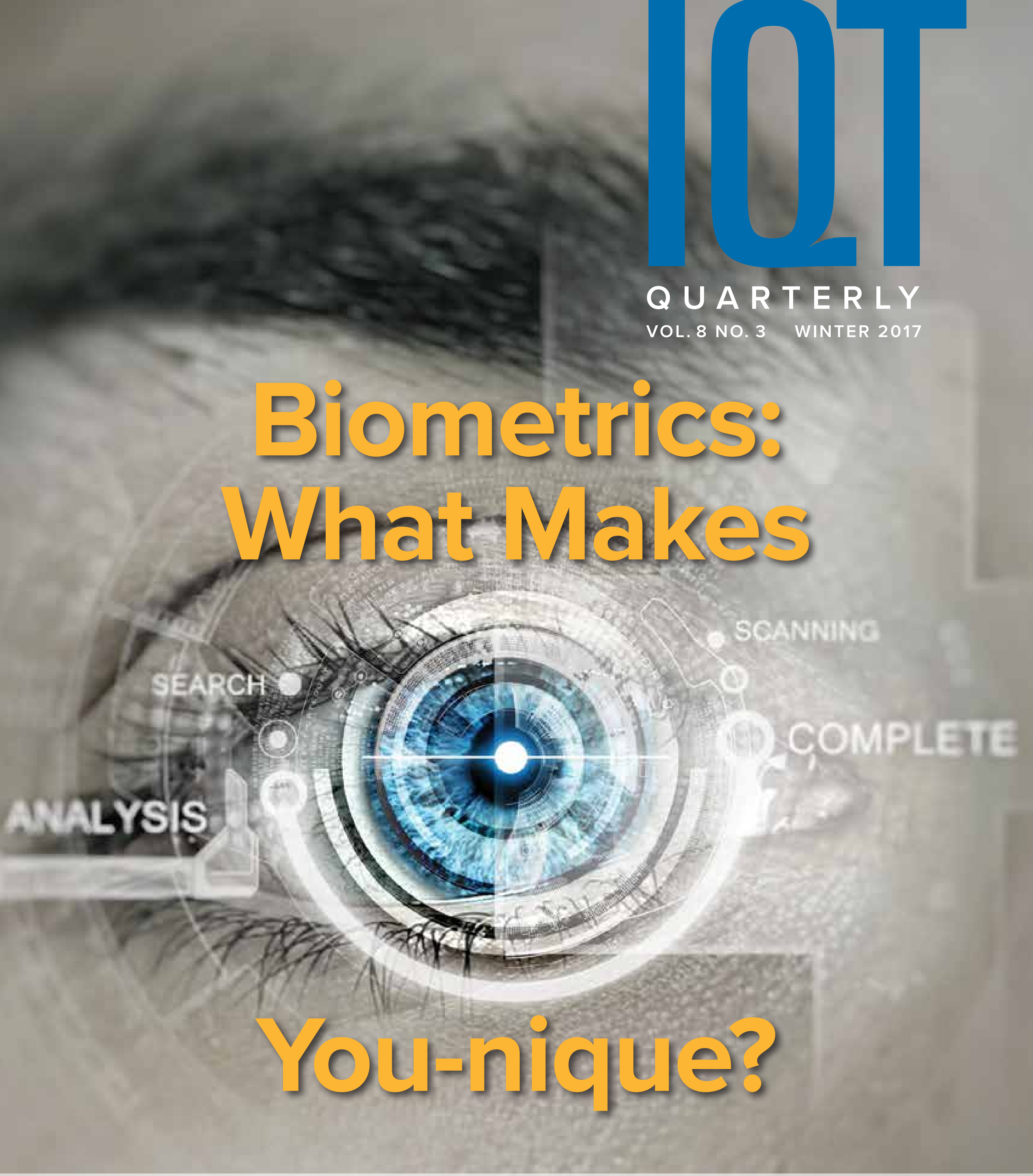
IQT
QUARTERLY
VOL. 8 NO. 3 WINTER 2017

Biometrics: What Makes You-nique?



IQT®
IN·Q·TEL

IQT Quarterly is a publication of In-Q-Tel, Inc., the strategic investor that serves as a bridge between the U.S. Intelligence Community and venture-backed startup firms on the leading edge of technological innovation. *IQT Quarterly* advances the situational awareness component of the IQT mission, serving as a platform to debut, discuss, and debate issues of innovation in the areas of overlap between commercial potential and U.S. Intelligence Community needs. For comments or questions regarding IQT or this document, please visit www.iqt.org, write to iqtquarterly@iqt.org, or call 703-248-3000. The views expressed are those of the authors in their personal capacities and do not necessarily reflect the opinion of IQT, their employers, or the Government.

EDITORIAL
IQT Quarterly, published by In-Q-Tel, Inc.
Editor-in-Chief: Melissa Hayes
Theme Editor: Karl Ni
Contributing Editors: Carrie Sessine and Christyn Zehnder
Printed in the United States of America

CONTENTS

VOL. 8 NO. 3

ON OUR RADAR

Biometrics: What Makes You-nique?

by Karl Ni

Colleagues in Emeryville who toiled on the movie *The Incredibles* have told me their favorite character introduction was in Edna: the diminutive costume and weapons designer of many a superhero. As she escorted the main character down the stairs to access her secret laboratory, she placed her palm on a reader, stared down an iris sensor, and in a low voice, she slowly, deliberately exaggerated, "Ed-na" into a microphone. Using biometrics is a common theme that many a spy movie has exploited to emphasize how secure a facility is. The hero's mission is usually to bypass these security measures, which according to James Bond or Ethan Hunt can be compromised rather crudely. In reality, it takes more than eye transplants or taped on fingerprints to fool any reliable system, and biometrics have become the most secure way to identify authorized access.

Right now, when you think of securing your valuables, you typically employ a password or a physical token like a key or keycard. These measures are designed to ensure that the person accessing the system is authorized. They want to make sure you are you.

It just so happens that there is an abundance of technology areas that have made it easy to hack systems protected under such measures. During the recent presidential campaign season, cyber crimes made front page news due to socially cracked passwords of the Democratic National Committee. As if external forces for abandoning passwords were not enough, 55 percent of us admit that we simply don't login because we couldn't remember our password and just as many people say they've abandoned purchases due to complications in the authentication process. Meanwhile, physical locks have spawned an entire subculture studying lock picking. Possession of lock picking tools is legal in most states, and illegal possession is only prosecutable if there is malicious intent, which is often difficult to prove. All the concerns were enough to make then-Prime Minister Stephen Harper switch systems in Canada, announcing that, "you can fake your name, you can fake your documents, but you can't fake your fingerprints". Indeed, biometrics have a case to be argued in favor of a system that identifies who you are rather than what you know or something you have…and what makes you more you than you?

To this end, a lot has changed already in our acceptance of biometrics. I grew up in an era that considered eye witness testimony incontrovertible. Now it's widely accepted that DNA, the blueprint of a human being and literal definition of someone's identity, is not only admissible but acknowledged to be more reliable than other forensic evidence. Around the time when most countries started to require fingerprinting for visa applications, biometrics is what got me into machine learning research. I couldn't publish academic articles if my paper submission didn't at least mention face detection and recognition. Since then, academics and federal laboratories have gone through full cycles of DARPA/IARPA face recognition funding (under FERET and JANUS) and datasets have been released with millions of people enrolled, ready for training algorithms. The advancements were enough for the Vice President of USAA Tom Shaw to say that "the password is dying."

Face recognition is just one quintessential example of research progress; the general field has been the Everest of artificial intelligence (AI) researchers everywhere. It's also one of the least publicized and most underappreciated fields of study. We seem to take it for granted when we get a phone call from the bank asking whether or not a transaction was ours. Many of us with iPhones have forgotten our passwords because we rely on the fingerprint analyzer. DNA is said to be the most reliable measurement in identifying innocent people that have found themselves in unfortunate situations in court. After the successful raid on the compound in Pakistan, few understand the importance and amount of work put forth in positively identifying Osama Bin Laden with DNA and face recognition. It's one of the most heavily invested technologies that has gone unnoticed because the movies make it seem so easy.

What Makes a Good Biometric

The term biometric can easily be broken down into its Greek atoms "bios", meaning life, and "metron" relating to measure. There are a variety of techniques satisfying the nomenclature, more than you might think. Besides irises and fingerprints, there are instances of biometrics related to behavior to include metrics like typing speed, your online social interactions, or the websites you frequent. Other examples that might be intuitive but don't come immediately to mind include the pulse in your veins, the geometry of your hands, an analysis of your gait, and your signature, not to be confused with handwriting (which is also a biometric). In fact, there's even a company (Tatt-C) that identifies tattoos.

Not only are such exemplars interesting and subtle, several systems add the permutation to assess multiple modalities at once, as in the case of the company Biometrica. Each illustration in the diversity of the concept of being a biometric are well-defined if they meet four major requirements:

1. **Universality**—you can measure it on virtually everyone
2. **Uniqueness**—only one person should own a particular set of measurements
3. **Permanence**—it doesn't change a lot over time
4. **Collectability**—You can actually measure it

The above four criteria must be satisfied, but on a more relevant note, a biometric's firm success intimately hinges on its robustness to countermeasures and loss. In building such capability, there is a lot of tech that goes into employing a single or combination of biometrics. After enrollment (where user data has been captured a priori), the technology can be broken up into standard steps of acquisition, preprocessing, feature extraction, matching, and database retrieval for any person to be identified.

Within each step, there can be a variety of ways in which they are performed. For example, the hardware acquisition of fingerprinting alone can vary from optical to capacitive to thermal to RF sensors. Taking a step back to the overall system, the assorted range of necessary innovation quickly grows in a list comprised of hardware sensors, database management, distributed networking and network security, computational scaling, and software and algorithmic improvements. It's no wonder that more than $24 billion is expected to be invested in biometrics by 2020 in areas like banking and healthcare.

Security and the Landscape

Of course, all of these advances don't come for free. Deployment of biometrics in both government and commercial sectors raises several questions related to privacy. Worries come from both voluntarily enrolled individuals and those whom biometrics are passively collected. Can the government or a civilian firm track me? Will others know about my medical conditions? Does the technology have the inclination to evolve and then be used for other functions? These are all valid concerns and need to be addressed, and several government agencies including National Science and Technology Board have promised to do so.

> “Getting ahead in this field will allow us to realize what’s next as well as determine where use cases can be augmented in every day applications, many of which biometrics have already taken center stage.”

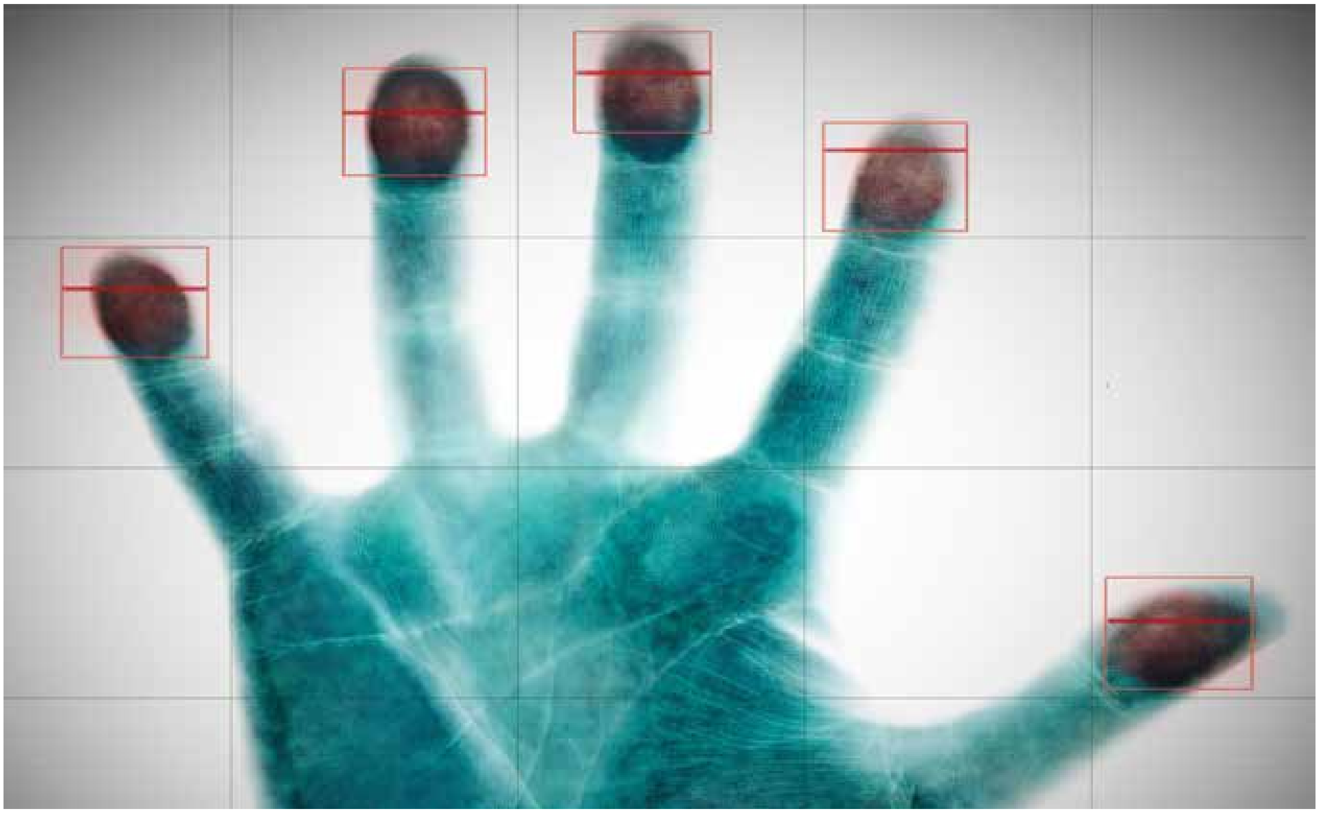
Fortunately, there are other applications that do not limit themselves to security and forensics. Car manufacturers adapt to the way you drive by identifying your driving profile. Tesla’s personalization prompted users to claim that it has crossed the line from minor perk “to an essential part of the ownership experience.” Furthering the personalization angle, Apple’s previous patent on fingerprint analysis suggested that their interest lay in personalization as well. These illustrations serve only to show the breadth that biometrics can contribute.

Here at In-Q-Tel, we realize the valuable role that biometrics has to play. Having surveyed the landscape of innovators in the United States, we understand that in a lot of ways, getting ahead in this field will allow us to realize what’s next as well as determine where use cases can be augmented in every day applications, many of which biometrics have already taken center stage. Closer to the valley, Lab41, one of the four IQT Labs, recently took on a real-world inspired project, where we concluded that automation technology can play a key role and save analysts countless hours in matching writers to their handwriting.

