A novel framework for smooth avatar-user synchronization in the metaverse

October 2 2023

State space f_n r_m

Action space r_m'

New state

New reward setting

Environment

Metaverse services

Allocated resource f_T for current service

$$f_T \geq \sum_{i=1}^{n} f_i , f_i \leq M_i$$

$$M_i \bmod f_i = 0$$

f_1 fps $\Delta(f_1)$ f_2 fps f_3 fps   ... f_N fps $\Delta(f_N)$

Contest Theory $\Delta(f_2)$ $\Delta(f_3)$ Tier-2 device
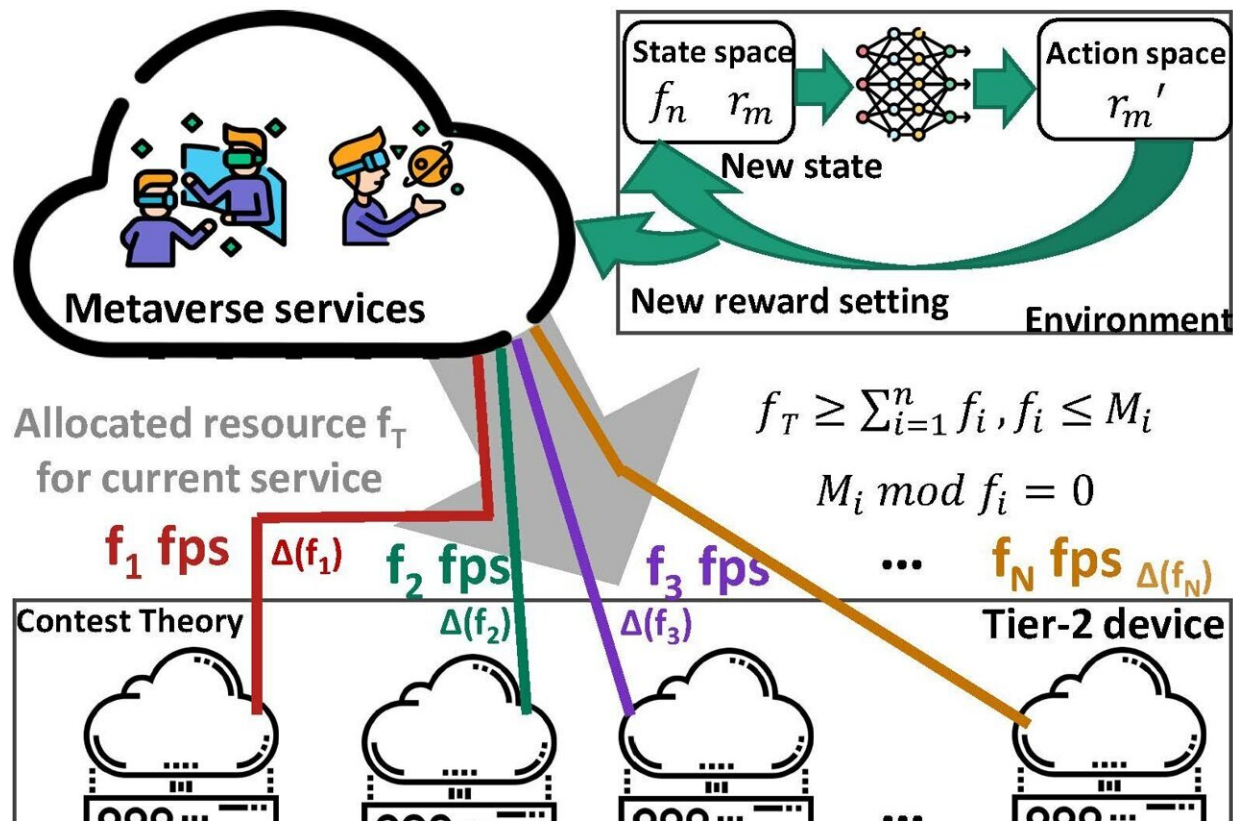
Credit: SUTD

Mobile phones, smartwatches and earbuds are some gadgets that we carry around without much thought. The increasingly digitized world sees a shrinking gap between human and technology, and many

researchers and companies are interested in how technology can be further integrated into our lives.

What if, instead of incorporating technology into our physical world, we assimilate ourselves into a [virtual environment](#)? This is what Assistant Professor Xiong Zehui from the Singapore University of Technology and Design (SUTD) hopes to achieve in his research.

