Project page

CVPR JUNE 17-21, 2024 SEATTLE, WA

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

Mathis Petrovich[1,2]* Or Litany[3,4] Umar Iqbal[3] Michael J. Black[2]
Gül Varol[1] Xue Bin Peng[3,5] Davis Rempe[3]

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

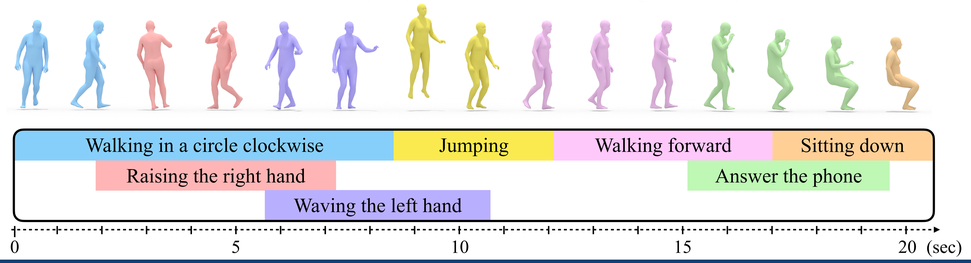
Introduction

Goal Fine-grained 3D Human Motion Generation

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



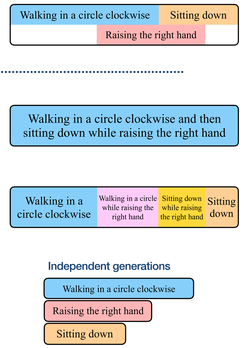
Comparison to baselines

Multi-Track timeline
Include both spatial and temporal compositions

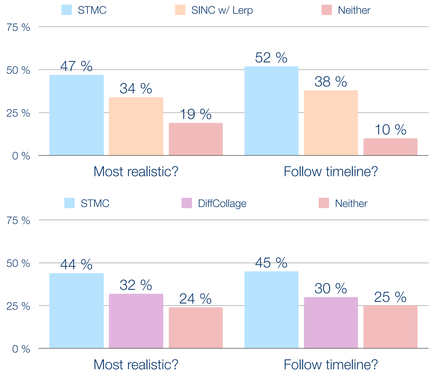
Single text input
Collapse the timeline into a text prompt

DiffCollage
Collapse the timeline into a single track timeline

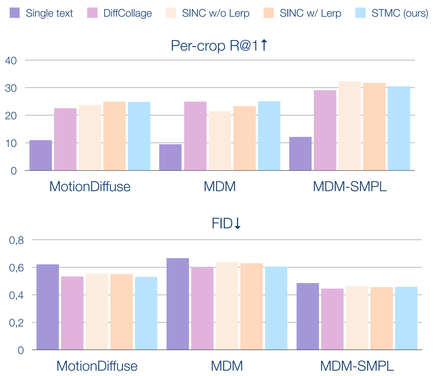
SINC
Combine the motions by body parts
w/o: raw output
w/ Lerp: smooth the transitions with linear interpolation



Perceptual study



Quantitative results

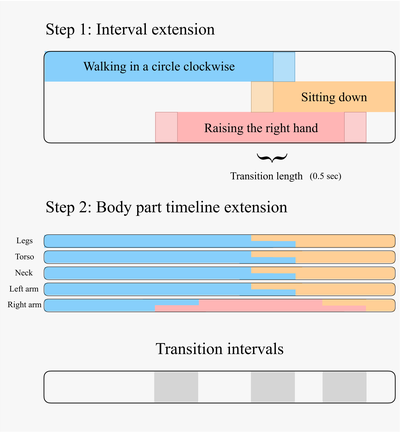


STMC: Spatio-Temporal Motion Collage

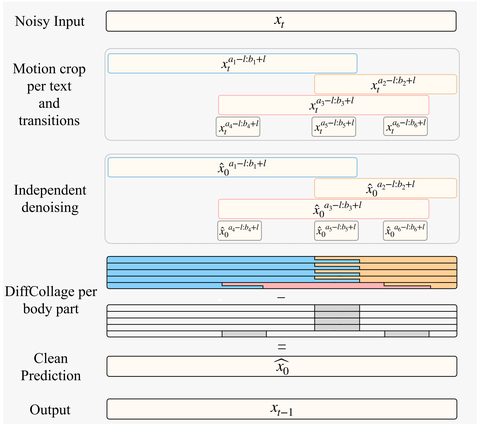
New **test-time** denoising method: can be **integrated** with any pre-trained motion diffusion model

(a) Body part timeline partitioning

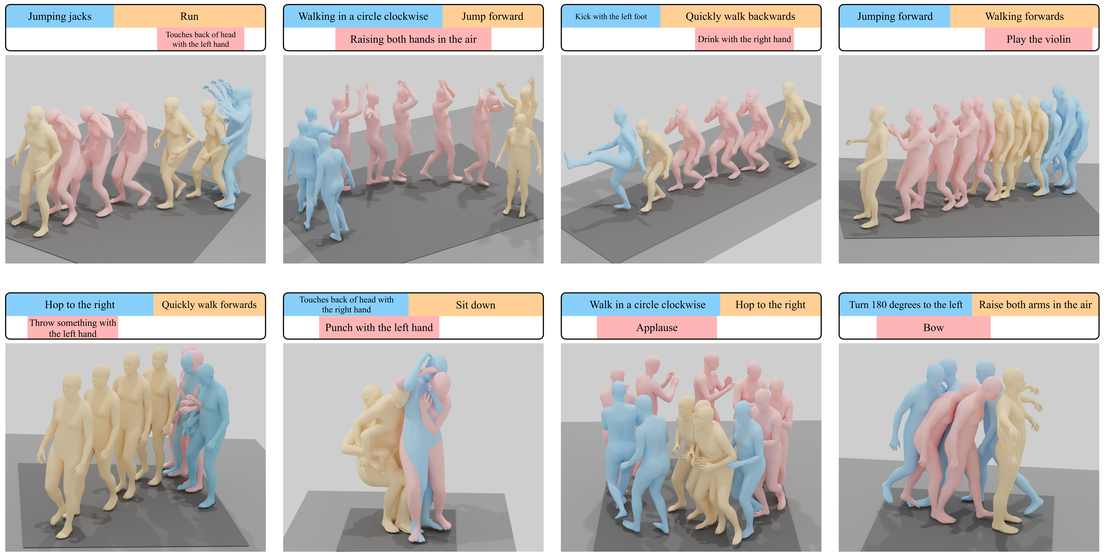
(b) Temporal stitching



(c) STMC denoising step t



Qualitative results



References

MotionDiffuse: Text-Driven Human Motion Generation with Diffusion Model, Zhang et al. arXiv 2022
SINC: Spatial Composition of 3D Human Motions for Simultaneous Action Generation, Athanasiou et al. ICCV 2023
DiffCollage: Parallel Generation of Large Content with Diffusion Models, Zhang et al. CVPR 2023
MDM: Human Motion Diffusion Model, Tevet et al. ICLR 2023

https://mathis.petrovich.fr/stmc

* Work done during an internship at NVIDIA

mathis.petrovich@enpc.fr