I hope you found our take on the growing field of biometrics useful. As analytics inevitably improve, the implications are that the matching between algorithms, software, and the best sensors can affect a readily employed and easily used package. Many businesses realize this and have begun to shift strategies, and coupled with multi-factor biometrics, the adoption rate has accelerated. The continuing migration of organizations to abandon the “good enough” conventional measures to newer and more convenient and secure biometric technology portends that it is more than just hype and intimate a trend that is indicative of its potential in the future. Q

***Dr. Karl Ni** is the senior data scientist at Lab41, an IQT Lab. His background is in statistical signal processing, machine learning, and computer vision. Prior to joining In-Q-Tel, he served as principle investigator for applied research programs with concentrations in RADAR, image processing, and social network analytics at MIT Lincoln Laboratory and Lawrence Livermore National Laboratory. Dr. Ni received his doctorate and masters of electrical and computer engineering at the UC San Diego and his BS at UC Berkeley in electrical engineering and computer science.*

A Look Inside



In this issue of the *IQT Quarterly*, "What Makes You-nique?", we venture into the growing world of biometric technologies, beyond the typical methods and applications most commonly seen. Karl Ni, Senior Data Scientist at IQT's Lab41, provided a foundation on the varying types of biometrics, including what constitutes a biometric, a look beyond the historically popular and most commonly utilized forms and functions, and a look at how biometrics are applied to and even shape the security landscape.

We then explore a less visual side of biometrics. Dr. Howard Lei of California State University discusses the wide history of speaker recognition as well as its recent advances and growing complexities.

Lab41 Data Scientists Brad H. and Patrick Callier delve into the field of handwriting recognition analysis. They begin with a look into the challenges faced in this area of biometric technology and go on to contemplate the anticipated direction and use in the future.

Rama Chellappa and a team of scholars from the University of Maryland embark on a technical exploration of facial recognition and analysis using deep learning, by taking a look at both the historical approaches and challenges as well as current methods.

Next, Stephen Elliott with the International Center for Biometric Research at Purdue University dives into a more holistic approach to biometrics, the Human Biometric System Interaction framework, looking at not one component of this technology, but rather a complete examination of biometric performance.

Finally, we conclude this issue with an article written by a team of authors from BehavioSec as they explore behavioral biometrics. This approach and technology examines the measurements of human behavior from a variety of sensors. Q

A Brief Overview of Speaker Recognition

by Howard Lei

Speaker recognition is the association of a speaker's identity to an acoustic utterance from characteristics of voices. In biometrics, a person's voice attributed by speaker recognition can be used to verify his or her identity for purposes of access control. Research in speaker recognition has a rich history dating back some four decades and has evolved massively in just a few years, owing to the complexity and difficulty of the problem.

A general pipeline of speaker recognition begins with a speech utterance that can contain any combination of voiced sounds from a speaker and can be of any duration. The voice can be spoken into any arbitrary voice-capturing device, e.g., a microphone, and in any acoustic environment. In the typical speaker recognition scenario, speaker models are constructed using utterances from a given speaker. Next, a test utterance spoken by an unknown speaker is evaluated against the speaker model to determine if the identity of the unknown speaker matches that from which the speaker model was constructed.

Instead of matching a single query to a large database of speakers, the speaker recognition task often determines whether two distinct utterances come from the same speaker or different speaker. One of the utterances can be

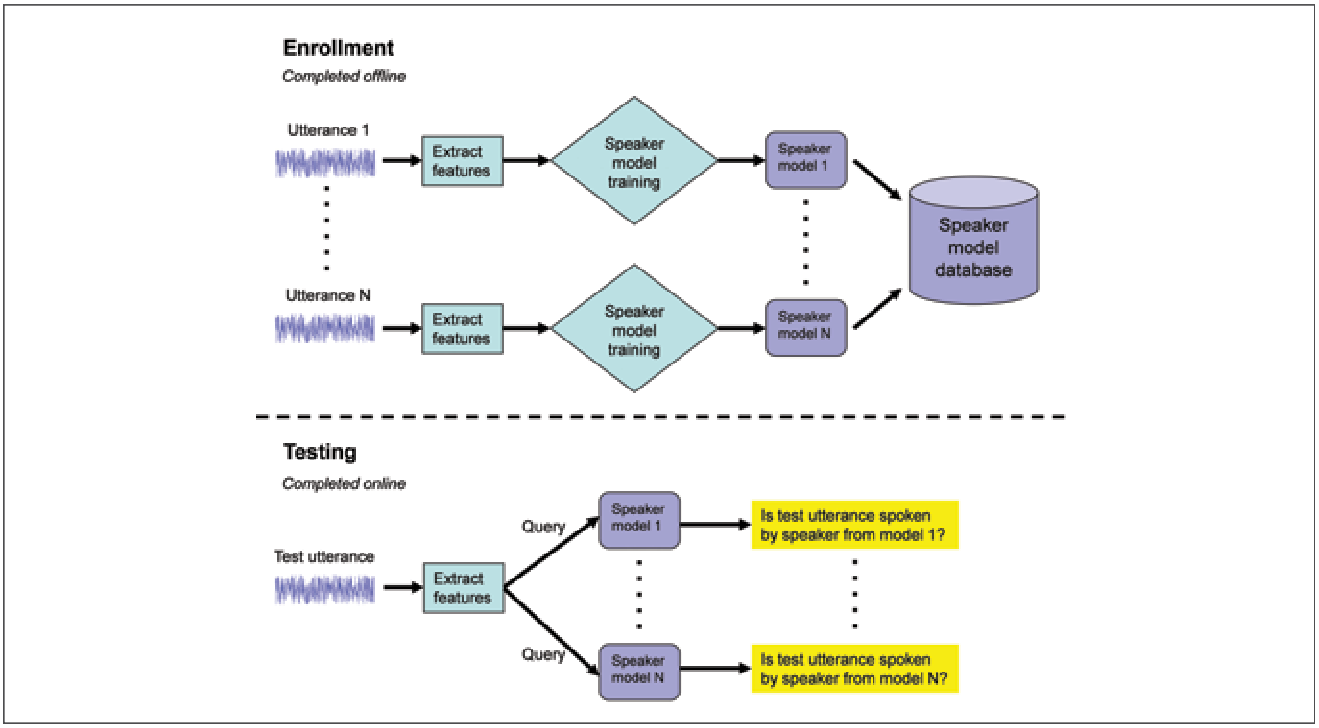


Figure 1 | The pipeline for the classical speaker recognition approach.

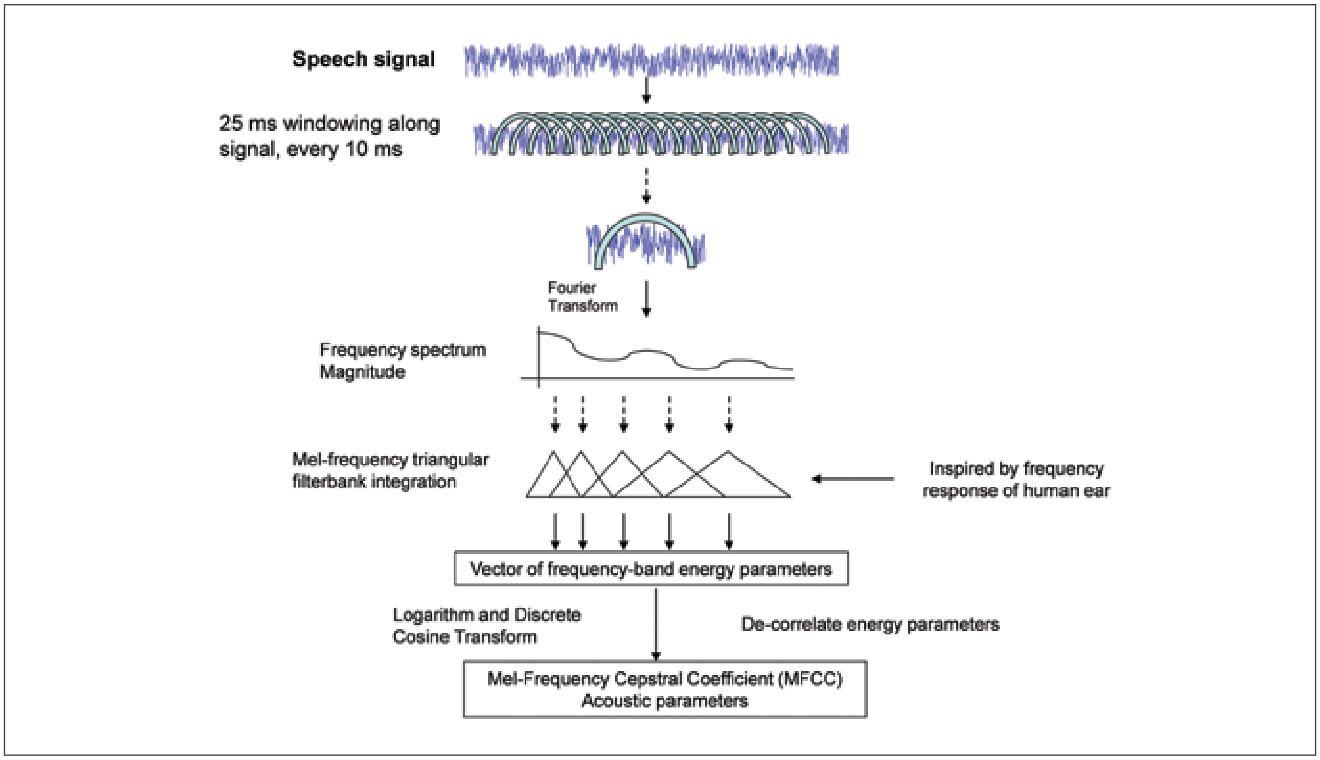

Speech signal
25 ms windowing along signal, every 10 ms
Fourier Transform
Frequency spectrum Magnitude
Mel-frequency triangular filterbank integration
Inspired by frequency response of human ear
Vector of frequency-band energy parameters
Logarithm and Discrete Cosine Transform
De-correlate energy parameters
Mel-Frequency Cepstral Coefficient (MFCC) Acoustic parameters

Figure 2 | Extraction of Mel-Frequency Cepstral Coefficient (MFCC) features.

thought of as representing the speaker model, while the other is considered to be the test utterance, whose speaker identity is unknown. This approach to speaker recognition is most commonly used by the community, and will be referred to for the remainder of this article.

The Fundamentals of Speaker Recognition Systems

Typical speaker recognition systems are considered to be classification systems based on machine learning algorithms. Because of the hundreds of thousands of hours of speech data a recognition system needs to process, considerable work has been invested into increasing the computational efficiency and accuracy of the systems. In certain speaker recognition tasks, such as the NIST Speaker Recognition Evaluation tasks[1], a speaker recognition system is typically asked to arrive at the same-speaker vs. different-speaker classification decisions for millions of pairs of speech utterances.

Figure 1 illustrates the classical speaker recognition system pipeline. Major components of the system include feature extraction, speaker model training, testing, and scoring. The following sections explain these components in detail.

Feature extraction for speaker recognition

Given a waveform speech utterance, the speaker recognition system must first convert the waveform into a set of parameters that can be used for speaker classification. Converting the waveform into the set of parameters is referred to as feature extraction, and the process typically requires a plethora of signal processing algorithms. The most common set of parameters used are the acoustic features, and the most popular acoustic features are Mel-Frequency Cepstral Coefficients (MFCCs)[2]. These features are first developed for automatic speech recognition, and have subsequently been found to perform well in speaker recognition. The MFCC features use information in logarithmically-spaced frequency bands of short-time speech spectra to match the logarithmically-spaced frequency responses of the human ear. The features are typically extracted on a frame-by-frame level, with 25 ms frames overlapping by 10 ms. Figure 2 illustrates the steps involved in MFCC feature extraction. Oftentimes, the temporal slope and accelerations of each acoustic feature vector component are used as well and augment the basic feature vectors. These coefficients are generally referred to as “delta” and “double-delta” coefficients.

The traditional GMM statistical modeling approach

From the early 1990s to the mid 2000s, the typical statistical approach used by speaker recognition systems to process the features and arrive at the overall classification decision is to model the distribution of features using Gaussian Mixture Models (GMMs)[3]. These models allow for the modeling of a wide range of feature distributions, with no prior knowledge of the distribution. The typical feature vector dimensions range from 40-60 dimensions, while a GMM can consist of up to 2,048 mixture components.

In the same-speaker vs. different-speaker classification scenario, features from one of the utterances are used to train a GMM. A measure of how well the features from the test utterance fit the distribution of the GMM is then computed. The higher the measure, the more likely the two utterances belong to the same speaker, and vice versa. Note that it is possible for a particular speaker to have multiple utterances of speech. Some tasks allow multiple utterances from a given speaker to be used to train the GMM, while other tasks allow only a single utterance to be used. Figure 3 below shows an example of the comparison between a test utterance and GMM models trained for three different speakers.

Advances in statistical modeling approaches

In the mid 2000s, one significant advance in the field of speaker recognition is the understanding that within-speaker variability contributes to a significant source of error in the classification decisions, and the development of algorithms to deal with this variability. The within-speaker variability can include differences in word usage across multiple utterances of a given speaker, and differences in the acoustic environment or recording conditions across the utterances.

Popular techniques in the mid to late 2000s that have been used to handle within-speaker variability include Nuisance Attribute Projection (NAP)[4], Joint Factor Analysis (JFA)[5], and i-vectors with probabilistic Linear Discriminant Analysis (pLDA)[6,7]. NAP is an algorithm that removes variability across utterances by training on multiple instances from the same speaker. JFA is an approach that decomposes the parameters of a speech utterance into a sum consisting of a speaker-independent component, a speaker-dependent component, and a component having to do with the recording conditions and/or acoustic environment. In the same-speaker vs. different-speaker classification problem, recognition can be performed using only the speaker-dependent component in the sum.

The i-vector technique attempts to create an "identity" vector for a given speaker while factoring in all sources of variability. The identity vector acts as a voice-print for the speaker. Using two utterances, a metric can be applied to give a measure of similarity between the i-vectors from the two utterances. The i-vector technique also allows for multiple post-processing algorithms to be applied in order to improve the final classification decision, one of the most effective techniques being probabilistic Linear Discriminant Analysis (pLDA). Figure 4 shows the high-level speaker recognition paradigm with the use of i-vectors. Figure 5 shows the details of the statistical extraction stage, and Figure 6 shows the details of the scoring stage. Note that Figure 2 showed the details of the feature extraction stage.

