ICCV 2025 Tutorial Proposal: Human Motion Generation

1 Introduction

Human motion generation is an important area of research with applications in virtual reality, gaming, animation, robotics, and AI-driven content creation. Generating realistic and controllable human motion is essential for creating interactive digital environments, improving character animation, and enhancing human-computer interaction. Recent advances in deep learning have made it possible to automate motion generation, reducing the need for expensive motion capture and manual animation.

Techniques such as diffusion models, generative masking, and variational autoencoders (VAEs) have been used to synthesize diverse and realistic human motion. Transformer-based models have improved the ability to capture temporal dependencies, leading to smoother and more natural movement. In addition, reinforcement learning and physics-based methods have helped create physically consistent and responsive motion, which is useful for applications like robotics and virtual avatars.

Human motion generation research spans across computer vision, computer graphics, and robotics. However, many researchers and developers may not be familiar with the latest advances and challenges in this area. This tutorial will provide an introduction to human motion generation, covering key methods, recent developments, and practical applications. We will also discuss open research problems and opportunities for future work. The tutorial will be divided into six parts: human motion generation basics, kinematic-based generation methods, physics-based generation methods, controllability of human motion generation, human-object/human/scene interactions, and open research problems.

2 Proposers (Detailed bio in Sec. 12)

Huaizu Jiang is an Assistant Professor at Northeastern University. *h.jiang@northeastern.edu*.
Chuan Guo is a Research Scientist at Snap Inc. *cguo2@snapchat.com*.
Lingjie Liu is an Assistant Professor at the University of Pennsylvania. *lingjie.liu@seas.upenn.edu*.
Zhiyang (Frank) Dou is a PhD student at the University of Hong Kong. *zhiyang0@connect.hku.hk*.
Yiming Xie is a PhD student at Northeastern University. *ymxyimingxie@gmail.com*.

3 Preference for full-day event

Given the wide range of topics to cover and the number of confirmed speakers, we prefer a **full-day event**, though a half-day event would also be acceptable.

4 Course description with list of topics to be covered, along with a brief outline and important details

The tutorial will cover the following topics of human motion generation.
Part I - Human motion generation basics: Theoretical foundations and best practices

- We will present basic concepts including motion representation, common approaches, and datasets in human motion generation. Human motion can be represented in different ways, including 3D joints, surface markers, and parametric models like SMPL, each serving different purposes in animation, robotics, and AI-driven applications. Common approaches include diffusion-based, generative masking, and GPT-based ones. Large-scale datasets such as HumanML3D, KIT, and MotionX, provide training and evaluation benchmarks.

Part II - Kinematic-based motion generation methods

- Kinematic-based methods focus on generating motion sequences without physical constraints, making them efficient and adaptable for animation and AI-driven applications. We will cover some milestone works/technologies in this direction, including VAEs learning compact motion embeddings, diffusion models refining noisy inputs into high-quality motion, and transformers modeling long-range dependencies. We will present a key task, text-to-motion generation, which takes in natural language descriptions and outputs human motion, allowing more intuitive and controllable motion synthesis for virtual avatars, gaming, and AI-driven content creation.

Part III - Physics-based human motion generation

- *Physical simulation & reinforcement learning*. Physics-based methods leverage physical simulation and reinforcement learning (RL) to generate realistic and physically plausible human motion. We will discuss the role of physics engines in motion synthesis and the integration of RL for learning control policies that enable natural, responsive, and adaptable character movements. We will also briefly introduce applications include biomechanical analysis, robotics, and interactive virtual environments.
- *Physics-based character animation*. Physics-based character animation ensures realism by adhering to physical laws. We will cover key techniques such as proportional-derivative controllers, deep RL, and optimal control for balance, agility, and motor skills. Imitation learning enables AI-driven characters to replicate human motion from demonstrations using motion capture and video data. We will discuss adversarial imitation learning, motion VAEs, and optimal control, along with challenges in stability, adaptability, and control robustness.

