Best way to detect 'deepfake' videos? Check for the pulse

October 27 2020

Frames from real videos (left) and deep fakes (right) are demonstrated for a small subset of the Deep Fakes Dataset. Credit: Binghamton University, Intel Corporation

With video editing software becoming increasingly sophisticated, it's sometimes difficult to believe our own eyes. Did that actor really appear in that movie? Did that politician really say that offensive thing?

Some so-called 'deepfakes' are harmless fun, but others are made with a more sinister purpose. But how do we know when a video has been manipulated?

Researchers from Binghamton University's Thomas J. Watson College of Engineering and Applied Science have teamed up with Intel Corp. to develop a tool called FakeCatcher, which boasts an accuracy rate above 90%.

FakeCatcher works by analyzing the subtle differences in skin color caused by the human heartbeat. Photoplethysmography (abbreviated as PPG) is the same technique used for a pulse oximeter put on the tip of your finger at a doctor's office, as well as Apple Watches and wearable fitness tracking devices that measure your heartbeat during exercise.

"We extract several PPG signals from different parts of the face and look at the spatial and temporal consistency of those signals," said Ilke Demir, a senior research scientist at Intel. "In deepfakes, there is no consistency for heartbeats and there is no pulse information. For real videos, the blood flow in someone's left cheek and right cheek—to oversimplify it—agree that they have the same pulse."

Working with Demir on the project is Umur A. Ciftci, a Ph.D. student at Watson College's Department of Computer Science, under Professor Lijun Yin's supervision at the Graphics and Image Computing Laboratory, part of the Seymour Kunis Media Core funded by donor Gary Kunis '73, LHD '02. It builds on Yin's 15 years of work creating multiple 3-D databases of human faces and emotional expressions. Hollywood filmmakers, video game creators and others have utilized the databases for their creative projects.

At Yin's lab in the Innovative Technologies Complex, Ciftci has helped to build what may be the most advanced physiological capture setup setup in the United States, with its 18 cameras as well as in infrared. A device also is strapped around a subject's chest that monitors breathing and heartrate. So much data is acquired in a 30-minute session that it requires 12 hours of computer processing to render it.

"Umur has done a lot of physiology data analysis, and signal processing research started with our first multimodal database," Yin said. "We capture data not just with 2-D and 3-D visible images but also thermal cameras and physiology sensors. The idea of using the physiology as

another signature to see if it is consistent with previous data is very helpful for detection."

Deepfakes found "in the wild" are many steps below the kind of quality that Yin's lab generates, but it means that manipulated videos can be much easier to spot.

"Considering that we work with 3-D using our own capture setup, we generate some of our own composites, which are basically 'fake' videos," Ciftci said. "The big difference is that we scan real people and use it, while deepfakes take data from other people and use it. It's not that different if you think about it that way.

"It's like the police knowing what all the criminals do and how they do it. You understand how these deepfakes are being done. We learn the tricks and even use some of them in our own data creation."

Since the FakeCatcher findings were published, 27 researchers around the world have been using the algorithm and the dataset in their own analyses. Whenever these kinds of studies are made public, though, there are concerns about telling malicious deepfake makers how their videos have been shown to be false, allowing them to modify their work to be undetectable in the future.

Ciftci is not too worried about that, however: "It's not going to be easy for someone who doesn't know much about the science behind it. They can't just use what's out there to make this happen without significant software changes."

Intel's involvement in the FakeCatcher research is connected to its interests in volumetric capture and augmented/virtual reality experiences. Intel Studios operates what Demir calls "the world's largest volumetric capture stage": 100 cameras in a 10,000-square-foot geodesic

dome that can handle about 30 people simultaneously—even a few horses once.

Future plans include volumetric-capture technology to be included in mainstream television shows, sports and augmented-reality applications, where the audience can immerse in any scene. Films in 3-D and VR also are in the works, with two VR projects recently premiering at the Venice Film Festival.

By compiling the FakeCatcher data and reverse-engineering it, Intel Studios hopes to make more realistic renderings that incorporate the kind of biological markers that humans with real heartbeats have.

"Intel's vision is changing from a chip-first company to putting AI, edge computing and data first," Demir said. "We are making a transformation to AI-specific approaches in any way we can."

Future research will seek to improve and refine the FakeCatcher technology, drilling further down into the data to determine how the deepfakes are made. That capability has many implications, including cybersecurity and telemedicine, and Yin also hopes for further collaborations with Intel.

"We're still in the brainstorming stage," he said. "We want to have an impact not only in academia but also to see if our research would have a role in industry."