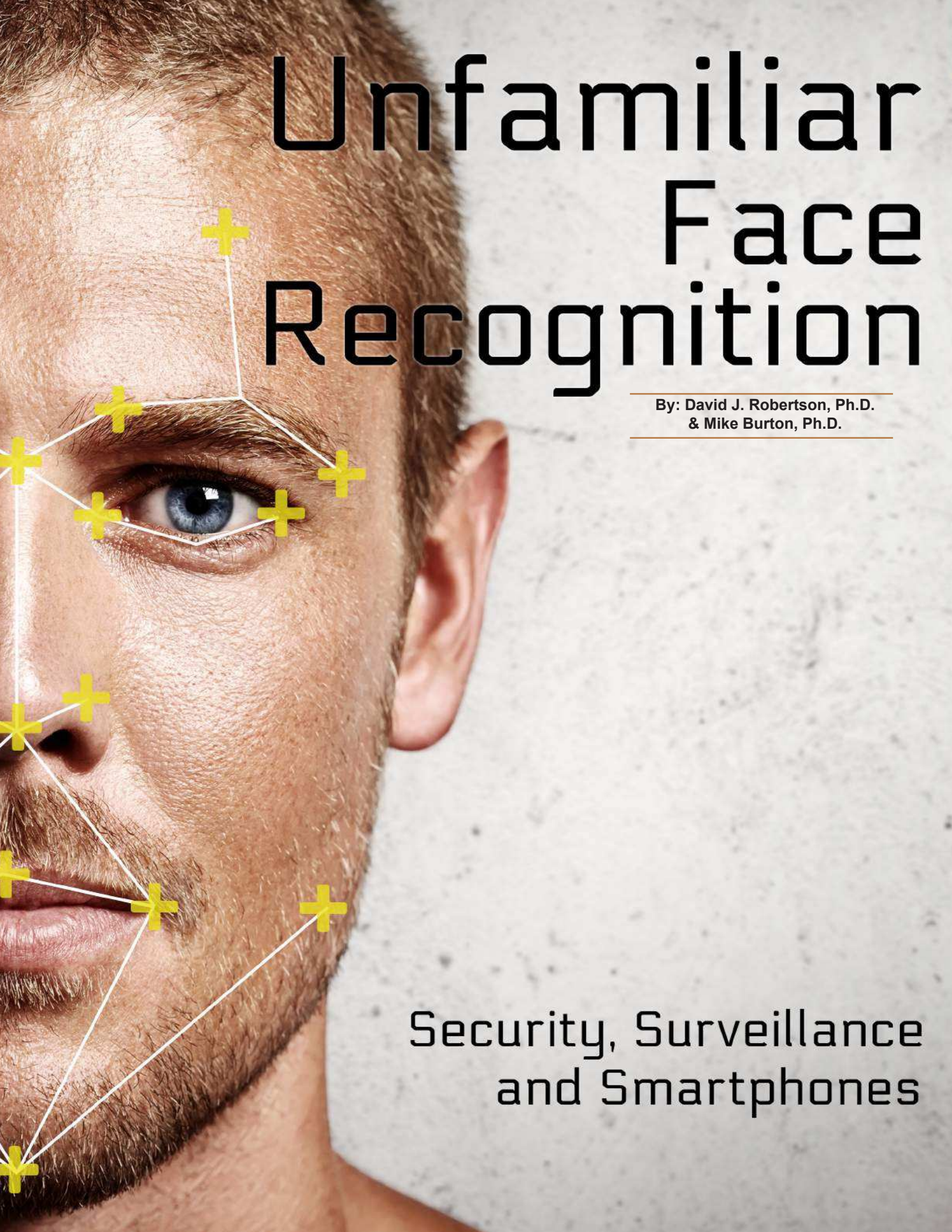

Unfamiliar Face Recognition

By: David J. Robertson, Ph.D.
& Mike Burton, Ph.D.

