TESTIMONY OF

Arun Abraham Ross
Site Director, NSF Center for Identification Technology Research
Professor, Department of Computer Science and Engineering
Michigan State University

BEFORE
The United States House of Representatives
House Committee on Science, Space, and Technology,
Subcommittee on Investigations & Oversight

TITLE: Privacy in the Age of Biometrics
DATE: June 29, 2022

Respected Chairman Foster, Ranking Member Obernolte, and Esteemed Members of the SubCommittee – thank you for the invitation to testify today. Biometrics is a valuable technology that has broad applications in a number of different domains. However, it is necessary to ensure that the privacy of individuals is not unduly compromised when their biometric data, such as face images and fingerprints, are used in an application. The purpose of this testimony is to communicate some of the ways in which the privacy of biometric data can be enhanced, thereby facilitating the responsible use of the technology. In this regard, the written testimony has the following parts: (a) introduction to biometrics; (b) enumeration of some of the privacy concerns related to biometrics, especially with respect to face recognition; (c) description of some technical methods that can be used to impart privacy to biometric data; and (d) recommendations to advance biometric technology for its use in a responsible and ethical manner. The views I express here are my own.

Introduction

The need for reliably determining the identity of a person is critical in a vast number of applications ranging from personal smartphones to border security; self-driving vehicles to e-voting; tracking child vaccinations to preventing human trafficking; crime scene investigation to personalization of customer service [1]. Biometrics, which entails the use of biological attributes such as face, fingerprints, iris, and voice for recognizing a person, is increasingly being used in several such applications [2]. For instance, many smartphones employ automated face or fingerprint recognition for unlocking and payment authentication purposes [3].

Take automated face recognition system as an example. Automated Face entails the comparison of two face images in order to assess the degree of similarity of dissimilarity between them [4]. This comparison results in a match score which is then used to render a “match” or “non-match” decision. A “match” would indicate that the two images are of the same person; a “non-match” would indicate that the images are of different people. Face recognition technology is currently either deployed or being considered in a number of applications including smartphone access [5], airport security [6], and contactless payment [7]. In addition, face recognition is being used to find missing children [8], combat human trafficking [9], and locate perpetrators of crimes [10]. In some countries, such as India [11], face recognition is being used in large-scale national ID card programs, either to identify an individual or to detect duplicate identities in a central database.

The increased use of biometric technology in consumer applications and law enforcement – especially those based on fingerprints and faces – is mostly driven by significant improvement in recognition accuracy of these systems over the past decade. For example, in a recent publication, the authors of the National Institute of Standards and Technology (NIST) report on Face Recognition Vendor Technology (FRVT) state the following [12]: "The major result . . . was that massive gains in accuracy have been achieved in the years 2013 to 2018 and these far exceed improvements made in the prior period, 2010 to 2013." In another report, the authors observed that there was a twenty-fold reduction (improvement) in false negatives[1] between 2014 and 2018 [13]. They go on to say, "The massive accuracy gains are consistent with an industrial revolution associated with the incorporation of convolutional neural network-based techniques into the prototypes." Indeed, the phenomenal rise of the paradigm of deep learning based on neural networks has radically changed the landscape of face recognition [14]. In addition, more recent face recognition algorithms have made rapid strides in overcoming differential performance across demographic groups [15, 16, 17].

Face Recognition and Privacy

Despite the steep and impressive improvement in face recognition accuracy and its benefits in many applications, the technology itself has come under attack in recent years [18]. For instance, some have claimed that the technology is "racist" and "sexist" [19]. Further, the inappropriate [20] and, in some cases, incorrect use of the technology by authorities [21] has heightened objections to using face recognition technology in public spaces. In addition, there are privacy concerns related to the technology: (a) the use of automated face recognition for covertly tracking individuals over time [22]; and (b) the extraction of sensitive attributes such as age, sex, race, and health cues from face images without the consent of the subjects [23].

Below, I summarize some of the key privacy concerns related to face recognition technology (FRT). While this is presented from the perspective of face recognition, they are related to other biometric modalities as well such as fingerprints, iris, voice, etc.

