Facial Recognition Technology (Part III): Ensuring Commercial Transparency & Accuracy

Introduction

Chairwoman Maloney, Ranking Member Jordan, and Members of the Committee, I am Chuck Romine, Director of the Information Technology Laboratory (ITL) at the Department of Commerce's National Institute of Standards and Technology (NIST). ITL cultivates trust in information technology and metrology through measurements, standards and testing. Thank you for the opportunity to appear before you today to discuss NIST's role in standards and testing for facial recognition technology.

Biometric and Facial Recognition Technology

Home to five Nobel Prizes, with programs focused on national priorities such as advanced manufacturing, the digital economy, precision metrology, quantum science, and biosciences, NIST's mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.

In the area of biometrics, NIST has been working with public and private sectors since the 1960s. Biometric technologies provide a means to establish or verify the identity of humans based upon one or more physical or behavioral characteristics. Examples of physical characteristics include face, fingerprint, and iris images. An example of behavioral characteristic is an individual's signature. Used with other authentication technologies, such as passwords, biometric technologies can provide higher degrees of security than other technologies employed alone. For decades, biometric technologies were used primarily in homeland security and law enforcement applications, and they are still a key component of these applications. Over the past several years, the marketplace for biometric solutions has widened significantly and today includes public and private sector applications worldwide, including physical security, banking and retail applications. According to one industry estimate, the biometrics technology market size will be worth $59.31 billion by 2025.[1] There has been a considerable rise in development and adoption of facial recognition, detection and analysis technologies in the past few years.

Face detection technology determines whether the image contains a face. Face analysis technology aims to identify attributes such as gender, age, or emotion from detected faces. Face recognition technology compares an individual's facial features to available images for verification or identification purposes. Verification or "one-to-one" matching confirms a photo matches a different photo of the same person in a database or the photo on a credential, and is commonly used for authentication purposes, such as unlocking a smartphone or checking a passport. Identification or "one-to-many" search determines whether the person in the photo has any match in a database and can be used for identification of a person.

Accuracy of face recognition algorithms is assessed by measuring the two classes of error the software can make: false positives and false negatives. A false positive means that the software wrongly considered photos of two different individuals to show the same person, while a false negative means the software failed to match two photos that, in fact, do show the same person.

NIST's Role in Biometric and Facial Recognition Technology

NIST responds to government and market requirements for biometric standards, including facial recognition technologies, by collaborating with other federal agencies, law enforcement, industry, and academic partners to:

- research measurement, evaluation, and interoperability to advance the

use of biometric technologies including face, fingerprint, iris, voice, and multi-modal techniques;
- develop common models and metrics for identity management, critical standards, and interoperability of electronic identities;
- support the timely development of scientifically valid, fit-for-purpose standards; and
- develop the required conformance testing architectures and testing tools to test implementations of selected standards.

NIST's work improves the accuracy, quality, usability, interoperability, and consistency of identity management systems and ensures that United States interests are represented in the international arena. NIST research has provided state-of-the-art technology benchmarks and guidance to industry and to U.S. Government agencies that depend upon biometrics recognition technologies.

Under the provisions of the National Technology Transfer and Advancement Act of 1995 (Public Law 104-113) and OMB Circular A-119, NIST is tasked with the role of encouraging and coordinating federal agency use of voluntary consensus standards in lieu of government-unique standards, and federal agency participation in the development of relevant standards, as well as promoting coordination between the public and private sectors in the development of standards and in conformity assessment activities. NIST works with other agencies to coordinate standards issues and priorities with the private sector through consensus standards developing organizations such as the InterNational Committee for Information Technology Standards (INCITS), Joint Technical Committee 1 of the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), the Organization for the Advancement of Structured Information Standards (OASIS), IEEE, the Internet Engineering Task Force (IETF), and other standards organizations such as the International Civil Aviation Organization (ICAO), and the International Telecommunication Union's Standardization Sector (ITU-T). NIST leads national and international consensus standards activities in biometrics, such as facial recognition technology, but also in cryptography, electronic credentialing, secure network protocols, software and systems reliability, and security conformance testing – all essential to accelerate the development and deployment of information and communication systems that are interoperable, reliable, secure, and usable.