Security, Surveillance and Smartphones

Familiar Face Recognition

A person's ability to recognize familiar faces across a wide range of viewing conditions is one of the most impressive facets of human cognition. As shown in Figure 1, it is easy to conclude, for a known individual, that each image in the set shows the same person (British Prime Minister David Cameron), despite a wide range of variation in viewing angle, physical appearance, camera and lighting. In fact, familiar face recognition performance is often at or near ceiling level, even when the images are of poor quality [1] or artificially distorted. [2] At first glance, the aptitude for familiar face recognition may suggest a similar level of expertise for the recognition of unfamiliar faces, thus the reliance on face-to-photo ID for identity verification. [3] This is not the case, as recent research shows people are surprisingly poor at recognizing new instances of an unfamiliar person.

The poor recognition of unfamiliar faces is a concern for the United States. Many preliminary screenings involve facial recognition by security agents. In order for this method to be effective, more robust training for security agents needs to be established. The Department of Defense utilizes facial and iris recognition technologies in order to eliminate human error in identifying persons of interest during surveillance operations. [4] DoD guidelines should be implemented by security agent guidance programs to ensure best practices in identification of potential threats.

Unfamiliar Face Recognition: Studies on the General Population

The Glasgow Face Matching Test [5] is a well-established measure of unfamiliar face recognition performance. As illustrated in Figure 2, this simple psychometric test requires participants to decide whether pairs of unfamiliar faces show two instances of the same person (taken seconds apart using different cameras) or two different people. The GFMT captures the real life

Figure 1. Ambient photos of the same face (U.K. Prime Minister David Cameron). All images used under Creative Commons or Open Government License. (Released)

1-1 matching situations encountered on a daily basis by passport control officers (i.e. matching a face to a passport photo) and military personnel (i.e. matching a face image captured from surveillance footage to images held on file). In the GFMT, viewers are not required to remember anything and can take as long as they like to make their decisions.

As seen in Figure 2, the GFMT trials use high quality front-facing images. Despite this, accuracy on this task is generally poor with error rates between 15 and 20 percent being the norm across hundreds of participants tested. Even the lower end of the estimate, 15 percent, is a non-trivial level of error in many circumstances. For example, millions of international passengers pass through airports each day. If this error rate were similar in professional groups (see next section) it would be unacceptably high.

Another striking example of poor performance in a test of unfamiliar face recognition was modeled on an old-fashioned police line-up scenario involving a series of one-to-ten unfamiliar face matching arrays. [6] The task was intended to emulate the best-case scenario for identifying images captured on a security video. As seen in Figure 3(a), participants were presented with a single high quality image of a 'suspect' taken from video footage and an array of 10 face images. Participants decided whether the suspect was present in the array, and if present, attempted to pick out the correct image. Despite using high quality images and video stills taken on the same day, in a similar pose and in optimal lighting conditions, error rates on this task were unacceptably high, at 30 percent on average. Performance dropped further with lower quality surveillance video. [7]

Researchers replicated this level of error using the same task and extended the findings to include the effects of race or ethnicity. [8] Using the same

Figure 2. An example of two trials from the Glasgow Face Matching Test. The top pair shows two instances of the same person (match trial), the bottom pair shows two different people (mismatch trial). (Released)