Speaker recognition system scoring

The performance of a speaker recognition system is often characterized by the Equal Error Rate (EER). A speaker recognition system generates one score for each comparison between a pair of utterances. Each comparison is called a trial. Higher scores should

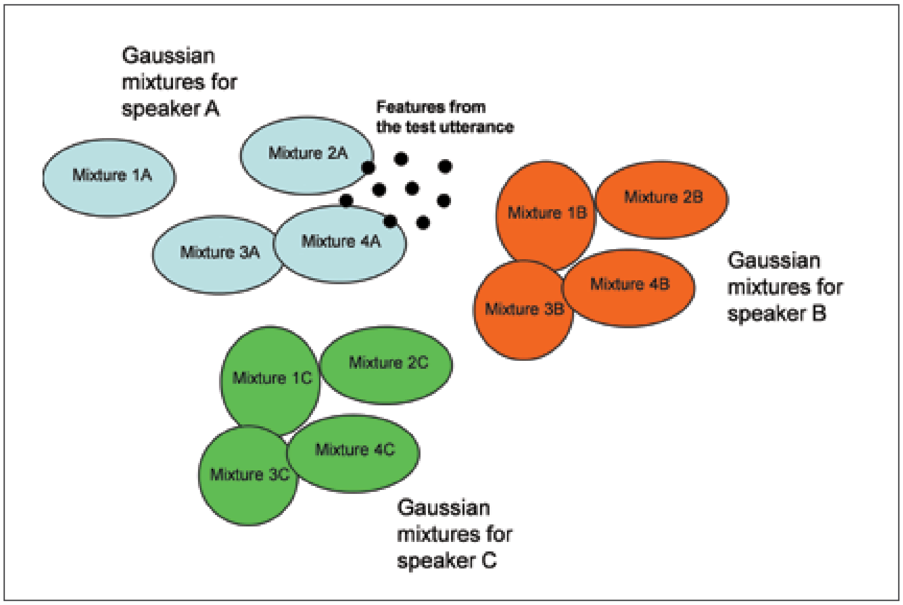

Gaussian mixtures for speaker A
Features from the test utterance
Mixture 1A
Mixture 2A
Mixture 3A
Mixture 4A
Mixture 1B
Mixture 2B
Mixture 3B
Mixture 4B
Gaussian mixtures for speaker B
Mixture 1C
Mixture 2C
Mixture 3C
Mixture 4C
Gaussian mixtures for speaker C

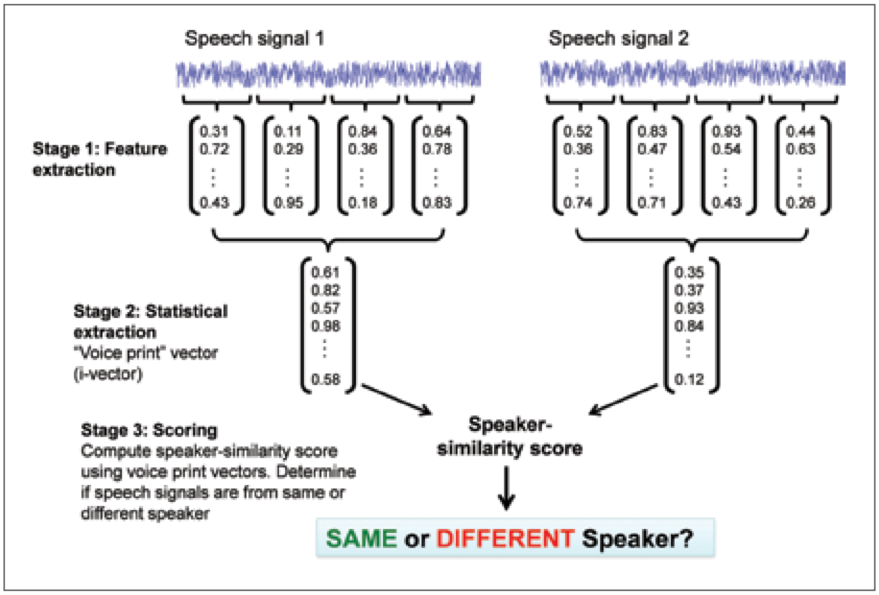
Figure 3 | GMM speaker modeling and classification of feature vectors.

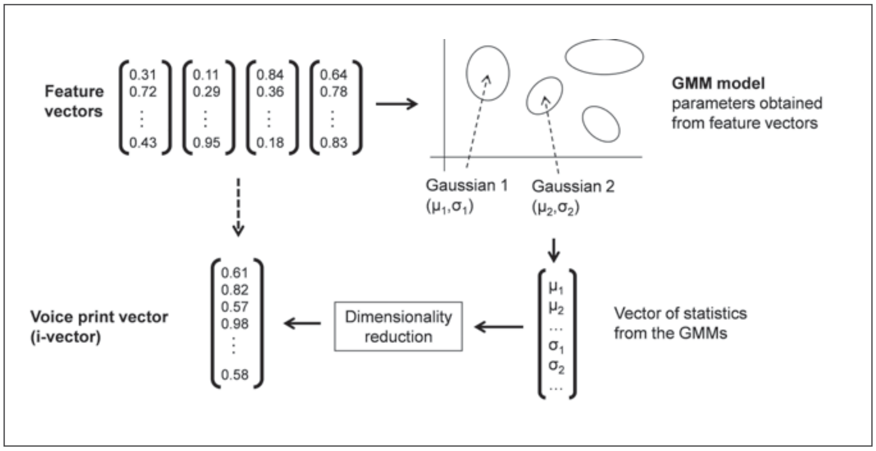
correspond to the trials with same-speaker comparisons, and lower scores should correspond to trials with different-speaker comparisons. A miss error occurs when trials with same speaker comparisons are classified as having different speakers, and a false alarm error occurs when trials with different speaker comparisons are classified as having the same speaker. The EER occurs at a scoring threshold where the miss error percentage equals the false alarm error percentage. A Detection Error Tradeoff (DET) curve (shown on the right of Figure 6) can be used to visualize the relationship between these two types of errors that a speaker recognition system makes.

Note that in Figure 6, an EER of 1.6 percent is observed in the DET curve. The total error percentage, consisting of both the false alarm and miss percentages, is twice the EER (e.g. 3.2 percent). This percentage suggests that given two speech utterances, the speaker recognition system would be able to correctly determine whether the speakers in the utterances are same or different 96.8 percent of the time. More recently systems have been able to produce the correct result ~100 percent of the time for cases where the utterances are noiseless telephone conversational speech, and sufficient training data is available.

NIST Speaker Recognition Evaluations

One of the driving factors for the development of techniques used in speaker recognition systems is the bi-annual Speaker Recognition Evaluations hosted by the National Institute of Standards and Technology (NIST). The latest evaluation is SRE16[8], which occurred in 2016. These evaluations require participants to run their speaker recognition systems on conversational speech recorded by the Linguistic Data Consortium (LDC)[9] with various recording devices. The





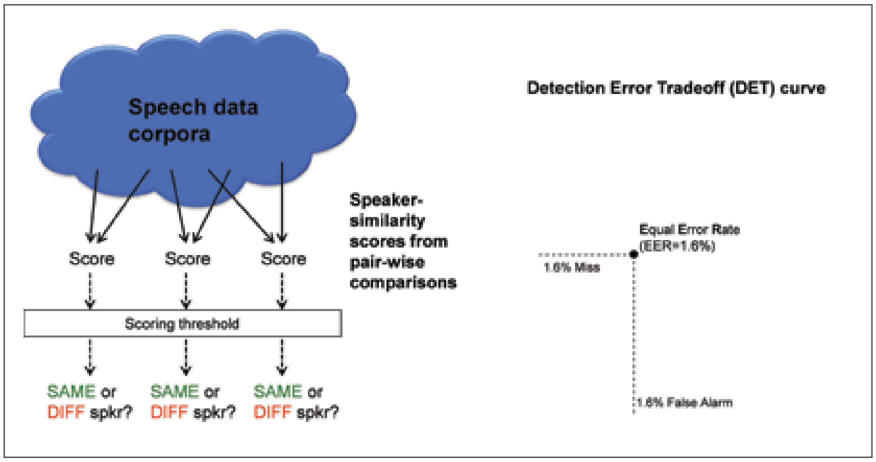


Figure 4 *(top)* | High-level overview of the i-vector speaker recognition paradigm.

Figure 5 *(middle)* | Details of the statistical extraction procedure for the i-vector speaker recognition paradigm.

Figure 6 *(bottom)* | Details of the scoring procedure speaker recognition systems.

> In the future, we can expect voice biometrics to become a more integral part of our daily lives.

earlier evaluations used only telephone conversational speech between two speakers, with each speaker speaking for about 2 to 2.5 minutes. Each speech utterance for a speaker would consist of only that person's speech in the telephone conversation. The later evaluations contained other speech styles of varying durations, along with more languages and recording environments from which speech data was collected. During each evaluation, systems are required to process a specified set of trials, where each trial requires a system to produce a speaker similarity score used for making the same-speaker vs. different-speaker determination. The i-vector technique has been one of the most successful techniques used by systems in the more recent evaluations. Not only does it handle the different forms of within-speaker variability well, but it is also computationally efficient and allows millions of trials to be processed in a timely fashion.

Future Challenges

A major challenge facing speaker recognition systems is that of handling speech in noisy recording environments, where the noise can come from both the voice-capturing device, as well as the acoustic background . The state-of-the-art approaches work well on conditions where the acoustic noise is limited and largely known, but performance decreases when the systems process speech recorded in more varying acoustic environments e.g., background music, voice, objects making sounds, etc.

Furthermore, one of the weaknesses of the i-vector approach is that it requires many hours of enrollment data to train various components of the system, and its performance suffers on speech of shorter duration. Finding effective approaches for speaker recognition on utterances of shorter duration is one of the ongoing efforts in the field. More recent efforts in speaker recognition have seen the use of Deep Belief Networks (DBNs) in conjunction with the i-vector approach, with varying degrees of success.

These challenges currently prevent speaker-recognition systems from being commonly used in our day-to-day lives, in the same way that speech-recognition systems like Siri are being used. Voice-based authentication in noisy environments is less reliable compared to other means of authentication, such as fingerprint (currently used in laptops) or DNA. Nevertheless, corporations like Nuance have developed voice-authentication solutions such as FreeSpeech[10] and VocalPassword[11]. These are systems used to verify a customer's identity by extracting the voice characteristics of the customer and comparing those characteristics to those stored in a database. ArmorVOX[12] also has developed a voice biometric engine, which is helpful for securing private information. Other corporations have developed systems based on voice biometrics as well, and in the future, we can expect voice biometrics to become a more integral part of our daily lives. **Q**

Dr. Howard Lei *completed his Ph.D. in Electrical Engineering and Computer Science at UC Berkeley in 2010, focusing on applied machine learning towards speaker recognition. He was a postdoctoral researcher at the International Computer Science Institute in Berkeley, CA, where he continued his work in speaker recognition, and engaged in multimedia analysis and classification. Dr. Lei began his work as an Assistant Professor in the School of Engineering at California State University, East Bay, in 2013, where he has been involved in teaching a variety of courses on computer software and hardware systems in the Computer Engineering program. He also has been involved in projects including the statistical prediction of medication demands for disease outbreaks, improving engineering education, and re-vamping curricula to introduce state-of-the-art software and hardware into his courses.*

Reference

1. Multimodal Information Group, National Institute of Standards and Technology (NIST). Speaker Recognition. Retrieved January 5, 2017 from https://www.nist.gov/itl/iad/mig/speaker-recognition.

2. S. Davis and P. Mermelstein, "Comparison of parametric representations of monosyllabic word recognition in continuously spoken utterances," Proceedings of ICASSP, 1980.

3. .A. Reynolds, T.F. Quatieri, and R. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," Digital Signal Processing, 10(3), pp. 19-41, 2000.

4. A. Solomonoff, W.M. Campbell, and I. Boardman, "Advances in channel compensation for SVM speaker recognition," in Proceedings of ICASSP, 2005.

5. P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, A Study of Interspeaker Variability in Speaker Verification," IEEE Transaction on Audio, Speech and Language, 16(5), pp. 980-988, July 2008.

6. N. Dehak, R. Dehak, P. Kenny, N. Brummer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," in Proceedings of Interspeech, 2009, pp. 1559-1562.

7. L. Burget, P. Oldřch, C. Sandro, G. Oldřej, M. Pavel, and N. Brümmer, "Discriminantly trained probabilistic linear discriminant analysis for speaker verification," in Proceedings of ICASSP, Brno, Czech Republic, 2011.

8. Multimodal Information Group, National Institute of Standards and Technology (NIST). Speaker Recognition Evaluation 2016. Retrieved January 5, 2017 from https://www.nist.gov/itl/iad/mig/speaker-recognition-evaluation-2016.

9. Linguistic Data Consortium (LDC), Current Projects. Retrieved January 5, 2017 from https://www.ldc.upenn.edu/collaborations/current-projects.

10. Nuance, Authentication via conversation, Retrieved January 5, 2017 from http://www.nuance.com/for-business/customer-service-solutions/voice-biometrics/freespeech/index.htm.

11. Nuance, Nuance VocalPassword. Retrieved January 5, 2017 from http://www.nuance.com/for-business/customer-service-solutions/voice-biometrics/vocalpassword/index.htm.

12. ArmorVOX, Retrieved January 5, 2015 from http://www.armorvox.com.

Writer Identification in Handwriting

by Brad H. and Patrick Callier

Forensic identification of the writers of handwritten documents is a rarefied vocation. Only a few dozen handwriting specialists exist nationwide, with a handful employed in the Intelligence Community. The reasons their craft is hard for humans overlap with the reasons it would be difficult to supplant them with computers.

The CSI Effect

Hollywood has had an impact on public perception of what is possible in technology. From scanning a crowd in a video while doing facial recognition, to the hacker that, after spending 10 seconds rattling on a keyboard, says, "I'm in." This is often often known as the CSI Effect. Even jurors have fallen victim, asking court-certified experts why they haven't run some test they saw on their favorite crime show[1]. The reality is that technology, specifically in the realm of biometrics, is often exaggerated in popular media.

Handwriting analysis falls into this category. There is no "enhance and compare" button for handwriting. It takes a skilled expert to compare two samples of handwriting and determine if they have common authorship. Many companies and academics have tried tackling this problem—for instance, in workshops at the IAPR Conference on Document Analysis and Recognition (ICDAR)[2]—but robust writer identification on real-world data remains elusive. Let's explore the reasons automating handwriting comparison is a road fraught with obstacles.

Why Handwriter Recognition Is Hard

The biggest hurdle in handwriter identification is the vast range of variability within a single writer's production. Consider the difference in how your writing looks when you use a ballpoint pen compared to a felt-tip pen. That difference goes beyond just the breadth of the stroke your pen creates; ballpoint pens leave more evidence of the pressure you use as you write, and of the flow of your hand through the stroke, both potential sources of information in forgery detection. Add on differences induced by the writing medium—lined or unlined paper? How far apart are the rule lines? Consider state of mind, how rushed the writer is, and even variability unattributable to any specific factor, and the range of written forms that a system would have to map to individual writers is quite vast.

Consider too that the major factors distinguishing two exemplars of handwriting are likely to be the content—what is actually written on the page—and the implement and medium. The points of invariance that distinguish an individual's writing are extremely uncommon. A document of hundreds of characters might only have a few points that tie it to other exemplars by the same writer. Those points of correspondence tend to be highly context-specific—for instance the height of a letter at the start of a connected sequence of letters in a cursive hand. And potential correspondences are highly defeasible. As one handwriting expert put it, they are "very important, unless they aren't."

As jaw-dropping as some recent advances in machine learning may be, most computer vision problems lack the degree of subtlety and difficulty posed by handwriter identification.

A small but dedicated community of research has arisen to take on the challenge of writer identification. They have pursued a variety of techniques, including tried-and-true machine learning algorithms and some of the more cutting-edge technologies associated with the "deep learning" craze. We will go through a few of these below to give a flavor of what this field looks like today and offer some suggestions about its future.

State of the Art: Contour Gradients

The state of the art in computer vision for writer identification makes use of a wide set of traditional techniques in computer vision. One top approach, from computer vision expert Rajiv Jain, engineers an automated computer vision pipeline that tries to mimic how forensic specialists work by hand.