Since 2010, NIST has organized the biennial International Biometric Performance Testing Conference. This series of conferences accelerates adoption and effectiveness of biometric technologies by providing a forum to discuss and identify fundamental, relevant, and effective performance metrics, and disseminating best practices for performance design, calibration, evaluation, and monitoring.

Facial Recognition Tests and Evaluations

For more than a decade, NIST biometric evaluations have measured the core algorithmic capability of biometric recognition technologies and reported the accuracy, throughput, reliability, and sensitivity of algorithms with respect to data characteristics, for example, noise or compression, and to subject characteristics, for example, age or gender. NIST biometric evaluations advance the technology by identifying and reporting gaps and limitations of current biometric recognition technologies. NIST evaluations advance measurement science by providing a scientific basis for "what to measure" and "how to measure." NIST evaluations also facilitate development of consensus-based standards by providing quantitative data for development of scientifically sound, fit-for-purpose standards.

NIST conducted the Face Recognition Grand Challenge (2004-2006) and Multiple Biometric Grand Challenge (2008-2010) programs to challenge the facial recognition community to break new ground solving research problems on the biometric frontier.

Since 2000, NIST's Face Recognition Vendor Testing Program (FRVT) has assessed capabilities of facial recognition algorithms for one-to-many identification and one-to-one verification.

Participation in FRVT is open to any organization worldwide. There is no charge for participation, and being an ongoing activity, participants may submit their algorithms on a continuous basis. The algorithms are submitted to NIST by corporate research and development laboratories and a few universities. As prototypes, these algorithms are not necessarily available as mature integrable products. For all algorithms that NIST evaluates, NIST posts performance results on its FRVT website and identifies the algorithm and the developing organization.

NIST and the FRVT program do not train face recognition algorithms. NIST does not provide training data to the software under test, and the software is prohibited from adapting to any data that is passed to the algorithms during a test.[2]

NIST provides technical guidance and scientific support for analysis and recommendations for utilization of facial recognition technologies to various U.S. government and law enforcement agencies, including the Federal Bureau of Investigation (FBI), Office of Biometric Identity Management (OBIM) at the Department of Homeland Security (DHS), Department of Homeland Security Science and Technology Directorate (DHS S&T), the Department of Homeland Security's U.S. Customs and Border Protection agency (DHS CBP), and the Intelligence Advanced Research Projects Activity (IARPA) at the office of the Director of National Intelligence.

Historically and currently, NIST biometrics research has assisted DHS. NIST's research was used by DHS in its transition to ten prints for the former US-VISIT program and NIST is currently working with DHS CBP to analyze performance impacts due to image quality and traveler demographics and provide recommendations regarding match algorithms, optimal thresholds and match gallery creation for its Traveler Verification Service program.

NIST Face Recognition Vendor Testing Program

NIST's Face Recognition Vendor Testing Program (FRVT) was established in 2000 to provide independent evaluations of both prototype and commercially available facial recognition algorithms. These evaluations provide the U.S. government with information to assist in determining where and how facial recognition technology can best be deployed. FRVT results also help identify future research directions for the facial recognition community.

The 2013 FRVT tested facial recognition algorithms submitted by 16 organizations, and showed significant algorithm improvement since NIST's 2010 FRVT test. NIST defined performance by recognition accuracy—how many times the software correctly identified the photo—and the time the algorithms took to match one photo against large photo data sets.

The 2018 FRVT tested 127 facial recognition algorithms from the research laboratories of 39 commercial developers and one university, using 26 million mugshot images of 12 million individuals provided by the FBI. The 2018 FRVT measured the accuracy and speed of one-to-many facial recognition identification algorithms. The evaluation also contrasted mugshot accuracy with that from lower quality images. The findings, reported in NIST Interagency Report 8238,[3] showed that massive gains in accuracy have been achieved since the FRVT in 2013, which far exceed improvements made in the prior period (2010-2013). The accuracy gains observed in the 2018 FVRT study stem from the integration, or complete replacement, of older facial recognition techniques with those based on deep convolutional neural networks. While the industry gains are broad, there remains a wide range of capabilities, with some developers providing much more accurate algorithms than others. Using FBI mugshots, the most accurate algorithms fail only in about one quarter of one percent of searches, and these failures are associated with images of injured persons and those with long time lapse since the first photograph. The success of mugshot searches stems from the new generation of facial recognition algorithms, and from the adoption of portrait photography standards first developed at NIST in the late 1990s.