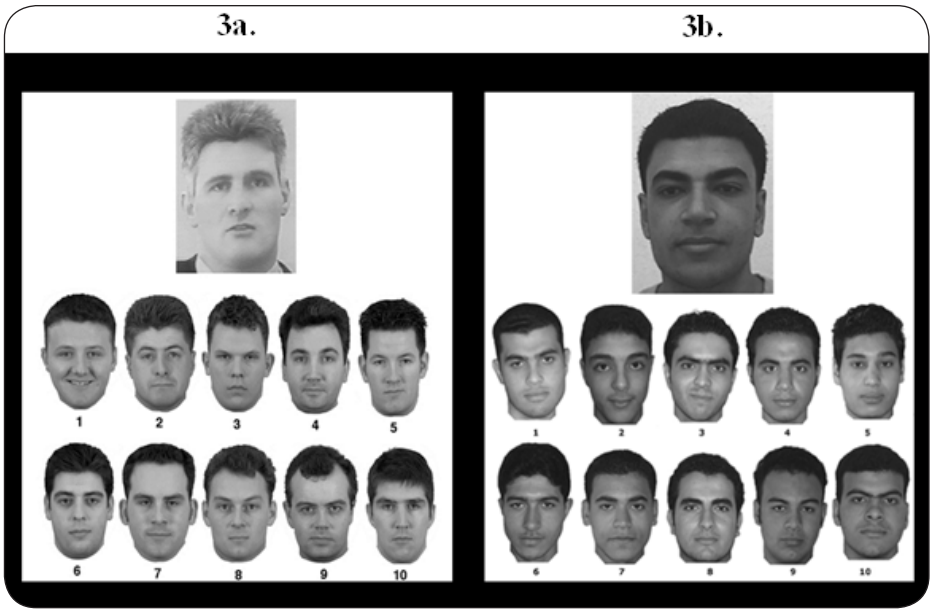

Figure 3(a). An example of a one-to-ten matching trial from Reference 6. In this example, the suspect pictured above the array is not present in the line-up below. Figure 3(b). An example of a one-to-ten matching trial from Reference 8. In this example, the suspect is present in the array, at position number 10. (Released)

matching task, [6] the researchers asked U.K. and Egyptian participants to complete the one-to-ten task using faces of people from the U.K. and from Egypt. [8] When participants performed the recognition test with faces from their own ethnic group (e.g., U.K. participant using U.K. face arrays), error rates were in line with those in earlier studies. [6,8] However, when participants were required to match faces from an opposite ethnic background, (e.g., U.K. participant using Egyptian face arrays; see Figure 3b) error rates rose to 40 percent. This is known as the 'other race effect' and is of particular importance considering a proportion of people attempting passport fraud or under military surveillance are likely to be unfamiliar other race individuals. Hence, accurate recognition of unfamiliar individuals for the purposes of detecting passport ID fraud or for military surveillance may be more difficult when the person is a foreign national.

The above studies provide insight into laboratory tests of unfamiliar face recognition. One criticism of these findings is that they focus on matching face photographs, whereas in the real world accurate matching involves real faces. Research shows, however, that matching a face photo to the face of a live person is just as difficult and error-prone as matching face images. For example, one study reported an error rate of 17 percent when participants were asked to match recent photos to live faces. [9] A later study reported more than 30 percent of participants made identification errors when asked to match recent high quality closed-circuit television images to a live defendant in a courtroom scenario study. [10] Unfamiliar face recognition is just as bad in real life as it is when matching two photos.

Unfamiliar Face Recognition: Studies on Specialists

The laboratory-based studies described above provide important insights into the accuracy of unfamiliar face recognition. However, these studies were performed using non-specialist viewers from the general population. While study results are informative, it is important to know whether people who carry out these tasks professionally are able to perform more accurately than untrained viewers.

A seminal study tested whether the inclusion of a face photo on credit cards would reduce fraud. [11] The study used real supermarket cashiers who routinely check photo-ID cards to prevent the sale of age-restricted goods, such as cigarettes and alcohol. Remarkably, it was reported that the retail staff accepted fraudulent photo credit cards (i.e. the photo did not depict the bearer) as genuine in 50 percent of trials. It is striking that this performance could be so poor, given that matching

Figure 4. Examples of the images used in Reference 1: A still from a CCTV video (left) and a high quality face photo (right). (Released)

faces to photo-ID cards is an important part of the job. While it could be argued that retail stores do not have the resources to provide training in photo-ID verification, this is not a criticism that can be leveled at studies involving police officers or passport officials, described next.

U.K. Police Officers

Researchers tested whether a group of U.K. police officers, with experience in forensic identification, would perform better than a group of untrained university students on a test of unfamiliar face recognition. [1] As seen in Figure 4, participants were required to view low quality video clips of individuals entering a building. They were told that they would later be asked to identify these people. The participants looked at high quality face photos and were asked to rate how confident they were that these individuals had been present in the video clips.