1. Linking identity across applications: Face images of an individual that are present in multiple platforms (e.g., social media profiles) can be linked using FRT. This means, if the individual had provided a pseudonym (or an alias) in one application (for the sake of privacy) and personally identifiable information (PII) in another application, their identity in the first application can be exposed by linking the face images across the two applications [24].
2. Deducing personal attributes without user consent: Rapid advances in the field of machine learning, especially deep learning, has led to the development of so-called attribute classifiers that can automatically extract information such as age, sex, race, and health cues from face images [25, 26]. The possibility of eliciting genetic information from facial images has also been demonstrated [27]. When these attributes are deduced without user consent, then the privacy of individuals can be breached.
3. Scraping face images from the web: A number of face datasets have been curated for research purposes by scraping publicly available face images from the web [28, 29, 30]. While most of these datasets do not have any identifiers (such as names) associated with the face images, concerns have been expressed about using these images for research purposes without user consent [31]. Furthermore, some commercial enterprises offer FRT services based on such curated datasets,

[1] A false negative is an error where the input face image is *not* matched with any of the face images in a database, despite the matching identity being present in the database.

wherein an anonymous face image can be linked to one or more face images in a dataset thereby potentially revealing the identity of the anonymous face [32, 33].

Privacy Enhancing Technology

In order to address these concerns, a number of different techniques have been developed over the years in order to impart privacy to biometric data [34]. A few of these techniques have been summarized below.

1. Homomorphic Encryption: This is a type of data encryption scheme where computations can be performed in the encrypted domain. Here, the raw biometric data of an individual (e.g., image of a person's face) is neither stored nor transmitted thereby mitigating the possibility of using the data for unspecified or unintended purposes [35, 36]. In other words, the original biometric data is never revealed.
2. Cancellable Biometrics: In this paradigm, the biometric data of an individual is intentionally distorted using a mathematical function. The distorted data can still be successfully used for biometric recognition purposes within a certain application; however, it pre-empts the possibility of linking the biometric data of an individual across applications. This is accomplished by using different mathematical transformations in different applications. This enhances the privacy of an individual, by making it difficult to connect a person's biometric data across applications [37]. This approach can also be used to "revoke" or "cancel" an individual's biometric data in an application.
3. Semi-adversarial Networks: More recent work has established the possibility of perturbing a face image in such a way that its biometric utility is retained, but the ability to extract additional attributes such as age, sex, race, etc. is obscured. This imparts what is known as "soft-biometric privacy" to face images. The perturbation itself is accomplished using neural networks known as semi-adversarial networks, or SANs [38, 39, 40]. In related work, researchers have sought to develop new face representation techniques that do not reveal sensitive attributes resident in a face image [41].
4. Synthetic Data: State-of-the-art face recognition techniques are based on deep neural networks which typically require massive amounts of training data. One of the purposes for scraping face images from websites is to obtain data for training face recognition algorithms before these algorithms are used for recognition purposes. One way to address this challenge is to use synthetically generated face images [42, 43] for training biometric recognition algorithms. Some researchers have shown that synthetically generated faces are nearly indistinguishable from real faces and are judged more trustworthy [44]. However, when GANS (i.e., Generative Adversarial Networks) are used to generate synthetic images, some of the existing concerns regarding privacy might persist. This is because GAN-based synthesizers also require a massive number of real face images for training purposes. Nevertheless, the use of synthetic data can alleviate some of the concerns related to training face recognition algorithms.
5. Privacy Sensors: Researchers are developing privacy-preserving cameras and sensors where the acquired images are not interpretable by a human and can only be used within a specific application [45]. Such cameras, when deployed in public spaces, can ensure that the acquired images are not viable for previously unspecified purposes.

Recommendations

Based on the aforementioned comments, I would like to offer a few comments regarding the responsible and ethical use of biometric technology as it pertains to privacy.

