The Design and Implementation of the Facial Recognition Vendor Test 2000 Evaluation Methodology

Duane M. Blackburn

Thesis submitted to the Faculty of
the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical Engineering

Douglas Lindner, Chair
Lynn Abbott
Doug Nelson

May 8, 2001
Blacksburg, Virginia

Keywords: Biometrics, Facial Recognition, Evaluation

The Design and Implementation of the Facial Recognition Vendor Test 2000 Evaluation Methodology
(Abstract)

Duane M. Blackburn

The biggest change in the facial recognition community since the completion of the FacE REcognition Technology (FERET) program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and significantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications?

Repeated inquiries from numerous government agencies on the current state of facial recognition technology prompted the DoD Counterdrug Technology Development Program Office to establish a new set of evaluations. The Facial Recognition Vendor Test 2000 (FRVT 2000), was co-sponsored by the DoD Counterdrug Technology Development Program Office, the National Institute of Justice, and the Defense Advanced Research Projects Agency, and was administered in May-June 2000.

The sponsors of the FRVT 2000 had two major goals for the evaluation. The first was a technical assessment of the capabilities of commercially available facial recognition systems. The sponsors wanted to know the strengths and weaknesses of each individual system, as well as obtain an understanding of the current state of the art for facial recognition.

The second goal of the evaluation was to educate the biometrics community and the general public on how to present and analyze results. The sponsors have seen vendors and would-be customers quoting outstanding performance specifications without understanding that these specifications are virtually useless without first knowing the details of the test that was used to produce the quoted results.

The Facial Recognition Vendor Test 2000 was a worthwhile endeavor. It will help numerous readers evaluate facial recognition systems for their own uses and will serve as a benchmark for all future evaluations of biometric technologies.

The FRVT 2000 evaluations were not designed, and the FRVT 2000 Evaluation Report was not written, to be a buyer’s guide for facial recognition. No one will be able to open the report to a specific page to determine which facial recognition system is best because there is not one system for all applications. The only way to determine the best facial recognition system for any application is to follow the three-step evaluation methodology described in the FRVT 2000 Evaluation Report and analyze the data as it pertains to each individual application.
This thesis explains the design and implementation of the FRVT 2000 evaluations, and discusses how the FRVT 2000 Evaluation Report met the author’s objectives for the evaluation.

Acknowledgments

The author would like to thank Dr. Doug Lindner, Dr. Lynn Abbott, and Dr. Doug Nelson for their willingness to serve as my committee, and for their review and feedback throughout the FRVT 2000 evaluations and the preparation of this document.

Table of Contents

Appendices (continued)

List of Figures

List of Tables

1.0 Introduction

The biggest change in the facial recognition community since the completion of the FacE REcognition Technology (FERET) program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and significantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications? This is one of the questions potential users most frequently asked the sponsors and the developers of the FERET program.

Although literature research had found several examples of recent system tests, none had been both open to the public and of a large enough scale to be completely trusted. This revelation, combined with inquiries from other government agencies on the current state of facial recognition technology, prompted the DoD Counterdrug Technology Development Program Office, the Defense Advanced Research Projects Agency (DARPA), and the National Institute of Justice (NIJ) to sponsor the Facial Recognition Vendor Test 2000 (FRVT 2000).

This thesis describes the work performed by the author as he led the development, evaluation, and reporting efforts throughout the FRVT 2000 evaluations. The evaluation methodology, along with previous related work, will first be explained. Results from the evaluations will then be explained and briefly analyzed. Finally, the appendices of this thesis will provide in-depth documentation of the numerous steps undertaken during FRVT 2000. The significance of this documentation will be explained throughout this thesis.

1.1 Overview of Biometrics

Biometrics are automated methods of recognizing a person based on physiological or behavioral characteristics. Common examples include:

- Fingerprint
- Hand Geometry
- Eye - Retinal
- Eye - Iris
- Facial Recognition
- Speaker
- Dynamic Signature

The operation of biometric devices can be explained via a three-step procedure.

