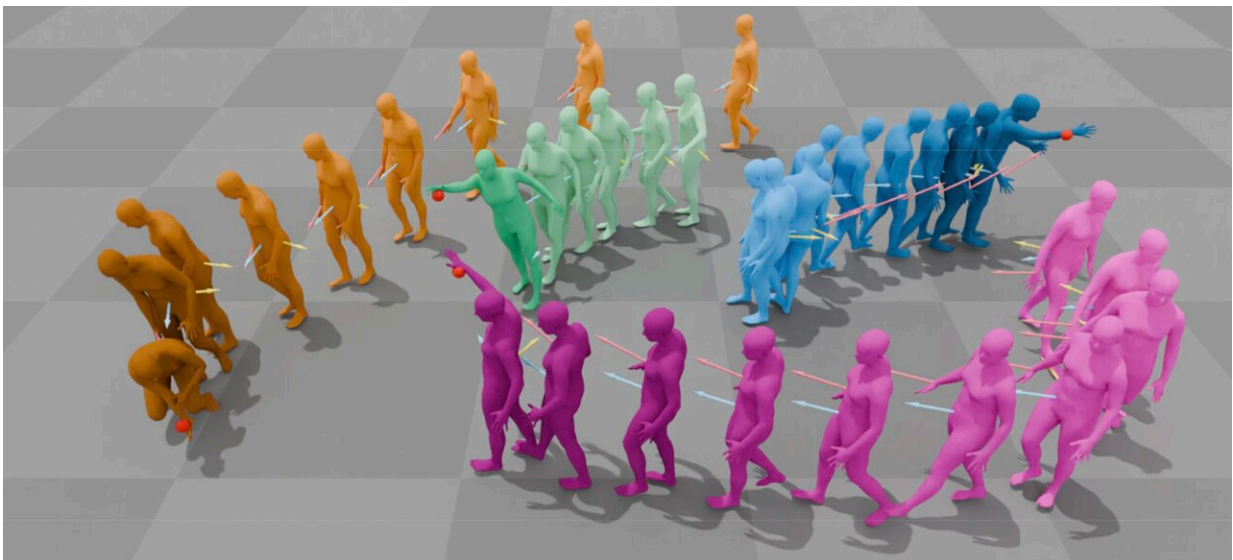


Data-driven model generates natural human motions for virtual avatars

May 30 2024, by Ingrid Fadelli



WANDR starts from an arbitrary body pose and generates precise and realistic human motions that reach a specified 3D goal (depicted as a red sphere). Employing a purely data-driven approach, WANDR is a conditional Variational Autoencoder guided by intention features (depicted arrows) that steer the human's orientation (yellow), position (cyan) and wrist (pink) towards the goal. WANDR can reach a wide range of goals even if they deviate significantly from the training data. Credit: Diomatari et al.

Humans can innately perform a wide range of movements, as this allows them to best tackle various tasks in their day-to-day life. Automatically reproducing these motions in virtual avatars and 3D animated human-

like characters could be highly advantageous for many applications, ranging from metaverse spaces to digital entertainment, AI interfaces and robotics.

Researchers at Max Planck Institute for Intelligent Systems and ETH Zurich recently developed WANDR, a new model that can generate natural human motions for avatars. This model, to be introduced in a paper presented at the Conference on Computer Vision and Pattern Recognition ([CVPR 2024](#)) in June, unifies different data sources under a single model to attain more realistic motions in 3D humanoid characters. The [paper](#) is also posted to the *arXiv* preprint server.

"At a high-level, our research aims at figuring out what it takes to create [virtual humans](#) able to behave like us," Markos Diomataris, first author of the paper, told Tech Xplore. "This essentially means learning to reason about the world, how to move in it, setting goals and trying to achieve them.

"But why go after this research problem? Fundamentally, we want to better understand humans, just like a neuroscientist would, and we are attempting this by following a 'try to build what you want to understand' philosophy."

The primary objective of the recent study by Diomataris and his colleagues was to create a model that would generate realistic motions for 3D avatars. These generated motions would allow the avatars to eventually interact with their [virtual environment](#), for instance reaching to grab objects.

"Consider reaching for a coffee cup—it can be as straightforward as an arm extension or can involve the coordinated action of our entire body," Diomataris said. "Actions like bending down, extending our arm, and walking must come together to achieve the goal. At a granular level, we

continuously make subtle adjustments to maintain balance and stay on course towards our objective."

