Preventing manipulation in automated face recognition

October 1 2019

Illustration of a face morphing attack. The original images on the left and right were morphed to create the fake image (center). Credit: Fraunhofer HHI

From unlocking smartphones to speeding up airport security checks: the use of automated face recognition for personal identification continues to grow. But this authentication method is vulnerable to morphing attacks: criminals can misuse it by melding two different facial images into one. A single passport featuring a photograph manipulated this way can then be used by two different people. Together with their partners, Fraunhofer research teams are developing a system that foils this type of

attack using machine learning methods.

Travelers who regularly visit the U.S. are used to being asked to look into a camera during passport inspection. The electronic photo is instantly compared with the photo stored in the biometric passport. In this biometric facial recognition process, a program captures the digital data of the live image and compares it with the data of the chip image in order to determine whether or not the individual facial characteristics in the photos match. Face recognition can also be used to unlock smartphones and tablets. This method is intended to lock out unauthorized third parties and restrict access to sensitive data. But the technology is vulnerable to targeted attacks, as a variety of tests have already demonstrated. "Criminals are capable of tricking face recognition systems—like the ones used at border control—in such a way that two people can use one and the same passport," says Lukasz Wandzik, scientist at the Fraunhofer Institute for Production Systems and Design Technology IPK in Berlin. Together with his colleagues at the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, HHI and other partners (see box), he is developing a process that identifies the image anomalies that occur during digital image processing in morphing processes. "In a morphing attack, two facial images are melded into a single synthetic facial image that contains the characteristics of both persons," Wandzik explains. As a result, biometric face recognition systems authenticate the identity of both persons based on this manipulated photo in the passport.

These attacks can take place for example before or during the process of applying for an ID document. In project ANANAS (from the German acronym for "Anomaly Detection for Prevention of Attacks on Authentication Systems Based on Facial Images"), the partners are focusing on this problem by analyzing and researching simulated imaging data. Here they apply modern image processing and machine learning methods, in particular deep neural networks designed explicitly

for processing image data. These complex networks consist of a large number of levels which are linked with one another in multilayer structures. They are based on connections between mathematical calculation units and imitate the neural structure of the human brain.

Preventing identity theft with neural networks

In order to test the processes and systems being developed, the project partners start by generating the data used to train the image processing programs to detect manipulations. Here different faces are morphed into one face. "Using morphed and real facial images, we've trained deep neural networks to decide whether a given facial image is authentic or the product of a morphing algorithm. The networks can recognize manipulated images based on the changes occurring during manipulation, especially in semantic areas such as facial characteristics or reflections in the eyes," explains Professor Peter Eisert, head of the Vision & Imaging Technologies department at Fraunhofer HHI.

LRP algorithms render AI predictions explainable

The neural networks make very reliable decisions on whether or not an image is genuine, with an accuracy rate of over 90 percent in the test databases created in the project. "But the real problem is much more that we don't know how the neural network makes the decision," says Eisert. Thus, in addition to the accuracy of the decision, the Fraunhofer HHI researchers are also interested in the basis for the decision. To answer this question they analyze the regions in the facial image which are relevant to the decision using LRP algorithms (Layer-Wise Relevance Propagation) that they developed themselves. This helps identifying suspicious areas in a facial image and to identify and classify artifacts created during a morphing process. Initial reference tests confirm that the algorithms can be used to successfully identify morphed images. The

LRP software labels the facial areas relevant to the decision accordingly. The eyes frequently provide evidence of image tampering.

The researchers also use this information to design more robust neural networks in order to detect the widest possible variety of attack methods. "Criminals can resort to more and more sophisticated attack methods, for example AI methods that generate completely artificial facial images. By optimizing our neural networks we're trying to stay one step ahead of the culprits and to identify future attacks," says the IT professor.

There is already a demonstrator software package including anomaly detection and evaluation procedures. It contains a number of different detector modules from the individual project partners that have been fused together. The interconnected modules apply different detection methods to find manipulations, generating an overall result at the end of the process. The objective is to integrate the software in existing face recognition systems at border checkpoints or to enhance the systems to include morphing components and thus to rule out falsification through corresponding attacks of this type.

Provided by Fraunhofer-Gesellschaft