1. The recognition accuracy of biometric systems, including face recognition, has significantly improved over the past decade. This improvement in accuracy should not be ignored; rather, it should be harnessed. Biometric technology is continually evolving and will be a valuable tool for recognizing individuals in applications where such a functionality is essential (e.g., border security, access control,

payment authentication). Therefore, it is necessary for supporting the development and deployment of the technology in a responsible and ethical manner.

2. A consortium consisting of researchers, practitioners, legal scholars, ethicists, policymakers, and end-users must be established in order to address privacy aspects of the technology in a systematic and comprehensive manner. Privacy should never be an afterthought; rather, it should be a factor that is prominently considered during the design and development of biometric recognition algorithms. However, a privacy-by-design paradigm requires a collaborative consortium where issues are raised and addressed during the development phase of the technology. Such a collaboration will lead to the development of effective biometric recognition algorithms, where the security and privacy of biometric data are enhanced. Further, privacy metrics must be better defined.
3. As discussed earlier, researchers have developed a number of techniques for imparting privacy to biometric data. However, the efficacy of these techniques must be rigorously evaluated in operational environments (i.e., for individual use cases). This requires the establishment of biometric test and evaluation centers and approved protocols for systematic and simultaneous evaluation of multiple biometric recognition algorithms.

In addition, it must be noted that academic researchers in biometrics, computer vision, and machine learning are becoming increasingly aware of the privacy and ethical implications of the technology they are developing. One such example is the research being conducted at the NSF Center for Identification Technology Research (CITeR). Privacy is no longer an afterthought; rather, the concept of privacy-by-design is being embraced by the broader academic research community [46]. In the context of biometrics this means that recognition accuracy is not the only metric being used to evaluate the overall performance of a biometric system. Rather, metrics related to security and privacy are also being increasingly considered. This shift in the research culture is remarkable and bodes well for the future of the technology.

References

1. Arun Ross, Sudipta Banerjee, Cunjian Chen, Anurag Chowdhury, Vahid Mirjalili, Renu Sharma, Thomas Swearingen and Shivangi Yadav, “Some Research Problems in Biometrics: The Future Beckons,” Proc. of 12th IAPR International Conference on Biometrics (ICB), June 2019.
2. Anil K. Jain, Karthik Nandakumar, Arun Ross, "50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities," Pattern Recognition Letters, Vol. 79, pp. 80 - 105, August 2016.
3. Anil K. Jain, Debayan Deb, Joshua J. Engelsma, "Biometrics: Trust, but Verify," in IEEE Transactions on Biometrics, Behavior, and Identity Science, October 2021.
4. Anil K. Jain, Arun Ross, K. Nandakumar, "Introduction to Biometrics", Springer Publishers, 2011. ISBN: 978-0-387-77325-4.
5. Apple Inc., “About Face ID advanced technology”, https://support.apple.com/en-us/HT208108, April 27, 2022. Accessed: 2022-06-26.
6. Elaine Glusac, “Your Face Is, or Will Be, Your Boarding Pass”, https://www.nytimes.com/2021/12/07/travel/biometrics-airports-security.html, January 11, 2022. Accessed: 2022-06-26.
7. Mastercard, “With a smile or a wave, paying in store just got personal”, https://www.mastercard.com/news/press/2022/may/with-a-smile-or-a-wave-paying-in-store-just-got-personal/, May 17, 2022. Accessed: 2022-06-26.
8. Anthony Cuthbertson, “Indian police trace 3,000 missing children in just four days using facial recognition technology”, https://www.independent.co.uk/lifestyle/gadgets-and-tech/news/india-police-missing-children-facial-recognitiontech-trace-find-reunite-a8320406.html, April 2018. Accessed: 2022-06-26.
9. Tom Simonite, “How facial recognition is fighting child sex trafficking”, https://www.wired.com/story/how-facial-recognition-fighting-child-sex-trafficking/, June 2019. Accessed: 2022-06-26.