Working with researchers from the Nanyang Technological University and the Guangdong University of Technology, this collaboration has so far yielded a preprint, "[Vision-based semantic communications for metaverse services: A contest theoretic approach](#)." The research will be presented at the [IEEE Global Communications Conference in December 2023](#).

The joint effort focused on the notion of the [metaverse](#)—a virtual reality (VR) universe where users can control avatars to interact with the virtual environment. In this world, people can meet others (through their avatars), visit virtual locations and even make online purchases. In a sense, the metaverse hopes to extend past the limits of our physical reality.

One challenge for mainstream adoption of metaverse services is the demand for real-time synchronization between human actions and [avatar](#) responses. "In the metaverse, avatars must be updated and rendered to reflect users' behavior. But achieving real-time synchronization is complex, as it places high demands on the rendering resource allocation scheme of the metaverse service provider (MSP)," explained Xiong.
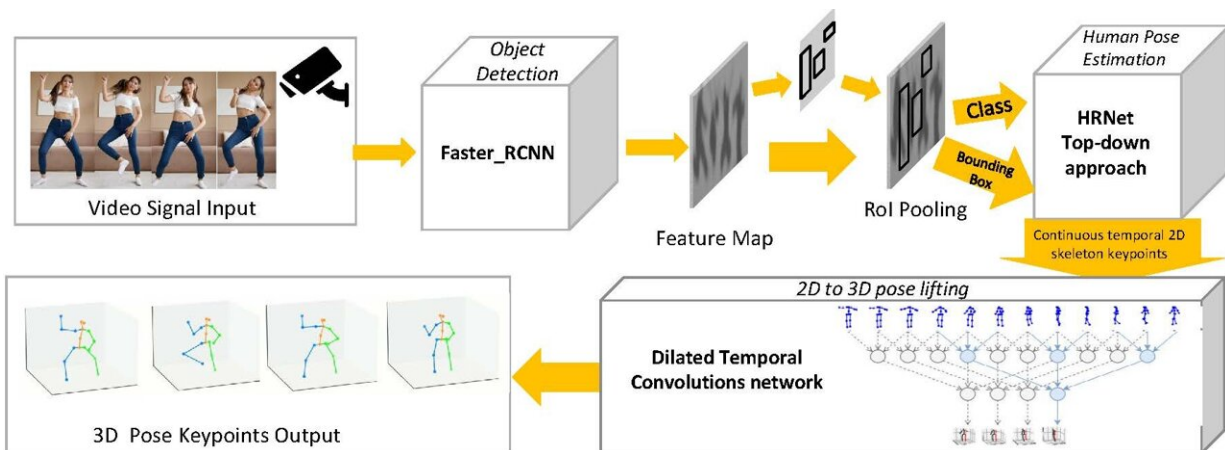
Credit: SUTD

MSPs take on an enormous burden, relaying gargantuan amounts of data between users and the server. The more immersive the experience, the larger the data payload. Individuals that perform fast actions, such as running or jumping, will be more likely to face a lapse in smoothness of their avatars, as the MSP struggles to keep up.

A common solution is to restrict the number of users in a single virtual environment, ensuring the MSP has sufficient resources, or bandwidth, to simulate all users regardless of activity. This is a highly inefficient approach as users who are standing still will be afforded excess resources that they do not need. Only users with large movements require constant updates to their avatar, and hence the surplus bandwidth. The problem then leaves the question hanging—how can resources be allocated without wastage?

Xiong and team proposed a novel framework to optimize resource allocation in MSPs, with the overall aim of ensuring a smooth and immersive experience for all users. The scheme uses a semantic

communication technique dubbed human pose estimation (HPE) to first reduce the information payload for users. Picking the most efficient distribution of resources among users was performed using contest theory, with user devices competing for just enough resources to simulate their avatars.

The first step for a seamless avatar-user interface requires efficient encoding of information to the MSPs. Consider a camera capturing the movements of a human to be translated into motions of their avatar. Each image captured by the camera is packed with redundant background information that is not useful for modeling the virtual characters.