Part IV - Controllability of human motion generation

- *Multi-modal motion control*. We will cover how motion can be controlled through diverse input modalities, including text, video, audio, etc. And how these information helps with controllable motion generation to align the output with users' inputs, supporting different applications.
- *Spatially-aware motion control*. Motion generation must consider spatial signals from the user and environment to ensure context-driven and realistic movements. We will discuss methods that integrate spatial constraints, terrain adaptation, and interaction-aware controls to improve realism and responsiveness in dynamic settings. We will primarily cover pure learning-based and optimization-based techniques for controllable motion generation.

Part V - Human-object/human/scene interaction

- *Human-object interactions with dynamic objects*. Human motion is often influenced by interactions with dynamic objects, such as catching, pushing, or carrying items. Modeling these interactions requires understanding human goals, object shape, and object affordance. In addition to the whole-body interaction, we will particularly introduce fine-grained object manipulation, such as grasping and tool use, requires precise modeling of hand-object interactions. We will introduce several representative methods, representations for hands, and benchmark datasets for human-object interaction.
- *Human-human interaction.* Modeling social interactions between humans, ranging from dyadic interactions to group dynamics, is critical for various applications. For two-person interactions, we will present actor-reactor generation, which synthesizes reactive motions in response to an actor's movements, and synchronized interaction generation, which simultaneously creates coordinated motions for both participants. We will also introduce recent advances on generation of coordinated movements among multiple participants. We will talk about benchmark datasets as well.
- *Human-scene interaction and motion planning*. Navigating and interacting with a scene requires planning motion based on environmental constraints, such as avoiding obstacles, sitting on chairs, or opening doors. Scene-aware motion generation relies on spatial reasoning to align human motion with affordances in the environment. We will covert common approaches and datasets used in human-scene interactions and scene navigation.

4.1 Tentative program outline

Morning

Welcome words by organizers [15 min] 8:45 - 9:00
Part I [30 min] 9:00 - 9:30
Part II [30 min] 9:30 - 10:00
Part III [60 min] 10:00 - 11:00
Coffee Break [15 min] 11:00 - 11:15
Part IV Part 1 [60 min] 11:15 - 12:15

Lunch Break [75 min] 12:15 - 13:30

Afternoon

Part IV Part 2 [60 min] 13:30 - 14:30
Coffee [15 min] 14:30 - 14:45
Part V [60 min] 14:45 - 15:45
Coffee [15 min] 15:45 - 16:00
Panel discussion (6 panelists) [60 min] 16:00 - 17:00

5 Subject/application area

Primary Subject: Human Motion Generation. **Secondary Subject Areas**: Motion Generation with Multi-Modal Input; Character Animation and Virtual Avatars; Human-Object/Scene/Human Interactions; Physics-Based Motion Generation; Embodied AI and Robotics.

6 List of tentative speakers (Detailed bio in Sec. 12)

- **Huaizu Jiang** (*confirmed*) is an Assistant Professor at Northeastern University. – *Opening*
- **Chuan Guo** (*confirmed*) is a Research Scientist at Snap. – *Part I/II*
- **Zhengyi Luo** (*confirmed*) is a final year PhD student at CMU. – *Part III*
- **Daniel Holden** is a Principal Animation Programmer at Epic Games. – *Part IV*
- **Libin Liu** is an Assistant Professor at Peking University. – *Part IV*
- **Gerard Pons-Moll** is a Professor at the University of Tübingen. – *Part V*
- **Lingjie Liu** (*confirmed*) is an Assistant Professor at University of Pennsylvania. – *Closing*

We will invite more speakers from different fields for our tutorial. Given the diverse nature of the research topics to be covered, having experts from different directions would enable high-quality lectures and thus improve the learning experience of the audience.

7 Expected audience and number of attendees

We anticipate this workshop will be of interest to researchers working on human motion generation, human-object/scene interaction, and humanoid robotics. Based on attendance at past CVPR 2024/2025 workshops, we expect 100 to 300 participants (Medium).

