Evaluating Facial Recognition Technology for Drug Control Applications

Mike Bone, NAVSEA Crane Division
300 Hwy 361, Crane, IN 47522
(812) 854-1141, Fax (812) 854-2655, bone_mike@crane.navy.mil

Dr. James L. Wayman*, Office of Research and Graduate Studies
San Jose State University, San Jose, CA 95192-0080
(408) 924-4037, Fax (408) 924-3818, biomet@email.sjsu.edu

Duane Blackburn, DoD Counterdrug Technology Development Program Office and the National Institute of Justice
17320 Dahlgren Road Code T43, Dahlgren, VA 22448
(540) 653-6062, Fax (540) 653-7471, blackburndm@nswc.navy.mil

Abstract

Automated facial recognition technology has seen many genuine advances in the last few years, but separating true progress from vendor claims has been difficult in this highly competitive market place. To this end, the DoD Counterdrug Technology Development Program Office, the National Institute of Justice, and the Defense Advanced Research Projects Agency recently sponsored Facial Recognition Vendor Test 2000 (FRVT 2000) [1] to evaluate commercially-available automated facial recognition systems. FRVT 2000 consisted of an extensive "technology evaluation" using mug-shot type photographs (similar in methodology to the widely respected Army Research Labs FERET tests [2-6]), and more limited access control and surveillance "scenario evaluations" using volunteers directly interacting with the systems. The recently published test results indicate the degree of suitability that commercial systems have for many applications of wide interest to the supply-side drug control community.

In this paper, we present an overview of the FRVT 2000 evaluation and examine case studies of potential drug control applications. We provide a requirements analysis for each application and show how the FRVT 2000 evaluation report could be used to determine if the current level of facial recognition technology is appropriate to meet those requirements.

Introduction

Facial recognition is the identification of humans by the unique characteristics of their faces [7-13]. Systems that automate this task appear to be promising for many law enforcement applications. When someone considers using automated facial recognition for an application, they often ask "Which system works best?" As it turns out, there is

* Under contract to the Combating Terrorism Technical Support Working Group (TSWG)

no simple answer to this question since there is no single performance measure that can be used to compare systems.

Some facial recognition vendors will quote performance statistics such as false acceptance (or alarm) rate (FAR) or false reject rate (FRR). FAR is the percentage of imposters wrongly accepted by the system while FRR is the percentage of valid users wrongly rejected by the system. The problem with using these figures is two-fold: 1) There is a trade off between FAR and FRR that depends on the security and throughput requirements of the system; 2) Both FAR and FRR are highly dependent upon the specifics of the application environment, including user motivation.

Some vendors may quote an equal error rate (EER) in a laboratory environment -- the error rates when the FAR is equal to the FRR. Again, the EER doesn't tell the whole story because it gives a performance measure for a single security/throughput requirement in a single environment, which will most likely not be that of a target application. An evaluation of a facial recognition system should consider the error rates over a range of settings and a range of environments.

Conceptually, verification is a one-to-one comparison where a user presents an identity along with a live biometric and the biometric system determines if the live template acquired by the sensor matches that stored under the claimed identity. Identification is a one-to-many comparison where a user presents only the live biometric with no claim to identity. The biometric system then searches an entire database of enrolled templates to find the identity of individuals who most closely match. In verification applications, error rates are dependent only upon the FAR/FRR setting. In identification applications, error rates depend upon both the FAR/FRR settings and the size of the searched database. This paper will discuss how this can be done using the FRVT 2000 evaluation report [1] and several case studies.

Evaluating Biometrics

To evaluate biometric systems, one must first have a good understanding of the intended application and the desired level of security [14-16]. With this information in hand, systems can be evaluated to determine which, if any, will meet that application's unique requirements. This can be accomplished by using a three-step evaluation protocol proposed in "An Introduction to Evaluating Biometric Systems" [17] that includes a technology evaluation followed by a scenario evaluation and an operational evaluation. This methodology applies to facial recognition as well as other biometrics [16,18].

A technology evaluation isolates and tests the face matching abilities of a facial recognition system without regard to other factors such as real-time performance, algorithm/camera interface, or database operations. At the heart of every facial recognition system is a matching algorithm that compares two images and gives a score telling the degree of difference between them. There are many other components that make up a complete system, but without a good matching algorithm, the system will have limited performance. Testing the matching component of a facial recognition system requires a large number of images for known individuals. The matching algorithm is used to compare each image to each of the other images and produce a score for each comparison. Since the true identity for each image is known, the performance of the matching algorithm can be evaluated. Systems that perform well in the technology evaluation can be selected for the next step, the scenario evaluation.

In a scenario evaluation, the facial recognition system is tested as a unit, rather than isolating one component (matching). The system is also tested in a manner that resembles the application it would eventually be used for. In addition to the matching algorithm, the database algorithms, algorithm/camera interaction, and system/subject interactions are tested. This gives a better understanding of how the overall system will perform in the real world, but is usually performed in controlled laboratory conditions with a small number of live test subjects. There are many possible applications for facial recognition, such as access control and mugshot searches, each with a unique set of requirements. The scenario evaluation must be designed to create the conditions that are relevant to the intended application. The top performers in the scenario evaluation can then advance to the operational evaluation.

In an operational evaluation, a facial recognition system is tested using a representative sample of the intended user population. This phase is conducted at the location where the system would ultimately be installed and gives the best indication of how it will perform in full operation. Operational evaluations typically last from several weeks to several months.

Once the three-step evaluation process (technology, scenario, operational) has been completed, one can be assured that they have chosen the proper system for their application.

FRVT 2000 Overview

In order to help determine which facial recognition system is best suited to a particular application, the DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), and the Defense Advanced Research Projects Agency (DARPA) sponsored FRVT 2000. This test consisted of a technology evaluation using 13,872 facial images and a limited scenario evaluation involving three live subjects. Vendors offering commercial facial recognition systems in the United States were invited to join the evaluation, and five vendors chose to participate. Testing took place in May/June 2000 and the evaluation report was made available on the web in February 2001 [1].

The technology evaluation for FRVT 2000 was named the Recognition Performance Test. For this part of the evaluation, each vendor was given a set of 13,872 images and asked to compare each image to each of the other images and report each comparison score in a pre-defined format.

The comparison scores of the 13,872 images can be divided into subsets based on differences in controlled collection conditions, such as imaging distance, illumination, facial pose angle, or expression. This allows for the execution of a broad range of “experiments” using the comparison data, such as assessing the effects of changes in pose angle or illumination. The comparison scores were processed “off-line” using software developed at the National Institute of Standards and Technology (NIST) to generate plots showing results for various test conditions.