> First, a sensor takes an observation. The type of sensor and its observation depend on the type of biometrics device used. This observation gives us a “Biometric Signature” of the individual.

Second, a computer algorithm normalizes the biometric signature so that it is in the same format (size, resolution, etc.) as the signatures on the system's database. The normalization of the biometric signature gives us a "Normalized Signature" of the individual.

Third, a matcher compares the normalized signature with the set (or sub-set) of normalized signatures on the system's database and provides a "similarity score" that compares the individual's normalized signature with each signature in the database set (or sub-set). What is then done with the similarity scores depends on the biometric system's application.

A glossary of biometric terms used throughout this thesis is available in Appendix A.

1.2 Need for the evaluation

The three sponsoring agencies decided to perform this evaluation for two main reasons. The first was to assess the capabilities of facial recognition systems that are currently available on the open market. The last formal government evaluation of facial recognition technology was the final series of evaluations during the FERET program in 1997. Those evaluations were performed primarily on university/research algorithms that were partially automatic. By 2000, these algorithms had migrated into becoming commercial, fully automated, facial recognition systems. There are three goals in accessing commercially available systems: to provide a comparison of commercially available systems, to provide an overall assessment of the state of the art in facial recognition technology, and to measure progress made since the conclusion of the FERET program.

Although there had not been any formal government evaluations of facial recognition technology since 1997, there were several smaller evaluations that were performed by varying entities. Unfortunately, these evaluations were flawed in one or more ways because of poor evaluation protocols, poor adherence to these protocols, non-existent documentation of the evaluations, and inherent bias by the entities performing the evaluations. This revelation provided the second purpose for performing this evaluation - to design, implement, and document a proper evaluation that would become the standard that all future evaluations will be compared against. This goal can be met by, for the first time, developing a thorough evaluation protocol, following the protocol, and then documenting the entire evaluation process rather than the industry standard approach of showing only the results This has numerous benefits. First, it allows others to understand the resources that would be required to run their own evaluation. Second, it sets a precedent of openness for all future evaluations. Third, it allows the community to discuss how the evaluation was performed and what modifications to the evaluation protocol could be made so that future evaluations are improved.

1.3 Author's Contribution

The author of this thesis assembled and led the FRVT 2000 Evaluation Team. A summary of tasks performed by the author are described below.

- **Establish evaluation team.** The author assembled a team of individuals to perform the FRVT 2000 evaluations. This team consisted of the author, who was the team leader, Mr. Mike Bone (NAVSEA Crane Division), Mr. Patrick Grother (NIST), and Dr. P. Jonathon Phillips (DARPA).
- **Write evaluation description document.** To start the project, the author wrote an evaluation description document that explained why we were performing the evaluation, how it was to be setup and performed, and what the final report should include. The author gave this document to the members of the evaluation team at the project onset so that we would all understand the entire project and could work together to achieve the goals that the author had set forth.
- **Gather images.** A large part of these evaluations was the gathering of images so that we could perform useful experiments. The author designed, setup, and ran the first image collection at Dahlgren and assisted in the second image collection at NIST. The image collection at NIST, as well as subsequent sessions for the DARPA HumanID program, have used the design and setup of the author's initial image gathering at Dahlgren.
- **Develop web site.** One of the key aspects of a successful evaluation is sharing information with those outside the project. To accomplish this, the author designed the FRVT 2000 website that was in place throughout the evaluations. This website was eventually replaced by another website the author developed that explains both the FRVT 2000 evaluations and the FERET program.
- **Announce evaluation to the public**. The author wrote an evaluation announcement and sent it to several discussion lists that have members who would be interested in the evaluations.
- **Release test plan and sample images**. The author made available to the participants a copy of the test plan and sample images before the evaluations took place.
- **Analyze results.** The author, as well as Mr. Grother and Dr. Phillips, analyzed results of various experiments for several months to determine if anomalies existed and to make sure the results we were to release were indeed correct.
- **Write FRVT 2000 Evaluation Report**. The author, with the exception of preparing the final version of the bar charts, ROC curves, and CMC curves, wrote the FRVT 2000 Evaluation Report.
- **Release report to the public**. After completing the FRVT 2000 Evaluation Report and obtaining permission to release to the public, the author placed the report on the FRVT 2000 website and announced it to the public using several methods.

