



Snapshots for Our Times: Face Recognition in 2005

Industry steps up with technological improvements and standards by which you can measure performance.

By Peter Kalocsai and Nick Imparato

Manufacturers and advocates of face recognition products have grown accustomed to fielding questions—and sometimes fending off attacks—about the technology. Isn't it a violation of personal privacy? Hasn't it been shown unreliable? Is it really necessary for security? These cultural and technological concerns have kept face recognition from realizing its full potential, but they may not hold it back much longer.

Privacy Concerns

The American Civil Liberties Union has demonstrated the most serious opposition to face recognition technology within the United States. Their main argument has been that widespread surveillance is likely to become increasingly invasive and abusive over time. They have frequently referred to some highly publicized airport installation fiascos

from several years back, as well as the infamous 2001 Super Bowl incident in Tampa, FL, buttressing their criticism by noting that since face recognition “does not even work,” it doesn't really increase safety and security.

This question of accuracy has been the biggest sticking point. Face recognition has worked fairly well in controlled environments, but has had serious problems in more realistic, uncontrolled settings. Orientation and lighting variations, and to some degree changes in expression, can produce markedly different facial images for the same person. Changes in hairstyle, facial hair and body weight, the effects of aging, and deliberate disguise can also hamper performance.

Given these problems, why choose face recognition? The truth is, face recognition is still the best passive and non-intrusive biometric available. It's also easy to use: An operator needs neither special hardware nor

expert skills to interpret the results.

In recent years, attacks on face recognition seem to be dampening. In part, this is due to the public's growing understanding and acceptance of video surveillance generally and familiarity with extensive surveillance efforts outside the U.S., particularly in venues like London.

Additionally, after well over a decade of gradual improvements, the technology has begun to show signs of maturation as it quietly gains ground by becoming more reliable.

New, Improved, Accurate

Face recognition advocates have actively pursued major developments that are now leading to increased accuracy.

3-D Representation. Faces are part of a 3-D world, so representing them with 3-D data makes intuitive sense. There are two main ways to do that.

In one alternative, structured near-visible infrared light is projected on a face. A 3-D representation is constructed based on the distortions of the pattern of light on the face. This technique is used by A4Vision.

An alternative method employs two or more cameras and builds a 3-D shape via correspondence analysis and triangulation methods. This is Geometrix's approach.

Both of these systems claim sub-millimeter accuracy. Because they need high resolution to achieve good results, they work best when the person is a couple of feet away from the capture device(s). The most attractive part of 3-D technologies is that they virtually eliminate the effects of orientation and illumination changes, the two major difficulties associated with traditional 2-D recognition techniques. On the other hand, 3-D systems are typically more expensive than their 2-D counterparts, and their computational load is quite a bit higher as well.

High-Resolution 2-D. Due to advances in sensor technology, today's cameras are able to acquire increasingly higher-resolution face images. You can expect better recognition results from a system that captures more details on a face. Although skin recognition technology

is supposed to work with mid-range cameras, higher resolution helps the analysis of skin patterns on the face.

The addition of skin pattern data can improve face recognition performance. That promise was most likely behind Identix's acquisition of DeLean Vision in 2004. Indeed, Identix claims that face recognition performance improves by 20 to 25% when skin pattern data is included.

Preprocessing. Recent advances in computer graphics and computer vision produce much better algorithms for automatic lighting and pose correction on 2-D face images. Such preprocessing methods detect and automatically correct for changes in illumination and pose before the face image is sent to the recognition system. By removing the external variations caused by chance effects, these corrections improve the system's ability to focus on meaningful, identity-based variations. Consequently, the recognition algorithm now has a significantly improved chance of coming up with the correct result.

Multiple Images. Another way to improve performance by using more data is to store multiple images of each person in the database. The hope here is that there is a better chance of

finding someone in a database when multiple matches are possible.

Some systems synthesize multiple images from a single shot. XID Technologies' predictive face synthesis technique simulates the effects of changing light, and orientation, as well as the addition of glasses, facial hair and other variable features, automatically producing several images from one 2-D picture.

