Study shows face recognition experts perform better with AI as partner

May 28 2018

Do these two faces show the same person? Trained specialists called forensic face examiners testify about such questions in court. A NIST study measuring their accuracy reveals the science behind their work for the first time. Credit: J. Stoughton/NIST

Experts at recognizing faces often play a crucial role in criminal cases. A

photo from a security camera can mean prison or freedom for a defendant—and testimony from highly trained forensic face examiners informs the jury whether that image actually depicts the accused. Just how good are facial recognition experts? Would artificial intelligence help?

A study appearing today in the *Proceedings of the National Academy of Sciences* has brought answers. In work that combines forensic science with psychology and computer vision research, a team of scientists from the National Institute of Standards and Technology (NIST) and three universities has tested the accuracy of professional face identifiers, providing at least one revelation that surprised even the researchers: Trained human beings perform best with a computer as a partner, not another person.

"This is the first study to measure face identification accuracy for professional forensic facial examiners, working under circumstances that apply in real-world casework," said NIST electronic engineer P. Jonathon Phillips. "Our deeper goal was to find better ways to increase the accuracy of forensic facial comparisons."

The team's effort began in response to a 2009 report by the National Research Council, "Strengthening Forensic Science in the United States: A Path Forward", which underscored the need to measure the accuracy of forensic examiner decisions.

The NIST study is the most comprehensive examination to date of face identification performance across a large, varied group of people. The study also examines the best technology as well, comparing the accuracy of state-of-the-art face recognition algorithms to human experts.

Their result from this classic confrontation of human versus machine? Neither gets the best results alone. Maximum accuracy was achieved

with a collaboration between the two.

"Societies rely on the expertise and training of professional forensic facial examiners, because their judgments are thought to be best," said co-author Alice O'Toole, a professor of cognitive science at the University of Texas at Dallas. "However, we learned that to get the most highly accurate face identification, we should combine the strengths of humans and machines."

The results arrive at a timely moment in the development of [facial recognition](#) technology, which has been advancing for decades, but has only very recently attained competence approaching that of top-performing humans.

"If we had done this study three years ago, the best computer algorithm's performance would have been comparable to an average untrained student," Phillips said. "Nowadays, state-of-the-art algorithms perform as well as a highly trained professional."

The study itself involved a total of 184 participants, a large number for an experiment of this type. Eighty-seven were trained professional facial examiners, while 13 were "super recognizers," a term implying exceptional natural ability. The remaining 84—the control groups—included 53 fingerprint examiners and 31 undergraduate students, none of whom had training in facial comparisons.

For the test, the participants received 20 pairs of face images and rated the likelihood of each pair being the same person on a seven-point scale. The research team intentionally selected extremely challenging pairs, using images taken with limited control of illumination, expression and appearance. They then tested four of the latest computerized facial recognition algorithms, all developed between 2015 and 2017, using the same image pairs.

Three of the algorithms were developed by Rama Chellappa, a professor of electrical and computer engineering at the University of Maryland, and his team, who contributed to the study. The algorithms were trained to work in general face recognition situations and were applied without modification to the image sets.

One of the findings was unsurprising but significant to the justice system: The trained professionals did significantly better than the untrained control groups. This result established the superior ability of the trained examiners, thus providing for the first time a scientific basis for their testimony in court.

The algorithms also acquitted themselves well, as might be expected from the steady improvement in algorithm performance over the past few years.

What raised the team's collective eyebrows regarded the performance of multiple examiners. The team discovered that combining the opinions of multiple forensic face examiners did not bring the most accurate results.

"Our data show that the best results come from a single facial examiner working with a single top-performing algorithm," Phillips said. "While combining two human examiners does improve accuracy, it's not as good as combining one examiner and the best algorithm."

Combining examiners and AI is not currently used in real-world forensic casework. While this study did not explicitly test this fusion of examiners and AI in such an operational forensic environment, results provide an roadmap for improving the accuracy of face identification in future systems.

While the three-year project has revealed that humans and algorithms use different approaches to compare faces, it poses a tantalizing question

to other scientists: Just what is the underlying distinction between the human and the algorithmic approach?

"If combining decisions from two sources increases accuracy, then this method demonstrates the existence of different strategies," Phillips said. "But it does not explain how the strategies are different."

More information: P. Jonathon Phillips el al., "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *PNAS* (2018). www.pnas.org/cgi/doi/10.1073/pnas.1721355115

Provided by National Institute of Standards and Technology