10. Ryan Lucas, “How a tip — and facial recognition technology — helped the FBI catch a killer”, https://www.npr.org/2019/08/21/752484720/how-a-tip-and-facialrecognition-technology-helped-the-fbi-catch-a-killer, August 2019. Accessed: 2022-06-26.
11. Ursula Rao and Vijayanka Nair, “Aadhaar: Governing with biometrics”, South Asia: Journal of South Asian Studies, 42(3):469–481, 2019.
12. Patrick Grother, Mei Ngan, and Kayee Hanaoka, “Face recognition vendor test (FRVT) Part 2: Identification”, NISTIR 8271 - Draft Supplement, July 2021.
13. Patrick Grother, Mei Ngan, and Kayee Hanaoka, “Ongoing face recognition vendor test (FRVT) part 2: Identification”, NISTIR 8238, 2018.
14. Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, and Lior Wolf, “Deepface: Closing the gap to human-level performance in face verification,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
15. Patrick Grother, Mei Ngan, Kayee Hanaoka, “Face recognition vendor test (FRVT) Part 3: Demographic effects”, NISTIR 8280, December 2019.
16. Michael McLaughlin and Daniel Castro, “The critics were wrong: NIST data shows the best facial recognition algorithms are neither racist nor sexist”, https://itif.org/publications/2020/01/27/critics-were-wrong-nist-datashows-best-facial-recognition-algorithms, January 2020. Accessed: 2022-06-26.
17. Jake Parker and David Ray, “What science really says about facial recognition accuracy and bias concerns”, https://www.securityindustry.org/2021/07/23/whatscience-really-says-about-facial-recognition-accuracy-and-bias-concerns/, July 2021. Accessed: 2022-06-26.
18. Clare Garvie, Alvaro Bedoya, and Jonathan Frankle, “The Perpetual Line-Up: Unregulated Police Face Recognition in America”, Georgetown Law, Center on Privacy & Technology, 2016.
19. Alessandra Malito, “Your facial recognition software may be racist and sexist”, https://www.marketwatch.com/story/your-facial-recognition-software-may-be-racist-and-sexist-2018-02-13, February 2018. Accessed: 2022-06-26.
20. Kashmir Hill, “Wrongfully accused by an algorithm”, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html, June 2020. Accessed: 2022-06-26.
21. Dan Robitzski, “Cops are using amazon’s facial recognition software wrong”, https://futurism.com/cops-amazon-facial-recognition, February 2019. Accessed: 2022-06-26.
22. Kai Strittmatter. We Have Been Harmonized: Life in China’s Surveillance State. Harper-Collins, 2020.
23. Antitza Dantcheva, Petros Elia, Arun Ross, "What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics," IEEE Transactions on Information Forensics and Security (TIFS), Vol. 11, No. 3, pp. 441 - 467, March 2016.
24. Alessandro Acquisti, Ralph Gross, Fred Stutzman, “Face Recognition and Privacy in the Age of Augmented Reality,” Journal of Privacy and Confidentiality, 2014. Available at SSRN: https://ssrn.com/abstract=3305312.
25. Yunlian Sun, Man Zhang, Zhenan Sun, Tieniu Tan, “Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(2): 332-351, 2018.
26. Arun Ross, Sudipta Banerjee, Anurag Chowdhury, “Deducing Health Cues from Biometric Data,” Computer Vision and Image Understanding (CVIU), Vol. 221, August 2022.
27. Yaron Gurovich, Yair Hanani, Omri Bar, Guy Nadav, Nicole Fleischer, Dekel Gelbman, Lina Basel-Salmon, Peter M. Krawitz, Susanne B. Kamphausen, Martin Zenker, Lynne M. Bird, Karen W. Gripp, “Identifying facial phenotypes of genetic disorders using deep learning,” Nature Medicine, Vol. 25, 2019.
28. Gary B. Huang, Manu Ramesh, Tamara Berg, Erik Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
29. Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, Evan Brossard, “The megaface benchmark: 1 million faces for recognition at scale,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
30. Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang, “Deep learning face attributes in the wild,” Proceedings of International Conference on Computer Vision (ICCV), December 2015.
31. Olivia Solon, “Facial recognition’s ’dirty little secret’: Millions of online photos scraped without consent,” https://www.nbcnews.com/tech/internet/facialrecognition-s-dirty-little-secret-millions-online-photos-scraped-n981921, March 2019. Accessed: 2022-06-26.