One measure of performance is the Receiver Operating Characteristic (ROC) curve shown in Figure 1. This plot shows the false alarm rate (FAR) on the horizontal axis and the probability of correct verification (computed and plotted as 1 – FRR) on the vertical axis. The upper left corner of the plot represents the ideal situation where all valid users are granted access and all imposters are denied access. Each of the plotted points indicate the FAR and FRR for a particular security setting on a hypothetical biometric system. The point where the plot crosses the diagonal line from the top left to the bottom right in the figure is the EER for the system. This is the point where the FAR equals the FRR. It can readily be seen that the EER gives only a single point on the plot. It doesn't give any information about the rest of the plot since there are an infinite number of plots that could possibly have this same EER.

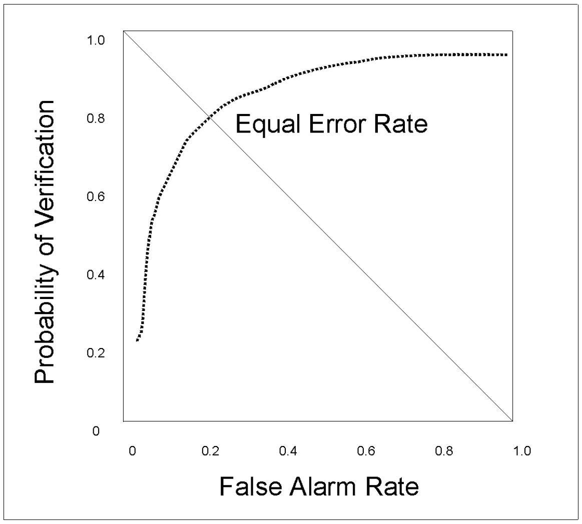

Figure 1: Receiver Operator Characteristic (ROC).

A second measure of performance is the "rank order statistics", displayed graphically as a "Cumulative Match Characteristic" (CMC) curve shown as Figure 2. The facial recognition research community has traditionally used the term "gallery" to mean the group (of size *m*) of enrolled facial images, and the term "probe" to indicate an unknown sample image to be compared to the entire gallery. The rank order statistics indicate the probability that the gallery image from the correct individual will be among the top *n* matches to a probe. This probability depends upon both *n* and *m*.

In Figure 2, the horizontal axis shows the rank, *n*, or the number of matches returned by the system against a gallery of size *m*. The vertical axis shows the probability that the correct match is included in the top *n* matches. For example, in Figure 2, there is a 70% chance that the correct person is included in the top 10 matches returned by the system with gallery size *m*, and an 80% chance that the correct person is included in the top 20 matches.

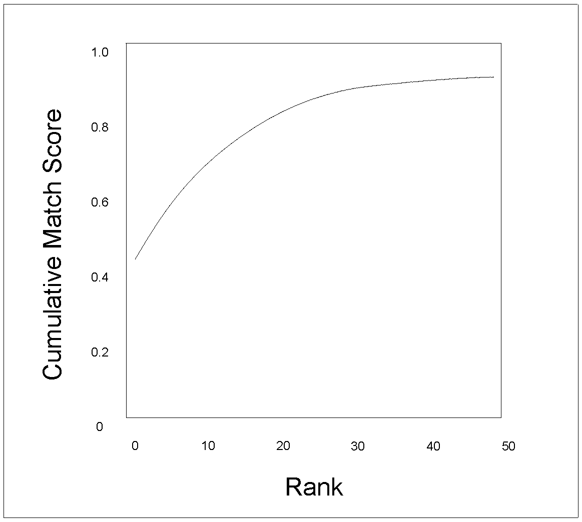

Figure 2: Cumulative Match Characteristic (CMC).

For each "experiment" performed on the comparison scores returned by each vendor, both ROC and CMC curves were calculated.

These curves give us understanding into the effects on system performance of variation of the test parameters, such as illumination, pose angle and facial expression, and allow a coarse level of performance prediction for applications similar to the test conditions.

In addition to the Recognition Performance Test, FRVT 2000 also included a limited example of a scenario evaluation entitled the Product Usability Test. An access control scenario was chosen using three live test subjects. Two different tests were performed: the “Old Image Database Timed Test” and the “Enrollment Timed Test”. For the “Old Image Database Timed Test”, vendors were given a set of 135 images acquired with a badge system developed by NAVSEA Crane. The subjects, who were included in the image set, performed trials walking towards the vendor systems starting at various distances. The time and correctness of the match were recorded for each trial with a limit of 10 seconds. The “Enrollment Timed Test” differed in that subjects were enrolled using the vendors’ specifications, generally involving multiple enrollment images, and the test subjects stood still rather than walked towards the system.

As a continuation of the FRVT 2000 effort, but not a part of the actual FRVT 2000 evaluations, an operational evaluation is being sponsored by NIJ that will take place at Prince George’s County (Maryland) Correctional Center. Based on the results of FRVT 2000, a facial recognition system from Visionics Corporation was selected for dual-location with an operator-assisted access control system. The combined “system” uses proximity badges and facial recognition technology to verify the identity of employees at the facility.

In the first phase of the evaluation, employees will use their proximity badges to enter the facility through a supervised entry area. When exiting the facility, they will use their proximity badges while the facial recognition system alerts an officer if a live surveillance camera image of the employee does not match the enrolled image of the employee within a certain threshold. The officer will use the information from the facial recognition system to help determine whether or not to open the electronically controlled door. This setup will help ensure that an inmate is not attempting to escape using an employee’s badge. Another benefit is the logging capabilities of the access control system that will allow correctional officers to determine which employees are inside the facility at any given time. This is an important safety feature that could be utilized in case of a fire or other emergency.

This operational evaluation will help determine how well facial recognition performs in a correctional environment and how well employees interact with the system. Based on the results of this phase, future phases may add the capability to monitor visitors entering the facility to identify former inmates and others that need to be watched more closely during visits.

Requirements Analysis

The first step in evaluating a facial recognition system for an application is gathering the requirements. These will vary widely among applications, and may even be quite different for the same application at different sites. Some of the things to consider are:

- Verification or identification system
- Size of the identification gallery
- Lighting conditions
- Overt or covert system
- Behavior of users

- Throughput rate
- Elapsed time since enrollment image
- Accuracy requirements

After determining the nature of the proposed application, it may be possible to find test results in the FRVT 2000 analysis under similar conditions. System performance can be estimated by using the FAR/FRR rates from the reported ROC curve [20].

In the remaining part of this paper, we will project the performance of some facial recognition implementations of possible interest to the supply-side drug control community.

Case Studies

The following case studies demonstrate the general process for analyzing facial recognition applications using a three-step evaluation protocol proposed in "An Introduction to Evaluating Biometric Systems" [17]. We use results of the FRVT 2000 technology and scenario evaluations to determine the level of performance that can be expected using the tested systems under the conditions used for testing. Individual vendor's scores will not be highlighted for these case studies, but should be studied by anyone performing a real analysis of the results for a specific application.

All images shown in the following cases were taken from the FRVT 2000 Evaluation Report [1]. Results from this report should be referred to rather than only the ones shown below when performing other case studies. Detailed information describing each graph shown below is also provided in the FRVT 2000 Evaluation Report [1].