An alternative way of using multiple images is to work with video data. Although the most recent government testing of face recognition systems, FVRT 2002, showed no advantage of video over still images, one might argue that if processed appropriately, more video data should eventually result in better performance for a face recognition system.

Standards Development

Shortly after passing the USA Patriot Act and the Enhanced Border and Visa Entry Reform Act 2002, the U.S. government turned to the International Organization for Standardization for guidance in the international standardization of biometric information. The American National Standards Institute was approached with similar requests on a domestic level.

ANSI's International Committee for

Table 1. Selected face recognition vendors

Vendor	Web Address	Technology
A4Vision	www.a4vision.com	3-D - Near visible structured IR light
Acsys Biometrics/AND	www.acsysbiometricscorp.com	2-D - Holographic/quantum neural technology
Cognitec	www.cognitec-systems.de	2-D - Feature extraction, working on 3-D
C-VIS	www.c-vis.com	2-D - Neural networks, facial landmarks
Dream Mirh	www.dreammirh.com/english/main.html	2-D - Template matching
Geometrix	www.geometrix.com	3-D - Multiple cameras, correspondence finding
HumanScan	www.humanscan.com	2-D - Synergetic computer classifier
Identix	www.identix.com	2-D - Local feature analysis, skin analysis, working on 3-D
Imagis	www.imagistechnologies.com	2-D - Descriptor array, spectral analysis
Neven Vision	www.nevenvision.com	2D - Wavelets, neural networks
OmniPerception	www.omniperception.com	2-D - Client-specific linear discriminant analysis
Pelco	www.pelco.com	2-D - Wavelets, facial landmarks, statistical analysis
Viisage	www.viisage.com	2-D - Originally PCA, now wavelets too, working on 3-D
VisionSphere	www.visionspheretech.com/menu.htm	2-D - Holistic feature code
XID Technologies	www.xidtech.com	2-D - Predictive face synthesis

Table 2. Various terminology used for describing accuracy and performance

Face Recognition Outcomes			
For people IN database		For people NOT IN database	
Portion of people MATCHED (correct)	Portion of people NOT matched (error)	Portion of people MATCHED (error)	Portion of people NOT matched (correct)
Identification Rate (IR)	False Reject(ion) Rate (FRR)	False Accept(ance) Rate (FAR)	Correct Rejection Rate (CRR)
Correct Match Rate (CMR)	False Non Match Rate (FNMR)	False Match Rate (FMR)	Correct Non Match Rate (CNMR)
Hit Rate (HR)	Miss Rate (MR)	False Alarm Rate (FAR)	Correct Reject Rate (CRR)
True Positive Rate (TPR)	False Negative Rate (FNR)	False Positive Rate (FPR)	True Negative Rate (TNR)
True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)
Sensitivity	1 - Sensitivity	1 - Specificity	Specificity
The sum of these two columns gives the probability of 1 (or 100%) ¹		The sum of these two columns also gives probability of 1 (or 100%)	

(Footnotes)

¹ The likelihood of the system to match or not match people in the database at a given threshold can either be expressed as probabilities or percentages. Since people who are in the database are either matched or not, the sum of these two outcomes provides the full 1 probability, or 100%. In other words, there is a 100% chance that people who are in the database will be matched or not matched. Of course, the same is true for people who are not in the database.

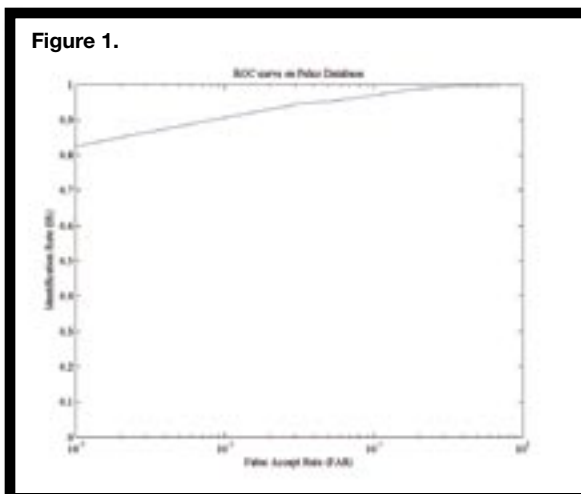
Information Technology Standards (INCITS) developed its face recognition format for data interchange last year, and ISO is to follow this year. Because of the quick growth in this area, INCITS has already decided to amend its standard to address 3-D face images. In addition, both organizations have already made progress in standardizing the biometric vocabulary.