By making these subtle adjustments, humans can produce fluid motions, integrating numerous smaller movements that converge towards a simple goal (e.g., placing a hand on a cup). Diomataris and his colleagues set out to teach a human [avatar](#) the same skills.

One approach to teach virtual agents new skills is reinforcement learning (RL), while another is to compile a [dataset](#) containing human demonstrations and then use it to train a machine learning model. These two approaches have different strengths and limitations.

"RL, in very simple terms, is learning skills through experience gathered from trial and error," Diomataris explained. "For our task, the agent would have to try all kinds of random motions in the beginning of its training until it manages to first stand properly, then walk, orient itself towards the goal, navigate towards it and finally reach it with its hand.

"This approach does not necessarily need a dataset, but it can require large amounts of computation as well as tedious design of rewards for the agent to prevent unnatural looking behaviors (e.g. prefer crawling instead of walking when moving)."

In contrast with RL, training models using datasets provides a virtual agent with richer information about a skill, rather than allowing it to figure this information out alone. While there are now various large datasets containing human motion demonstrations, very few include reaching motions, which the team also wished to replicate in avatars.

"Prioritizing motion realism, we chose to learn this skill from data," Diomataris said. "We present a method that is able to leverage both big datasets with various general motions and smaller datasets that specialize

in humans reaching for goals."

Diomataris and his colleagues first designed a training objective that is agnostic to the existence of goal labels. This key step allowed WANDR to learn general navigation skills from larger datasets, while still using the labeled data it attained from smaller datasets.

"WANDR is the first human motion generation model that is driven by an active feedback loop learned purely from data, without any extra steps of reinforcement learning (RL)," Diomataris said. "What is an active feedback loop? WANDR generates motion autoregressively (frame-by-frame). At each step, it predicts an action that will progress the human to its next state."

WANDR's predictions of avatar actions are conditioned by time and goal dependent features, which the researchers define as "intention." These features are re-computed at every frame, acting as a feedback loop that guides an avatar in reaching a given goal using its wrist.

"This means that, similarly to a human, our method constantly adjusts the actions taken trying to orient the avatar towards the goal and reach for it," Diomataris said. "As a result, our avatar is able to approach and reach moving or sequential goals even though it has never been trained for something like that."

Existing datasets containing goal-oriented reaching human motions, such as [CIRCLE](#), are scarce and do not contain enough data to allow models to generalize across different tasks. This is why RL has so far been the most common approach to training models for reproducing human motions in avatars.

"Inspired by the paradigm of behavioral cloning in robotics, we propose a purely data-driven approach where during training a randomly chosen

future position of the avatar's hand is considered as the goal," Diomataris said.

"By hallucinating goals this way, we are able to combine both smaller datasets with goal annotations such as CIRCLE, as well as large scale like [AMASS](#) that have no goal labels but are essential to learning general navigational skills such as walking, turning etc."

WANDR, the model developed by Diomataris and his colleagues, was trained on data from different datasets and sources. By appropriately mixing data from these sources, the model produces more natural motions, allowing an avatar to reach arbitrary goals in its environment.

"So far, works that study motion generation either use RL, or completely lack the element of online adaptation of motion," Diomataris said.

"WANDR demonstrates a way to learn adaptive avatar behaviors from data. The 'online adaptation' part is necessary for any real time application where avatars interact with humans and the real world, like for example, in a virtual reality video game or in human-avatar interaction."

In the future, the new [model](#) introduced by this team of researchers could aid the generation of new content for videogames, VR applications, animated films, and entertainment, allowing human-like characters to perform more realistic body movements. As WANDR relies on various data sources and datasets with human motions are likely to grow over the next decades, its performance could soon improve further.

"Right now, there are two major pieces missing that we plan on researching in the future," Diomataris added. "Firstly, avatars need to be able to leverage large and uncurated datasets of videos to learn to move and interact with their virtual world and in addition to this, they need to

have the ability to explore their virtual world and learn from their own experiences.

"These two directions represent the fundamental means that humans also acquire experience: by taking actions and learning from their consequences, but also observing others and learning from their experience."

More information: Markos Diomataris et al, WANDR: Intention-guided Human Motion Generation, *arXiv* (2024). [DOI: 10.48550/arxiv.2404.15383](https://doi.org/10.48550/arxiv.2404.15383)

© 2024 Science X Network

Citation: Data-driven model generates natural human motions for virtual avatars (2024, May 30) retrieved 20 June 2024 from <https://techxplore.com/news/2024-05-driven-generates-natural-human-motions.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.