Case 1: Watch-listed Person Appears at Baggage Claim Counter

Suppose we have a "watchlist" of mugshots and wish to be notified if any of those persons appear at the Baggage Claim counter of an airline. At the counter, we can take multiple images under controlled lighting of each individual as they claim their baggage. Individuals may be notified that they are being imaged, but can be expected to be "indifferent" to the imaging process, meaning that they cannot be relied on to face the camera squarely and to pose. Memorization of a dynamic mugshot watchlist of considerable size is probably beyond the capability of human inspectors, so we would like to narrow down the manual search to 10 images. Using the identification results from the FRVT 2000 technology evaluation, we can determine the expected performance of several systems, considering the effects of varying probe and gallery conditions.

Compression: Depending on the imaging equipment used, the images acquired at the baggage claim area (probes) may be stored in compressed jpeg format. Figures 3—6 show that if a person is in a watchlist of 1196 gallery images, there is an 80% chance that he will show up in the top 10 matches returned by a facial recognition system if compression up to 30:1 is used and there is no other variability between the "watchlist" mugshots and the acquired image. Figure 7 shows that at 40:1 compression, there is a 78% chance of finding the subject in the top 10 matches.

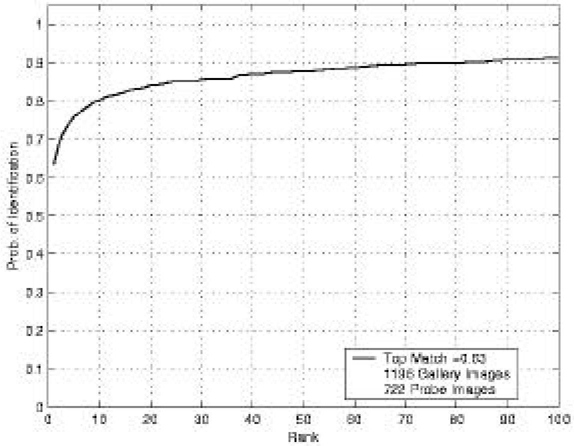

Figure 3: Best identification scores for FRVT 2000 compression experiment C0. Uncompressed gallery and probe images. FRVT 2000 Evaluation Report image M-1.

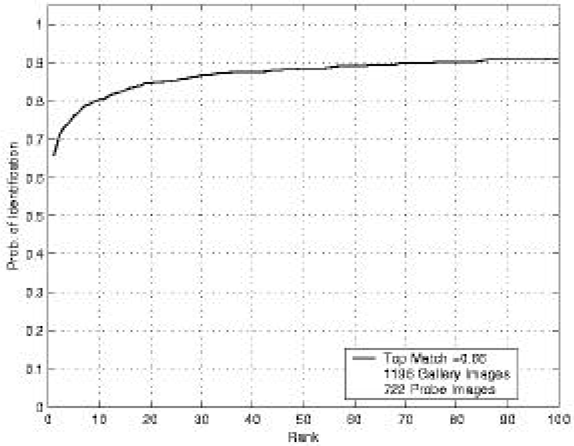

Figure 5: Best identification scores for FRVT 2000 compression experiment C2. Uncompressed gallery images, probes compressed 20:1. FRVT 2000 Evaluation Report image M-3.

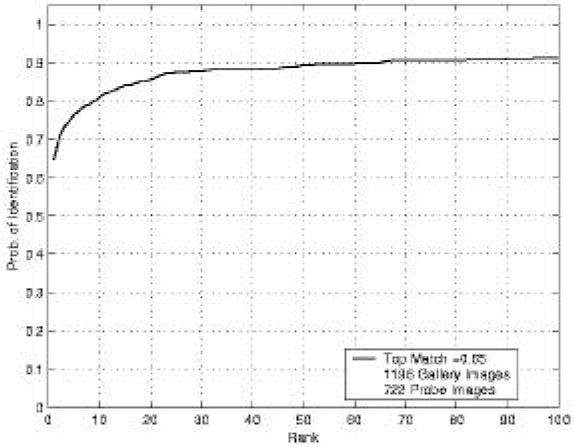

Figure 4: Best identification scores for FRVT 2000 compression experiment C1. Uncompressed gallery images, probes compressed 10:1. FRVT 2000 Evaluation Report image M-2.

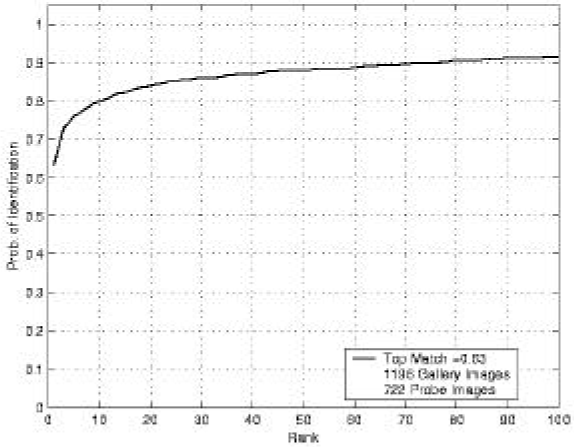

Figure 6: Best identification scores for FRVT 2000 compression experiment C3. Uncompressed gallery images, probes compressed 30:1. FRVT 2000 Evaluation Report image M-4.

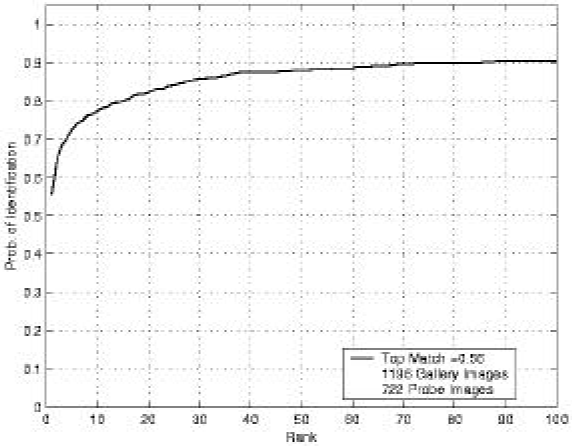

Figure 7: Best identification scores for FRVT 2000 compression experiment C4. Uncompressed gallery images, probes compressed 40:1. FRVT 2000 Evaluation Report image M-5.

Distance: The distance between the camera and subjects at the baggage claim area may not be the same as that used for the watchlist images. Figures 8—10 show the results of varying the distance for a watchlist of 185 images. The gallery images for these figures were taken indoors with a digital camera placed 1.5m from the subjects while the probes were taken indoors with a video camera at distances of 2, 3, and 5m. If we wish to examine the top 10 matches, we can expect a 60% chance of correct identification at a distance of 2m, 50% at 3m, and 30% at 5m, again assuming that there are no other factors varying between the "watchlist" mugshot and the acquired image.