In HPE, the computer is tasked to identify humans as the object, and highlight only the skeletal joints. Based on the joints, the algorithm can reconstruct a simple stickman-like model that can be sent to the MSPs. This caricature then guides the MSPs to model the actions taken by the avatar. In the research, Xiong and team managed to reduce the data overhead by a million-fold, from megabytes to bytes.
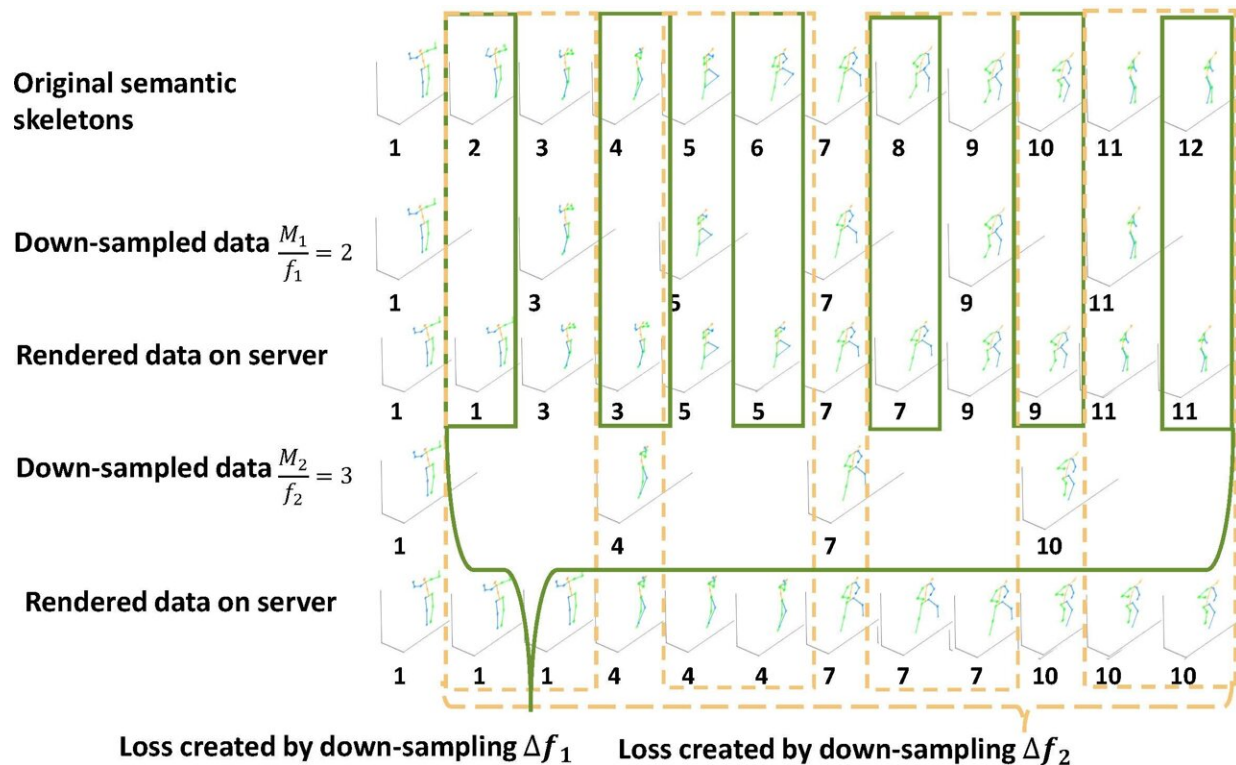
Credit: SUTD

With this massive savings in bandwidth, the team then turned to modeling interactions between the MSPs and the network of users using contest theory. In this approach, users (or rather, their devices) are competitors fighting for the resources of the MSP. The algorithm seeks to minimize the latency across all users over a fixed amount of available resources. At the same time, the individual devices decide on their own update rates, depending on the actions taken by the user.

To test for lag, the algorithm measures the differences in the avatar position with different update rates. Users that face lag will have large discrepancies between their HPE stickmen and their avatars. At the same time, the MSP's resources are treated as an award given out to

competitors that performed well without lag.

However, each user still needs to be able to accurately deduce the right amount of resources to request from the MSP. Faced with the complexity of the task, the team turned to using machine learning. A [neural network](#), dubbed the deep Q-network (DQN), optimizes the resources distributed. Under this framework, the team effort yielded a 66% improvement in lag across all users, compared to traditional methods.

Xiong is optimistic for the future of the metaverse, citing health care, education, and marketing as potential areas that could benefit from metaverse services. He said, "Some developments or advancements that I'm most looking forward to include integrating cutting-edge technologies such as generative AI and VR, as well as the growth of global, digital, and virtual economies. It will be exciting to see how these advancements shape the future of the metaverse."

The findings are published on the *arXiv* preprint server.