8 Format: Specify the intended format in advance - in-person or mixed in-person attendance. Please describe how you will support virtual attendance if the workshop is partially virtual

We plan to hold the tutorial primarily in person. We will provide video recordings to support asynchronous learning experience.

9 List of citations and/or URLs to relevant publications and/or products by the organizers, and to other relevant related work

Action2motion, HumanML3D, Momask, CASE, EMDM, CBIL, TLControl, OmniControl, Smoodi, MARDM

10 A description of how this proposal relates to tutorials/short courses appearing at ICCV, CVPR, and ECCV within the last three years;

This proposal relates to the HuMoGen workshops in CVPR 2024/2025, but focus more on providing educational sessions on a particular topic, method, or tool.

11 Links to a few previous recorded talks given by the presenters (if available)

Chuan Guo: AMII AI Seminar Series 2023; Computer Vision Meetup 2023
Daniel Holden: HuMoGen CVPR Workshop 2024;SCA 2022; MPCRL 2021
Gerard Pons-Moll: Vision & Graphics Seminar at MIT, 2022;LHMP 2022;TUM AI Lecture Series
Libin Liu: Asiagraphics Webinar

12 Description of and/or links to any planned materials or resources to be distributed to attendees

HumanML3D; MDM; DeepMimic; AMP; PhysicsVAE; Actor; OmniControl; TLControl; Tech Blog1: Animation Quality; Tech Blog2: BVH View; AI4Animation: Deep Learning for Character Control [Github].

Biographies

Proposers

- **Huaizu Jiang**. Huaizu Jiang is an Assistant Professor in the Khoury College of Computer Sciences at Northeastern University. He has broad research interests in computer vision, computational photography, natural language processing, and machine learning. His long-term research aims to teach machines to develop visual intelligence in a manner analogous to humans. In the short term, his research goal is to create smart visual perception tools to improve people's life experiences of using cameras. Prior to joining Northeastern University, he was a Postdoc Researcher at Caltech and a Visiting Researcher at NVIDIA. He obtained my Ph.D. from UMass Amherst, where he was fortunately advised by Prof. Erik Learned-Miller. When pursuing my Ph.D., he did joyful internships at NVIDIA Research and Facebook AI Research (FAIR). He got my M.E. and B.E. degrees from Xi'an Jiaotong University in 2009 and 2012, respectively.

- **Chuan Guo**. Chuan Guo is a Research Scientist at Snap Inc., specializing in generative AI for digital human performance and character animation. His research focuses on 3D avatar animation, motion synthesis and stylization, and human-scene/object interaction. He earned his Ph.D. from the University of Alberta, advised by Prof. Li Cheng. Chuan has numerous publications in top-tier venues such as IJCV, CVPR, ICCV, and ICLR. He is also a co-organizer of the Human Motion Generation (HuMoGen) Workshop at CVPR 2024-2025.

- **Lingjie Liu**. Lingjie Liu is the Aravind K. Joshi Assistant Professor in the Department of Computer and Information Science at the University of Pennsylvania, where she lead the Penn Computer Graphics Lab and she is also a member of the General Robotics, Automation, Sensing & Perception (GRASP) Lab. Previously, she was a Lise Meitner Postdoctoral Research Fellow at Max Planck Institute for Informatics. She received my Ph.D. degree at the University of Hong Kong in 2019. Her research interests are at the interface of Computer Graphics, Computer Vision, and AI, with a focus on Neural Scene Representations, Neural Rendering, Human Performance Modeling and Capture, and 3D Reconstruction. She is especially excited about exploring a new genre of 3D reconstruction and rendering algorithms for human characters and general scenes, which combine classical computer graphics pipelines with deep learning techniques.

- **Zhiyang (Frank) Dou**. Zhiyang (Frank) Dou is a Ph.D. student in Computer Graphics Group at The University of Hong Kong, supervised by Prof. Wenping Wang and Prof. Taku Komura. He received my B. Eng. degree with honors at Shandong University, advised by Prof. Shiqing Xin. He is currently a visiting student in the Department of Computer and Information Science at the University of Pennsylvania, where he works with Prof. Lingjie Liu in the Graphics Lab and the GRASP Lab. He also collaborates closely with Prof. Cynthia Sung in the Department of Mechanical Engineering and Applied Mechanics at UPenn. His research interests include character animation, geometric modeling and processing, simulation, computer graphics, and human behavior modeling&analysis.