Table 2 organizes the various terms and acronyms that one might come across on various company Web sites, in white papers or in other publications that reference face recognition performance. INCITS is adopting the following terms: False Accept Rate (FAR) and False Reject Rate (FRR) in its Biometric Performance Testing and Reporting Standard, currently under development. The ISO will likely do the same.

How to Evaluate Systems

Since the standardization of testing regimes for biometric systems is still developing, an objective evaluation and comparison of face recognition products can be quite a challenge. Manufacturers and vendors are typically exuberant about the performance of their own systems, but that doesn't always make for quality, accurate assessments.

Government Testing. The U.S. Department of Defense has been evaluating face recognition systems since 1993, and its tests have



Left: Figure 1. The Receiver operating characteristic curve of the Pelco Face Recognition System on the Pelco database. Used with permission.

close attention to the FVRT test results and use them as guidelines in deciding which vendor to partner with.

ROC Curves and Other Metrics. Probably one of the most important graphs to describe the performance of a face

recognition system is its receiver operating characteristic (ROC) curve.

The curve plots the FAR of the system against its identification rate (IR) at various threshold levels, but essentially it could be thought of as showing how the system trades off FAR for FRR. Note that if we used 1-to-1 matching we would talk about verification rate (VR) instead of IR. For better viewing, you will often see the FAR data plotted on a logarithmic scale on the X-axis, as shown in Figure 1. The ROC is an accepted standard for demonstrating the performance of a recognition system, but plotting the FAR and FRR values directly together at various threshold levels can elucidate their complementary nature

become the most reliable indicators of system viability. Both research enterprises and commercial organizations around the world have been invited to take part in the Face Recognition Grand Challenge administered from May 2004 through July 2005. The FRGC will be followed by the Face Recognition Vendor Test 2005, scheduled for August/September. These evaluation venues will perform tests on the same databases for all participants in order to achieve fair and objective comparisons. FVRT 2005 could become the most important industry shaping event for several years ahead. Companies wanting to integrate face recognition solutions into their products pay

The ROC is an accepted standard for demonstrating the performance of a recognition system, but plotting the FAR and FRR values directly together at various threshold levels can elucidate their complementary nature

become the most reliable indicators of system viability. Both research enterprises and commercial organizations around the world have been invited to take part in the Face Recognition Grand Challenge administered from May 2004 through July 2005. The FRGC will be followed by the Face Recognition Vendor Test 2005, scheduled for August/September. These evaluation venues will perform tests on the same databases for all participants in order to achieve fair and objective comparisons. FVRT 2005 could become the most important industry shaping event for several years ahead. Companies wanting to integrate face recognition solutions into their products pay

more readily, as shown in Figure 2. At lower threshold levels there is virtually no false reject error (meaning everyone is matched), but the false accept error is high (meaning many are matched who were not supposed to be).

At high threshold levels the pattern reverses: Now you see the false accept rate dropping sharply, but at the same time the false reject rate is increasing. It is becoming harder and harder to match even legitimate enrollees.

The green curve, total error rate (TER), shows the sum of these two error types (FAR and FRR) at increasing threshold levels. If the cost of making both types of error is the same, then the system's minimum overall error can be found at the lowest point of the green curve (threshold level 12 in this example). Note, however, that for a low-security situation, such as a public cafeteria, one might set the threshold a bit lower than that in order to minimize false rejections. At the same time, for a nuclear facility, one might set the threshold higher to eliminate false accepts as much as possible.