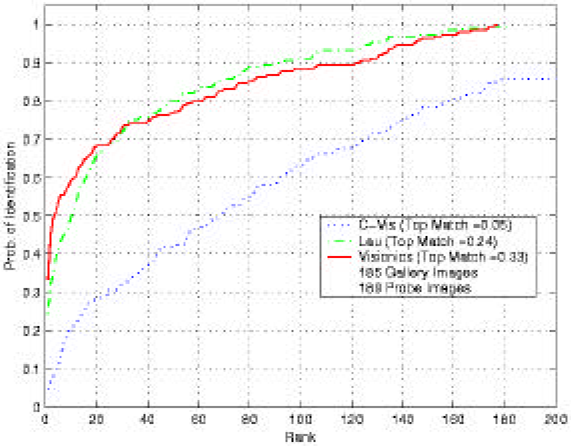

Figure 8: Identification scores for FRVT 2000 distance experiment D1. Gallery images taken with camera distance 1.5m, probe images taken with camera distance 2m. FRVT 2000 Evaluation Report image M-12.

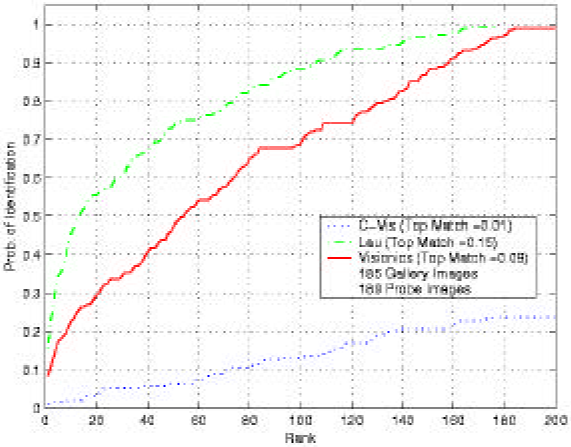

Figure 9: Identification scores for FRVT 2000 distance experiment D2. Gallery images taken with camera distance 1.5m, probe images taken with camera distance 3m. FRVT 2000 Evaluation Report image M-13.

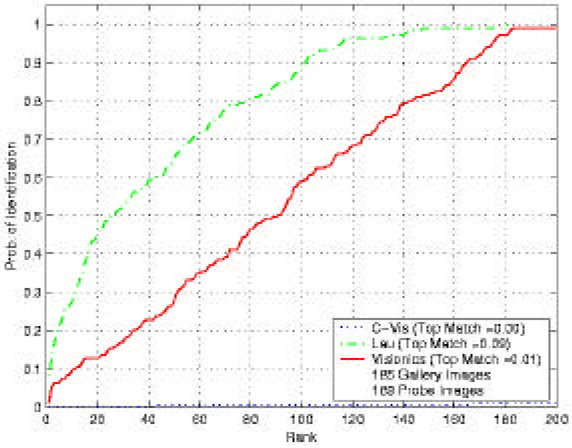

Figure 10: Identification scores for FRVT 2000 distance experiment D3. Gallery images taken with camera distance 1.5m, probe images taken with camera distance 5m. FRVT 2000 Evaluation Report image M-14.

Expression: We wouldn't expect persons at the baggage claim to have the same facial expressions as those of the mugshot images in the watchlist. Figures 11 and 12 show the results of varying expression for a watchlist of 224—225 images. The figures show that we can expect at least a 95% chance of finding the correct person in the top 10 matches if expression is the only variable factor.

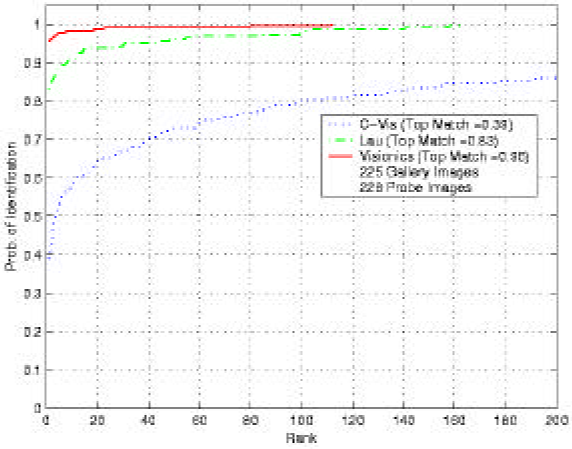

Figure 11: Identification scores for FRVT 2000 expression experiment E1. Regular expression used for gallery, alternate expression used for probes. FRVT 2000 Evaluation Report image M-19.

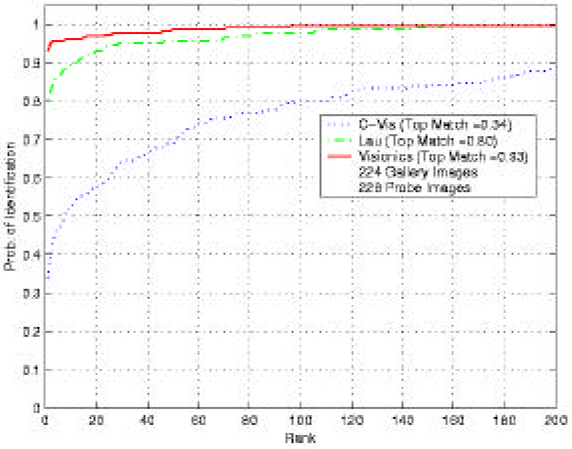

Figure 12: Identification scores for FRVT 2000 expression experiment E2. Alternate expression used for gallery, regular expression used for probes. FRVT 2000 Evaluation Report image M-20.

Illumination: Although the lighting can be controlled at the baggage claim area to some degree, it may be unrealistic to use the same type of photographic flood lamps used for mugshots. Figure 13 shows the results of using overhead fluorescent lights for the probe images and standard mugshot lighting for 227 gallery images. The figure shows that we can expect a 95% chance of finding the correct person in the top 10 matches if illumination alone varies between mugshot and acquired image.

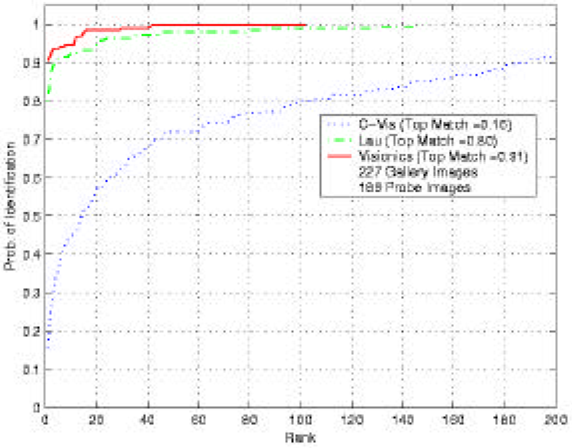

Figure 13: Identification scores for FRVT 2000 illumination experiment I1. Mugshot lighting used

for gallery, overhead fluorescent lighting used for probes. FRVT 2000 Evaluation Report image M-21.