Jain's approach first decomposes a corpus of documents into segments corresponding to individual letters or small groups of letters and clusters them together using unsupervised learning. The result is a "pseudoalphabet" of visually similar letter forms. This allows direct comparison between specific locations in separate exemplars, similar to how a forensic expert might isolate individual letters or digraphs and contrasts them with a canonical representation. For example, a connected cursive form of the word *of* might show up as its own pseudoletter across multiple documents, allowing for a direct comparison between exemplars that happen to use this form.

Each segment in a document is described by a grid of "contour gradients," a way of representing the orientations of penstrokes across different parts of a segment. Jain's algorithm compares each pair of exemplars in a corpus by matching segments according to which pseudoletter class they correspond to and finds the closest pair of segments across exemplars for each such pseudoletter class. It then computes the average difference between these closest-matched pairs of segments and uses that to create an overall measure of similarity between two exemplars. If *of* were the only pseudoletter shared between two exemplars, this would amount to finding the closest match between two instances of *of* and using their similarity as a measure of the similarity between the exemplars themselves.

This approach rapidly zooms in on potential points of comparison between exemplars, but it is not particularly tolerant of noisy or poor-quality images of documents. It performs less well on real-world data, which typically comes along with many types of noise. To do better on writer identification in the presence of noise, Lab41 turned to deep learning for help.

Deep Learning

Deep learning has been successfully applied to many image-related problems. Though handwriting and deep learning haven't had a long courtship, there have been recent significant advancements. In August 2013, Alex Graves impressed the world with his Recurrent Neural Network (RNN) that generates handwriting[3].

You simply type any sentence and choose a handwriting style, and the RNN produces an online handwriting sample. This was a major step forward in computer vision, but it didn't directly address the problem of matching exemplars by writer.

A year later, Stefan Fiel and other scholars used a convolutional neural network (CNN) to capture features of writers and compare them across a corpus of document features[4]. A CNN is a wise choice, because it uses a small sliding window to move across (convolve) the entire image and gather features. If the sliding window is large enough, it can beautifully capture transitions from one letter to the next, the spacing between letters, and the edges and curves of handwritten characters. The paper produces excellent results on the ICDAR 2013 Competition on Writer Identification[5]. But real-world handwritten documents are often much messier than those found in academic datasets. They can have lines, coffee stains, watermarks, stamps, and many more types of noise that make feature extraction difficult.

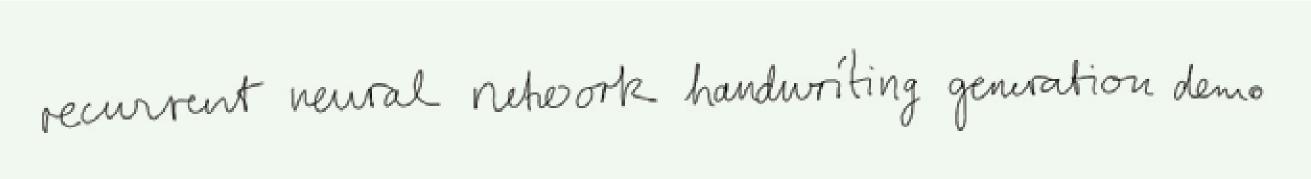
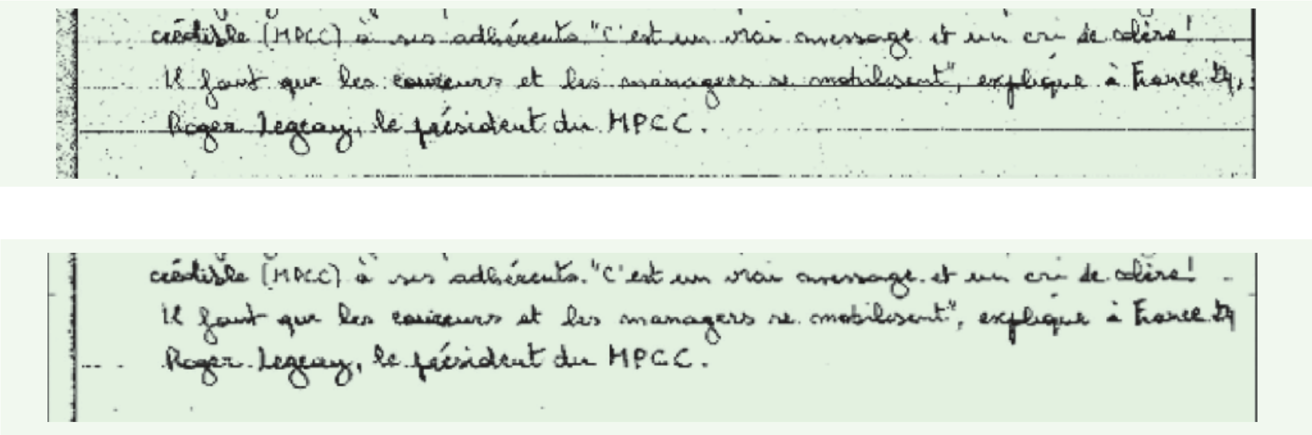


Figure 1



crédible (MPCC) à ses adhérents. "C'est un vrai message et un cri de colère! Il faut que les coureurs et les managers se mobilisent", explique à France 3, Roger Legeay, le président du MPCC.

crédible (MPCC) à ses adhérents. "C'est un vrai message et un cri de colère! Il faut que les coureurs et les managers se mobilisent", explique à France 3, Roger Legeay, le président du MPCC.

Figure 2 | *(top)* Noisy handwriting image, including notebook lines and scanner artifacts. *(below)* Denoised handwriting image, with greatly reduced noise and most notebook lines removed.

The noisy nature of documents is what inspired Lab41's project D*Script. D*Script focuses on combining the strengths of both deep learning and traditional computer vision techniques that have been used for writer identification.

The first part of the system D*Script proposed is a denoising autoencoder (DAE). A DAE is a neural network architecture that takes normal inputs, adds some random noise to them, and then tries to reconstruct the original noiseless input. The DAE in D*Script has a similar job, but the added noise is not random. Lines, watermarks, and many more types of noise are added to clean document images. D*Script's DAE is tasked with removing this noise. It learns by comparing its reconstruction with the original noiseless image.

The contour gradient method can then be used to take the cleaned-up image and extract handwriting for comparison to other documents in the database. For further information and code, please visit our Lab41 GitHub page[6].

The Future of the Art

There is much room for improvement and innovation when it comes to automating handwriting authorship adjudication. There are several avenues that look like promising paths forward. One such avenue is based on research by Diederik Kingma[7]. The model in his work learns how to generate handwriting and style. Given an example of a character, it uses its knowledge of all of the

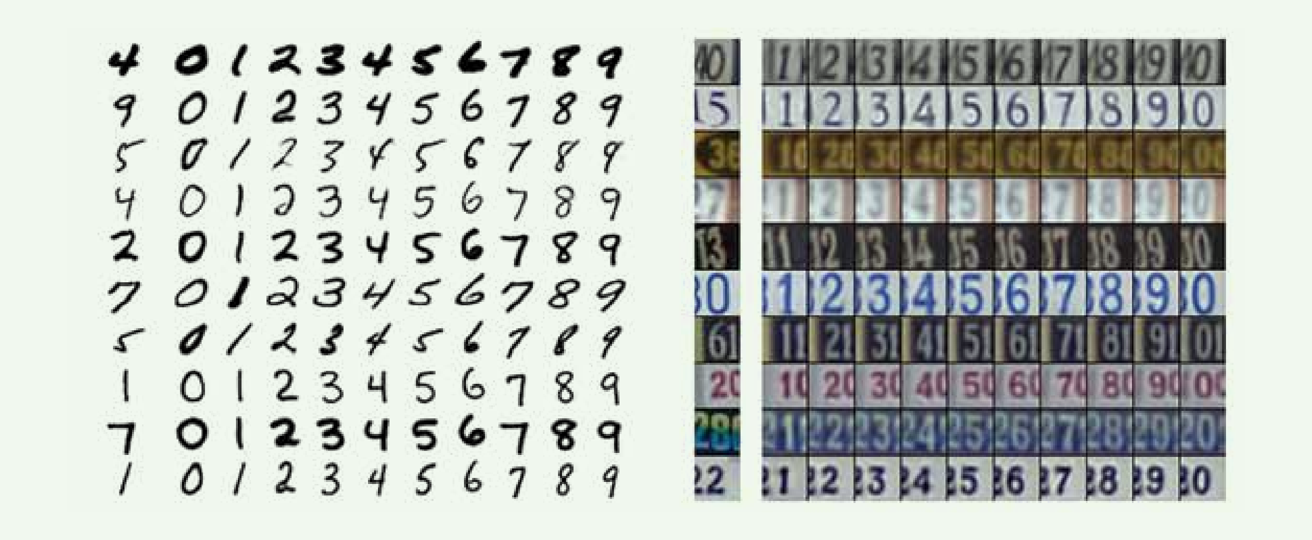


Figure 3 | Handwritten digits *(left)* and house numbers from the Google Street View dataset *(right)*. The numbers on the left-hand column of each set come from the real world. The digits 0-9 are generated by the model in the style of the real-world sample.

other characters and styles it has seen in the dataset to create "analogical fantasies" of other characters.

This means we can generate handwriting not in our data set. For example, if the number "7" was never captured from a certain writer, the model can imagine how the writer may write it. Not just one "7", but infinitely many slight variations of it, thus generating the author's possible range of variation. The downside, of course, is that this paves the way for children to create handwritten excuse notes to their teacher that their parents never wrote.

The potential for this kind of high-tech spoofing offers a lesson: handwritten communication—for some a profound symbol of the slow death of old communication technologies—is part and parcel of a world where AI, machine learning, and high-tech communication coexist. Writer identification poses challenges unique in the greater handwriting analysis and computer vision worlds, and we expect to see a lot of advancement in this area in the years to come. **Q**

Brad H. *is a Data Scientist at Lab41, specializing in deep learning. He graduated from the University of Utah with a BS in mathematics and a Masters in statistics. He has been involved in building algorithms for biometrics for over three years. When he is not rock climbing or spending time with his family, Brad enjoys watching the lights flicker at home as he applies deep learning models to financial trading strategies.*

Patrick Callier *is a Data Scientist and a Linguist at Lab41. He graduated from Stanford University with a BA, obtained his Ph.D. at Georgetown University, and finished his PostDoc at Stanford University in Computational Linguistics. Prior to joining In-Q-Tel in 2015, Patrick became an Insight Data Science Fellow, developing analytics for user inactivity on Twitter. His interests include prototyping deep learning algorithms and other machine learning techniques for natural language processing and computer vision to develop solutions in applied problem areas.*

References

1. https://apps.americanbar.org/litigation/committees/trialevidence/articles/winter spring2012-0512-csi-effect-jurors.html
2. http://u-pat.org/ICDAR2017/
3. http://www.cs.toronto.edu/~graves/handwriting.html
4. http://www.caa.tuwien.ac.at/cvl/wp-content/uploads/2014/12/fiel-caip2015.pdf
5. http://users.iit.demokritos.gr/~louloud/ICDAR2013WriterIdentificationComp/
6. https://github.com/Lab41/d-script
7. https://arxiv.org/pdf/1406.5298.pdf

What Can a Single Deep Learning Algorithm Say about a Face?

by Rama Chellappa, Rajeev Ranjan, Jun-Cheng Chen, Swami Sankaranarayanan, and Carlos D. Castillo

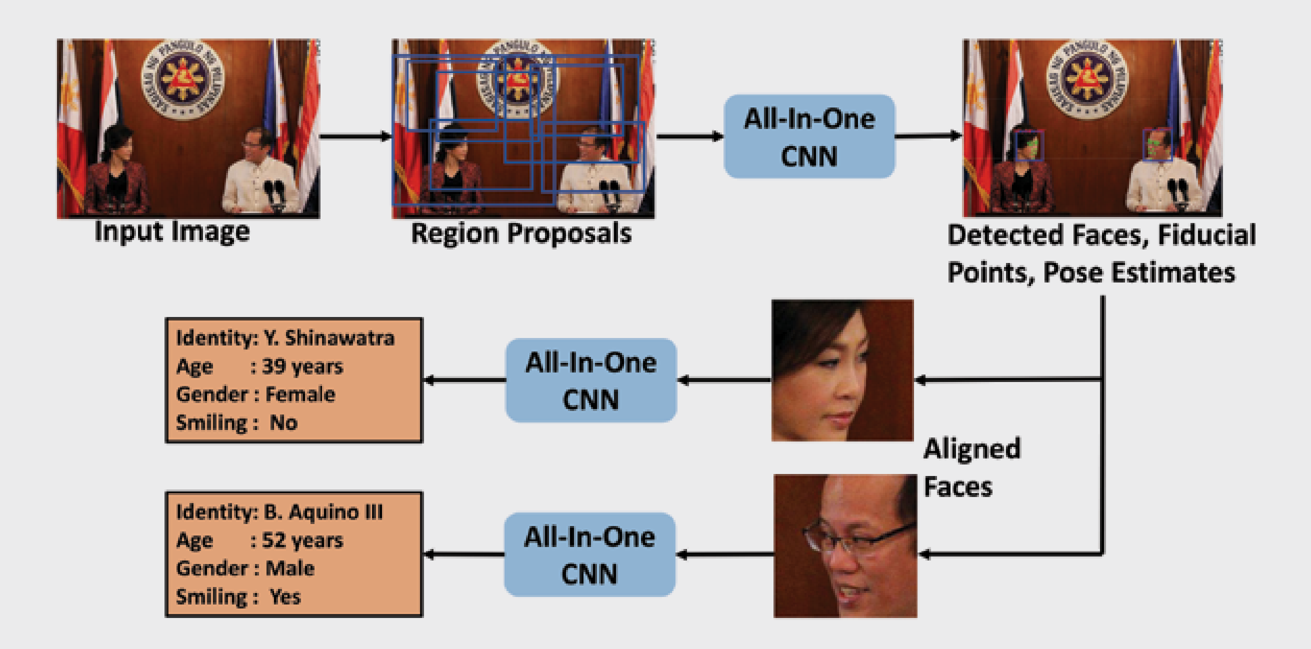
Facial analytics is a challenging problem in computer vision and has been actively researched for over two decades[36]. The goal is to extract as much information as possible from a face, such as location, pose, gender, age, emotion, etc., that will be useful in surveillance, human-computer interaction (HCI), smart cars, and other applications. In the past, different methods have been designed for extracting gender, pose, ID and so on. The integration of deep networks and multi-task learning is making it possible to extract facial analytics using a single network that shares features across multiple tasks and domains.

Historically, two major tasks have dominated the facial analysis literature: face identification and face verification. Face identification aims to identify the subject identity of a query image or video from a set of enrolled persons in the database. On the other hand, face verification, given two images or videos, determines whether they belong to the same person. Since the early 1990s, numerous algorithms have been shown to work well on images and videos that are collected in controlled settings. However, the performance of these algorithms often degrades significantly on images that have large variations in pose, illumination, expression, aging, and occlusion.