The 2019 FRVT quantified the accuracy of face recognition algorithms for demographic groups defined by sex, age, and race or country of birth, for both one-to-one verification algorithms and one-to-many identification search algorithms. NIST conducted tests to quantify demographic differences for 189 face recognition algorithms from 99 developers, using four collections of photographs with 18.27 million images of 8.49 million people. These images came from operational databases provided by the State Department, the Department of Homeland Security and the FBI. Previous FRVT reports[4] documented the accuracy of these algorithms and showed a wide range in accuracy across algorithms. The more accurate algorithms produce fewer errors and can therefore be anticipated to have smaller demographic differentials.

NIST Interagency Report 8280,[5] released on December 19, 2019, quantifies the effect of age, race, and sex on face recognition performance. It found empirical evidence for the existence of demographic differentials in face recognition algorithms that NIST evaluated. The report distinguishes between false positive and false negative errors, and notes that the impacts of errors are application dependent.

I will first address one-to-one verification applications. There, false positive differentials are much larger than for false negatives and exist across many, but not all, algorithms tested. Across demographics, false positives rates often vary by factors of 10 to beyond 100 times. False negatives tend to be more algorithm-specific, and often vary by factors below 3. False positives might present a security concern to the system owner, as they may allow access to impostors. False positives may also present privacy and civil rights and civil liberties concerns such as when matches result in additional questioning, surveillance, errors in benefit adjudication, or loss of liberty. False positives are higher in women than in men and are higher in the elderly and the young compared to middle-aged adults. Regarding race, we measured higher false positive rates in Asian and African American faces relative to those of Caucasians. There are also higher false positive rates in Native American, American Indian, Alaskan Indian and Pacific Islanders. These effects apply to most algorithms, including those developed in Europe and the United States. However, a notable exception was for some algorithms developed in Asian countries. There was no such dramatic difference in false positives in one-to-one matching between Asian and Caucasian faces for algorithms developed in Asia. While the NIST study did not explore the relationship between cause and effect, one possible connection, and area for research, is the relationship between an algorithm's performance and the data used to train the algorithm itself.

I will now comment on one-to-many search algorithms. Again, the impact of errors is application dependent. False positives in one-to-many search are particularly important because the consequences could include false accusations. For most algorithms, the NIST study measured higher false positives rates in women, African Americans, and particularly in African American women. However, the study found that some one-to-many algorithms gave similar false positive rates across these specific demographics. Some of the most accurate algorithms fell into this group. This last point underscores one overall message of the report: Different algorithms perform differently. Indeed all of our FRVT reports note wide variations in recognition accuracy across algorithms, and an important result from the demographics study is that demographic effects are smaller with more accurate algorithms.

A general takeaway from these studies is that, there is significant variance between the performance facial recognition algorithms, that is, some produce significantly fewer errors than others. Consequently, users, policy makers, and the public should not think of facial recognition as either always accurate or always error prone.

NIST Face in Video Evaluation Program

The Face in Video Evaluation Program (FIVE) assessed the capability of facial recognition algorithms to correctly identify or ignore persons appearing in video sequences. The outcomes of FIVE are documented in NIST Interagency report 8173,[6] which enumerates accuracy and speed of facial recognition algorithms applied to the identification of persons appearing in video sequences drawn from six different video datasets. NIST completed this program in 2017.

Human Factors: Facial Forensic Examiners

NIST is researching how to measure the accuracy of forensic examiners matching identity across different photographs. The study measures face identification accuracy for an international group of professional forensic facial examiners working under circumstances approximating real-world casework. The findings, published in the proceedings of the National Academy of Sciences,[7] showed that examiners and other human face "specialists," including forensically trained facial reviewers and untrained super-recognizers, were more accurate than the control groups on a challenging test of face identification. It also presented data comparing state-of-the-art facial recognition algorithms with the best human face identifiers. The best machine performed in the range of the best-performing humans, who were professional facial examiners. However, optimal face identification was achieved only when humans and machines collaborated.