Another good indicator of system performance is the Equal Error Rate (EER), shown by the arrow in Figure 2, where the FAR is equal to the FRR, or graphically, where the FAR and FRR curves cross one another. The minimum of the total error rate, $\min(\text{TER})$, does not necessarily equal $2 \times \text{EER}$, although these numbers are usually close to each other.

As Figure 2 shows, practically any system can show perfect or close to perfect performance on FAR or FRR data separately, but these numbers are only meaningful when presented together. In other words, ask either for the minimum total error of the system, $\min(\text{TER})$, or FAR and FRR data presented at the same

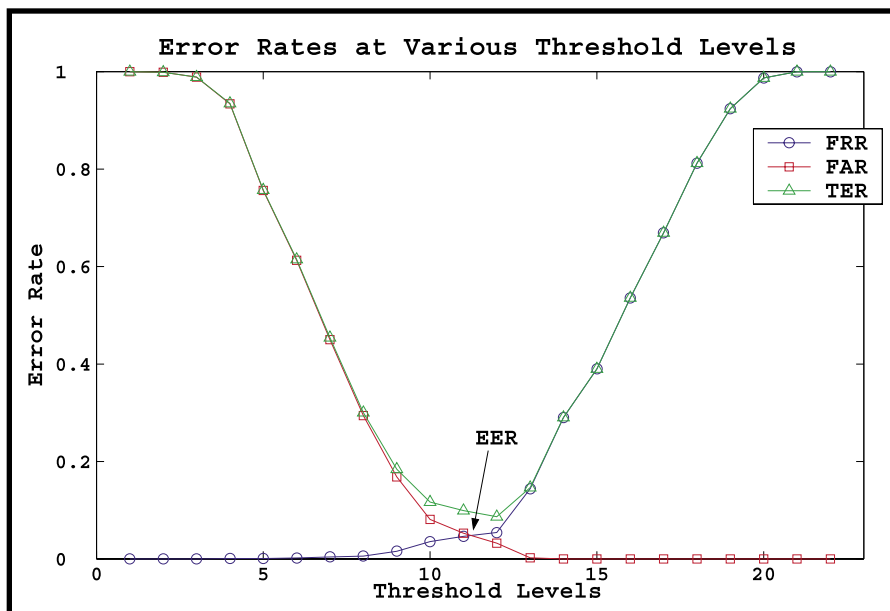


Figure 2: False Accept Rate (FAR), False Reject Rate (FRR) and their total (TER) is plotted for 21 threshold levels. Pelco data, used with permission.

threshold level. Alternatively, the EER value is a threshold-independent good overall indicator of performance.

Equally important is the quality and the size of the database that was used to test the system. The FRVT 2002 results show that face recognition performance drops linearly with the logarithm of the database size.

Bad quality or uncontrolled imagery can also dramatically reduce performance. Indeed, this is one of the reasons so much effort is being invested to standardize picture taking and image quality.

There is a host of technological advances accelerating the improvement in face recognition accuracy. Just as important, public attitudes are becoming less negative—the specter of DNA identification, for example, makes face recognition look less intimidating. The policy efforts, such as the development of rigorous standards,

also underscore the likelihood that face recognition has now gained the traction needed to become an enduring security tool. As deployments increase, and as face recognition increasingly becomes a component in multi-modality solutions, revenues and investor confidence will grow, providing the resources for further research and development. The innovations that are developed, in turn, will deepen performance, and the cycle will repeat itself. For some time ahead, in other words, industry dynamics will move the entire cycle upwards. **STD**

Quick Hits

American Civil Liberties Union: www.aclu.org

Face Recognition Vendor Test 2002: www.frvt.org/FRVT2002/documents.htm

Face Recognition Grand Challenge: www.frvt.org/FRGC

Face Recognition Vendor Test 2005: www.frvt.org/FRVT2005/default.aspx



Peter Kalocsai, PhD. (pkalocsai@pelco.com), is a senior research scientist at Pelco's R&D department, focusing on face and object recognition, intelligent video applications and statistical analysis.



Nicholas Imparato, PhD. (imparato@usfca.edu), is a professor at the School of Business, University of San Francisco, a Research Fellow at the Hoover Institution, Stanford University and an industry advisor.