Unconstrained Face Recognition

An automatic face recognition system typically consists of the following components: (1) face detection, (2) facial landmark detection to align faces, (3) feature representation and (4) metric learning to identify a subject's identity or to determine two faces from the same identity or not. Face detection determines whether and where a face is located in an image. Facial landmark detection aligns faces into canonical coordinates in order to robustly compare queries to enrolled faces. Feature representations distill the salient and discriminative information of a face into a fixed set of numerical values. Finally, metric learning is a procedure by which we can compare the feature representations. For decades, the large number of published papers have gone through generations of change and progress, and generally have followed this pipeline.

Despite significant progress, the performance of conventional systems has not been adequate for deployment. Fortunately, over the last five years, methods based on deep convolutional neural networks (deep CNNs) have shown impressive performance improvements for object recognition[13, 28] and object/face detection[8, 18]. In addition, face recognition systems based on deep CNNs[26, 17] have yielded performance surpassing human recognition accuracy. One of the most well-known in the community is "The Labeled Faces in the Wild" (LFW) dataset[9]. This has been made possible due to the availability of large annotated datasets, a better understanding of the non-linear mapping between input images and class labels as well as the affordability of GPUs.



Identity: Y. Shinawatra
Age : 39 years
Gender : Female
Smiling : No

Identity: B. Aquino III
Age : 52 years
Gender : Male
Smiling : Yes

Figure 1 | Overview of the All-in-One CNN system for face recognition and facial analysis.

A Different Perspective

When a human looks at a face in an image, he or she can detect where the face is, gender, rough pose, age, expressions, etc. When machines are designed to perform these tasks, they are often designed as independent algorithms solving each of these tasks. However, one can design a deep network that can simultaneously accomplish all of the tasks by sharing the deep features and exploiting the relationships among these tasks. This approach is called multi-task learning, which learns to optimize different targets for different tasks using the same underlying program. One can view such a network as akin to a group of students getting together to study for an exam, complementing each other's strengths, with the overall goal of everyone getting high marks. The conjecture is that learning multiple facial analytics tasks simultaneously results in superior performance of each individual task.

With that in mind, a comprehensive system for facial analytics that can simultaneously perform face detection, face alignment, face identification/verification and extract other details such as 3D head position and angle, gender, smile, and age using a single deep learning algorithm appears to make the most sense. We will take you through an "All-in-One" deep CNN approach that employs a multi-task learning framework that exploits the synergy among different domains and tasks for face recognition. Such an approach has been demonstrated to be superior on the challenging IARPA Benchmark A dataset (IJB-A) and is an effective system.

All-in-One CNN for Facial Analytics

The All-in-One deep CNN[20] is a single CNN model for simultaneous face detection, landmark localization, face recognition, 3D head pose estimation, smile detection, facial age estimation, and gender classification. The All-in-One CNN architecture is shown in Figure 1. We start with the previously trained face identification CNN from Sankaranarayanan *et al*[25]. This network is used as a backbone network for training the face identification task and shares parameters with other face-related tasks. The central tenet for this design choice is that a CNN pre-trained on face identification task provides better initialization for a generic face analysis task, since the filters retains discriminative face information.

As shown in Figure 1, the tasks are then divided into two groups: 1) subject-independent tasks which include face detection, facial landmark localization and visibility, pose estimation and smile prediction, and 2) subject-dependent tasks which include age estimation, gender prediction and face recognition.

Since no single large dataset is available with all the annotations for face bounding box, fiducial points, pose,

whether you're at the highest
or the most modest level
Shinzo Abe
Japanese Prime Minister
SPD
BENAZIR BHUTTO 18-10-2007
IN PLANE AT DUBAI TAKING WITH KTN NEWS
rtve.es

Figure 2 | The IJB-A dataset is the new dataset after LFW to push the development of next-generation face recognition system and contains faces in large variations of pose, illumination, image quality, occlusion, etc.

gender, age, smile and identity information, we train multiple CNNs with respect to task-related datasets D_i, and share the parameters among them. In this way, the shared parameters adapt to the complete set of domains $(D_1, D_2, \cdots, D_d)$ instead of fitting to a task-specific domain. Additionally, the total number of training samples increases to roughly one million, which is advantageous for training deep CNNs. At test time, these sub-networks are fused together into a single All-in-One CNN. Table 1 lists the different datasets used for training our All-in-One CNN, along with their respective tasks and sample size. The complete network is trained end-to-end using a software package developed by UC Berkeley called *Caffe*[10].

Face Detection, Key-points Localization and Pose Estimation: These tasks are trained using the Annotated Facial Landmarks in the Wild (AFLW)[12] dataset. The algorithm that we employ is similar to HyperFace, a single CNN model for simultaneous face detection, landmark localization, pose estimation, and gender classification. HyperFace consists of three modules. The first generates class independent region-proposals from the given image and scales them to an appropriate resolution. The second module is a CNN that takes in the resized candidate regions and determines whether or not it is, indeed, a face. If a region is a face, the network then predicts facial landmarks locations, 3D head pose and gender information. The third module is a post-processing step that iteratively proposes candidate regions and removes duplicate detections and boosts the face detection score while improving the performance of individual tasks.

Dataset	Facial Analysis Task	# Training Samples
CASIA [35]	Identification, Gender	490,356
MORPH [22]	Age, Gender	55,608
IMDB+WIKI [23]	Age, Gender	224,840
Adience [14]	Age	19,370
CelebA [15]	Smile, Gender	182,637
AFLW [12]	Detection, Pose, Fiducials	20,342
Total		**993,153**

Table 1 | Datasets used for training.

The AFLW dataset is a large-scale, real-world dataset used for landmark localization. We randomly select 1000 images from the AFLW dataset for testing, and use the remaining images for training. We use a selective search[29] algorithm to generate region proposals for faces from an image. Regions with considerable overlap with the ground truth bounding box are considered positive examples whereas other regions are used as negative examples for training the detection task.

Gender Recognition and Smile Detection: Gender and smile classification are binary classification problems. The training images are first aligned using facial key-points which are either provided by the dataset or computed using HyperFace. Then, a simple machine learning loss function is used to measure the likelihood of the gender as male or female and presence of a smile.

Facial Age Estimation: We formulate the age estimation task by making the CNN learn to predict the facial age from a face image. It has been shown that apparent age estimation can be well-modeled when the standard deviation of age is known in a dataset. Unfortunately, outliers cause such models to converge slowly and perform poorly. The algorithm that we propose to remedy this dichotomy is to use a weighted combination of two loss functions using the standard deviation information as well as training example pairs that have the predicted age and the ground-truth age. The procedure is to first initialize and slowly fit the model with the known standard deviation of the annotated age value.

Face Recognition: We use 10,548 subjects from the CASIA dataset[35] to train the face identification task. The images are aligned using HyperFace before passing them through the network. The neural network is then optimized to find the best parameters that minimize the error between truth labels and predictions, a *loss function*. The final overall loss L is the weighted sum of individual loss functions, i.e. $L = \sum_{t=1}^{8} \lambda_t L_t$,where L_t is the loss and λ_t is the loss-weight corresponding to task *t*. These loss-weights are chosen empirically from the data.

For the testing stage, we deploy a two-stage process as shown in Figure 1. In the first stage, we use the selective search to generate region proposals from a test image, which are passed through our all-in-one network to obtain the detection scores, pose estimates, fiducial points and their visibility. We also use separate processes to filter out non-faces and improve fiducials and pose estimates. For the second stage, we use the obtained fiducial points to align each detected face to a canonical view using the similarity transform. The aligned faces, along with their flipped versions are passed again through the network to get the smile, gender, age, and identity information.

Face Identification/Verification on the IJB-A dataset

We present the results of the proposed All-in-One CNN for face recognition task on the challenging IARPA Janus Benchmark A (IJB-A)[11]. The receiver operating characteristic curves (ROC) and the cumulative match characteristic (CMC) scores are used to evaluate the performance of different algorithms for face verification. The ROC curve measures the performance in the verification scenarios, and the CMC score measures the accuracy in closed set identification scenarios.

The IJB-A dataset contains 500 subjects with 5,397 images and 2,042 videos split into 20,412 frames. Sample images and video frames from the datasets are shown in Figure 2. The IJB-A evaluation protocol consists of verification (1:1 matching) over 10 splits. Each split contains around 11,748 pairs of templates (1,756 positive and 9,992 negative pairs) on average. Similarly, the identification (1:N search) protocol also consists of 10 splits, which are used to evaluate the search performance. In each search split, there are about 112 gallery templates and 1,763 probe templates (i.e. 1,187 genuine probe templates and 576 impostor probe templates). The training set contains 333 subjects, and the test set contains 167 subjects without any overlapping subjects. Ten random splits of training and testing are provided. Unlike LFW[9] and YTF[32] datasets, which only use a sparse set of negative pairs to evaluate the verification performance, the IJB-A divides the images/video frames into gallery and probe sets so that all the available positive and negative pairs are used for the evaluation. Also, each gallery and probe set consist of multiple templates. Each template contains a combination of images or frames sampled from multiple image sets or videos of a subject. In contrast to LFW and YTF datasets, which only include faces detected by the Viola Jones face detector[30], the images in the IJB-A and JANUS CS2 contain extreme pose, illumination, and expression variations. These factors essentially make IJB-A a challenging face recognition dataset[11].

For face detection and facial landmark localization tasks, we present the sample results on the IJB-A dataset in Figure 3. The results demonstrate that All-in-One CNN is able to detect and localize facial landmarks for face in large pose, illumination, and facial age variations. This attributes to the multi-task learning over various face datasets. In addition, the All-in-One CNN can get reliable estimate for facial age, gender, 3D head pose, and smile.

We present the identification/verification results of the proposed approach for the IJB-A dataset in Table 2 (page 21). Besides using the average feature representation, we also perform media averaging, which first averages the features combing the same media (image or video) and then further averages the media features to generate the final feature representation followed by Triplet Probabilistic Embedding [25].

Table 2 summarizes the scores (i.e., both ROC and CMC numbers) produced by different face identification/verification methods on the IJB-A dataset. We compare

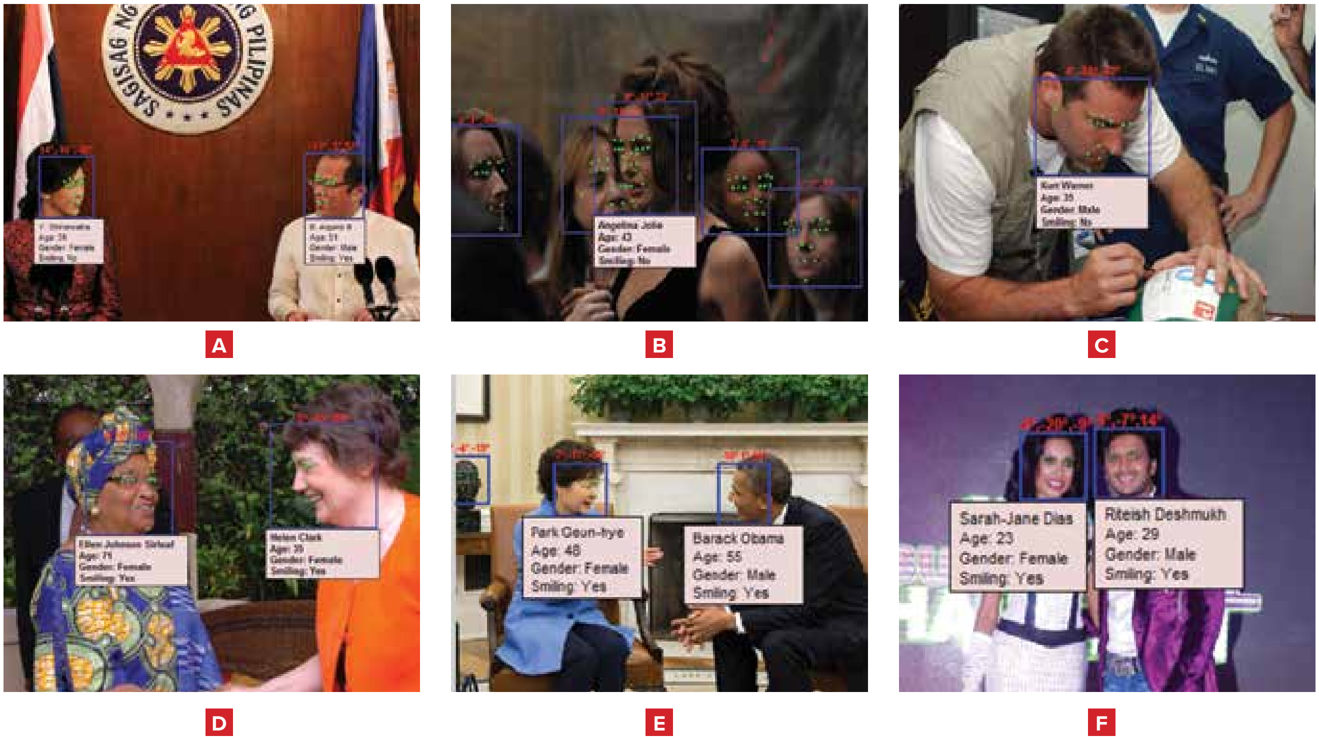


Figure 3 | Sample results of the all-in-one CNN for the IJB-A dataset with detected face bounding boxes, fiducial points, and identity along with 3D head pose, gender, smile, and facial age estimation. Although the algorithm predicts identity, age, gender and smile attributes for all the faces, we show them only for subjects that are present in the IJB-A dataset for better image clarity.

the results with DCNN$_{bl}$ (bilinear CNN[24]) DCNN$_{pose}$ (multi-pose DCNN models[2]), Neural Aggregation Network for Vidoe Face Recognition[34], DCNN$_{3d}$[16], template adaptation (TP)[7], DCNN$_{tpe}$[25], and those reported recently by NIST where JanusB-092015 achieved the best verification results, and JanusD-071715 the best identification results[1]. From the ROC and CMC scores, we see that All-in-One CNN achieves good performances for face identification/verification tasks. This can be attributed to the fact that the DCNN model does capture face variations over various face dataset and generalizes well to a new dataset. In addition, the proposed approach achieves better and comparable face identification/verification than without applying any fine-tuning procedures using the training dataset as Chen (et al) did to boost their performances[4, 5]. We conjecture that with better-detected face bounding boxes and fiducial points from All-in-One CNN, we can reduce the false alarms caused by face detection and perform better face alignment to mitigate the domain shift between the training and test set. On the other hand, TP adapted the one-shot similarity framework[33] with linear support vector machine for set-based face verification and trained the metric on-the-fly with the help of a pre-selected negative set during testing. Although TP achieved significantly better results than other approaches, it takes more time during testing than the proposed method since our metric is trained on-line and requires much less time for testing than TP.

Run Time

We implemented our All-in-One network on a machine with eight CPU cores and GTX TITAN-X GPU. It takes an average of 3.5s to process an image. The major bottleneck for speed is the process of generating region proposals and passing each of them through the CNN. The second stage of our method takes merely 0.1s of computation time. We are currently working on faster implementations of face detection algorithms.