The results of the study showed that the police officers performed very poorly on this task, and in fact, did no better than the group of students. Hence, any training the officers had in forensic identification appeared to provide no assistance when it came to recognizing new instances of an unfamiliar face.

Australian Passport Officials

A recent major study of unfamiliar face recognition in Australian passport officers, conducted in collaboration with the Australian Department of Foreign Affairs, asked 30 passport officers to decide whether a passport photo matched the face of a person standing in front of them. [12] The study found passport officials incorrectly accepted a fake passport photo as genuine in 14 percent of trials. Interestingly, the findings also showed no relationship between employment duration/experience and accuracy on this task, as illustrated in Figure 5. That is, those who had 20 years' experience at the passport office were no more likely to be accurate than those who had just started working with the service.

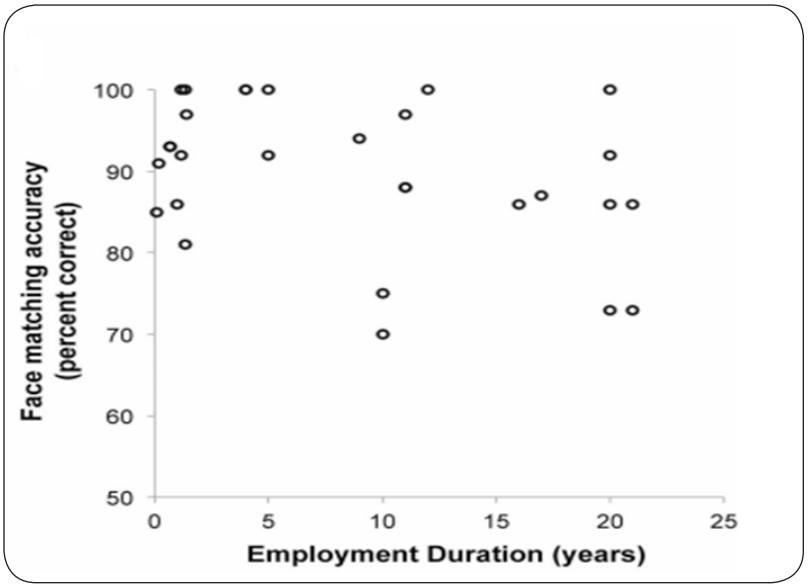

Figure 5. Unfamiliar face recognition accuracy (live face-passport photo) for Australian passport office staff presented as a function of employment duration. [12] (Released)

As the figure shows, there were very large individual differences between officers. In fact, this is the standard finding—some people consistently perform better at face tasks than others. Once again, the important point here is, in another specialist group, professional training and experience does not appear to lead to more accurate unfamiliar face recognition.

While the incidence of attempted passport fraud is very low, researchers recently showed that this, in itself, leads to an additional source of error. [13] In a phenomenon known as the low prevalence effect, researchers found that detecting unfamiliar mismatch face pairs was much poorer when they occurred on 10 percent of trials in comparison to 50 percent of trials. These findings suggest that a lack of familiarity with travelers and the low incidence of attempted passport fraud present major security risks at border control points.

How Can We Improve Unfamiliar Face Recognition Performance?

Based on the experimental evidence outlined above, it is clear that unfamiliar face recognition is a difficult task which is highly prone to error. This is the case regardless of whether the task involves photos or real faces, and whether it involves professional or untrained viewers. For the foreseeable future, reliance on face recognition for ID verification will continue at many border control points and in military operations. Consequently, recent research focused on ways to improve human performance in this domain.

Selection

As seen in Figure 5, researchers showed that recognition performance did not relate to occupational experience or years in employment; some individuals were simply better than others at this task. [12] This finding suggests that using an established test, such as the GFMT, [5] could help select high performing people for jobs in which accurate unfamiliar face recognition is a critical component of the job.

Paired Decision Making

Research shows that performance on recognition tests can be improved when participants work together and come to a judgment in pairs. [14] Across four experiments, the study tested unfamiliar face recognition individually (pre-test), as pairs (paired-test) and again individually (post-tests). The authors report that participants were more accurate when they made their judgments in pairs than individually. Furthermore, those who started with low performance showed a lasting benefit of having worked in pairs, suggesting that this type of procedure may be a particularly effective training method.

