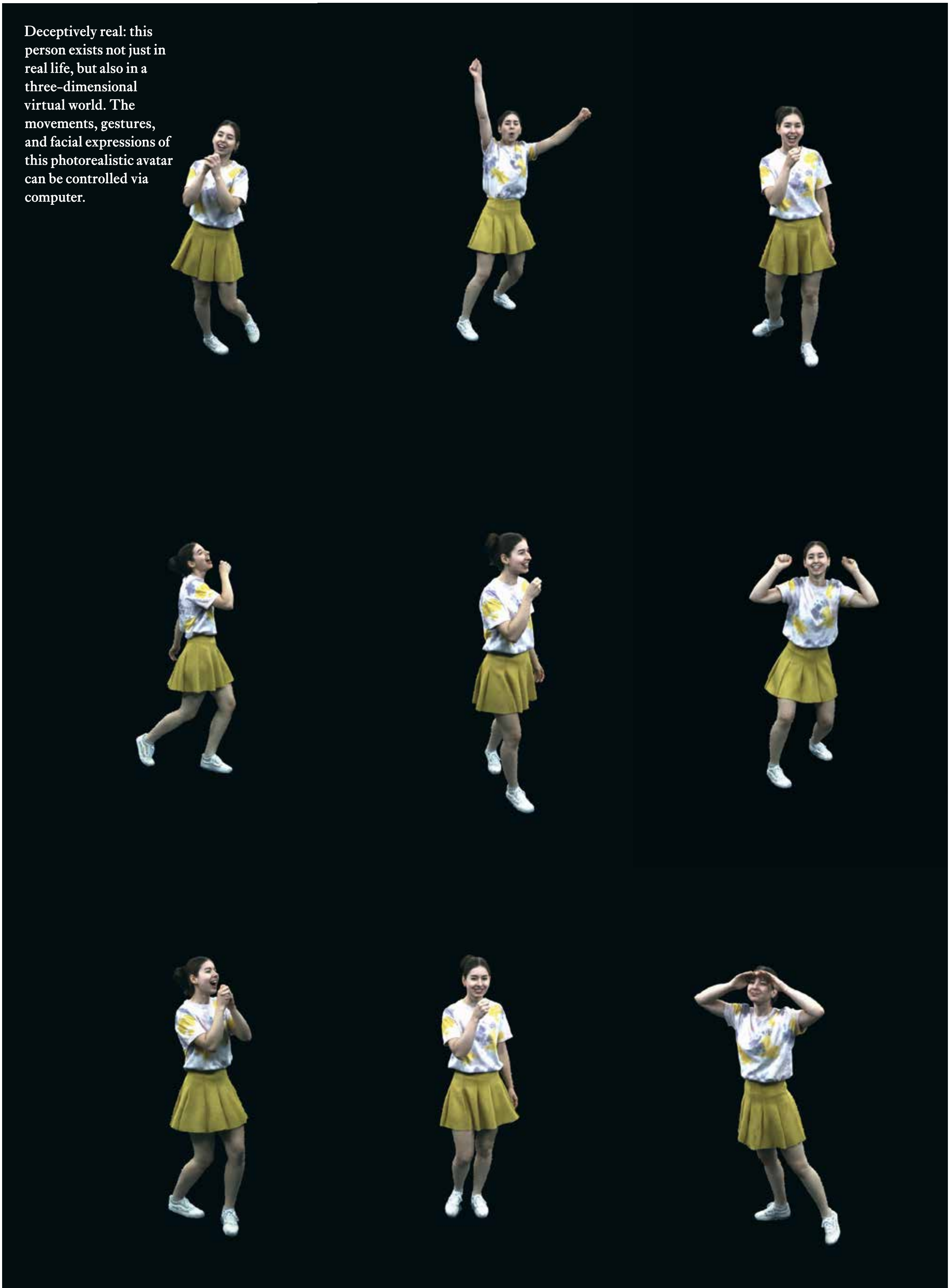

Deceptively real: this person exists not just in real life, but also in a three-dimensional virtual world. The movements, gestures, and facial expressions of this photorealistic avatar can be controlled via computer.

PHOTO: HENDRIK JUNKAWITSCH, GUOXING SUN, HEMING ZHU, CHRISTIAN THEOBALT, AND MARC HABERMANN. 2025. EVA: EXPRESSIVE VIRTUAL AVATARS FROM MULTI-VIEW VIDEOS. IN PROCEEDINGS OF THE SPECIAL INTEREST GROUP ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES CONFERENCE PAPERS (SIGGRAPH CONFERENCE PAPERS '25). ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, NY, USA, ARTICLE 147, 1–11.

AVATARS AT EYE LEVEL

TEXT: KATJA ENGEL

Conferences and business meetings generate substantial CO_2 emissions when participants travel by plane. Digital conferences with realistic avatars in virtual space could offer a solution. To make this work seamlessly with a headset when working from home, researchers at the Max Planck Institute for Informatics in Saarbrücken are refining the digital measurement of people – and using it to train artificial intelligence.

In 1977, the first *Star Wars* film took the world by storm and presented a vision of a new kind of communication technology in the process. The droid R2-D2, resembling a dustbin on wheels, replays a hologram inside the modest home of Jedi Obi-Wan Kenobi on the planet Tatooine. A bluish, flickering image of Princess Leia – Luke Skywalker's sister – appears, projected onto a stone pedestal. With a desperate expression, she pleads for Obi-Wan's help in the fight against Darth Vader's Empire: "Help me, Obi-Wan Kenobi. You're my only hope!" says a tinny voice coming from the robot. When the film premiered in cinemas, this form of communication was pure science fiction – and even video calls on two-dimensional screens were, at the time, possible only with considerable effort.

Even today, nearly 50 years later, holographic messages have not become a reality. In 2018, researchers in the United States managed to mimic the technology made famous by Star Wars: in their volumetric, three-dimensional laser projections, people appear to flicker in space – just like in the film. Yet the illusion falls short. The projections are smaller than life-sized and look distinctly artificial.

Researchers at the Max Planck Institute for Informatics in Saarbrücken are taking a different approach. They are working on telepresence via avatars in virtual spaces. The group led by →

Christian Theobalt, director at the Institute and head of the Department of Visual Computing and Artificial Intelligence, aims to capture people as precisely as possible in order to recreate them digitally. Using artificial intelligence and just a few cameras mounted on a headset, they generate a photorealistic three-dimensional likeness of the person wearing it. Multiple users of the technology can then meet in virtual conference rooms and feel as if they are physically present – hence the term telepresence. "If I want to achieve something like telepresence, the avatar model has to be generated extremely quickly in real time," explains Theobalt. Access to virtual spaces is already possible today through VR and AR headsets – short for virtual reality and augmented reality. The former are fully immersive, allowing users to step entirely into a simulated environment. The latter, still the subject of ongoing research, are partly transparent and give the impression that three-dimensional figures are moving within the real world. The scene from Star Wars would therefore also have worked if everyone present had been wearing an AR headset.

There are already three-dimensional avatars that allow people to meet in virtual spaces, for example, in gaming or at certain conferences. But most of these are only semi-realistic, with a cartoonish look. "Between reality and the digital image, there is a gap known as the uncanny valley," explains Marc Habermann, group leader in Theobalt's team. Humans can immediately tell when an avatar is only almost real, and that near-realism quickly creates discomfort. That's the crux of the problem and why many video-conferencing platforms stick with cartoon-style avatars. "But we've now managed to leave that uncanny valley behind by creating super-photorealistic avatars of real people," says Habermann. At the Max Planck Institute, researchers are leading the way in developing full-body avatars, while also keeping pace with major industry players such as Meta when it comes to photorealistic faces.

"Considering today's technology, it's actually surprising how long we've been stuck in the medium of video calls," says Habermann. Video conferencing became mainstream during the Covid-19 pandemic, if not even before that, yet many complain about how exhausting it can be. Making use of three-dimensional, highly realistic personal avatars, telepresence in meetings or conferences could provide a fresh approach. While video calls only transmit faces and voices, telepresence allows participants to interact virtually through facial expressions, gestures, and body language. This creates a more immersive and personal experience, especially when one's own avatar appears photorealistic.

SUMMARY

A photorealistic virtual self that can be created easily from home – that's the goal of researchers at the Max Planck Institute for Informatics.

The technology is also used in the film industry, but the effort there is considerable. To allow people to create a full-body avatar using only a VR headset with built-in cameras and sensors, the scientists are training a complex AI model.

In the future, such avatars could make it possible to hold international meetings in virtual spaces that are far more climate-friendly than in-person gatherings, while still conveying a strong sense of presence and liveliness.

To achieve this, the Saarbrücken-based Max Planck team first teaches an AI to understand the human body – its appearance and movements. A photorealistic three-dimensional representation of an individual is then generated based on the model that is created in this process, explains Theobalt.

The first stage in training the AI is the "Green Room". It's as spacious as half a handball court, and the walls are draped with green fabric from which 120 high-resolution cameras protrude. They capture every angle of a person's movements, whether they're doing push-ups, playing golf, or walking. Computers analyze the images, extract the motion, and simultaneously calculate the skeletal structure in real time, which is then displayed on the person in the video as colored lines. For real-time motion tracking, the researchers use software from The Captury, a startup spun out of Theobalt's group in 2013. Six hours of recording generates a petabyte of raw data, which is hundreds of times more than the storage capacity of a typical consumer hard drive.

The neural network behind the AI generates a three-dimensional representation of a human body from the motion data – a model that moves like the test subject but doesn't yet look like them. A convincing photorealistic avatar also requires natural shadows, skin that reflects light correctly, and clothing that folds realistically. The as-yet-untrained AI body model learns all of this step by step, by repeatedly being confronted with the person's real appearance.

As Realistic as in *Star Wars*

This takes place in another room, the Lightstage. The 3.5-meter-high dome is surrounded by darkness. Inside, 13,000 individually controlled LEDs shine on the test subject from every angle, staging them as if they were standing amid the colorful light patterns of the stained-glass windows of Grace Cathedral in San Francisco. "We can also simulate how light falls in a deep urban canyon or a plain conference room," says Marc Habermann, Scientific Manager of the Real Vir-

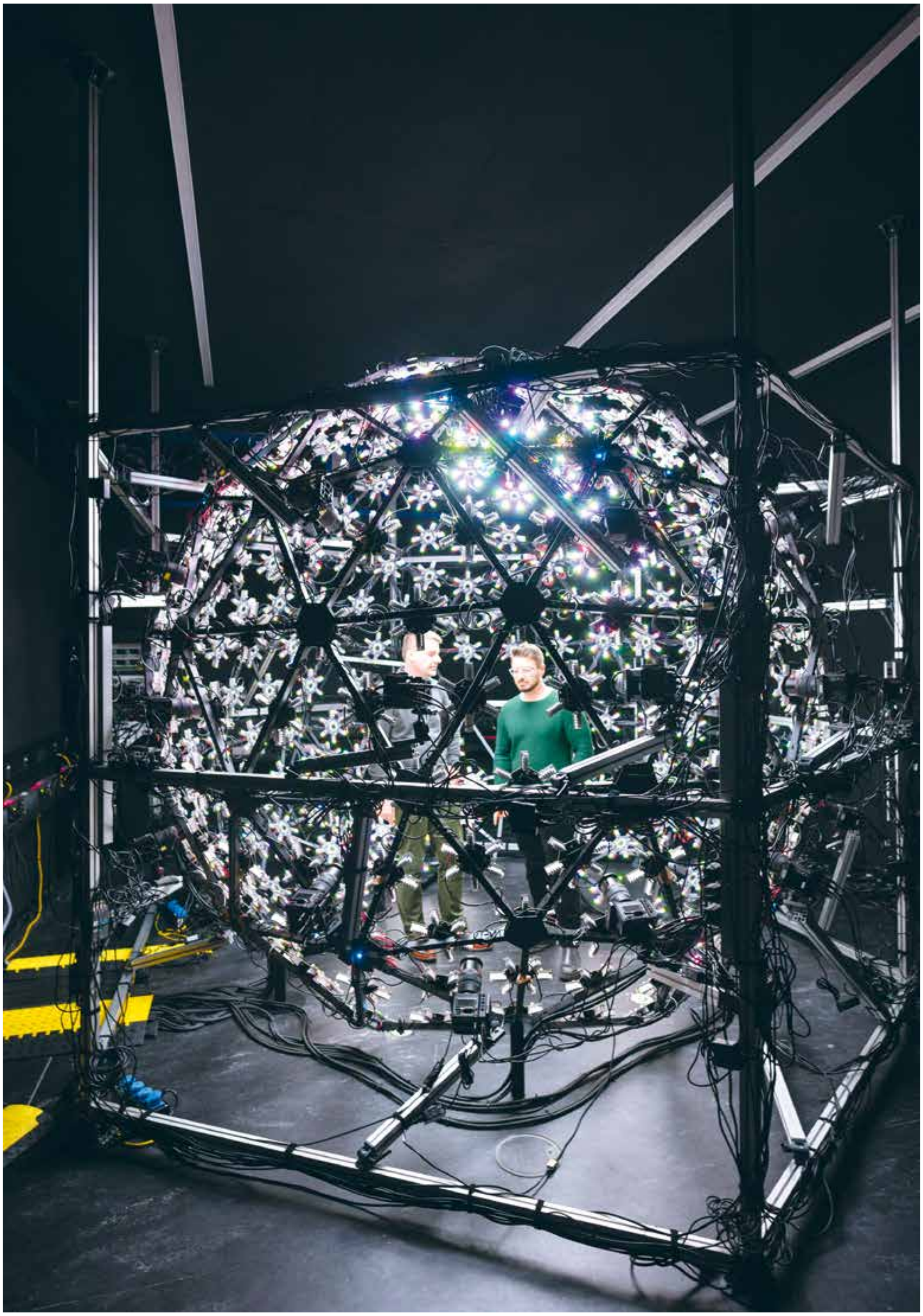

PHOTO: OLIVER DIETZE / MAX PLANCK INSTITUTE FOR INFORMATICS

A world of their own: at the push of a button, this Lightstage immerses Christian Theobalt (left) and Marc Habermann in one of hundreds of preset lighting scenarios – whether inside Notre Dame or in a brightly lit conference room.

tual Lab, which houses the Lightstage. In total, the Saarbrücken team has more than 500 interior lighting simulations alone, with even more for outdoor scenarios. Alongside the LEDs, 40 high-resolution cameras are mounted in the dome. They capture the person from every angle, including how skin, hair, and clothing reflect the light. Using all of this information, the AI's neural network learns to dress the avatar in photorealistic textures that take into account the material properties of the clothing. This method of AI training is not new and is used in parts of the film industry. Under the umbrella of CGI, or computer-generated imagery. Lightstages are employed to give digitally created superheroes in the Marvel movies – or the towering characters in

→

IMAGE: GLEN REES

Cheese! Members of the international organization Astronomers for Planet Earth gather in January 2025 in the auditorium of the virtual conference center – specially designed for their anniversary meeting – for a group photo. On the right of the image an exhibition area can be accessed in the Spatial app.

Avatar – a look that resembles the real actors. On film sets, their real movements are captured using a marker suit, a full-body suit with reflective markers tracked and transferred to the digital model. A striking example of how far the technology had already come nine years ago can be found in *Star Wars*. The first film, released in 1977, is Episode IV of the saga. *Rogue One*, released in 2016, is set just a few weeks before Episode IV from 1977. Because the original actor who played Tarkin, the commander of the Death Star, had passed away in 1994, he had to be digitally recreated for Rogue One. The team used archival video footage, a highly realistic plaster model of Tarkin, and the body of a live actor, who – equipped with a marker suit and filmed in a Lightstage – was transformed into the spitting image of Tarkin. But: “It takes months to complete a digital film character, even with the highly equipped visual effects pipeline of the film industry,” says Christian Theobalt.

For telepresence to work from the home office, the entire process of creating the avatar has to happen at the snap of a finger. “At the moment, a person would theoretically still need to come to us and have their body scanned in the Lightstage,” says Habermann. The ultimate goal, however, is a generic AI model that has learned from countless scans of different bodies under various lighting conditions and can be applied to anyone. Once this fully trained generic model exists, all that’s needed to create a virtual full-body self is the model and a VR or AR headset equipped with just a few cameras and sensors.

Some of the cameras on these headsets point forward, while the down-facing cameras capture the lower part of the body, and an IMU (inertial measurement unit) measures movement by tracking acceleration and rotation. The cameras detect how a person’s forearms, parts of the torso, and feet move, as well as the distance to the ground. In the future, this data should be sufficient for the generic avatar model not only to look like the real person but also to replicate their movements live in the virtual space. “The model has even learned to infer, from the motion of an arm, how and where the virtual avatar’s clothing will fold,” says Habermann.

The Problem of Deepfakes

Beyond telepresence, animated film scenes, or more realistic gaming, the technology has other potential applications. “Our research isn’t aimed at any single use case,” emphasizes Theobalt. These include remote diagnostics and telemedicine, where movement patterns are important. Avatar technology could also change how people interact with AI. In turn, the AI could communicate as an avatar using gestures and facial expressions. “If I ask ChatGPT to explain a yoga pose, I get a detailed text description,” says Habermann.

"Wouldn't it be much easier if ChatGPT appeared as an avatar and demonstrated the pose for me?"

As exciting as these applications sound, it's equally important to understand the risks. "What once required a 40-person CGI team in the film industry could soon happen on someone's living room couch," says Marc Habermann. "We are at a threshold where the technology is becoming increasingly accessible." Once a universal avatar model exists, the path to deepfakes isn't far. "That's why our doctoral researchers attend ethics classes to learn how to handle the technology responsibly," says Theobalt. There is also a dedicated research group at the Max Planck Institute for Informatics that studies why an AI makes the decisions it does. Theobalt's team has even used this approach to detect deepfakes. In this method, they work backward through the steps that lead to a generic avatar model and analyze why the avatar looks the way it does. Perhaps in the future, virtual conference organizers will also need to check "digital watermarks" for participants, ensuring that the person behind an avatar is truly who they claim to be.

If a fully secure and virtual meeting room becomes a reality, conferences and business meetings could become significantly more climate-friendly: air travel to global congresses or international corporate conferences produces a substantial amount of greenhouse gas emissions. Leonard Burtscher, an astronomer and advisor on climate and environmental policy at the Munich Environmental Institute, together with other researchers, calculated the difference using the example of the five-day annual meeting of the European Astronomical Society. When the 2019 conference was held in person in Lyon, roughly 3,000 times more CO_2 was emitted than the following year, when the conference took place entirely online due to the Covid-19 pandemic.

According to a subsequent study by Andrea Gokus from Washington University in St. Louis, which also involved Knud Jahnke from the Max Planck Institute for Astronomy in Heidelberg, each participant at one of the 362 astronomy conferences analyzed was responsible for an average of around one ton of climate-damaging greenhouse gases. Even when taking into account the intensive training of a generic avatar model, a conference of individual avatars would likely have a significantly smaller CO_2 footprint than a conference where all participants travel by plane. After all, the computational power required to train the model is only needed once.

And three-dimensional virtual conferences can be fun, as Leonard Burtscher knows well. He immediately recalls a situation: "We had a group meeting around a virtual campfire on a virtual island. Unlike Zoom meetings, I still clearly remember the campfire, the nighttime setting, a palm tree swaying in the wind – and I still associate it with the discussion we had there," says Burtscher. So if an in-person meeting isn't possible – because travel would be too cumbersome or environmentally harmful – meeting as avatars can also make sense from a psychological perspective. Even in virtual reality, participants develop a sense of connection to the meeting place. "When they think back to it later, it helps them remember the conversations that took place," says Simone Kühn, Director of the Center for Environmental Neuroscience at the Max Planck Institute for Human Development in Berlin. She studies how the physical environment affects the individual and the brain.

But that's not all: many people know how exhausting video conferences can be, with participants crammed into a checkerboard of video tiles. "A well-designed virtual space with realistic avatars, on the other hand, can be stimulating," says Kühn. "Accurate representations of natural environments can even reduce stress." So why not let your personal avatar take a short break and stroll along a virtual sandy beach – conveniently placed right outside the meeting room door?

←

An avatar at home: the cameras on a VR headset (left) capture the movements of the body, hands, and feet, transmitting them in real time to the computer-generated photorealistic avatar in the virtual, three-dimensional space (right).

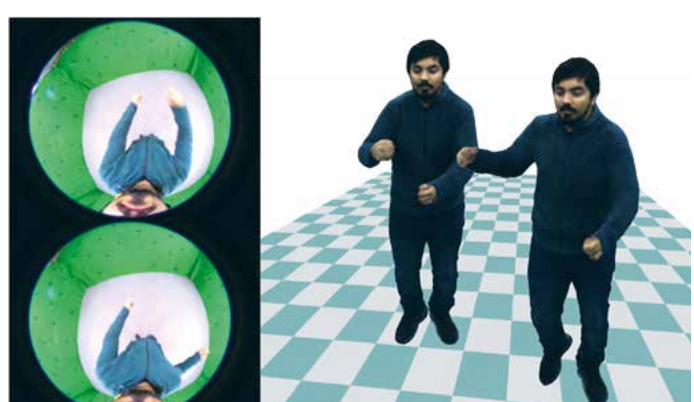

IMAGE: JIANCHUN CHEN, JIAN WANG, YINDA ZHANG, ROHIT PANDEY, THABO BEELER, MARC HABERMANN, AND CHRISTIAN THEOBALT. 2024. EGOAVATAR: EGOCENTRIC VIEW-DRIVEN AND PHOTOREALISTIC FULL-BODY AVATARS. IN SIGGRAPH ASIA 2024 CONFERENCE PAPERS (SA '24). ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, NY, USA, ARTICLE 102, 1–11.