CREATING A
Face
FROM A
DNA Sample

**By: Ryan Tokola
& Michael Leuze, Ph.D.**

Introduction

The Department of Defense (DoD) continually seeks new and emerging methods to improve biometrics for enhanced security and identification. Biometric facial recognition technology analyzes collected information and establishes unique facial characteristics to compare against watch lists of suspected or known terrorists for identification purposes.

Some existing technologies use DNA to identify demographic characteristics, such as gender and ethnic background, and create a face prediction consistent with these general attributes. This approach is of limited use for identification as the predicted faces do not account for the individual's specific genetic information.

In contrast, the ongoing research at the Department of Energy's Oak Ridge National Laboratory (ORNL) focuses on interpreting portions of DNA that govern the unique facial structure of each person.

Traditional facial recognition biometrics techniques use cameras or sensors to capture an image before comparing it to a previously established database of known suspects in order to identify unknown persons. [1] Likewise, DNA samples are typically useful in biometrics only when ref-

erenced back to a pre-existing database of identified DNA.

Researchers at ORNL are moving the idea of creating a face from a DNA sample closer to reality with a novel method that assesses facial scans and compares them to genetic markers.

The ability to create a facial image based solely on small amounts of genetic material, without the need for pre-existing images of the individual or extra equipment, provides a huge advancement in current biometric capabilities.

ORNL's DNA2Face project demonstrates that predicting a person's facial structure from their genetic data, such as a small blood or tissue sample, is possible. The advanced identification possibilities brought forth by this technology will benefit the DoD community as well as law enforcement, national security and the intelligence community.

DNA2Face Project

The goal of the initial DNA2Face project was to determine whether a significant association exists between DNA and unique facial features. In order to demonstrate the association, ORNL researchers created a statistical method to represent the 3D shape of faces.

The researchers' novel biometric approach first defined a reference, average face, mapped with tens of thousands of 3D points. The reference face was then aligned with a similarly mapped facial scan. Comparing the two created correspondence vectors (CV) — essentially a measure of the differences between the reference face and the scanned face.

Once the CVs were established, the researchers applied the statistical technique, principal component analysis (PCA), which transformed each CV into a relatively small set of numbers, providing a very compact representation of the face.

The researchers then performed a genome-wide association study (GWAS) using DNA samples obtained from the subjects of facial scans. A GWAS is a statistical approach to finding genomic variations that are associated with a particular trait. The results of the GWAS showed several strong connections between DNA and facial features by comparing PCA scores to DNA markers. [2]

The researchers found 30 mutation locations with genome-wide significance and more than 5,000 locations with suggestive significance. In comparison, an earlier study without ORNL's statistical method found only five regions with genome-wide significance. [3]

"These days if you get a genetic sample from a crime scene, basically all you can do is apply it to some databank of samples, and we don't have a lot of those. But if you can take that genetic material and make a face prediction then you can run it against a database of faces, which we have a lot more of," notes Principal Investigator Ryan Tokola of ORNL's Imaging, Signals and Machine Learning Group.

Recognition and Prediction

Using facial recognition technologies is not a new concept for the DoD. From 1993 to 1997, the DoD's Face Recognition Technology (FERET) program's primary mission was to develop automatic face recognition capabilities able to assist security, intelligence and law enforcement personnel in performing their duties. [4]

The ability to identify an individual from genetic material would enable the DoD to identify extremists from remaining touch DNA or any genetic material left behind. Furthermore, the ability to create a facial image without a database of known persons could allow the military to track and identify extremists operating near deployed forces.

With the rise of lone-wolf terrorism, it is impossible to keep a comprehensive record of individuals who may commit acts of terrorism. The ability to identify suspects with only genetic material available will assist officials in tracking and apprehending persons of interest.

The DNA2Face method could also be deployed to identify unknown remains of service members killed in action. Additionally, during special operation missions, the DNA2Face technology could provide target confirmation after military operations.

The Details

As technology reduces the time and cost required to sequence DNA and identify genotypes, researchers are better able to produce massive databases containing human genomics information. This allows portions of genetic code to be correlated to specific physical traits, or phenotypes.

Utilizing vectors to define facial characteristics allows efficient comparison of genetic variations from different individuals with a variety of traits. Employing a GWAS allows for the rapid scanning of the genome to pinpoint markers and genetic variations for the same feature, such as facial shape. The GWAS array used in the study consisted of more than 700,000 single-nucleotide polymorphisms, or small variations in the genome. [5]

In order to minimize cost and to analyze many faces to identify correlations between phenotype and genotype, researchers used pre-existing data sets via the Face-Base Consortium. [6] Phenotype datasets provided 3D face meshes in a Wavefront format. Each of the faces were oriented facing the –y direction with the origin of the coordinate system located either behind or in front of the facial plane.

"Because the coordinate points are very close together, and there are tens of thousands of them, it results in a very dense representation of the face," Tokola explained.

Demographic characteristics such as ethnicity, gender and age were taken into account for the subjects included in the research. The most significant cluster of variations was found on chromosome 3 on the RAF1 gene, which is known to affect craniofacial shape. Researchers suggest analyzing select portions of the face could result in the discovery of more genes that directly influence facial shape.

Conclusion

Researchers at ORNL have demonstrated a strong association between genetic data and facial morphology. Going forward, a much larger dataset will be required to account for genetic variations in order to accurately predict facial shape and other finer features.