Multi-Photo ID

A different approach to improving unfamiliar face matching focuses on the ID document itself. The selection of photos for passports is a complex process,

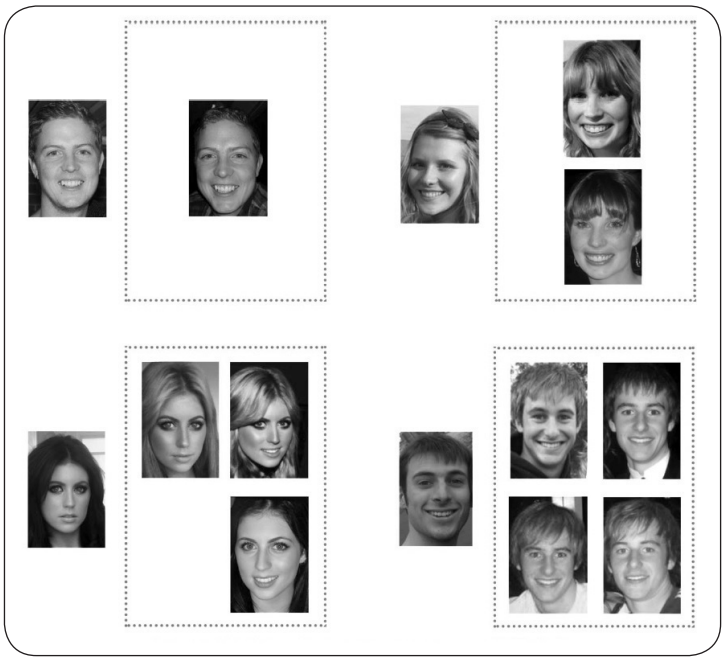

Figure 6. An example of the face matching arrays from reference 12. A single face is presented to the left, the array of photos to the right. Participants decide if the face on the left matches the photo(s) on the right. The example arrays showing one and three photos show the same person (match trials), while arrays showing 2 and 4 photos show different people (mismatch trials). (Released)

and one that differs between countries. Despite rather strict rules about the format required for these photos, people often comment that their IDs look rather unlike them. What modern research makes clear is that a single instance of a person can never form a true representation of their appearance. One suggestion is that the key to improving unfamiliar face recognition is learning how an individual varies across a naturally occurring set of instances. In other words, familiarity is shorthand for learning an individual's idiosyncratic variation in appearance. [15] It would be possible to achieve this for photo-ID by increasing the number of photos required on an ID document. Researchers reported that unfamiliar face matching performance significantly improved when mock ID cards contained two, three or four photo arrays (see Figure 6). [16] So, a relatively small increase in the number of photos contained on an ID card could reduce the error found with single image identity documents.

Automatic Face Recognition

Due to recent technological advances, there has been an increase in the use of automatic face recognition systems at airports, in military surveillance operations and even in social media (e.g., Facebook). Using the example of implementing electronic facial recognition gates (e-gates) at airports across the world, the expectation is these machines would remove the level of human error reported in the above passport officers' study. However, there is evidence to suggest that machine recognition systems are not providing the level of performance claimed by the developers. [17] Indeed, a recent report by the U.K. Inspectorate of Borders questioned the level of security provided by the e-gates.

The report cited an example in which a married couple was able to accidentally swap their passports and still make it through the e-gate system. Despite these systems performing well in benchmark tests, they routinely perform poorly in real world settings. [18] In the next section, we outline research which shows that current machine recognition performance can improve without any alterations to the algorithm.

