Synthesizing and Adapting Virtual Humans in 3D Environments

Paul Robert Herzog | Bosch

Acquisition, Curation & Refinement of Real Data

Introduction

In the realm of synthetic data generation, creating lifelike virtual humans has seen remarkable advancements enabling the power of infinite labeled data generation. Our key ingredients in this context are:

- **Procedural 3D human** generation using a Blender Addon “HumGen 3D” [1] and encoding of geometry [4] suited for gradient based intersection minimization
- Realistically **posing humans** obeying intricate 3D environments modelled as a signed-distance-field [2]
- An **automated** (neural) **rendering pipeline** based on blender [3]

Sphere-Proxy Human

Human meshes are unsuited for volumetric intersection tests and internally we use a representation based on a set of spheres, that:

- *is well-suited for „roundish“ bodies,*
- *allows for efficient (self-)intersection tests,*
- *has same number of spheres per human,*
- *can be fit to implicit surfaces modeled as signed distance fields [4].*

We trained a variational auto-decoder as in [4] based on 800 generated human meshes, which outputs a set of spheres in rest pose. The spheres are linked to the skeleton „bones“ and can be kinematically transformed (see Fig. 1).

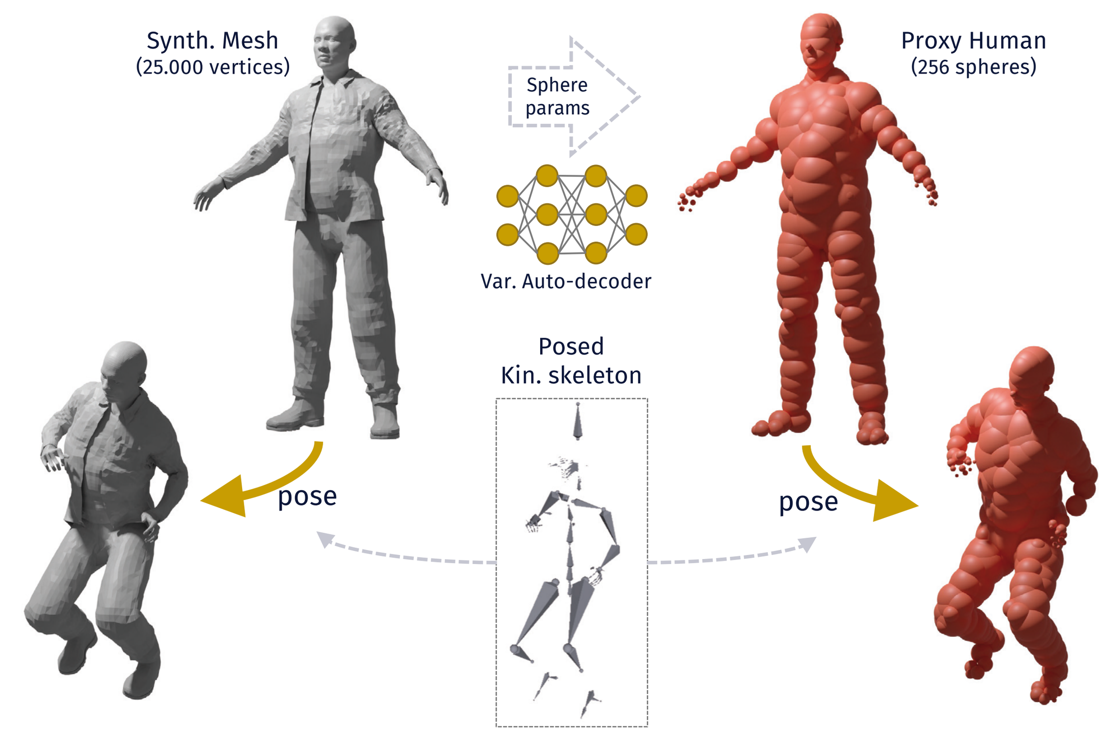

Figure 1: Human Proxy encoding with a var. auto-decoder (© Bosch)

Multi-Human Posing in Static 3D Environments

We formulate it as minimization task with four jointly optimized losses penalizing if articulated humans

1. *intersect the static 3D environment,*
2. *self-intersect body parts,*
3. *intersect other articulated humans,*
4. *exhibit unrealistic poses.*