32. Kashmir Hill, “The secretive company that might end privacy as we know it”, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html, January 2020. Accessed: 2022-06-22.
33. Kashmir Hill, “A Face Search Engine Anyone Can Use Is Alarmingly Accurate”, https://www.nytimes.com/2022/05/26/technology/pimeyes-facial-recognition-search.html, May 26, 2022. Accessed: 2022-06-22.
34. Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuiper, Walter J. Scheirer, Arun Ross, Peter Peer, Vitomir Štruc, “Privacy-Enhancing Face Biometrics: A Comprehensive Survey,” IEEE Transactions on Information Forensics and Security, Vol. 16, pp. 4147-4183, DOI: 10.1109/TIFS.2021.3096024, 2021.
35. Vishnu Boddeti, “Secure Face Matching Using Fully Homomorphic Encryption,” IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2018.
36. Marta Gomez-Barrero, Emanuele Maiorana, Javier Galbally, Patrizio Campisi, and Julian Fierrez, “Multi-biometric template protection based on Homomorphic Encryption,” Pattern Recognition, vol. 67, pp. 149–163, 2017.
37. Vishal M. Patel, Nalini K. Ratha, and Rama Chellappa, “Cancelable Biometrics: A review,” IEEE Signal Processing Magazine, vol. 32, no. 5, pp. 54– 65, 2015.
38. Vahid Mirjalili, Sebastian Raschka, Anoop Namboodiri, Arun Ross, "Semi-Adversarial Networks: Convolutional Autoencoders for Imparting Privacy to Face Images," Proc. of 11th IAPR International Conference on Biometrics (ICB), February 2018.
39. Vahid Mirjalili, Sebastian Raschka, Arun Ross, “FlowSAN: Privacy-Enhancing Semi-Adversarial Networks to Confound Arbitrary Face-Based Gender Classifiers,” IEEE Access, Vol. 7, No. 1, pp. 99735 - 99745, December 2019.
40. Vahid Mirjalili, Sebastian Rashcka, Arun Ross, “PrivacyNet: Semi-Adversarial Networks for Multi-attribute Face Privacy,” IEEE Transactions on Image Processing, Vol. 29, pp. 9400 - 9412, September 2020.
41. Aythami Morales, Julian Fierrez, Ruben Vera-Rodriguez, Ruben Tolosana, “SensitiveNets: Learning Agnostic Representations with Application to Face Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 6, pp. 2158–2164, 2021.
42. Erroll Wood, Tadas Baltrusaitis, Charlie Hewitt, Sebastian Dziadzio, Matthew Johnson, Virginia Estellers, Tom Cashman, Jamie Shotton, “Fake It Till You Make It: Face analysis in the wild using synthetic data alone”, IEEE/CVF International Conference on Computer Vision (ICCV), October 2021.
43. Alon Shoshan, Nadav Bhonker, Igor Kviatkovsky, Gerard Medioni, “GAN-Control: Explicitly Controllable GANs,” IEEE/CVF International Conference on Computer Vision (ICCV), October 2021.
44. Sophie J. Nightingale and Hany Farid, “AI-synthesized faces are indistinguishable from real faces and more trustworthy,” Proceedings of the National Academy of Sciences, Vol. 119, 2022.
45. Jeffrey Byrne, Brian DeCann, Scott Bloom, “Key-Nets: Optical Transformation Convolutional Networks for Privacy Preserving Vision Sensors,” 31st British Machine Vision Conference (BMVC), September 2020.
46. Julien Bringer, Hervé Chabanne, Daniel Métayer, Roch Lescuyer, “Reasoning about Privacy Properties of Biometric Systems Architectures in the Presence of Information Leakage,” 18th International Conference on Information Security, 2015.