Media: For this application, we'll assume that digital media is used for capturing the watchlist images as well as the images at the baggage claim area. Therefore, media should not be an important consideration here.

Pose: For the mugshot images in the watchlist, it can be expected that persons will be facing directly toward the camera. But when capturing images at the baggage claim area, persons may have their heads turned at different angles to the camera. Figures 14—17 show the results of varying the pose angle for the probe images while comparing with 200 gallery images taken with subjects directly facing the camera. The figures show that we can expect about a 99% chance of finding the correct person in the top 10 matches with a pose angle of 15 or 25 degrees. However, the probability drops to 88% for 40 degrees and 58% for 60 degrees. Again, we assume that pose is the only varying factor.

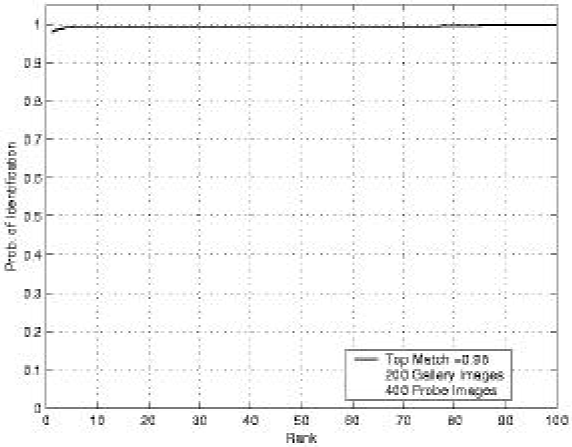

Figure 14: Best identification scores for FRVT 2000 pose experiment P1. Subjects faced directly toward camera for gallery, turned 15 degrees for probes. FRVT 2000 Evaluation Report image M-6.

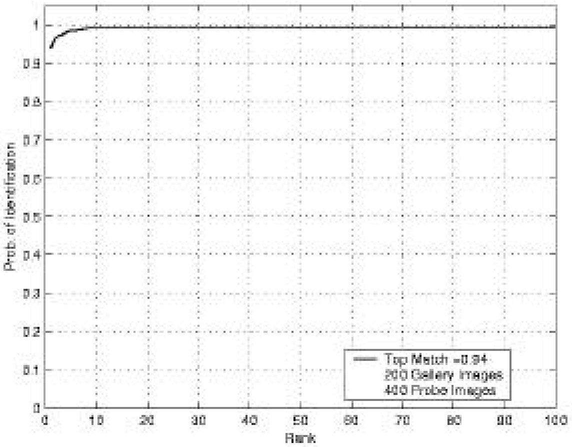

Figure 15: Best identification scores for FRVT 2000 pose experiment P2. Subjects faced directly toward camera for gallery, turned 25 degrees for probes. FRVT 2000 Evaluation Report image M-7.

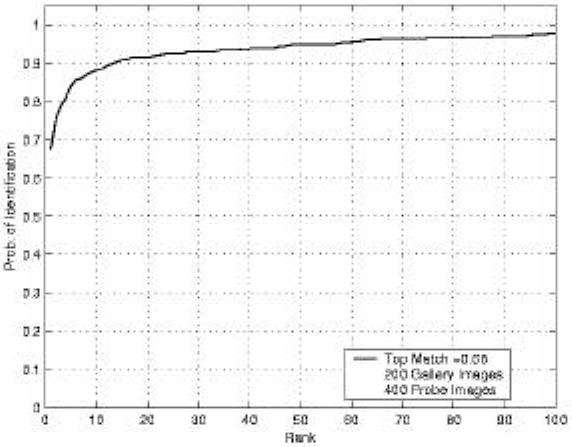

Figure 16: Best identification scores for FRVT 2000 pose experiment P3. Subjects faced directly toward camera for gallery, turned 40 degrees for probes. FRVT 2000 Evaluation Report image M-8.

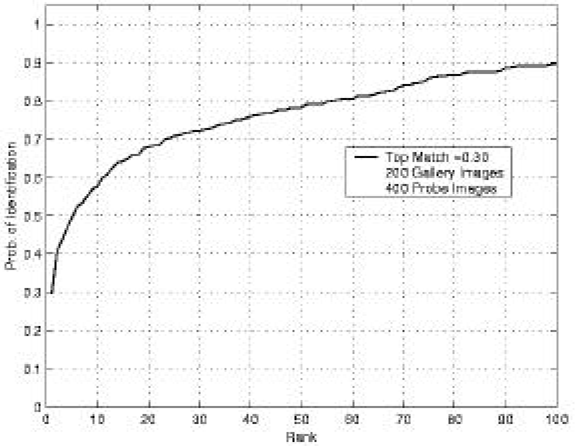

Figure 17: Best identification scores for FRVT 2000 pose experiment P4. Subjects faced directly toward camera for gallery, turned 60 degrees for probes. FRVT 2000 Evaluation Report image M-9.

Resolution: Another variation that may occur between the mugshot images and those captured at the baggage claim area is the resolution of the image, quantified as the number of pixels between the centers of the eyes. Resolution can vary due to the resolution of the imaging device or the distance between the camera and subject. Figures 18—21 show the results of varying the eye separation of probes while comparing to a gallery of 101 images with eye separation ranging from 88 to 163 pixels. For probe eye separations of 60, 45, 30, or 15 pixels, we can expect a 98% chance of finding the correct person in the top 10 matches.

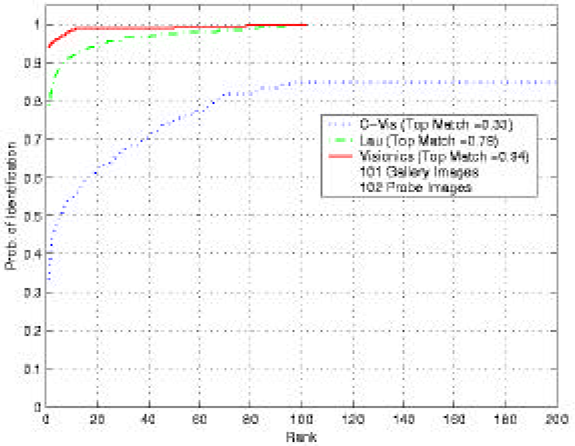

Figure 18: Identification scores for FRVT 2000 resolution experiment R1. Gallery eye separation ranged from 88 to 163 pixels, reduced to 60 pixels for probes. FRVT 2000 Evaluation Report image M-27.