Given a pose $x \in S^{3K}$ (K is number of joints) loss 1 estimates by how much the body intersects the surrounding 3D environment (eq.1) modeled as a signed distance field (sdf):

$$\operatorname*{argmin}_{\boldsymbol{x}} \sum_i \max\big(0, sdf(FKLBS(\boldsymbol{x}, p_i))\big); \;\; p_i \in p_{skin} \quad (eq. 1)$$

where p_{skin} represents points inside the 3D human proxy in rest pose and FKLBS stands for **F**orward **K**inematic with **L**inear **B**lend **S**kinning. Loss 2 and 3 are trivial to compute for spheres and loss 4 is based on [2] and penalizes the distance to a learned **human pose manifold**. Fig. 2 shows an example before and after 50 steps of this joint optimization.

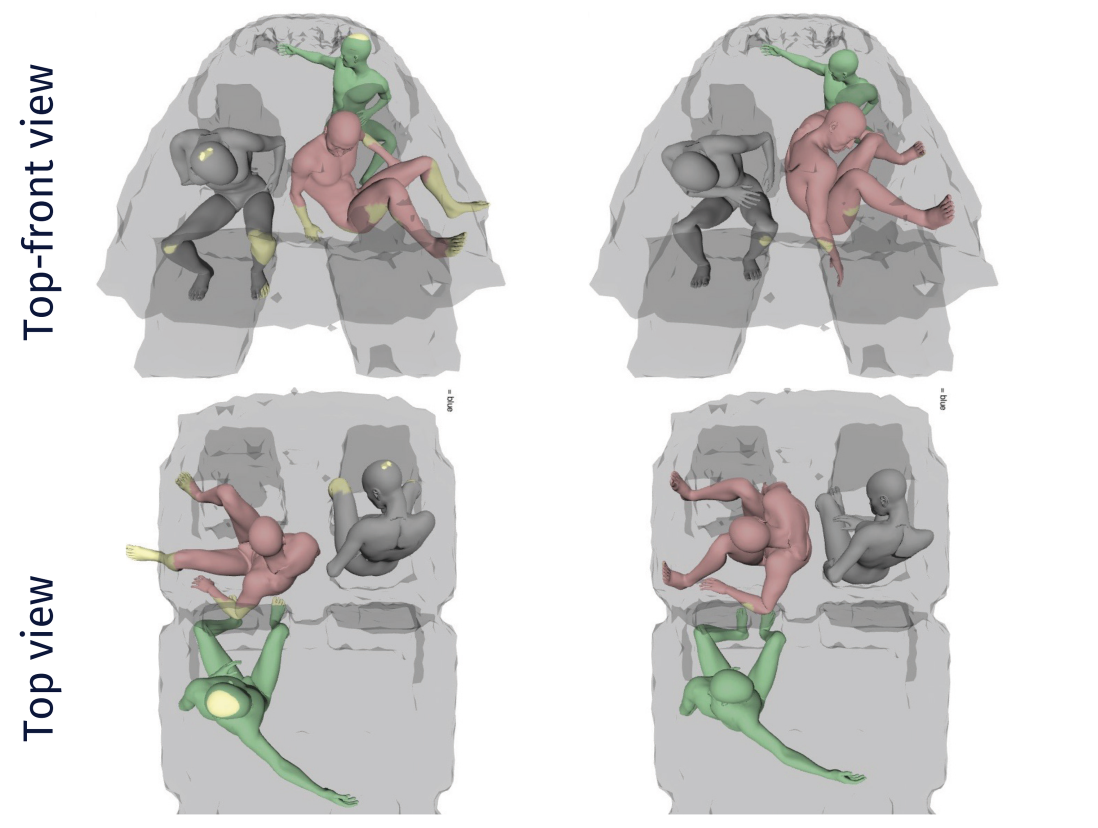

Figure 2: Three virtual humans in non-standard poses placed initially (left) and after fitting (right) inside a vehicle; yellow mesh parts are intersecting the vehicle, which is modeled implicitly, only iso-surface at signed distance = 0 is visualized transparently in grey (© Bosch)

Show-Case – Humans in Vehicles

Using our pipeline [3] we generated a synthetic image dataset (e.g. Fig. 3) consisting of random people in vehicles and trained a human 3D pose estimator, which improved performance on real test data by 30% after being finetuned on ca. 4000 real images.

