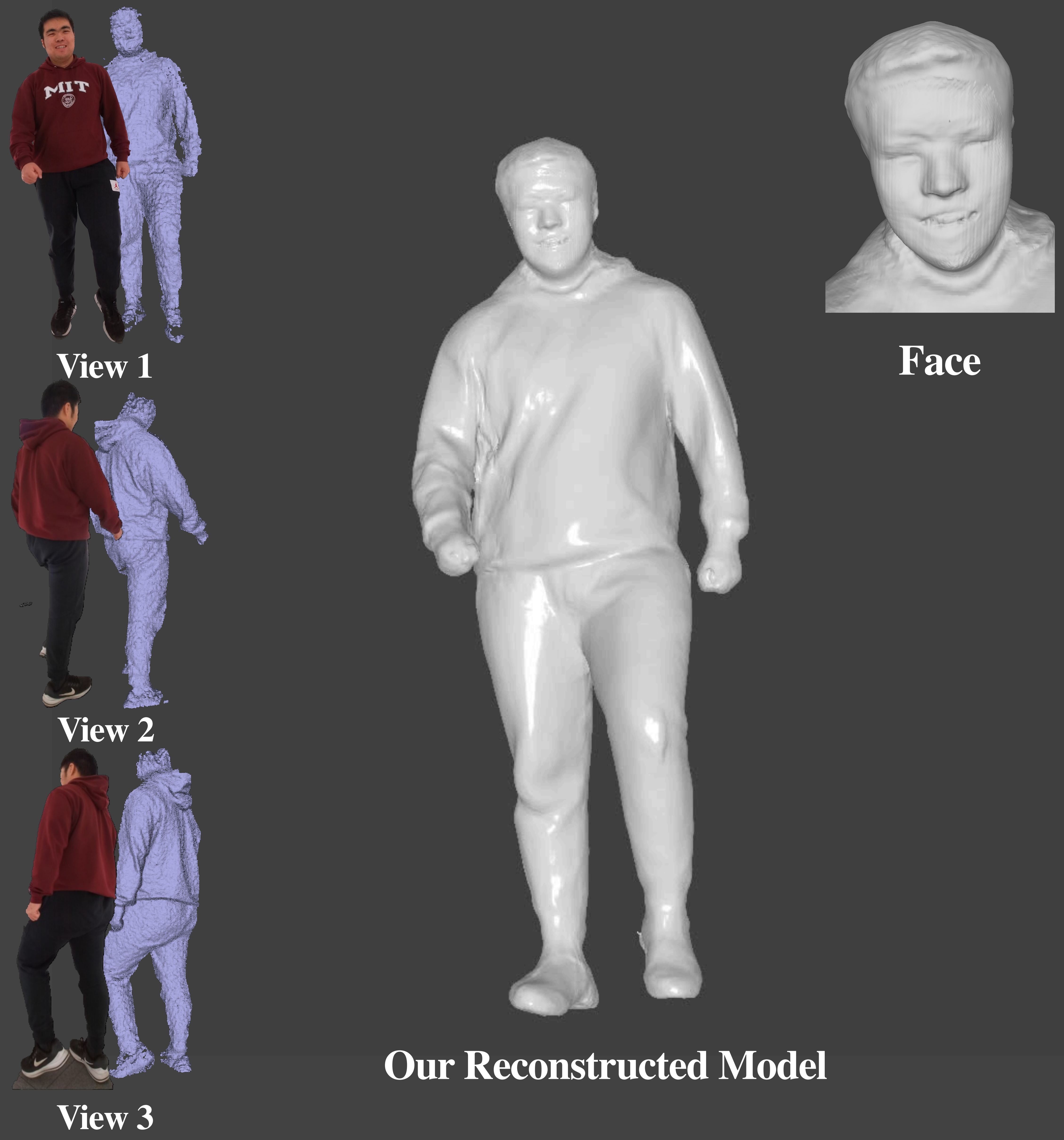
View 3

Our Reconstructed Model

Face

Demos


View 1
View 2
View 3
Face
Our Reconstructed Model


MIT
View 1
View 2
View 3
Face
Our Reconstructed Model

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you for watching !