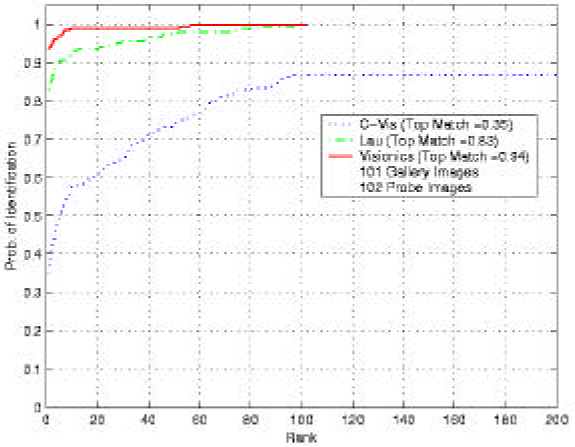

Figure 19: Identification scores for FRVT 2000 resolution experiment R2. Gallery eye separation ranged from 88 to 163 pixels, reduced to 45 pixels for probes. FRVT 2000 Evaluation Report image M-28.

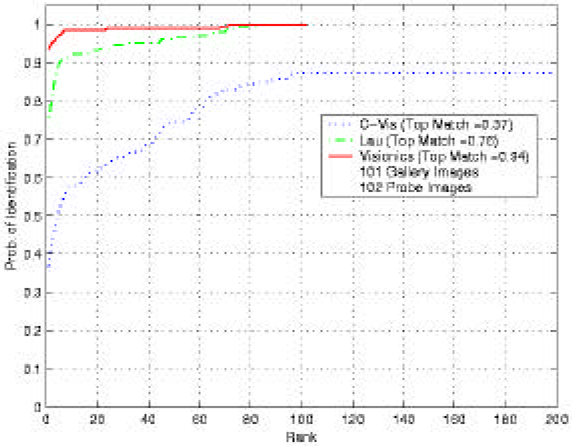

Figure 20: Identification scores for FRVT 2000 resolution experiment R3. Gallery eye separation ranged from 88 to 163 pixels, reduced to 30 pixels for probes. FRVT 2000 Evaluation Report image M-29.

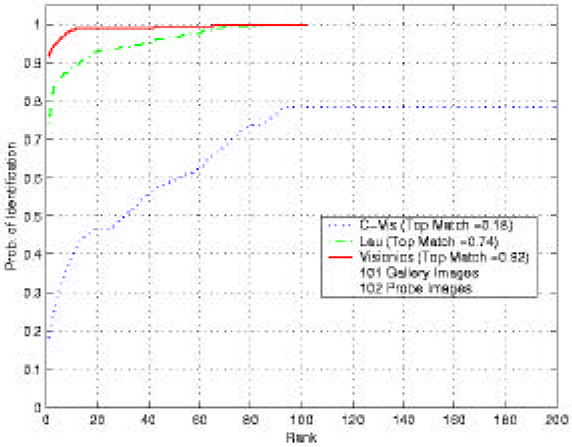

Figure 21: Identification scores for FRVT 2000 resolution experiment R4. Gallery eye separation ranged from 88 to 163 pixels, reduced to 15 pixels for probes. FRVT 2000 Evaluation Report image M-30.

Temporal: By the time a watchlist person makes it to the baggage claim area, some time may have elapsed since the mugshot image was taken. Figures 22 and 23 show the results for a gallery of 1,196 images taken up to 1,031 days before the probe images. The figures show that we can expect an 80% chance of finding the correct person in the top 10 matches. Figures 24—26 show the results for gallery sizes of 226—227 images with various lighting conditions taken up to 13 months after the probe images. We can expect a 75—80% chance of finding the correct person in the top 10 matches under these conditions.

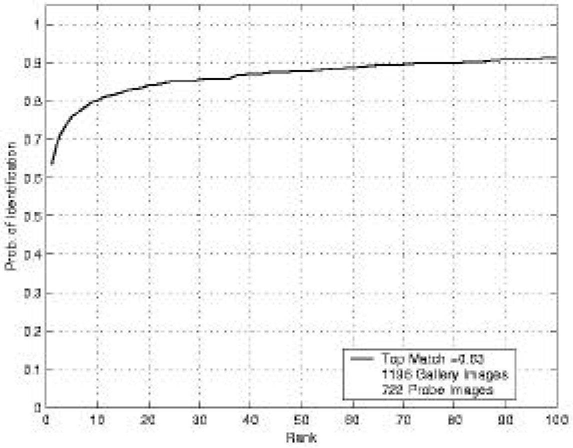

Figure 22: Best identification scores for FRVT 2000 temporal experiment T1. Gallery images were taken 0 to 1,031 days after the matching probes. FRVT 2000 Evaluation Report image M-10.

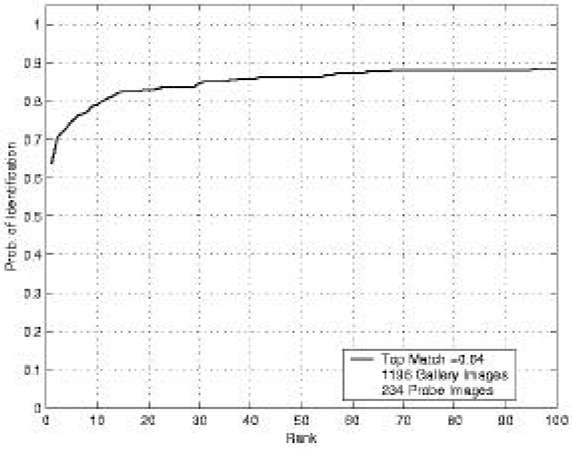

Figure 23: Best identification scores for FRVT 2000 temporal experiment T2. Gallery images were taken 540 to 1,031 days before the matching probes. FRVT 2000 Evaluation Report image M-11.

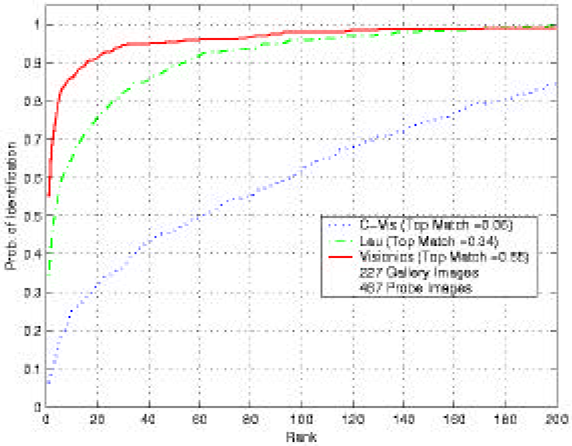

Figure 24: Identification scores for FRVT 2000 temporal experiment T3. Gallery images (Mugshot lighting) were taken 11 to 13 months after the matching probes. FRVT 2000 Evaluation Report image M-31.

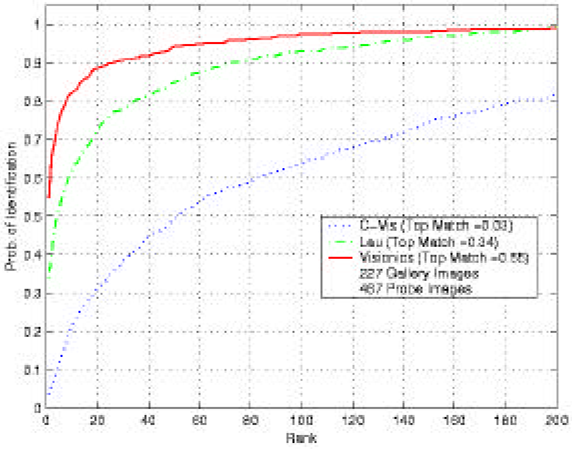

Figure 25: Identification scores for FRVT 2000 temporal experiment T4. Gallery images (FERET lighting) were taken 11 to 13 months after the matching probes. FRVT 2000 Evaluation Report image M-32.