Conclusion

In this article, we presented a multi-task CNN-based method for simultaneous face detection, face alignment,

IJB-A-Verif	[31]	JanusB [1]	JanusD [1]	DCNN$_{bl}$ [24]	NAN [34]	DCNN$_{3d}$ [16]
FAR=1e-3	0.514	0.65	0.49	-	0.785	0.725
FAR=1e-2	0.732	0.826	0.71	-	0.897	0.886
FAR=1e-1	0.895	0.932	0.89	-	0.959	-
IJB-A-Ident	**[31]**	**JanusB [1]**	**JanusD [1]**	**DCNN$_{bl}$ [24]**	**NAN[34]**	**DCNN$_{3d}$ [16]**
Rank-1	0.820	0.87	0.88	0.895	-	0.906
Rank-10	-	0.95	0.97	-	-	0.977
IJB-A-Verif	**DCNN$_{pose}$ [2]**	**DCNN$_{fusion}$[5]**	**DCNN$_{tpe}$[25]**	**DCNN$_{ours}$**	**DCNN$_{ours+tpe}$**	**TP [7]**
FAR=1e-3	-	0.76	0.813	0.787	**0.823**	-
FAR=1e-2	0.787	0.889	0.9	0.893	0.922	**0.939**
FAR=1e-1	0.911	0.968	0.964	0.968	**0.976**	-
IJB-A-Ident	**DCNN$_{pose}$ [2]**	**DCNN$_{fusion}$[5]**	**DCNN$_{tpe}$[25]**	**DCNN$_{ours}$**	**DCNN$_{ours+tpe}$**	**TP [7]**
Rank-1	0.846	0.942	0.932	0.941	**0.947**	0.928
Rank-10	0.947	0.988	0.977	0.988	**0.988**	0.986

Table 2 | Results on the IJB-A dataset. The TAR of all the approaches at FAR=0.1, 0.01 and 0.001 for the ROC curves (IJB-A 1:1 verification). The Rank-1, Rank-5, and Rank-10 retrieval accuracies of the CMC curves (IJB-A 1:N identification). We report average and standard deviation of the 10 splits for all-in-one CNN, DCNN$_{ours}$ and after Triplet Probabilistic Embedding, DCNN$_{ours+tpe}$. All the performance results reported in [1], JanusB (JanusB-092015), Janus D (JanusD-071715), DCNN$_{bl}$ [24], DCNN$_{3d}$, DCNN$_{fusion}$ [5], [16], NAN [34], DCNN$_{pose}$ [2], DCNN$_{tpe}$ [25], and TP [7].

pose estimation, gender and smile classification, age estimation and face verification and recognition. This work demonstrates that subject-independent tasks benefit from multi-task learning and network initialization from face recognition task. Experimental results demonstrate that the performance of the proposed system on the IJB-A dataset is comparable to other state-of-the-art approaches and much better than COTS and GOTS matchers. This clearly suggests that MTL helps in learning robust feature descriptors.

Given sufficient number of annotated data and GPUs, DCNNs have been shown to yield impressive performance improvements. Still many issues remain to be addressed to make the DCNN-based systems robust and practical, such as reducing reliance on large training data sets, handling data bias and degradation in training data, incorporating domain knowledge, reducing the training time when the network goes deeper and wider, and building the theoretical foundations to understand the characteristics and behaviors of DCNN models[5].

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Prof. Rama Chellappa *is a Distinguished University Professor and a Minta Martin Professor in Engineering at the University of Maryland. He is also affiliated with the Center for Automation Research and UMIACS (Permanent Member) and is serving as the Chair of the ECE department. His current research interests span many areas in image processing, computer vision and pattern recognition. Prof. Chellappa is a recipient of an NSF Presidential Young Investigator Award and four IBM Faculty Development Awards. He received the K.S. Fu Prize from the International Association of Pattern Recognition (IAPR). He is a recipient of the Society, Technical Achievement and Meritorious Service Awards from the IEEE Signal Processing Society and the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society. At UMD, he received college and university level recognitions for research, teaching, innovation and mentoring of undergraduate students. He was recognized as an Outstanding ECE by Purdue University and as a Distinguished Alumni by the Indian Institute of Science. Recently, he received the inaugural Leadership Award from the IEEE Biometrics Council. He is a Fellow of IEEE, IAPR, OSA, AAAS, ACM, AAAI and holds six patents.*

Rajeev Ranjan *is a Research Assistant at University of Maryland. He is pursuing Ph.D. from the department of Electrical and Computer Engineering UMD, under the supervision of Prof. Rama Chellappa. His current research interests include face detection and recognition, computer vision, deep learning and statistical pattern recognition. He is a recipient of UMD Outstanding Invention of the Year award, 2015 in the category of Information Science. He received the Jimmy Lin Award for Invention from the ECE department UMD. He is a recipient of A. James Clark School of Engineering Distinguished Graduate Fellowship 2014. He obtained his Bachelor's degree in Electronics and Electrical Communication Engineering from Indian Institute of Technology Kharagpur, India in 2012. He has authored several research papers in reputed conferences and won prizes in many industrial design contests. He worked as a research intern at NVIDIA in 2016, and as an engineer in Cisco Systems, Bangalore during 2012–2014. He was a visiting research student at University of Western Ontario, London in 2011 and at Raman Research Institute, Bangalore in 2010.*

Swami Sankaranarayanan *is a Ph.D. candidate in the ECE department at the University of Maryland, advised by Prof. Chellappa. Before beginning his Ph.D. study, he completed his B.Tech from National Institute of Technology, Trichy in India and Masters in Computer Science from Delft University of Technology in the Netherlands. He has been a research intern at several places including the computer vision labs at INRIA, Sophia Antipolis in France and GE Global Research, Niskayuna NY. His current research interests lie in Face Analysis, Deep Learning and Adversarial Learning. He is affiliated as a student member to the IEEE Computer Science Society and the Society for Industrial and Applied Mathematics (SIAM).*

Jun-Cheng Chen *(Ph.D., UMD, 2016) is a postdoctoral research fellow at the University of Maryland Institute for Advanced Computer Studies (UMIACS). Prior to that, he pursued Ph.D. under the supervision of Dr. Rama Chellappa. His current research interests include computer vision and machine learning with applications to face recognition and facial analysis. He was a recipient of ACM Multimedia best technical full paper award, 2006.*

Carlos D. Castillo *(Ph.D., UMD, 2012) is a postdoctoral research associate at the University of Maryland Institute for Advanced Computer Studies (UMIACS). Prior to that, he pursued a Ph.D. under the supervision of Dr. David Jacobs. He was recipient of the best paper award at the International Conference on Biometrics: Theory, Applications and Systems (BTAS) 2016. His current research interests include face detection and recognition, stereo matching and machine learning.*

References

1. National institute of standards and technology (NIST): IARPA Janus benchmark-a performance report.
2. W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassne, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajana, R. Nevatia, and G. Medioni. Face recognition using deep multi-pose representations. In IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.
3. D. Chen, X. D. Cao, F.Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In IEEE Conference on Computer Vision and Pattern Recognition, 2013.
4. J.-C. Chen, R. Ranjan, A. Kumar, C.-H. Chen, V. M. Patel, and R. Chellappa. An end-to-end system for unconstrained face verification with deep convolutional neural networks. In IEEE International Conference on Computer Vision Workshop on ChaLearn Looking at People, pages 118{126, 2015.
5. J.-C. Chen, R. Ranjan, S. Sankaranarayanan, A. Kumar, C.-H. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa. An end-to-end system for unconstrained face verification with deep convolutional neural networks. CoRR, abs/1605.02686, 2016.
6. J.-C. Chen, S. Sankaranarayanan, V. M. Patel, and R. Chellappa. Unconstrained face verification using Fisher vectors computed from frontalized faces. In IEEE International Conference on Biometrics: Theory, Applications and Systems, 2015.
7. N. Crosswhite, J. Byrne, O. M. Parkhi, C. Stau_er, Q. Cao, and A. Zisserman. Template adaptation for face verification and identification. arXiv preprint arXiv:1603.03958, 2016.
8. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pages 580{587, 2014.

9. G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition, 2008.

10. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Ca_e: Convolutional architecture for fast feature embedding. In ACM International Conference on Multimedia, pages 675{678, 2014.

11. B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In IEEE Conference on Computer Vision and Pattern Recognition, 2015.

12. M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies, 2011.

13. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classi_cation with deep convolutional neural networks. In Advances in Neural Information Processing Systems, pages 1097{1105, 2012.

14. G. Levi and T. Hassner. Age and gender classi_cation using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops, June 2015.

15. Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In IEEE International Conference on Computer Vision, pages 3730{3738, 2015.

16. I. Masi, A. T. Tran, J. T. Leksut, T. Hassner, and G. Medioni. Do we really need to collect millions of faces for e_ective face recognition? arXiv preprint arXiv:1603.07057, 2016.

17. O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. British Machine Vision Conference, 2015.

18. R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. In IEEE International Conference on Biometrics: Theory, Applications and Systems, 2015.

19. R. Ranjan, V. M. Patel, and R. Chellappa. HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition, March 2016.

20. R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa. An all-in-one convolutional neural network for face analysis. arXiv preprint arXiv:1611.00851, 2016.

21. R. Ranjan, S. Zhou, J-C. Chen, A. Kumar, A. Alavi, V. M. Patel, and R. Chellappa. Unconstrained age estimation with deep convolutional neural networks. In IEEE International Conference on Computer Vision (ICCV) workshop on ChaLearn Looking at People (ChaLearn LaP), 2016.

22. K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In International Conference on Automatic Face and Gesture Recognition, pages 341{345, April 2006.

23. R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In IEEE International Conference on Computer Vision Workshop on ChaLearn Looking at People, pages 10{15, 2015.

24. A. RoyChowdhury, T.-Y. Lin, S. Maji, and E. Learned-Miller. One-to-many face recognition with bilinear cnns. In IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

25. S. Sankaranarayanan, A. Alavi, C. Castillo, and R. Chellappa. Triplet probabilistic embedding for face veri_cation and clustering. arXiv preprint arXiv:1604.05417, 2016.

26. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. arXiv preprint arXiv:1503.03832, 2015.

27. K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In British Machine Vision Conference, volume 1, page 7, 2013.

28. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. arXiv preprint arXiv:1409.4842, 2014.

29. J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154{171, 2013.

30. P. Viola and M. J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137{154, 2004.

31. D. Wang, C. Otto, and A. K. Jain. Face search at scale: 80 million gallery. arXiv preprint arXiv:1507.07242, 2015.

32. L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 529{534. IEEE, 2011.

33. L. Wolf, T. Hassner, and Y. Taigman. The one-shot similarity kernel. In International Conference on Computer Vision, pages 897{902. IEEE, 2009.

34. J. Yang, P. Ren, D. Chen, F. Wen, H. Li, and G. Hua. Neural aggregation network for video face recognition. arXiv preprint arXiv:1603.05474, 2016.

35. D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. arXiv preprint arXiv:1411.7923, 2014.

36. W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. ACM Computing Surveys, 35(4):399{458, 2003.

Human Biometric System Interaction (HBSI) — A Complementary Approach to Examining Biometric Performance

By Stephen Elliott

Biometrics are the ability to recognize individuals based on their physiological or behavioral characteristics, such as fingerprints, face, iris, and voice[1]. However, this definition belies the nature of that interaction with the individual, and the purposes for that interaction. While the biometric component is the ability to recognize an individual, in many cases that component is functioning in a much broader system of systems, whose purpose it is to carry out the authentication, security, admission, and other activities. It is the broad-based nature of biometrics — the fact that authentication exists on some platforms, from on the phone (fingerprint, iris, voice, multi-factor authentication), to large-scale customs and border protection (CBP) applications that make the development of biometrics such an intriguing story for the better part of 20 years.

That said, understanding and counteracting the limitations of such systems in the context of their deployment is less well understood, and that is, in part, due to the limitations of the traditional metrics that we use to evaluate the performance of the system. Biometric performance metrics examine performance in a series of trade-offs — typically, but not uniformly, as a compromise between false accepts rates (FAR) and false reject rates (FRR). These metrics alongside the operational threshold make the determination of an established set of performance metrics for the biometric system. We use these metrics to determine, in some part, the success or failure of the biometric system, or to compare one biometric system with another.

Challenges in a Biometric System of Systems

Over the past 15 years, researchers at the International Center for Biometric Research (ICBR) at Purdue University in collaboration with other institutions, have examined the role of the interaction of the human with that biometric system, within the context of the larger system in which the biometric component operates in. Moreover, as systems become more complex, the interaction has evolved into a framework of interaction models that provide additional information in determining the performance of the biometric system, in the context of that operation. As such, these metrics complement the traditional biometric performance metrics. Others too, have also examined the role of the user and their interaction with the biometric sensor. Researchers at the National Institute of Standards and Technology (NIST) have made several contributions in this field[2-9], as has Coventry who first discussed the usability of biometrics within the context of an ATM[10], followed by two other articles[11,12].

At the most fundamental level, the user in an overt biometric system will interact with a biometric sensor, such as a fingerprint sensor, and either do it correctly or incorrectly[14]. This determination of a correct or incorrect presentation is impacted by several factors that include how the individual presents their fingerprint to the sensor, their ability to concentrate during the task, and the environment in which they are doing the task. The determination of what causes the correct or incorrect presentation is also of interest to integrators of biometric systems. Take the scenario of a CBP control booth for immigration purposes, where several biometric modalities are deployed — fingerprint, face, and in this example, iris. The success of the biometric system in this case relies on the appropriate level of quality of the biometric being presented, which directly impacts the performance. In this scenario, the user must interact with more than just the biometric component; they have to communicate with the border control officer, present their passports, completed paperwork and answer questions. Those additional interactions drive the performance of the overall biometric system as well. Alongside the border, the booth scenario is the automated border security gate (ABC), which relies on the individual interacting with the system without prompting. When comparing the performance of these two systems, additional contextual information has to be provided for a valid assessment to be completed.

> “Thus, the interaction of one individual with one biometric sensor has now evolved into the interaction of many individuals with many sensors in a much larger system, which becomes quite complicated...”

Thus, the interaction of one individual with one biometric sensor has now evolved into the interaction of many individuals with many sensors in a much larger system, which becomes quite complicated to disassemble the various attributes that can be impacting the biometric system performance.

Interaction Framework

The Human Biometric Sensor Interaction (HBSI) framework was born out of this need to understand the impact of the human within a biometric system. Initial studies nearly a decade ago were based on a standalone fingerprint sensor, whose interaction was a swipe across a very thin silicon sensor, at a uniform speed. The research team at the ICBR noticed various issues with interaction — and were not the first group to do this. As mentioned, others have also examined the role of the user and their interaction with the biometric sensor. Researchers at NIST as well as Coventry made early contribution to this field[2-9], specifically related to the usability of a biometric system within the context of customer acceptance[13]. Customer acceptance can also be related to the HBSI metrics. However, at Purdue, issues of interaction across a broad range of modalities, and classified particular aspects of the interaction into correct, or incorrect, presentations have been examined.