- **Yiming Xie**. Yiming Xie is currently a Ph.D. candidate at Khoury College of Computer Sciences, Northeastern University supervised by Prof. Huaizu Jiang. He received my Bachelor's degree in 2019 from Zhejiang University (ZJU), advised by Prof. Xiaowei Zhou. He is a recipient of the 2024 Apple Scholars in AI/ML PhD fellowship.

Speakers

- **Chuan Guo**. See bio above.

- **Daniel Holden**. Daniel Holden is a programmer and occasional writer currently working as a Principal Animation Programmer at Epic Games and doing research mainly on Machine Learning and

Character Animation. His interests are Computer Graphics & Animation, Machine Learning, Game Development, Theory of Computation, and Programming Languages.

- **Gerard Pons-Moll**. Gerard Pons-Moll is a Professor at the University of Tübingen endowed by the Carl Zeiss Foundation, at the department of Computer Science. He is also core faculty at the Tübingen AI Center and head of the Emmy Noether independent research group "Real Virtual Humans", senior researcher at the Max Planck for Informatics (MPII) in Saarbrücken, Germany, and faculty at the IMPRS-IS (International Max Planck Research School - Intelligent Systems in Tübingen) and faculty at Saarland Informatics Campus. His research lies at the intersection of computer vision, computer graphics and machine learning – with special focus on analyzing people in videos, and creating virtual human models by "looking" at real ones. His research has produced some of the most advanced statistical human body models of pose, shape, soft-tissue and clothing (which are currently used for a number of applications in industry and research), as well as algorithms to track and reconstruct 3D people models from images, video, depth, and IMUs. His work has received several awards including the prestigious Emmy Noether Grant (2018), a Google Faculty Research Award (2019), a Facebook Reality Labs Faculty Award (2018), and recently the German Pattern Recognition Award (2019), which is given annually by the German Pattern Recognition Society to one outstanding researcher in the fields of Computer Vision and Machine Learning. In 2020 he received a Snap-Research gift. His work got Best Papers Awards BMVC'13, Eurographics'17, 3DV'18 and CVPR'20 and has been published at the top venues and journals including CVPR, ICCV, Siggraph, Eurographics, 3DV, IJCV and PAMI. He served as Area Chair for ECCV'18, 3DV'19, SCA'18'19, FG'20, ECCV'20. He will serve as Area Chair for CVPR'21, IJCAI'21 and 3DV'20.

- **Zhengyi Luo**. Zhengyi Luo is a final year PhD student at Carnegie Mellon University's Robotics Institute, School of Computer Science, advised by Prof. Kris Kitani. He earned my bachelor's degree from University of Pennsylvania in 2019, where he worked with Prof. Kostas Daniilidis. His research has been supported by Qualcomm Innovation Fellowship and the Meta AI Mentorship Program. His research interest lies at the intersection of vision, learning, and robotics. He is working on topics including human pose estimation, human-object interaction, human motion modeling etc. Through his research, he wants to create methods that effectively interpret spatial-temporal sensory input and build a representation of the 3D world to reason about the interactions between agents and the physical environment. On the application side, he is excited about humanoid robots and AR/VR.

- **Libin Liu**. Libin Liu is an assistant professor at the School of Intelligence Science and Technology, Peking University. Before joining Peking University, he was the Chief Scientist of DeepMotion Inc. He was a postdoctoral research fellow at Disney Research and the University of British Columbia. He received my Ph.D. in computer science and B.S. degree in mathematics and physics from Tsinghua University. He is interested in character animation, physics-based simulation, motion control, and related areas such as reinforcement learning, deep learning, and robotics. He put a lot of work into realizing various agile human motions on simulated characters and robots.