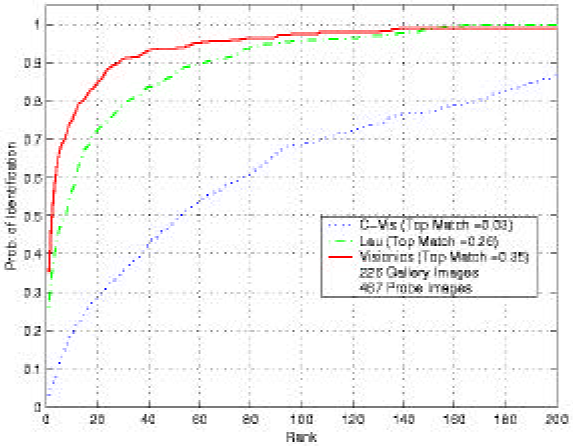

Figure 26: Identification scores for FRVT 2000 temporal experiment T5. Gallery images (Overhead lighting) were taken 11 to 13 months after the matching probes FRVT 2000 Evaluation Report image M-33.

If the results shown above suggest that the expected performance from the technology evaluation meets user requirements, a scenario evaluation can be used to further analyze the performance achievable for this application. This scenario is similar to the “indifferent” portion of the “Old Image Database Timed Test” of FRVT 2000. In the best result over all the tested vendors in the “Old Image Database Timed Test”, indifferent subjects were identified as the top match one-third of the time from a database of 135 subjects. If this proves to be acceptable performance, the authors recommend performing a more application-specific scenario evaluation followed by an operational evaluation on candidate systems.

Case 2: Persons obtaining travel documents are matched against those disembarking the conveyance

Suppose we have a mode of transportation where persons register to board then disembark some hours later at another location. We may wish to make certain that each of the disembarking passengers was the one registered for a particular boarding pass. We can notify travelers that their speedy handling will require cooperation with the imaging procedures during both registration and disembarkation. We wish to flag imposters 99% of the time, so the desired FAR is 1%. With these requirements in mind, we can use the verification results from the FRVT 2000 technology evaluation to analyze the expected performance.

Compression: The effects of compression may be a performance factor here as it was in the previous example. However, verification scores for compression experiments were not part of FRVT 2000.

Distance: In this example, it should be possible to ensure that the camera to subject distance for disembarkation is the same as that used for registration. Therefore, distance should not be a concern here.

Expression: Passengers can be expected to have different facial expressions during registration and disembarkation. Figures 27 and 28 show the results for gallery sizes of 224—225 images where facial expressions differ between probe and gallery images. We can expect a 5% FRR (1 valid passenger in 20 will be denied automatic access, requiring "exception handling") to correspond with an FAR of 1%.

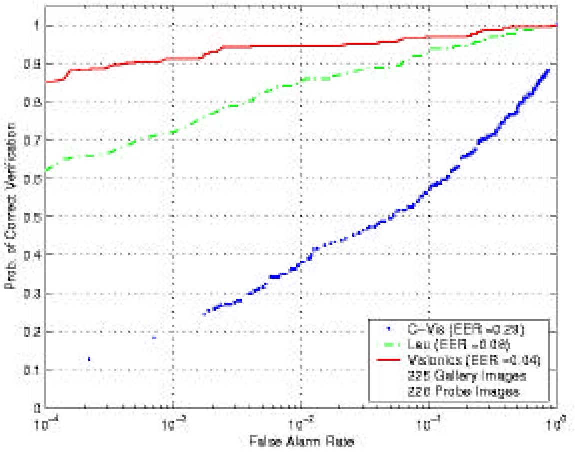

Figure 27: Verification scores for FRVT 2000 expression experiment E1. Regular expression used for gallery, alternate expression used for probes. FRVT 2000 Evaluation Report image M-41.

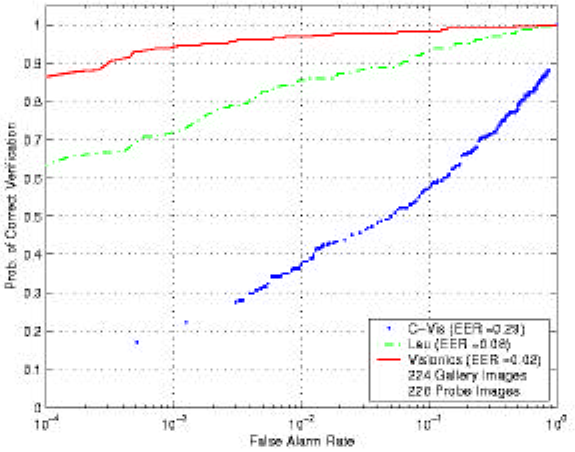

Figure 28: Verification scores for FRVT 2000 expression experiment E2. Alternate expression used for gallery, regular expression used for probes. FRVT 2000 Evaluation Report image M-42.

Illumination: It may not be possible to ensure that the lighting conditions are the same for registration and disembarkation, so illumination variations should be considered. Figures 29—31 show the results for galleries of size 129—227 images using probes with different lighting conditions. If we desire a FAR of 1%, we can expect a 5% FRR unless we are forced to use outdoor lighting at disembarkation, in which case we can expect the FRR to increase to 45% (1 valid passenger in 2 requiring “exception handling”).

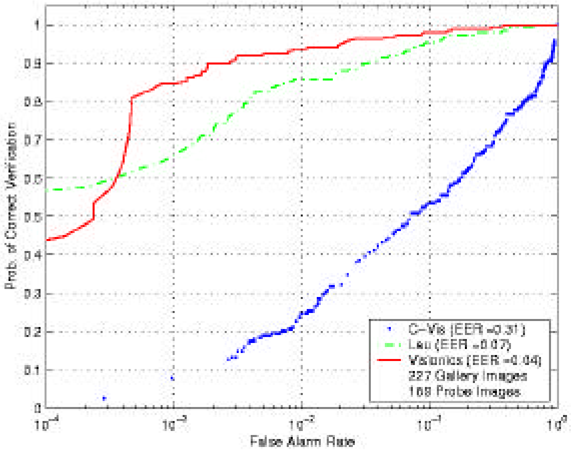

Figure 29: Verification scores for FRVT 2000 illumination experiment I1. Mugshot lighting used for gallery, overhead fluorescent lighting used for probes. FRVT 2000 Evaluation Report image M-43.