Voluntary Consensus Standards

When properly conducted, standards development can increase productivity and efficiency in government and industry, expand innovation and competition, broaden opportunities for international trade, conserve resources, provide consumer benefit and choice, improve the environment, and promote health and safety.

In the U.S., most standards development organizations are industry-led private sector organizations. Many voluntary consensus standards from those standard development organizations are appropriate or adaptable for the government's purposes. OMB Circular A-119 directs the use of such standards by U.S. government agencies, whenever practicable and appropriate, to achieve the following goals:

- eliminating the cost to the Federal Government of developing its own standards and decreasing the cost of goods procured and the burden of complying with agency regulation;
- providing incentives and opportunities to establish standards that serve national needs, encouraging long-term growth for U.S. enterprises and promoting efficiency, economic competition, and trade; and
- furthering the reliance upon private sector expertise to supply the Federal Government with cost-efficient goods and services.

Examples of NIST Consensus Standards Development Activities

ANSI/NIST-ITL – The ANSI/NIST-ITL standard for biometric information is used in 160 countries to ensure biometric data exchange across jurisdictional line and between dissimilar systems. One of the important effects of NIST work on this standard is that it allows accurate and interoperable exchange of biometrics information by law enforcement globally and enables them to identify criminals and terrorists. NIST’s own Information Technology Laboratory is an American National Standards Institute (ANSI)-accredited standard development organization. Under accreditation by ANSI, the private-sector U.S. standards federation, NIST continues to develop consensus biometric data interchange standards. Starting in 1986, NIST has developed and approved a succession of data format standards for the interchange of biometric data. The current version of this standard is ANSI/NIST-ITL 1: 2015, Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information.[8] This standard continues to evolve to support government applications including law enforcement, homeland security, as well as other identity management applications. Virtually all law enforcement biometric collections worldwide use the ANSI/NIST-ITL standard. NIST biometric technology evaluations in fingerprint, face, and iris have provided the government with timely analysis of market capabilities to guide biometric technology procurements and deployments.

ISO/IEC Joint Technical Committee 1, Subcommittee 37 (JTC1/SC37) - Biometrics

From the inception of the ISO Subcommittee on Biometrics in 2002, NIST has led and provided technical expertise to develop international biometric standards in this subcommittee. Standards developed by the Subcommittee on Biometrics have received widespread international and national market acceptance. Large international organizations, such as the ICAO for Machine Readable Travel Documents and the International Labour Office (ILO) of the United Nations for the verification and identification of seafarers, specify in their requirements the use of some of the international biometric standards developed by this subcommittee.

Since 2006, JTC1/SC37 has published a series of standards on biometric performance testing and reporting, many of which are based on NIST technical contributions. These documents provide guidance on the principles and framework, testing methodologies, modality-specific testing, interoperability performance testing, access control scenarios, and testing of on-card comparison algorithms for biometric performance testing and reporting. NIST contributes towards the development of these documents and follows their guidance and metrics in its evaluations, such as the FRVT.

Conclusion

NIST is proud of the positive impact it has had in the last 60 years on the evolution of biometrics capabilities. With NIST's extensive experience and broad expertise, both in its laboratories and in successful collaborations with the private sector and other government agencies, NIST is actively pursuing the standards and measurement research necessary to deploy interoperable, secure, reliable, and usable identity management systems.

Thank you for the opportunity to testify on NIST's activities in facial recognition and identity management. I would be happy to answer any questions that you may have.

[1] https://www.grandviewresearch.com/industry-analysis/biometrics-industry

[2] The process of training a face recognition algorithm (or any machine learning algorithm) involves providing a machine learning algorithm with training data to learn from. The training data shall contain the correct answer, which is known as ground-truth label, or a target. The learning algorithm finds patterns in the training data that map the input data attributes to the target and builds a machine-learning model that captures these patterns. This model can then be used to get predictions on new data for which the target is unknown.

[3] https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8238.pdf

[4] Part 1:
https://www.nist.gov/system/files/documents/2019/11/20/frvt_report_2019...
and
Part 2:
https://www.nist.gov/system/files/documents/2019/09/11/nistir_8271_2019...

[5] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[6] https://www.nist.gov/publications/face-video-evaluation-five-face-recog...

[7] https://www.pnas.org/content/115/24/6171

[8] https://www.nist.gov/publications/data-format-interchange-fingerprint-f...