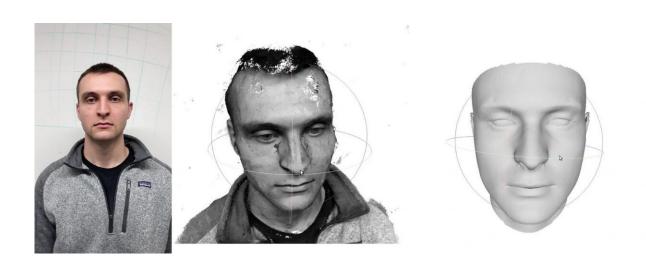# Smartphone videos produce highly realistic 3-D face reconstructions

April 1 2020, by Byron Spice

In a 3-D face reconstruction process developed at Carnegie Mellon University, smartphone video of a person, left, is analyzed to produce an imperfect model of the face, middle. Deep learning is then combined with conventional computer vision techniques to complete the reconstruction, right. Credit: Carnegie Mellon University

Normally, it takes pricey equipment and expertise to create an accurate 3-D reconstruction of someone's face that's realistic and doesn't look creepy. Now, Carnegie Mellon University researchers have pulled off the feat using video recorded on an ordinary smartphone.

Using a smartphone to shoot a continuous video of the front and sides of

the face generates a dense cloud of data. A two-step process developed by CMU's Robotics Institute uses that data, with some help from deep learning algorithms, to build a digital reconstruction of the face. The team's experiments show that their method can achieve sub-millimeter accuracy, outperforming other camera-based processes.

A digital face might be used to build an avatar for gaming or for virtual or augmented reality, and could also be used in animation, biometric identification and even medical procedures. An accurate 3-D rendering of the face might also be useful in building customized surgical masks or respirators.

"Building a 3-D reconstruction of the face has been an open problem in computer vision and graphics because people are very sensitive to the look of facial features," said Simon Lucey, an associate research professor in the Robotics Institute. "Even slight anomalies in the reconstructions can make the end result look unrealistic."

Laser scanners, structured light and multicamera studio setups can produce highly accurate scans of the face, but these specialized sensors are prohibitively expensive for most applications. CMU's newly developed method, however, requires only a smartphone.

The method, which Lucey developed with master's students Shubham Agrawal and Anuj Pahuja, was presented in early March at the IEEE Winter Conference on Applications of Computer Vision (WACV) in Snowmass, Colorado. It begins with shooting 15-20 seconds of video. In this case, the researchers used an iPhone X in the slow-motion setting.

"The high frame rate of slow motion is one of the key things for our method because it generates a dense point cloud," Lucey said.

The researchers then employ a commonly used technique called visual

simultaneous localization and mapping (SLAM). Visual SLAM triangulates points on a surface to calculate its shape, while at the same time using that information to determine the position of the camera. This creates an initial geometry of the face, but missing data leave gaps in the model.

In the second step of this process, the researchers work to fill in those gaps, first by using deep learning algorithms. Deep learning is used in a limited way, however: it identifies the person's profile and landmarks such as ears, eyes and nose. Classical computer vision techniques are then used to fill in the gaps.

"Deep learning is a powerful tool that we use every day," Lucey said. "But deep learning has a tendency to memorize solutions," which works against efforts to include distinguishing details of the face. "If you use these algorithms just to find the landmarks, you can use classical methods to fill in the gaps much more easily."

The method isn't necessarily quick; it took 30-40 minutes of processing time. But the entire process can be performed on a smartphone.

In addition to face reconstructions, the CMU team's methods might also be employed to capture the geometry of almost any object, Lucey said. Digital reconstructions of those objects can then be incorporated into animations or perhaps transmitted across the internet to sites where the objects could be duplicated with 3-D printers.

Provided by Carnegie Mellon University