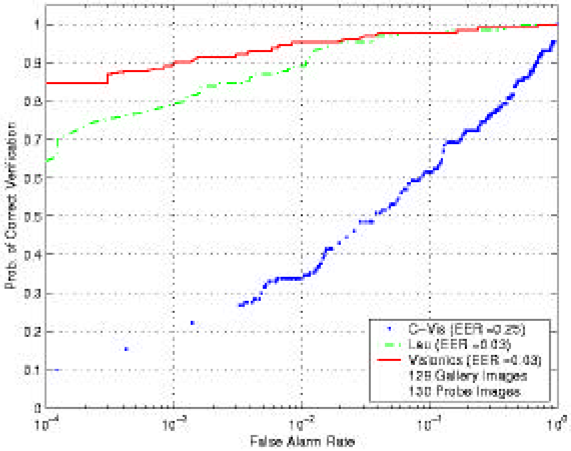

Figure 30: Verification scores for FRVT 2000 illumination experiment I2. Mugshot lighting used for gallery, single flood lamp used for probes. FRVT 2000 Evaluation Report image M-44.

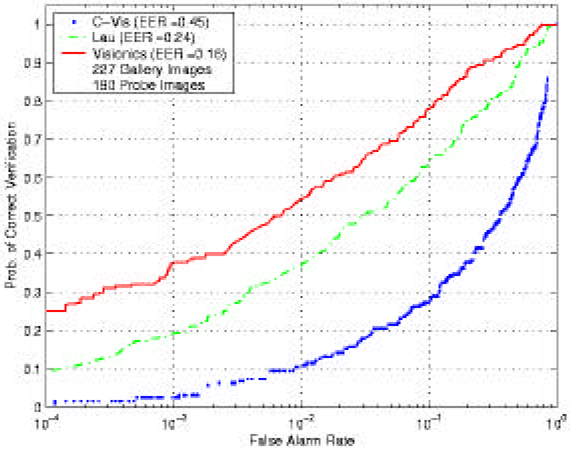

Figure 31: Verification scores for FRVT 2000 illumination experiment I3. Mugshot lighting used for gallery, outdoor lighting used for probes. FRVT 2000 Evaluation Report image M-45.

Media: We can assume that the same media is used for capturing registration and disembarkation images, so media should not be a performance factor for this example.

Pose: The pose angle should be similar for registration and disembarkation, so pose should not be a performance factor for this example

Resolution: Resolution should not be a factor here for the reasons stated in the distance analysis.

Temporal: Registration and disembarkation should take place without a significant time difference, so temporal effects should not be relevant here.

Assuming this performance is acceptable, we can look at the results of the “Enrollment Timed Test” which is similar to this example (assuming same camera system and lighting). In the verification portion of this scenario evaluation, two vendors successfully matched subjects in all but one trial (where both systems failed to acquire probe images) using a gallery of 135 images with no false matches.

If this proves to be acceptable performance, the authors recommend performing a more application-specific scenario evaluation followed by an operational evaluation on candidate systems.

Conclusions

In this paper, we have presented a portion of the results from the Facial Recognition Vendor Test 2000 and discussed the use of facial recognition in applications of potential interest to the supply-side drug control community. We have shown that, at the current state of the art, facial recognition technology can perform many tasks valuable to our community when human assistance is available to make final matches from the list of top candidates or to resolve verification errors.

Choosing which facial recognition system to use for each application is not a trivial task. Those tasked to investigate facial recognition for a specific application should follow the three step process (technology evaluation, scenario evaluation, and operational evaluation) that was illustrated in the case studies section of this paper.

References

[1] D. Blackburn, M. Bone, and P. J. Phillips. "Facial Recognition Vendor Test 2000 Evaluation Report." February, 2001. Available online at: http://www.dodcounterdrug.com/facialrecognition/FRVT2000/documents.htm.

[2] P.J. Phillips, *et al*, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *Proc. IEEE Conf.on Comp.Vis.and Patt. Recog.*, San Juan, Puerto Rico, June 1997

[3] S.A. Rizvi, *et al*, "The FERET Verification Testing Protocol for Face Recognition Algorithms", NIST, NISTIR 6281, October 1998

[4] P.J. Phillips, *et al*, "The FERET Evaluation" in H. Wechsler, etal (eds) Face Recognition: From Theory to Applications (Springer-Verlag, Berlin, 1998)

[5] P.J. Phillips, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms", *Image and Vision Computing Journal*, Vol. 16, No.5, 1998, pg. 295-306

[6] P.J. Rauss, *et al*, "FERET (Face-Recognition Technology) Recognition Algorithms", *Proc. of ATRWG Science and Technology Conference*, July 1996

[7] A. Pentland and T. Choudhury, "Face recognition in smart environments", *IEEE Computer*, Vol 33, No. 2, February, 2000, pg. 50-55.

[8] L.D. Harmon, M.K. Khan, R. Lasch, and P.F. Ramig, "Machine Recognition of Human Faces", *Pattern Recognition*, Vol. 31, No. 2, 1981, pg. 97-110

[9] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey", Pattern Recognition, Vol. 25, 1992, pg. 65-77

[10] R. Chellappa, C.L. Wilson, S. Sirohey, "Human and machine recognition of faces: A survey", *Proc. IEEE*, Vol. 83, No.5, 1995, pg. 705-740

[11] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces", Journ. Optical Soc. Am, Vol. 4, 1987, pg. 519-524

[12] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journ. of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pg. 71-86

[13] J. Zhang, Y. Yan, M. Lades, "Face recognition: Eigenface, elastic matching and neural nets", *Proc. IEEE*, Vol. 85, No. 9, 1997, pg. 1423-1436

[14] J.L. Wayman, "Fundamentals of biometric authentication technologies", to appear in *Int. Journ. of Imaging and Graphi*cs, Vol. 1, No.1, 2001

[15] J.L. Wayman, "Technical Testing and Evaluation of Biometric Identification Devices" in A. Jain, *et al* (eds.) Biometrics: Personal Security in Networked Society, (Kluwer Academic Press, 2000)

[16] United Kingdom Biometric Working Group, "Best Practices in Testing and Reporting Biometric Device Performance", version 1.0, online at www.afb.org.uk/bwg/bestprac10.pdf

[17] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. An Introduction to Evaluating Biometric Systems. *IEEE Computer*, Vol 33, No. 2, February, 2000, pg. 56-63.

[18] D. Maio, D. Maltoni,, J. Wayman and A. Jain, "FVC2000: Fingerprint Verification Competition 2000", *Proc. 15th International Conference on Pattern Recognition*, Barcelona, September 2000, ", available on-line at www.csr.unibo.it/research/biolab

[19] P.J. Philips, "On performance statistics for biometric systems", Proc. AutoID'99, Summit, NJ, 28-29 October, 1999, pg. 111-116

[20] J.L. Wayman, "Error Rate Equations for the General Biometric System", *IEEE Robotics and Automation*, Vol. 6, No. 1, March 1999, pg. 35-48