This predictive capability would be a significant step forward in the application of biometrics to several security and intelligence-related problems. With additional research and technological advances such as high-powered computing, the ability to map a complete face with genetic material is possible.

The ability to recreate faces based on DNA sequences would revolutionize the way unknown individuals are tracked and monitored. This technology could also benefit border and airport security.

Additionally, this technology could be used by the DoD to confirm and identify extremists and positively identify service members killed in action. ∎

Biometrics Timeline

1858: **First systematic capture of hand images for identification purposes is recorded.** Sir William Herschel records a handprint on the back of a contract for each worker to distinguish employees from others who might claim to be employees when payday arrived. [1]

1904: **NY State Prisons begin using fingerprints.** All Bertillon Units in the state's prison system begin collecting fingerprints in addition to Bertillon measurements. [2]

1960s: **Facial recognition becomes semi-automated.** Woodrow W. Bledsoe develops first semi-automatic face recognition system. The system required the administrator to manually locate features such as eyes and ears on photographs. [3]

1970s: **Facial recognition moves forward.** Goldstein, Harmon and Lesk used 21 specific subjective markers such as hair color and lip thickness to further automate face recognition. Measurements were still manually located. [3]

1988: **Law enforcement runs composite drawing against mugshot database.** A division of the Los Angeles County Sheriff's Department began using composite drawings or video images of a suspect to conduct a database search of digitized mugshots. [3]

1993 to 1997: **DoD's FERET Program.** Working with the Defense Advanced Research Projects Agency, FERET's mission was to develop automatic face recognition capabilities that could be employed to assist security, intelligence and law enforcement personnel in the performance of their duties. [4]

2000: **Face Recognition Vendor Test (FRVT) developed.** The goal of FRVT was to compare competing techniques for performing facial recognition. [5]

2014: **FBI launches Interstate Photo System (IPS).** Using facial recognition, the IPS provides a way to search millions of criminals' photos and generates a list of ranked candidates as potential investigative leads for authorized agencies. [6]

Ongoing: **Researchers at ORNL develop DNA2Face project.** The DNA2Face project uses DNA sequences from blood or tissue samples in order to create a rendering of a face.

Timeline References

1. Vacca, J. (2007). Biometric technologies and verification systems (1st ed., p. 187). Amsterdam: Butterworth-Heinemann/Elsevier (accessed May 8, 2017).
2. New York State. (n.d.). The American system of fingerprint classification. Retrieved from http://www.criminaljustice.ny.gov/ojis/history/am_sys.htm (accessed May 8, 2017).
3. JADE Learning. (n.d.). Face biometrics. Introduction and history. Retrieved from https://jadelearning.com/jadecc/courses/UNIVERSAL/NEC05.php?imDif=6489.0 (accessed May 8, 2017).
4. National Institute of Standards and Technology. (n.d.). The Facial Recognition Technology (FERET) database. Retrieved from http://www.itl.nist.gov/iad/humanid/feret/feret_master.html (accessed May 8, 2017).
5. National Institute of Standards and Technology. (n.d.). Face Recognition Vendor Test (FRVT) 2000. Retrieved from https://www.nist.gov/itl/iad/image-group/face-recogntion-vendor-test-frvt-2000 (accessed May 8, 2017).
6. FBI. (n.d.). Next Generation Identification (NGI). Retrieved from https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics/ngi (accessed May 8, 2017).

References

1. 1. Woodward, J. D., Jr., Horn, C., Gatune, J., & Thomas, A. (2003). Biometrics: A look at facial recognition. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/documented_briefings/DB396.html (accessed May 1, 2017).
2. 2. National Human Genome Research Institute. (n.d.). Genome-wide association studies. Retrieved from https://www.genome.gov/20019523/ (accessed May 1, 2017).
3. 3. Liu, F., Lijn, F. V., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., . . . Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLOS Genetics, 8(9). doi:10.1371/journal.pgen.1002932
4. 4. National Institute of Standards and Technology. (n.d.). The Facial Recognition Technology (FERET) database. Retrieved from http://www.itl.nist.gov/iad/humanid/feret/feret_master.html (accessed May 8, 2017).
5. 5. Genetics Home Reference. (n.d.). What are genome-wide association studies? Retrieved from https://ghr.nlm.nih.gov/primer/genomicresearch/gwastudies (accessed May 1, 2017).
6. 6. Boehnen, C., Peters, T., & Flynn, P. J. (2009). 3D signatures for fast 3D face recognition. In M. Tistarelli & M. Nixon (Eds.), Advances in Biometrics. ICB 2009. Lecture Notes in Computer Science, 5558, 12-21. Springer Berlin Heidelberg. Retrieved from https://link.springer.com/chapter/10.1007/978-3-642-01793-3_2 (accessed May 1, 2017).

Ryan Tokola received his M.S. in electrical engineering from the University of Michigan in 2013. He received his B.S. in studio art from Whitman College in 1997 and a B.S. in electrical engineering from The Ohio State University in 2010. He is a research and development staff member at Oak Ridge National Laboratory (ORNL) specializing in machine learning, computer vision and signal processing, with applications in biometrics and national security.

Michael Leuze, Ph.D., is the task leader for computational biomolecular modeling and bioinformatics at ORNL His research interests involve computational genomics on a massive scale and include the analysis of metagenomic samples, sequence-based approaches to understanding cancer, and the development of intelligent storage systems for organizing, managing and analyzing genomic data.