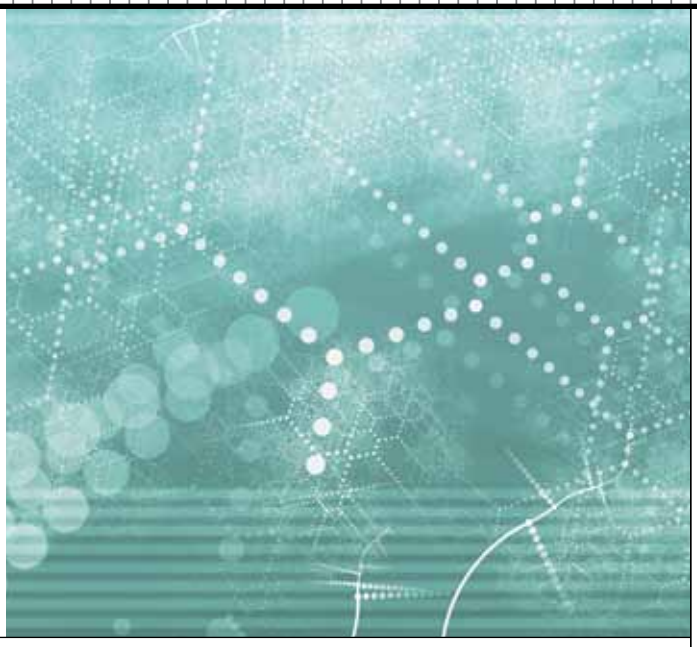
Improving Face Recognition Technology



P. Jonathon Phillips,
National Institute of Standards and Technology

US-government-sponsored evaluations and challenge problems have helped spur more than two orders of magnitude improvement in face recognition system performance.

Since 1993, the error rate of automatic face recognition systems has decreased by a factor of 272. The reduction applies to systems that match people with face images captured in studio or mugshot environments, like those shown in Figure 1. In Moore's law terms, the error rate decreased by one-half every two years.

Two key factors in achieving this reduction have been periodic independent evaluations and challenge problems sponsored by numerous US government agencies.

INDEPENDENT EVALUATIONS

Independent evaluations provide an unbiased assessment of face recognition system performance.

Eighteen years ago, the US government began conducting a series of such evaluations (http://face.nist.gov). The first evaluations were part of DARPA's Face Recognition Technology (FERET) program, which ran from 1993 to 1997. The Face Recognition Vendor Test (FRVT), conducted by the National Institute of Standards and Technology (NIST) and sponsored by numerous other government agencies, followed in 2002, with a second FRVT in 2006. The most recent evaluation, administered by NIST last year, was the Multiple Biometric Evaluation (MBE) 2010.

These evaluations assessed the performance of the technology underlying automatic face recognition.

CHALLENGE PROBLEMS

In concert with these periodic evaluations, the US government sponsored a series of three challenge problems. The purpose of the challenges was to assist researchers in academia and industry to develop and advance face recognition technology. The first was part of the FERET program. The second was the Face Recognition Grand Challenge (FRGC), which ran from 2004 to 2005, and the third was the 2008-2009 Multiple Biometric Grand Challenge (MBGC)—both the FRGC and MBGC were conducted by NIST.

A face recognition challenge problem involves three stages. First, the sponsoring government agency makes available a large corpus of face images to researchers to develop new algorithms. Second, the sponsoring agency creates a common set of experiments for the data. Third, researchers report performance on the common set of experiments.

By reporting results on a common set of experiments, researchers know where they stand relative to other researchers. This in turn unleashes the competitive spirit of researchers to make technological innovations.

The FRGC illustrates the design and impact of these challenge problems. The FRGC aimed to reduce the error rate of the FRVT 2002 results by a factor of 10. To achieve this goal, NIST provided the research community with a corpus of 40,000 face images organized into four key experiments. These experiments focused on recognizing still images and 3D face scans.

With more than 20 participants, the FRGC successfully galvanized the

Published by the IEEE Computer Society

research community into advancing face recognition technology, as confirmed by the FRVT 2006 evaluation.

FACE RECOGNITION PERFORMANCE

Why separate evaluations and challenge problems? The core of a face recognition system is a pattern recognition algorithm. Part of the development procedure is training the algorithm on face images. Testing the system using a sequestered set of images ensures that the resulting performance measures are unbiased.

Face recognition consists of two basic tasks. In a *verification* task, a person claims an identity and the system acquires a face image and compares it with the enrolled image of the claimed identity; the system either accepts or rejects the claim. In an *identification* task, the system returns a list of people whose faces are most similar to the face presented to the system.

Progress in face recognition is measured by verification task performance, which is characterized by two errors. A *false reject* occurs when the system does not accept a valid claim—for example, that I am Jonathon Phillips. The error rate associated with false rejects is the false reject rate (FRR). A *false accept* occurs when the system accepts an invalid claim—for example, that I, Jonathon Phillips, am Cary Grant. The error rate associated with false accepts is the false accept rate (FAR).