If you take the border control scenario again and examine just one biometric modality — iris recognition, the correct presentation would be if the user is standing in the appropriate location in the volume, and looking at the camera. An incorrect presentation will be if the user is not watching the camera for whatever reason. Automatic classification of this interaction enables a series of metrics to be collected that quantify the potential errors or issues within that iris subsystem. This example illustrates the first level of the framework and comprises the following definitions. For incorrect presentations, a defective interaction (DI) occurs when a user makes an incorrect presentation that is not detected by the biometric system[15]. In some cases, the biometric system does not recognize that the user is trying to present a biometric sample. This can occur in an iris collection environment if the user is looking in the wrong direction so that the system does not detect their eyes.

A concealed interaction (CI) occurs when an incorrect presentation is detected by the biometric system but is not classified correctly as an error[15]. A CI is a mistake that, despite being caused by the user, is accepted into the biometric system as a successfully processed sample. CI's are a critical error because they should be rejected by the biometric system but are not. An example of a CI would be in fingerprint recognition if the user is supposed to place their right index finger but uses their left index finger instead. The system will accept the sample as long as minimum quality and minutiae are met but cannot differentiate between the two fingers.

A false interaction (FI) is an incorrect presentation that is detected by the biometric system, but unlike a CI, is correctly handled as an error[15]. In an FI, the biometric system is performing as expected and rejecting the incorrect presentation from being processed. Upon detection of an FI, the biometric system may respond to the user with feedback such as an error message or allow for a retry.

From the perspective of a correct presentation, the framework returns the following metrics — a failure to detect (FTD) is a proper presentation made by the user that is not detected by the biometric system[15]. The result of an FTD is the same as a DI, but in this case, the fault lies in the biometric system, rather than the user. The user will have correctly presented their biometric sample but the system does not detect it, and the state will remain unchanged. An example of this is in fingerprint recognition if the user successfully places their right index finger but due to the system error, it does not detect that any placement has occurred.

A failure to process (FTP) is a correct presentation made to the biometric system that encounters an error when the system processes it. Due to this processing error, the biometric template is not created, and the sample is not saved to the database. Reasons for this error include system processes such as segmentation, feature extraction, or quality control[15]. This occurs in fingerprint recognition if the system requires a certain number of minutiae points or a quality level to be accepted. Although the biometric

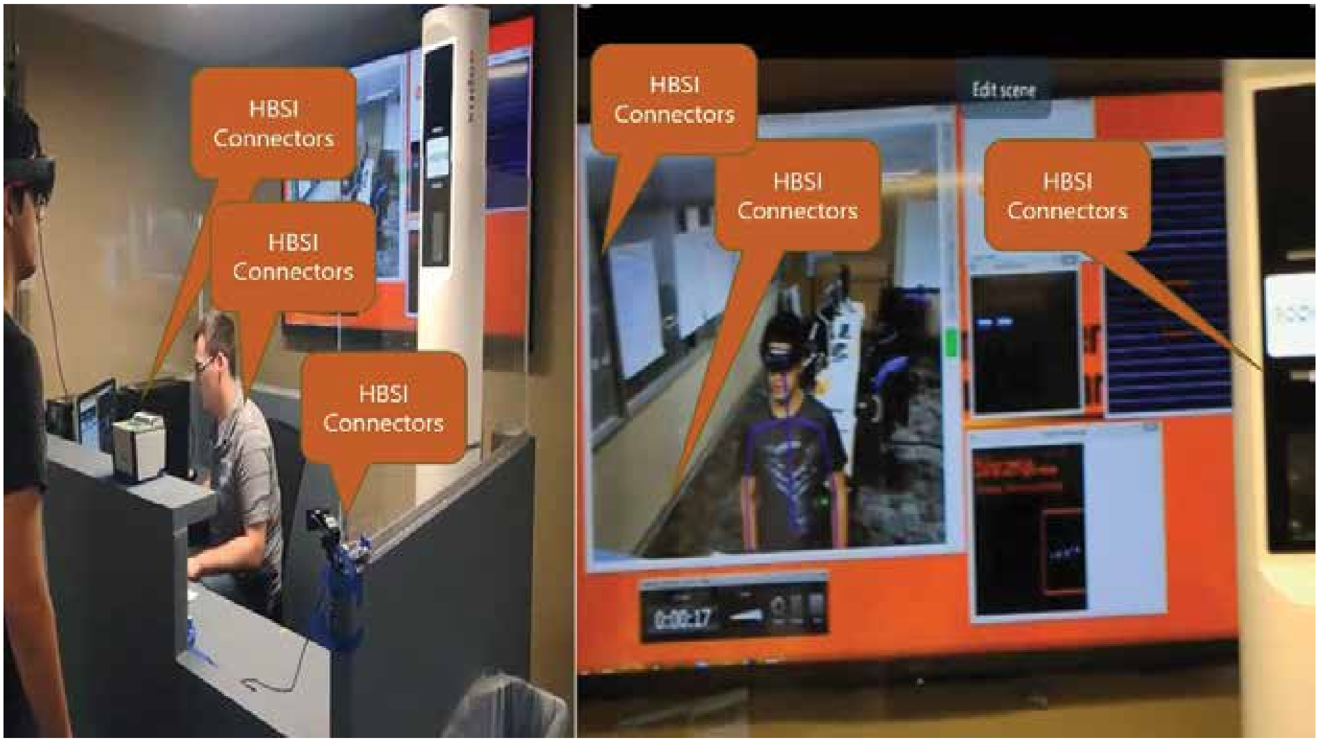


Figure 1 | Framework of the HBSI Connectors

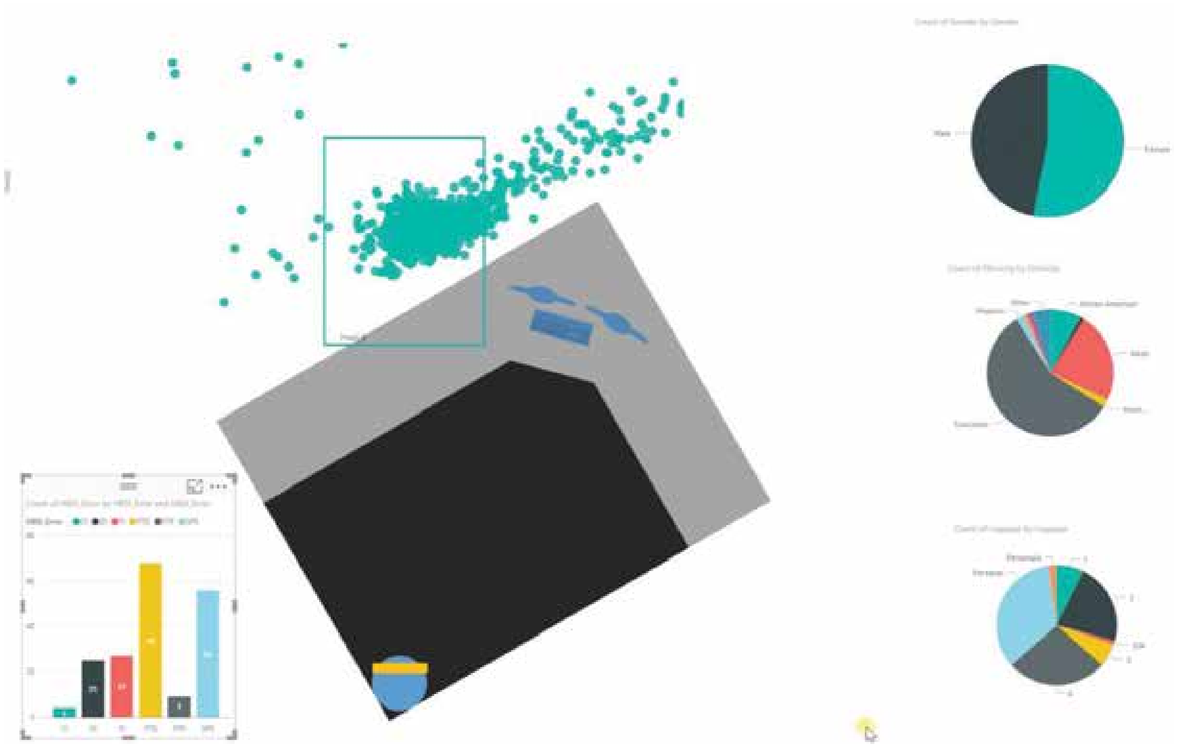


Figure 2 | Bird's eye view of dashboard.

sample was presented correctly, a characteristic of the fingerprint such as age or temperature does not meet the system's tolerances and is rejected.

A successfully processed sample (SPS) is a correct presentation that is detected by the biometric system and successfully processed as a biometric sample. The biometric sample meets system specifications, allowing for the template to be created or the sample to be saved to the database. An example of an SPS occurs in fingerprint recognition when a user correctly places their right index finger which meets the biometric system's requirements and is subsequently saved to the database. The SPS rate is calculated by the total number of SPSs divided by the total number of attempts.

These terms also work for behavioral biometrics, even though behavioral biometrics like signature and voice add in continuous streams of data — how long you sign or intonation of your voice. Researchers at the University of Kent, Canterbury contributed to the development of the behavioral model. As biometrics exist in a larger system — for example, border control, the token HBSI model was developed to take into consideration the passport. In the original model, we make the assumption that the user was a genuine actor, and in reality, this is not always the case. Thus, an additional framework was constructed for the impostor. This structure morphs from the particular interaction task of the user and the biometric system to that of the entire system in which the biometric system is performing and impacted in.

Moreover, in building the framework, the ICBR has constructed connectors that provide additional context to the scenario. Such sensors include environmental data from IoT sensors, system data from previous interactions (such as image quality, interaction data, user behavior, and characteristics) to provide more contextual information so that the system can assess anomalies in the performance of the biometric system. An example of the integration is shown in Figure 1 — where the connectors are illustrated across the multi-modal biometric system.

Furthermore, these connectors provide real-time analytics to dashboards so that individuals can be assessed across a broad range of metrics. In the scenario above, the ICBR has integrated the following sensors — environmental sensors; body posture sensors using the Kinect; throughput and timing sensors, biometric devices including 10-print, iris and face; as well as audio recordings including conversational feedback using commercially available cognitive services that breakout cognitive insights, speech sentiment (negative, positive and neutral) of each speaker, biometric image quality (ISO/IEC JTC 1 Image Quality Standards for Face, Iris, and Fingerprint), augmented with face emotion algorithms (neutral, happiness, sadness, anger, disgust, fear and surprise). Figure 2 on the previous page is an image of the dashboard illustrating the booth from bird's eye view, with gender, ethnicity demographics, and in this case luggage characteristics (rollaway, shoulder bag), and on the bottom right-hand side, the human biometric sensor interaction errors. All of these are automatically classified as individuals enter the booth and start a transaction.

Therefore, this framework highlights the integration of many different sensors and algorithms to provide better context in the performance of the biometric system. Q

__Dr. Stephen Elliott__ is an associate professor with an appointment in Technology, Leadership and Innovation at Purdue University, where he has been a member of the faculty since 2001. He is the Director of the International Center for Biometric Research (ICBR). Dr. Elliott serves on advisory and standards boards and has been honored for his teaching, research, and work in the voluntary consensus standards. He has published 100 articles on biometrics, ranging from chapters, journal articles, and conference proceedings, and has given numerous talks on biometric research. Dr. Elliott is also active in teaching and learning activities at Purdue University. He received many awards, including the INCITS Chairman's Award in 2011, IEC Young Professional in 2011, and outstanding tenured faculty member in 2010. Dr. Elliott has also co-edited some international and national standards and was a U.S. delegate to ISO/IEC SC 37 Biometric Standards Committee. Dr. Elliott serves on the ANSI Committee on Education, and as a member of the INCITS Executive Board.

Reference

1. E. Kukula and S. Elliott, "Implementing ergonomic principles in a biometric system: a look at the human biometric sensor interaction (HBSI)," *Carnahan Conf. Secur. …*, pp. 86–91, 2006.
2. R. Micheals, B. Stanton, M. Theofanos, and S. Orandi, "A Taxonomy of Definitions for Usability Studies in Biometrics." NIST, Gaithersburg, p. 9, 2006.
3. M. Theofanos, S. Orandi, R. Micheals, B. Stanton, and N. Zhang, "Effects of Scanner Height on Fingerprint Capture." National Institute of Standards and Technology, Gaithersburg, p. 58, 2006.
4. M. Theofanos, B. Stanton, R. Micheals, and S. Orandi, *Biometric Systematic Uncertainty and the User*. Ieee, 2007, pp. 1–6.
5. M. Theofanos, B. Stanton, S. Orandi, R. Micheals, and N.-F. Zhang, "NISTIR 7403 Usability Testing of Ten-Print Fingerprint Capture NISTIR 7403 Usability Testing of Ten-Print Fingerprint Capture," Gaithersburg, MD, 2007.
6. M. Theofanos, R. Micheals, J. Scholtz, E. Morse, and P. May, "Does habituation affect fingerprint quality?," in *Conference on Human Factors in Computing Systems: CHI '06*, 2006, pp. 1427–1432.
7. M. F. Theofanos, B. Stanton, Y. Choong, and R. Micheals, "Usability Testing of an Overlay to Improve Face Capture," *Int. Organ.*, 2009.
8. M. Theofanos, B. Stanton, C. Sheppard, R. Michels, J. Libert, and S. Orandi, "NISTIR 7540 - Assessing Face Acquisition," National Institute of Standards and Technology, Gaithersburg, MD, Sep. 2008.
9. M. Theofanos, B. Stanton, S. Orandi, R. Micheals, and N. Zhang, "Usability Testing of Ten-Print Fingerprint Capture." NIST, Gaithersburg, p. 56, 2007.
10. L. Coventry, A. De Angeli, G. Johnson, and P. McCabe, "Biometric Verification at a Self Service Interface," in *Proceedings of the British Ergonomic Society Conference*, 2003, pp. 247–252.
11. L. Coventry, G. I. Johnson, T. McEwan, and C. Riley, "Biometrics in Practice: What Does HCI Have to Say?," *HumanComputer Interact. – INTERACT 2009*, vol. 5727/2009, pp. 920–921, 2009.
12. L. Coventry, A. De Angeli, and G. Johnson, "Usability and biometric verification at the ATM interface," *Proc. SIGCHI Conf. Hum. factors Comput. Syst.*, pp. 153–160, 2003.
13. A. S. Patrick, "Usability and Acceptability of Biometric Security Systems," Ottawa, Canada.
14. M. Brockly, R. Guest, S. Elliott, and J. Scott, "Dynamic Signature Verification and the Human Biometric Sensor Interaction Model," in *45th annual IEEE International Carnahan Conference on Security Technology*, 2011, pp. 253–258.
15. S. J. Elliott and E. P. Kukula, "A Definitional Framework for the Human-Biometric Sensor Interaction Model," 2010, p. 76670H–76670H–8.

Perfect, Just The Way You Are

By Ingo Deutschmann and Neil Costigan

As humans, we are taught that we are the weakest link when it comes to IT security. Many attempts have been made to remove the human factor from the security equation, but no one has succeeded. If we look at the security we're used to in our devices and services, it is based on thinking from the 1970s, where a binary 'yes' or 'no' at login made more sense. In our always-on culture, that kind of thinking is no longer adequate. Equally, adding extra steps can be a good way to boost security, but also gets in the way of user experience. It is ironic, then, that the human factor, the so-called 'weakest link' can be the solution to the security challenge, simply by humans behaving normally.

