Stochastic Scene-Aware Motion Prediction

Mohamed Hassan[1], Duygu Ceylan[2], Ruben Villegas[2], Jun Saito[2], Jimei Yang[2], Yi Zhou[2], and Michael Black[1]

Max Planck Institute for Intelligent Systems, Tübingen, Germany[1]   Adobe[2]

Goal

Enable virtual humans to *navigate* and *interact* with cluttered scenes.

Contribution

- Synthesizing *diverse* motion styles in real-time.
- Modeling plausible goal locations and orientations on target objects.
- Enable *navigation* in cluttered scenes using explicit path planning.
- A *new MoCap dataset* with diverse human-scene interactions.

Method

- **MotionNet:** Autoregressive cVAE that generates the pose in a sequential manner.
- **GoalNet:** cVAE that predicts the goal position and orientation on the desired object.
- **Path Planning Module:** Use the A-star algorithm to compute a collision-free path to the goal.

MotionNet

- The weights of the Prediction Network are computed by blending a number of expert weights.
- The blending weights are predicted by the Gating Network.

MotionNet Data

- Existing MoCap datasets are dominated by locomotion, rarely capture the scene, and lack diversity.
- Our dataset has about *100 minutes* of motion covering diverse styles of *sitting, lying down, walking, running,* and *idling*.
- Track the body and the object.
- Efficient augmentation pipeline.
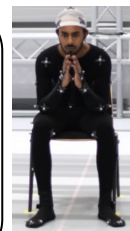
GoalNet
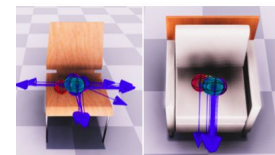
Generated Goals

Guided Motion

Path Planning
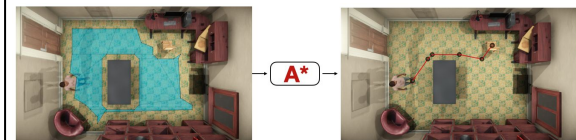
Quantitative Evaluation

Metric	Sit		Carry	
	SAMP	NSM	SAMP	NSM
Precision PE (cm) ↓	**15.97**	16.95	**4.58**	4.72
Precision RE (deg) ↓	5.38	**2.32**	1.78	**1.65**
Execution Time (sec) ↓	12.93	**10.26**	13.29	**12.82**
FD ↓	6.20	**4.21**	10.17	**7.31**
Diversity ↑	**0.44**	0.0	**0.26**	0.0
Penetration (%) ↓	**3.8**	8.11	**3.62**	8.45