Improving Automatic Face Recognition

No machine recognition system currently emulates the level of recognition accuracy found in humans when dealing with familiar faces. As noted above, familiar viewers are able to recognize photos of someone they know despite changes in lighting, camera characteristics, pose, expression and age. Unfamiliar face recognition in humans and machines is thought to be so poor precisely because these systems do not have

Figure 7. Individual images of the same person can look very different. Averaging these together produces a stable image, which will match a much wider range of the user's face and improve security. (Released)

Figure 9. Images of the six different locations used in Reference 27 to test the effectiveness of face averages. (Released)

access to ways in which an individual's appearance varies. Previous research suggests the brain may become familiar with a person by storing many different instances of them in memory. [19] However, in contrast, another study proposed that people may store an abstract representation of a person's face which retains identity specific information and discards irrelevant information which varies across images (e.g., lighting etc.). [20] This abstract representation, essentially a morph of different images, is called a face average and an example can be seen in Figure 7.

In an effort to track travel of non-U.S. citizens to the United States, the U.S. Customs and Border Protection uses biometric technology at the border to capture face and iris scans of all individuals entering or exiting the country. [21] Non-U.S. citizens traveling to the United States also require a visa that implements biometric identifiers, such as fingerprint scans, as well as a digital photo, unless they are part of the Visa Waiver Program. [22] This will allow the CBP to utilize a database instead of facial recognition, which is prone to human error. Facial recognition alone has proven to be an unreliable source and with the new data the CBP is collecting, it will be able to track persons of interest and foreign travel to ensure safety. [23]

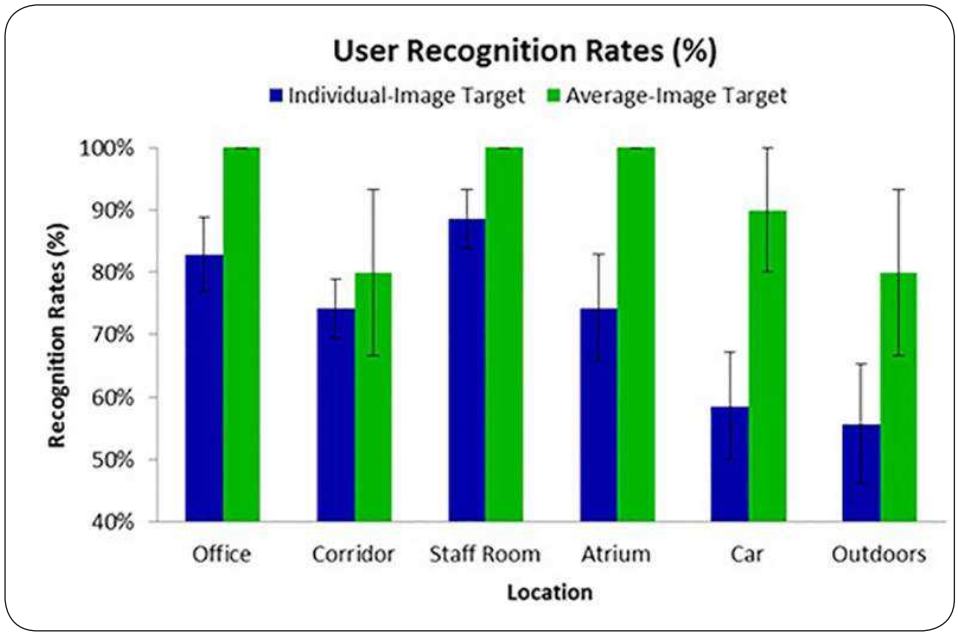

Figure 8. Shows the mean automatic face recognition accuracy when storing a person's face average (green) compared to an individual image of them (blue). Performance is shown for each of the six testing locations. (Released)

To protect sensitive information, as well as the homeland, federal agencies issue identification cards containing biometric data to federal employees and contractors. [24,25] This process makes it more difficult for unauthorized individuals to access secure facilities. Additionally, the use of stolen or counterfeited badges no longer poses a threat without the matching biometrics of the personnel. A multimodal method of detection is the most robust.

Research showed that face averages could be used to improve an automatic face recognition system. [20] This study assessed the performance of an online version of the then industry stan-

dard recognition system, FaceVACS. This system contained a large database of celebrity face photographs, (more than 30,000 images and more than 3,000 celebrities) which varied considerably in illumination, pose, facial expression, age and image quality. Users would upload a photo and the system would return the closest matching image in its database.

