A. Supplemental

A.1. Skeleton/headset alignment

We use the headset's built-in calibration tool to align its forward direction with the forward direction of the motion capture volume. We also calibrate the mocap system and the headset such that they measure the same floor height. This reduces the alignment problem to finding a planar offset $\vec{o}$ (we do not modify the skeleton's height) that is used to align the livestreamed skeleton to the actor in VR. To achieve this, we assume that the midpoint of the eyes (measured by the headset),

$$\vec{r}_e = \frac{\vec{r}_{\text{left eye}} + \vec{r}_{\text{right eye}}}{2},$$

lies along a line that has the direction of the head bone's local forward vector and contains the midpoint of the head bone's top face ($\vec{f}_h$ and $\vec{r}_{ht}$, both respectively measured by the mocap system). We can write this as a linear system with 3 constraints and 3 unknowns,

$$\vec{r}_e = \vec{r}_{ht} + \lambda \vec{f}_h + \vec{o},$$

where

$$\vec{o} = \begin{bmatrix} o_x \\ o_y \\ 0 \end{bmatrix},$$

that we solve during calibration.

B. High resolution result images

For the complete versions of the images in Fig. 7, see Fig. 9 and Fig. 10, respectively.

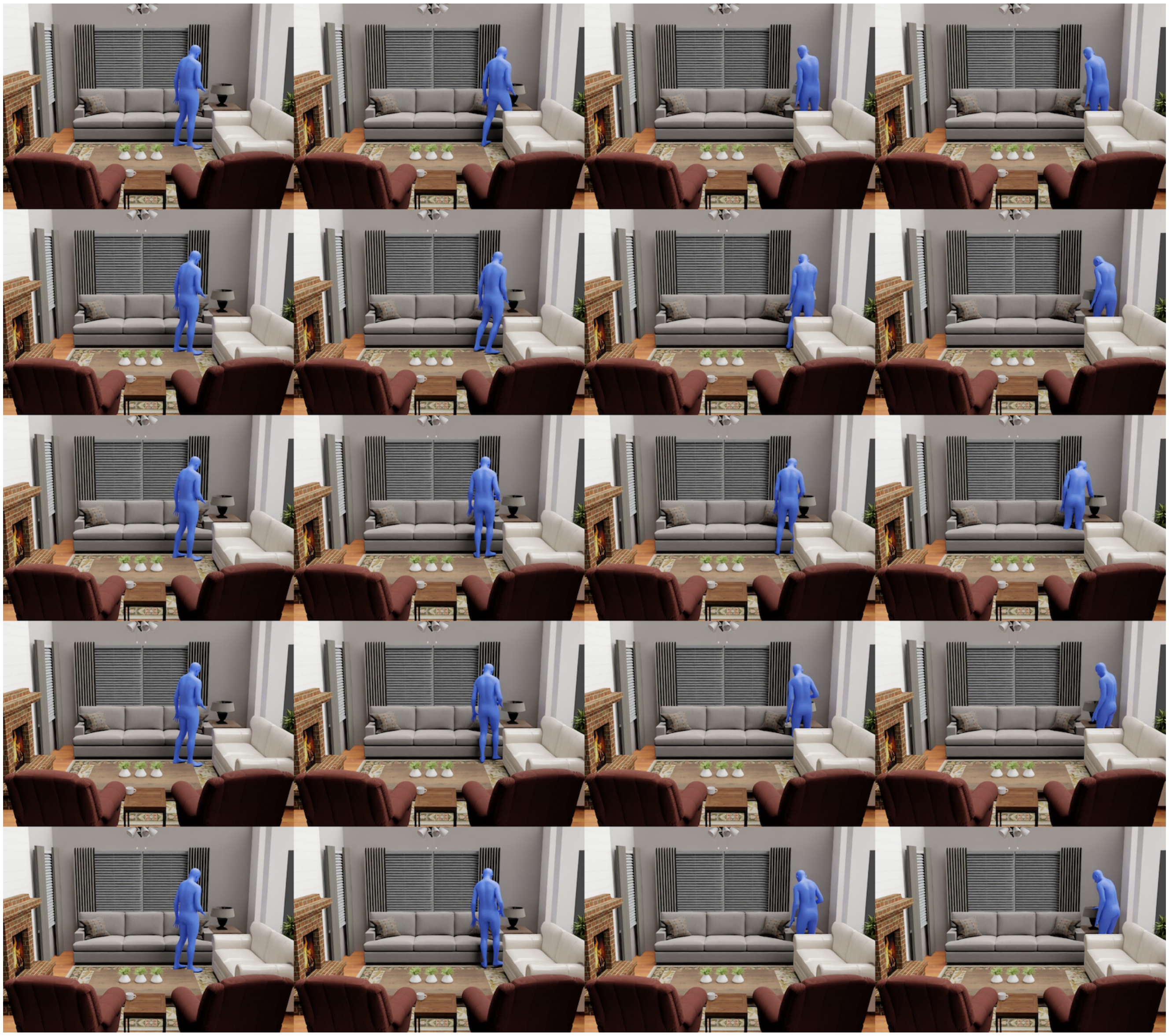

Figure 9. Comparison of the i) ground truth motion (top row), and outputs generated by ii) GOAL (second row), iii) our model without scene information (third row), iv) scene information using pointnet (fourth row), v) scene information using BPS (fifth row). We see that the sequence generated by GOAL fails to achieve the objective, and that introducing scene information reduces collisions.

