Introduction

Multi-Track Timeline Control for Text-Driven 3D Human Motion Generation

Mathis Petrovich1,2, Or Litany3,4, Umar Iqbal1, Michael J. Black2, Güler Varol1, Xue Bin Peng5,6, Davis Rempe7

1 IMAGINE, École des Ponts ParisTech, France
2 Max Planck Institute for Intelligent Systems, Tübingen
3 NVIDIA
4 Technion
5 Simon Fraser University

Problem

Single text prompt as input lacks fine-grained control:
• Defining precise durations for parts of the motion
• Composing multiple actions

Solution

Multi-track timeline of multiple prompts as input: fine-grained input interface

STMC: Spatio-Temporal Motion Collage

New test-time denoising method: can be integrated with any pre-trained motion diffusion model

Quantitative results

Perceptual study

Most realistic?

Follow timeline?

MotionDiffuse

MOM

MDM

MDM-SMP

SINC

DiffCollage

Motioncrop

Quantitative results

References

MotorDiffuse: Text-Driven Human Motion Generation with Diffusion Models, Zhang et al. CVPR 2022
DiffCollage: Parallel Generation of Large Content with Diffusion Models, Zhang et al. CVPR 2023
MDM: Human Motion Diffusion Model, Tse et al. EMMCVPR 2023
MDM-SMP: Single-Model Motion Compositional Perceptual Study