When the researchers uploaded individual images of different celebrities, they found the system would only return another picture of the correct identity on 54 percent of occasions [26] – a rather poor level of performance. However, when celebrities' face averages were uploaded, the system achieved perfect levels of performance (100 percent accuracy). This provided clear evidence that using a person's face average, rather than an instance of them, could dramatically improve machine recognition without an alteration to the algorithm.

While some research established the face average advantage using photo-to-photo matching, [26] a more recent study assessed whether the advantage remained when dealing with the recognition of real faces. [27] This study used the 'face-unlock' security system in the Samsung Galaxy smartphone. This system allows users to store an image of their face in the phone's memory, which can be used to recognize the user and allow access as an alternative to passcodes. Researchers reported that when they stored a person's face average rather than an instance of them, real face recognition rates were significantly increased, sometimes to perfect levels of performance as seen in Figure 8. [27] Impressively, this effect held across a variety of everyday environmental conditions, including outdoors, as seen in Figure 9.

In summary, face averages have been shown to improve machine recognition performance both with face photos and real faces. This technique could be particularly important to police or military operations, which are required to match a suspect or surveillance target to face photos held on a large database. Additionally, if face averages replaced individual photos on passports, e-gate security systems may perform better than they currently do.

Conclusions

This article presents experimental evidence, which shows that unfamiliar face recognition is a difficult task and one that is highly prone to error. Despite this, reliance on face photos for identity verification in domains with direct implications for national security continues. Research has identified several ways of improving human and machine face recognition. Other forms of biometric security, such as the iris scan, have the potential to replace face photo-ID in the future. However, many agencies around the world prefer to use faces for identification, possibly because their use is relatively unobtrusive, and appears natural. This article shows the limitations of this reliance.

The difficulty in recognizing unfamiliar faces is a security concern for the United States and requires increased use of biometric data for more accurate identification. This is especially critical when identifying individuals privy to sensitive information, non-citizens entering or leaving the United States, as well as criminals or terrorists. Incorrect identification of these individuals leaves the United States vulnerable.

The DoD fields an Automated Biometric Identification System to house and share biometric data. Originally developed for Operation Iraqi Freedom and Operation Enduring Freedom, ABIS is used to identify persons of interest that have a prior criminal history and are deemed as potential threats to U.S. forces. [28] In order to enable intelligence sharing amongst federal agencies, information within this system is shared with interagency partners like the Federal Bureau of Investigation and external partners that include the CBP and U.S. Citizenship and Immigration Services.

References

1. Burton, A., Wilson, S., Cowan, M., & Bruce, V. (1999). Face Recognition in Poor-Quality Video: Evidence From Security Surveillance. *Psychological Science*, 10, 243-248. doi: 10.1111/1467-9280.00144
2. Hole, G., George, P., Eaves, K., & Rasek, A. (n.d.). Effects of geometric distortions on face-recognition performance. *Perception*, 31, 1221-1240. doi: 10.1068/p3252
3. Ritchie, K., Smith, F., Jenkins, R., Bindemann, M., White, D., & Burton, A. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition*, 141, 161-169. doi: 10.1016/j.cognition.2015.05.002
4. Face Recognition. (2006, Aug. 7) Biometrics.gov. Retrieved from http://

