Final Report: 3D Virtual Try-On Using SMPLify-X on MPV3D Dataset

# 1. Introduction

Virtual try-on systems are gaining momentum in fashion and e-commerce, enabling users to preview clothing in 3D before purchase. However, building a reliable system using only 2D keypoints (without depth or RGB input) remains challenging. In this project, we propose a minimal-input 3D body reconstruction pipeline that leverages SMPLify-X optimization on sparse keypoints to produce SMPL-X-based 3D human meshes.  
  
Our pipeline is designed to work with the MPV3D dataset and targets applications such as AI-powered virtual dressing rooms and personalized avatar generation.

# 2. Contributions

- Developed an end-to-end 3D human mesh reconstruction pipeline using only 2D keypoints  
- Integrated preprocessing modules tailored for MPV3D format  
- Applied SMPLify-X with VPoser priors for anatomically plausible pose estimation  
- Aligned body mesh output for potential downstream applications (e.g., garment fitting, animation)

# 3. Dataset

We use the MPV3D dataset, which provides monocular images, 2D keypoint annotations (25-joint OpenPose format), and clothing images. Each sample includes a clothing item image and a person image wearing that garment.

# 4. Methodology

## 4.1 Preprocessing Pipeline

To enable SMPLify-X optimization, we performed the following preprocessing steps on the raw MPV3D data:  
  
- Image Preparation  
  → Person image resized to 320×512 and placed in /image  
  → Garment image cropped or background removed (via thresholding or remove.bg) → /cloth  
- Human Segmentation  
  → Applied 2D parsing models (e.g., SCHP) on the person image → /image-parse  
  
A person in a blue pants and a floral top

AI-generated content may be incorrect.

Figure : Segmented body

- Pose Estimation  
  → Ran OpenPose on each image to extract 2D body, hand, and facial keypoints → /pose  
- Data Finalization  
  → Preprocessing utility script aligned and resized all components, producing palm masks and edge gradients necessary for body-cloth alignment  
  
A white silhouette of a person

AI-generated content may be incorrect.

Figure : Masked clothing

## 4.2 3D Body Mesh Fitting via SMPLify-X

- Input: Person image with corresponding 2D keypoints  
- Model: SMPL-X parametric body model with shape β and pose θ  
- Prior: VPoser used as a latent constraint to keep poses natural and realistic  
- Optimization: SMPLify-X minimized reprojection error between 3D joint projections and 2D OpenPose keypoints. The process iteratively refined global orientation, shape, and joint angles.  
  
- Output: A fitted 3D body mesh (SMPL-X topology), including hands and facial landmarks, saved in .pkl/.obj format.  
  
Full shot of a mannequin

AI-generated content may be incorrect.

Figure : Generated 3D SMPL Mesh

# 5. Results and Evaluation A comparison of a person's body

Figure : image and SMPL model (or SMPL 3d mesh)  
  
  
A mannequin wearing a white shirt and leggings

AI-generated content may be incorrect.

Figure : Overlaying of 3D Mesh on the original image

# 6. Related Work Comparison

Comparison of our method with other state-of-the-art approaches:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Input | 3D Accuracy | Clothing Ready | Real-Time |
| PARE | RGB | Medium | No | ✅ |
| VIBE | Video | High | No | ❌ |
| TailorNet | RGB+Cloth | High | Custom | ❌ |
| Ours (SMPLify-X) | 2D Keypoints | High | Mesh-Ready | ⚠️ Near-RT |

# 7. Challenges Faced

- Ambiguity from Sparse Keypoints  
 → Mitigated using VPoser latent constraints and pose smoothing  
- Body mesh overshooting or underfitting certain limbs  
 → Handled by multi-stage optimization and manual tuning  
- Coordinate system alignment between 2D keypoints and SMPL-X  
 → Resolved through calibration heuristics and OpenPose scaling adjustments  
- Inference Speed  
 → Reduced through mesh decimation and fewer iterations during testing

# 8. Clothing Warping Overview

Clothing warping is a crucial post-reconstruction step in virtual try-on systems. It involves deforming or adapting 3D garment meshes so they conform to a posed human body mesh, typically generated by a model like SMPL-X.  
  
This stage addresses the challenge of transferring garments from a canonical rest pose (T-pose) to match the dynamic pose and shape of the reconstructed human mesh. Key techniques include:  
  
- Mesh Registration: Aligning garment vertices to the target body mesh using closest-point mapping or energy minimization.  
- Linear Blend Skinning (LBS): If skinning weights are available, garments can be animated using the same skeletal transformations as the human mesh.  
- Collision Handling: Ensures the garment doesn’t intersect with the body mesh. Some pipelines apply post-correction or use cloth simulation in Blender.  
- Cloth Fitting & Refinement: Further adjustment can be done via deformation transfer, Laplacian mesh editing, or physics-based simulation to enhance realism.  
  
Regardless of the chosen approach, successful clothing warping ensures that the final overlayed garments preserve their identity while following the pose and shape of the reconstructed body.

# 9. Conclusion

This project presented a modular, hybrid 3D virtual try-on pipeline combining SMPLify-X-based body reconstruction with downstream garment integration components. The contribution centered on enabling reconstruction from 2D sparse keypoints and proposed a robust process adaptable to real-world settings. Future work includes improving clothing-body interaction realism and supporting multi-person scenes.