MEASURES OF PROGRESS

Figure 2 quantifies progress in face recognition from the FERET evaluations through the MBE 2010. Improvement is reported at five key milestones. For each milestone, the FRR at a FAR of 1 in 1,000 is given for a representative state-of-the-art algorithm.

The 1993 milestone is a retrospective implementation of Matthew Turk and Alex Pentland's eigenface algorithm, which was partially automatic (it required providing eye coordinates). The FRR was 0.79—that is, the system rejected 79 out of 100 valid claims.

The 1997 milestone of an FRR of 0.54 is from the final FERET evaluation. The 1993 and 1997 results are reported on the FERET dataset. These results show that under the FERET program, algorithm technology simultaneously progressed from partially to fully automatic and the error rate declined by approximately one-third.

Beginning with the FRVT 2002, the evaluations switched to a benchmark dataset provided by the US Department of State (DoS). The FERET and DoS datasets are comparable, because both have similar resolution

Figure 1. An example of two images used in evaluating face recognition systems. The images are of the same person collected on different dates in a studio environment.

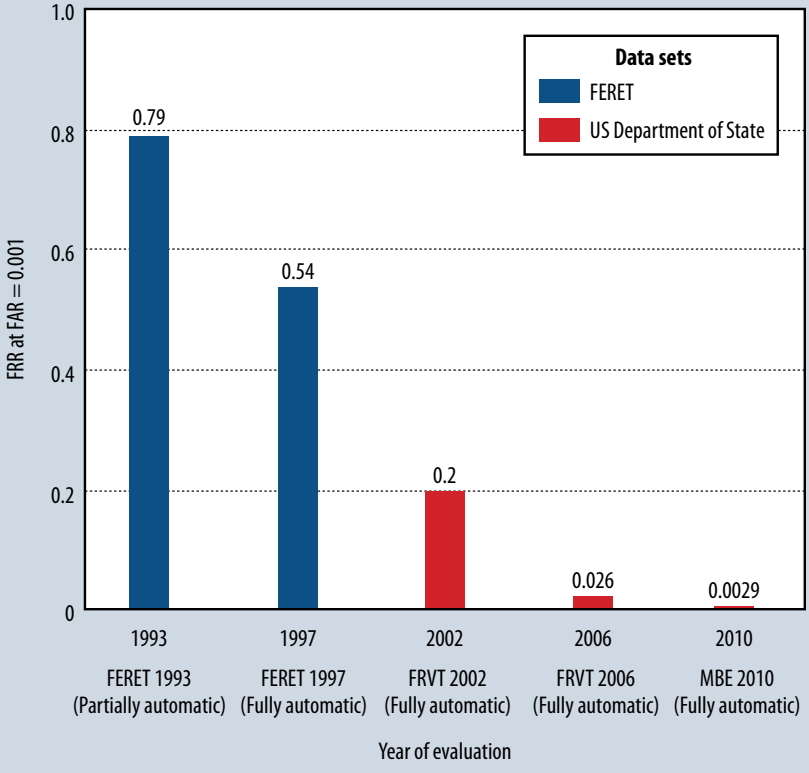


Figure 2. Reduction in the false reject rate (FRR), at a false accept rate (FAR) of 1 in 1,000, of state-of-the-art face recognition algorithms as reported in the FERET, FRVT 2002/2006, and MBE 2010 evaluations.

and produce similar results using a baseline algorithm. The FRVT 2002 achieved an FRR of 0.2, and the FRVT 2006 an FRR of 0.026.

The MBE 2010 reported an FRR of 0.0029—the top system rejected only 29 out of 10,000 valid claims.

The US-government-sponsored evaluations and challenge problems were a key factor in this rapid decline in the error rate over 17 years. The FERET, FRGC, and MBGC challenge problems provided the research community with large datasets and challenge problems designed to spur the development of new face recognition algorithms. The FERET, FRVT 2002/2006, and MBE 2010 evaluations provided academia, industry, and government with performance benchmarks and unbiased assessments of the efficacy of this technology.

As impressive as these results are, they do not imply that automatic face recognition is a solved problem. Performance degrades when attempting to recognize faces in images collected outside of studio or mugshot environments.

With the active support of NIST and many other US government agencies, researchers are expanding their efforts to develop algorithms capable of matching faces in images taken in less controlled conditions (such as outdoors), nonfrontal face images, and faces in video sequences.

P. Jonathon Phillips *is an electronics engineer at the National Institute of Standards and Technology, Gaithersburg, Maryland. Contact him at jonathon@nist.gov.*

Editor: Karl Ricanek Jr., director of the Face Aging Group at the University of North Carolina Wilmington; ricanekk@uncw.edu