www.biometrics.gov/Documents/facerec.pdf (accessed January 12, 2016).
5. Burton, A. M., White, D., and McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42(1) 286 - 291. Retrieved from http://homepages.abdn.ac.uk/m.burton/pages/BWM2010.pdf (accessed January 12, 2016).
6. Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A., & Miller, P. (1999). Verification of face identities from images captured on video. J*ournal of Experimental Psychology: Applied*, 5(4), 339-360. doi: 10.1037/1076-898X.5.4.339
7. Henderson, Z., Bruce, V., and Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, 15(4), 445-464. doi: 10.1002/acp.718
8. Megreya, A. M., White, D., and Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology*, 64(8), 1473-1483. doi: 10.1080/17470218.2011.575228
9. Megreya, A. M., and Burton, A .M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, 14, 364-372. doi: 10.1037/a0013464
10. Davis, J. P., and Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, 23(4), 482-505. doi: 10.1002/acp.1490
11. Kemp, R. I., Towell, N., and Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology,* 11(3). 211–222. doi: 10.1002/(SICI)1099-0720(199706)11:33.3.CO;2-F
12. White, D., Kemp, R.I., Jenkins, R., Matheson, M., and Burton, A.M. (2014). Passport Officers' Errors in Face Matching. *PLoS One*, 9(8). doi: 10.1371/journal.pone.0103510
13. Papesh, M. H., and Goldinger, S. D. (2014). Infrequent identity mismatches are frequently undetected. *Attention, Perception and Psychophysics*, 76(5), 1335–1349. doi: 10.3758/s13414-014-0630-6
14. Dowsett, A. J., and Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, 106(3), 433-445. Retrieved from http://eprints.whiterose.ac.uk/85532/1/Pairs_Paper_inpress.pdf (accessed January 12, 2016).
15. Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, 66(8), 1467-1485. doi: 10.1080/17470218.2013.800125
16. White, D., Burton, A.M., Jenkins, R., and Kemp, R. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, 20, 166-173. doi: 10.1037/xap0000009
17. Jenkins, R., and Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London*, B, 366(1571), 1671-1683. doi: 10.1098/rstb.2010.0379
18. Jenkins, R., and White, D. (2009). Commercial face recognition doesn't work. *Bioinspired Learning and Intelligent Systems for Security*, 2009. BLISS'09. Symposium on, 43-48. IEEE. doi: 10.1109/BLISS.2009.22.
19. Longmore, C. A., Liu, C. H., and Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 77–100. doi: 10.1037/0096-1523.34.1.77
20. Burton, A. M., Jenkins, R., Hancock, P. J., and White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, 256–284. doi: 10.1016/j.cogpsych.2005.06.003
21. CBP to Begin Biometric Entry/Exit Testing at Otay Mesa Port of Entry. (2015, December 10). Retrieved from https://www.cbp.gov/newsroom/local-media-release/2015-12-10-000000/cbp-begin-biometric-entryexit-testing-otay-mesa-port (accessed January 12, 2016).
22. Safety and Security of U.S. Borders: Biometrics. (2015) U.S. Department of State. Retrieved from https://travel.state.gov/content/visas/en/general/border-biometrics.html (accessed January 12, 2016).
23. CBP to Begin Biometric Entry/Exit Testing at Otay Mesa Port of Entry. (2015, December 10). Retrieved from https://www.cbp.gov/newsroom/local-media-release/2015-12-10-000000/cbp-begin-biometric-entryexit-testing-otay-mesa-port (accessed January 12, 2016).
24. Biometrics Frequently Asked Questions. (2006). National Science and Technology Council's Subcommittee on Biometrics. Retrieved from http://biometrics.gov/Documents/FAQ.pdf (accessed January 12, 2016).
25. Personal Identity Verification (PIV) Credential. (2015). U.S. General Services Administration. Retrieved from http://fedidcard.gov/ (accessed January 12, 2016).
26. Jenkins, R., and Burton, A.M. (2008). 100% accuracy in automatic face recognition. *Science*, 319(5862), 435. doi: 10.1126/science.1149656
27. Robertson, D. J., Kramer, R. S. S., and Burton, A. M. (2015). Face averages enhance user recognition for smartphone security. *PLoS One*, 10(3). doi: 10.1371/journal.pone.0119460
28. DoD Automated Biometric Identification System (ABIS). (2013). Army Programs. Retrieved from http://www.dote.osd.mil/pub/reports/FY2013/pdf/army/2013dodabis.pdf (accessed January 12, 2016).

David J. Robertson, Ph.D., is a post-doctoral research scientist in Mike Burton's face research group. His work focuses on assessing and improving unfamiliar face recognition in professional contexts (see www.facevar.com).

Mike Burton, Ph.D., is a professor of Psychology and a specialist in face recognition research. His current work, supported by major grants from the ERC and ESRC, is focused on improving our understanding of face recognition in both theoretical and applied contexts.