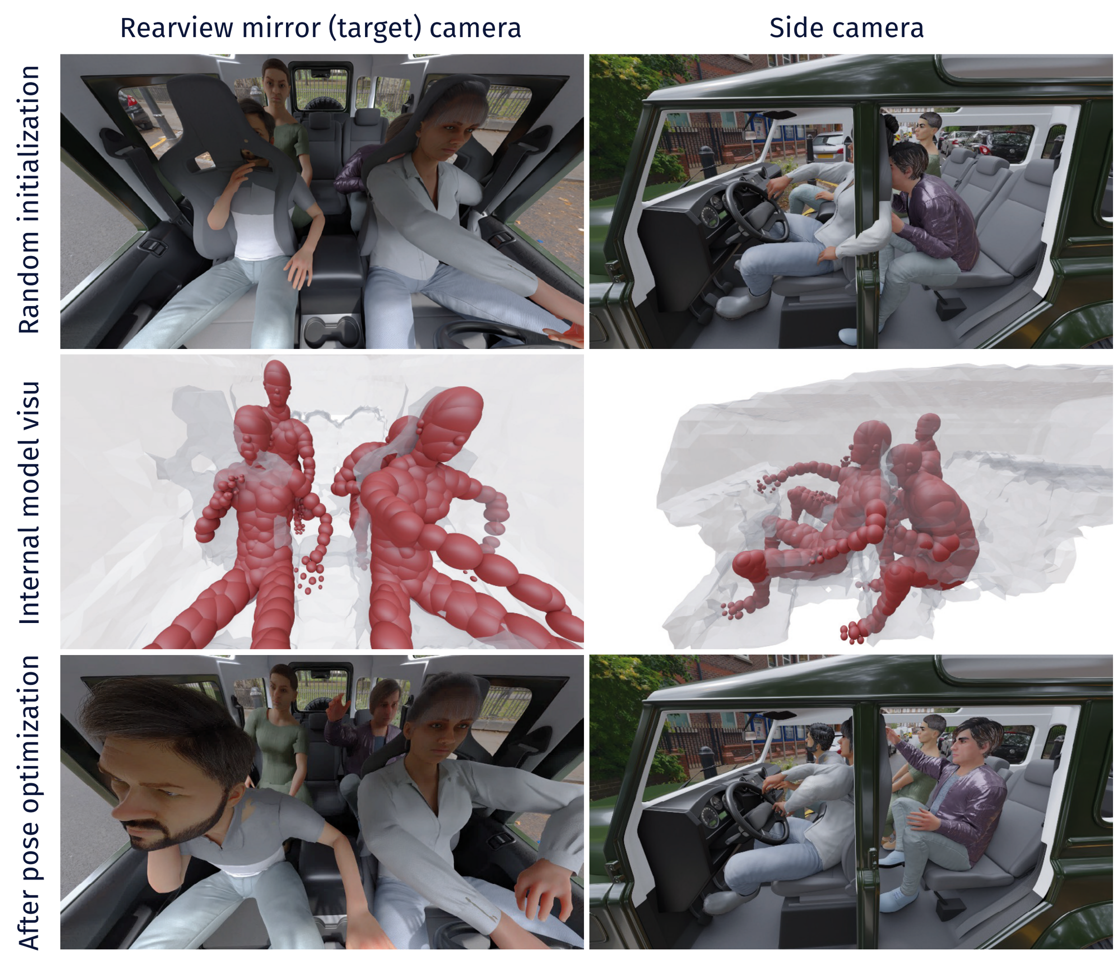

Figure 3: One example of virtual human creation and pose adaptation in a 3D vehicle environment; down left is the final image (© Bosch)

References:

[1] Human Generator, Blender Addon for Procedural Generation of Virtual Humans, https://www.humgen3d.com

[2] Tiwari et al., Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields, ECCV 2022

[3] Image Render Setup, a Rendering Automation System for Blender, https://github.com/boschresearch/image-render-setup

[4] Hao et al., DualSDF: Semantic Shape Manipulation using a Two-Level Representation, CVPR 2020

Partners

External partners

For more information contact:

paulrobert.herzog@de.bosch.com

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag