

# Tracking humans in 3-D with off-the-shelf webcams

July 5 2017

---



Computer scientists have developed a system that requires only a web camera to capture a person's movements digitally in 3-D. Credit: Oliver Dietze

Many applications require that people and their movements are captured digitally in 3-D in real-time. Until now, this was possible only with

expensive systems of several cameras, or by having people wear special suits. Computer scientists at the Max Planck Institute for Computer Science have developed a system that requires only a single video camera. It can even estimate the 3-D pose of a person acting in a pre-recorded video, for instance a YouTube video.

"This lets you capture [video](#) with your cell phone out in the Alps and do body tracking. Doing this in 3D, in real-time and just with a [camera](#) like the one on your mobile device—that is a big leap," reports Dushyant Mehta, PhD student in the Graphics, Vision and Video Group headed by Professor Christian Theobalt at the Max Planck Institute for Informatics in Saarbruecken (MPI).

Together with his colleagues, he developed a software system that needs only a conventional camera to digitally capture a person, along with their movements, in real-time. "So far, several video cameras, or a so-called depth camera as in the Kinect, have been necessary for this task," explains Srinath Sridhar, also a researcher in the Graphics, Vision and Video Group.

The new system is based on a [neural network](#) which researchers call a "[convolutional neural network](#)", or CNN for short, that is often associated with the term "deep learning". The MPI researchers have developed a new method to calculate the three-dimensional pose of the person from the two-dimensional information of the video streams with the aid of a neural network.

A short video on their website, produced by the scientists, shows what this looks like. A researcher juggles with clubs in the back of a room, while in the foreground a monitor shows the corresponding video recording. The figure of the researcher is here superimposed by a simplified, red stick figure. Another 3D view shows the motion from the side, showing that, for the first time, the full 3D pose is captured in real-

time. No matter how fast or how far the researcher moves or extends his or her limbs, the stick figure makes the same movements in 3D, just like the more fleshed-out virtual character version in the virtual space, shown on another monitor to the left.

The researchers call their system "VNect". The system both predicts both the 3D pose of the person in the image and localizes the person in the image. This allows the system to avoid wasting computations on image regions which don't contain a person. The neural network of the system is trained using tens of thousands of annotated images during the machine learning process. The system provides 3D pose information in terms of joint angles, which can easily be used to control virtual characters.

"VNect makes 3D body pose tracking for virtual reality of computer games accessible to a wider audience because they don't need to have Kinect or other cameras available, don't need to wear special suits, and can just use webcams which are more readily accessible," says Mehta and adds, "It also enables new experiences in first-person virtual reality." Besides this interactive character control, VNect is the first system which can also be used to estimate the 3D pose of a person in community videos such as those provided on the online platform YouTube. Christian Theobalt continues: "There are many other applications possible, from Human-Computer-Interaction to Human-Robot Interaction to Industry 4.0, where man and robot work together in a factory. Also think about autonomous driving, where the car may in the future estimate the full articulated motion of people from a color camera to assess their behavior."

But VNect still has its limitations. The accuracy of the pose estimation is a bit lower than the accuracy obtained with multi-camera or marker-based pose estimation. It gets into trouble if the face of the person is occluded, the motions are too fast or the poses are too far away from the

trained set of poses. Occlusion by multiple persons is a problem, too.

Nevertheless, Sridhar is sure that the technology will further mature and be able to handle increasingly more complex scenes, so that it can be used in everyday life.

**More information:** [gvv.mpi-inf.mpg.de/projects/VNect/](http://gvv.mpi-inf.mpg.de/projects/VNect/)

Provided by Saarland University

Citation: Tracking humans in 3-D with off-the-shelf webcams (2017, July 5) retrieved 25 April 2024 from <https://phys.org/news/2017-07-tracking-humans-d-off-the-shelf-webcams.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.