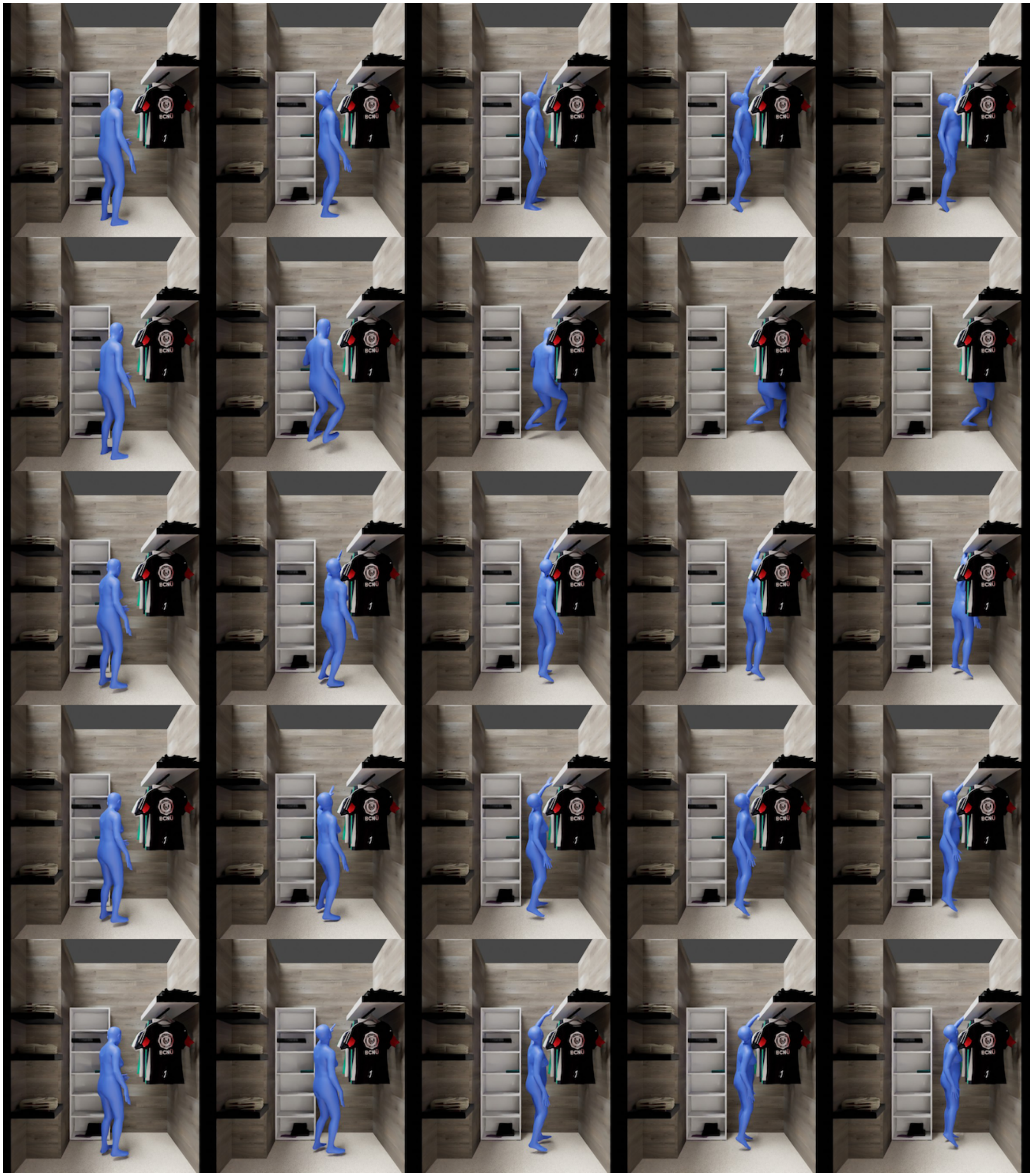

Figure 10. Comparison of the i) ground truth motion (top row), and outputs generated by ii) GOAL (second row), iii) our model without scene information (third row), iv) scene information using pointnet (fourth row), v) scene information using BPS (fifth row). We see that the sequence generated by GOAL fails to achieve the objective, and that introducing scene information reduces collisions.