Big Data Machine Learning Biometrics

To find the beginning of the story we need to travel all the way to the Arctic Circle in northern Sweden to Luleå University of Technology. In 2006, an undergraduate behavioral biometrics project, with help from the university's innovation team resulted in three students founding the spin-out, BehavioSec. The idea was uncomplicated, while the technology, the algorithm, was not. Would normal end user interaction with a device or keyborad be enough to verify the identity of a human being? Are we that unique?

Behavioral Biometrics

Human gestures can be repeated in ways that may look similar to the naked eye, however when they are measured by a behavioral algorithm, they look totally distinct. The way a person holds, swipes, or types on a screen or keyboard is a source of data for user authentication and verification. Behavioral biometrics technology doesn't measure just one gesture, but a whole range of data inputs, with a high level of accuracy and precision, and can do so throughout a user session. This new capability, to be able to continuously authenticate an end user, not just at login, is intriguing to a wide range of organizations, as they see a solution that can protect against account takeover, identity theft, and even internal fraud.

Behavioral biometrics is the measurement of human behavior to verify the identity of a person.

BehavioSec collects data from a wide range of behaviors and actions, such as key flight, accelerometer, gyroscope, hit zone, and others. The combination of these data sources creates a sophisticated AI that has proven to be commercially viable.

Different types of sensors provide different, but complementary mechanisms to profile a user:

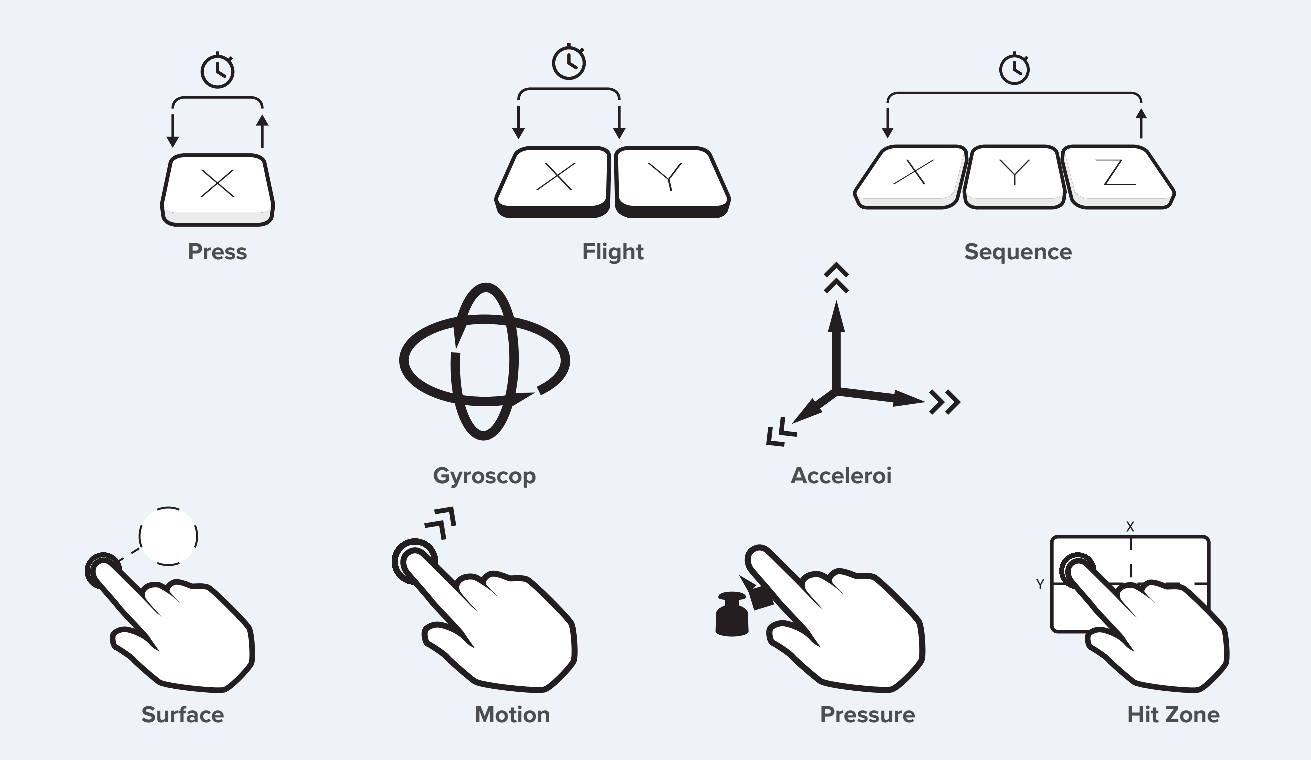


Figure 1 | BehavioSec collects data from a wide range of behaviors and actions, such as key flight, accelerometer, gyroscope, hit zone, and others. The combination of these data sources creates a sophisticated AI that has proven to be commercially viable.

- Keyboard entry timings look at very accurate measurements of when a key was pressed, when it was released, how long between pressing each key and the sequence in which they were pressed
- Mouse provides sequential 2D coordinates and thus a speed of deflection and movement pattern
- Gyroscope and Acclerometer provide 3D coordinate sequences, and thus velocities and patterns for the actual movement of a mobile device
- Surfaces provide the coordinates and measurements for 2D movements, as well as detailed and valuable information on how much pressure a user exerts, and exactly where they press.

Learning by Doing

BehavioSec's first set of customers came from commercial banking and implemented the solution into their online banking sites, primarily for Fraud Forensics. It gave them the ability to, in real-time, look at a specific session and see behavior profiling down to each keyboard stroke or mouse movement. One early lesson was that the mindset of security is that it is all about replacing the old, rather than designing the new. Behavioral biometrics is disrupting the way to look at the authentication process, but the prospects wanted a solution to replace the old non-functional silver bullet, "the password".

Even though the mindset is changing to favor the new risk-based approach, countless hours have been spent on educating senior security executives where the industry is going.

DARPA

In 2011 DARPA published a Broad Agency Announcement for research in Active Authentication, novel ways of validating the identity of the person at the console that focus on the unique aspects of the individual. BehavioSec was chosen as the only non-U.S. vendor to participate in the succesful project to continuously verify the end user with behavioral biometrics.

The early success with banks in Northern Europe was based on the BehavioSec machine-learning algorithm where the the big data number crunching was executed on a server. BehavioSec proposed the technical challenge to DARPA to re-engineer the solution to work offline on the device itself, an autonomous AI. Since this initial DARPA project BehavioSec has proposed and delivered the same solution not only in desktop environments but in mobile and handset devices and a new contract will continue in 2017 to explore new ideas in behavioral biometrics. One factor that has worked in our favor is Moore's law; it has been kind to us as the computing power on a smartphone has skyrocketed.

The Power of Choice

The modern end user of today has high expectations on user friendliness, and they know that they are in a power position to get what they want. Whenever end users are offered a choice they will act with brutal decisiveness: One BehavioSec client operates as an identity provider for banks with a combined user base of 7 million. When they started offering strong authentication with a mobile app supported by our behavioral biometrics technology, they saw an exponential growth in usage from 3-4 transactions a month to 20-25.

This highlights the potential for user experience successes and how the disruption of financial services is already in progress.

Risk Based Authentication

Product, customer, and end-user experience teams are continuously working to decrease friction in order to meet the high expectations of busy, multi-tasking users. Adaptive, dynamic, layered security helps you to create authentication processes that align with these expectations.

Fewer than 30 percent of us log out of our accounts when we're finished using a service. Our mobile apps are especially vulnerable, now that social media services also act as identity providers, and will soon be entering the payments space. Security needs aren't all the same, even within these individual services. For example, checking your bank balance is not as risky as carrying out a large transfer or changing account details.

The Right Level of Security at the Right Time

BehavioSec analyzes every session from start to finish, continuously profiling behavioral patterns. The system creates a profile match score based on a range of factors by comparing it to stored results. Is this person typing as they normally do? Are they in a recognized location, using their usual device? This is monitored throughout the session so that security is an ongoing process, not merely a step.

> "Product, customer and end-user experience teams are continuously working to decrease friction, in order to meet the high expectations of busy, multi-tasking users."

From Yes or No to "If This then That"

As a user interacts during a session, the similarity score is fed into your risk engine, and your security or fraud team determines what happens next. If the score is high, the system allows the user through. If it's not high enough, that's when you can add further steps using the other layers in your system. If the score is very low, your system can log the user out completely.

If Not You, then Who?

BehavioSec already has proven to successfully verify that the right person requested access. The holy grail of fraud prevention is to be able to transform end user behavior to narrow down the group of known users who are the prime suspect for a potential fraud. This is accomplished by efficient machine learning capabilities and applying artificial intelligence to user profiles. Our user profile's level of sophistication enables BehavioSec to find the needle in the haystack.

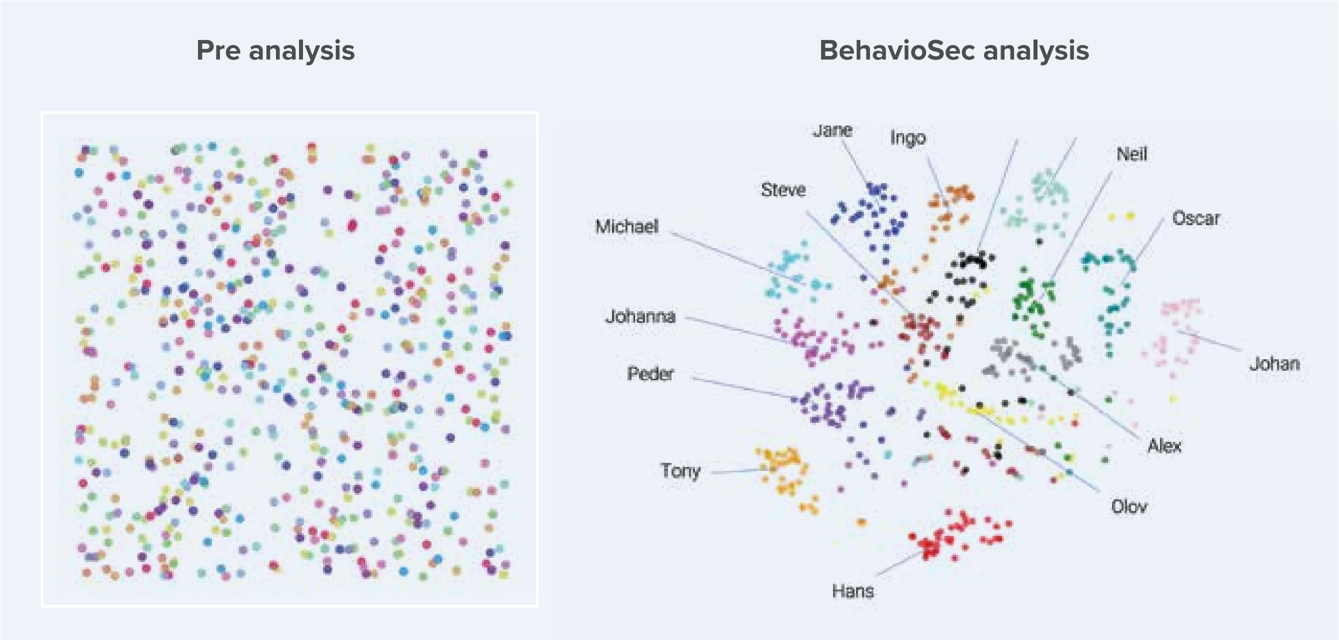
How to Make an Impact

The BehavioSec R&D unit has stayed in Luleå and is now a neighbor to Facebook's first European datacenter. The remote location has the potential to be a limiting factor on innovation and knowledge transfer, but for BehavioSec it has worked in our favor.

As the solution is software only and the technology is based on standards, BehavioSec has become an API company by heart and today the Luleå R&D hub is running projects and implementations across the globe.

Entering Mainstream

Apart from the research completed with DARPA, the early adopters of our technology are the commercial



Pre analysis

BehavioSec analysis

Figure 2 | The BehavioSec scatter plots above show before and after BehavioSec algorithm analasis. Left: clusters of processed end user data from 15 people typing one same password where each dot represents one session. The end user behavior profile cluster is a result of a transformation of 22 dimensions that is simplified into 2 dimensions.

banks that have the challenge of meeting the needs of their services towards an 'always-on' paradigm. This innovative trend was first embraced by financial technology providers ("FInTech"), and is now being followed by other verticals in a movement independent of geographical location.

Behavioral biometrics will soon be mainstream and utilized by most service providers online. BehavioSec is conducting pilots at handset manufactures, payments providers, e-learning platforms, customer relationship managers (CRM's) and others; the list is growing by the week. The thing that these verticals have understood is that the success of a service is through the ambition of their end-users to have an efficient customer journey. Security has to be smart by design, and recognize differences in your normal user behavior: It should be able to learn that you are perfect, just the way you are! **Q**

Dr. Neil Costigan *is Principal Investigator for DARPA BAA 12-06 Phase 1 and BAA 13-12 Phase 2. Dr. Constigan is a software development professional with over 15 years' experience. Dr. Costigan is CEO at BehavioSec. Former background as VP R&D for Security products at Gemplus, will add into the team his knowledge in security software development as well as his experience from the Swedish start-up company Celo Communications. Neil also adds his academic credibility as Ph.D. in computer security from the University of Dublin.*

Ingo Deutschmann *is a security professional with more than 15 years' experience in development, consulting and product services. Deutschmann is the Business Development Director DACH at BehavioSec. Former background as General Manager Germany at Gemplus, he will add to the team his knowledge in security software development as well as his experience from the Swedish start-up company Celo Communications and German DEH GmbH, where he was responsible for the R&D operations. Ingo was co-developer of the hardware antivirus solution ExVira. He is a Mathematician from the University of Jena, and holds worldwide patents for a smart card reader.*

From the IQT Portfolio

The *IQT Quarterly* examines trends and advances in technology. IQT has made a number of investments in innovative technologies, and several companies in the IQT portfolio are garnering attention for their unique solutions.

Brainspace™

Brainspace is focused on a singular mission: creating machine learning that accelerates human leaning. Brainspace's revolutionary approach to processing information helps surface insights and avenues of inquiry that are difficult to find with any other solution. Brainspace has been an IQT portfolio company since June 2016.

www.brainspace.com

SNAPDNA

SnapDNA enables DNA analysis in a fraction of the time and with far greater specificity, convenience, accessibility and affordability than current tests. This simple concept is enabled by highly defensible, proprietary technology and demonstrated with some of the most stringent DNA tests. SnapDNA has been an IQT portfolio company since November 2012.

www.snapdna.com

FUEL3D®

Fuel3D is a developer of advanced 3D scanning systems and solutions. Originally developed for the medical imaging sector, Fuel3D technology has been adapted for the broader 3D market, with the goal of bringing the benefits of point-and-shoot 3D imaging to consumers, professionals and businesses. The technology combines photometric stereo imaging with stereoscopic imaging to produce a single 3D image. Fuel 3D has been an IQT portfolio company since December 2014.

www.fuel-3d.com

IQT®

IN·Q·TEL