The FRVT 2000 Evaluation Report, described in this thesis, is the first truly open and completely documented evaluation of biometric technologies. The actual results of the evaluation are of secondary importance. The true impact of this work will be improvements in future evaluations of biometric technology as others will be pressured to perform their evaluations to the standard shown in the FRVT 2000 Evaluation Report.

2.0 Review of Prior Work

The author talked with numerous government agencies and several members of the biometrics community, including facial recognition vendors, to determine if this evaluation should be, and how it should be, performed. The overwhelming response was to proceed with the evaluation. Government agencies and the biometrics community wanted to know if the facial recognition vendors could live up to their claims, which systems performed best in certain situations, and what further development efforts would be needed to advance the state of the art for other applications. Unofficially, the vendors wanted to have an evaluation to prove that they had the best available product. Everyone cited the FERET program because it is the de facto standard for evaluating facial recognition systems, but they also stressed the need to have a live evaluation. The author took this information and analyzed different methods to evaluate facial recognition systems. Three items had a profound effect on the development of the FRVT 2000 evaluation methodology:

- An Introduction to Evaluating Biometric Systems, P. J. Phillips, A. Martin, C. L. Wilson, M. Przybocki, IEEE Computer, February 2000, p. 56-63.
- The FERET program.
- A previous scenario evaluation of a COTS facial recognition system.

2.1 An Introduction to Evaluating Biometric Systems

The author was part of the pre-publication review team for an article eventually published as "An Introduction to Evaluating Biometric Systems" in IEEE Computer, Feb 2000. The author studied this article in great depth and used it's ideals to develop the FRVT 2000 evaluation methodology.

The first ideal presented in the article is that successful evaluations must be administered by independent groups that will not reap any benefits should one system outperform the other. If this is not the case, conflicts of interest, even if only perceived, will doom the evaluation. Sometimes these "independent groups" may actually be funded by vendors or entities with alliances to particular vendors.

The second ideal is to use test data that has not been previously seen by any of the systems. Over time, system developers will learn the properties of previously seen test data and can tune their systems for maximum performance.

A third ideal is that the evaluation itself must not be too easy, nor too difficult. If the evaluation is too easy, all of the systems will perform well and will be grouped together at one end of the capabilities spectrum. If the evaluation is too difficult, none of the systems will perform well and will be grouped together at the other end of the capabilities spectrum. In either case, the evaluation will fail to produce results that will distinguish one system from another.

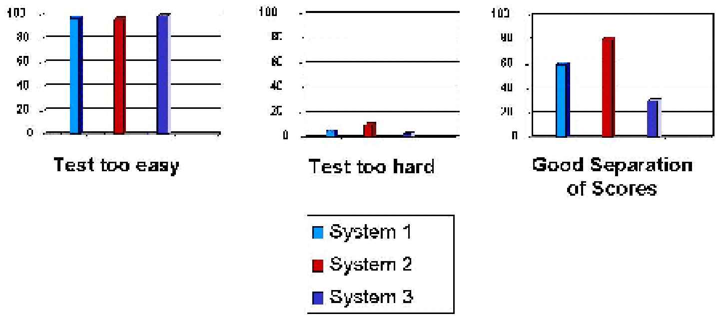


Figure 1 - Difficulty of Evaluation

A fourth ideal is that the evaluation itself must be repeatable and made available to the technical and practitioner communities. This requires extensive documentation about all phases of the evaluation including the gathering of test data, the evaluation protocol, the testing procedures, performance results, and examples of the test data. There are two reasons to document all of these phases. The first is so that the technical and practitioner communities accept the validity of the evaluation. More vendors will be willing to participate in an evaluation if the process is described beforehand. They will know that those performing the evaluation understand what they are doing, and thus will not be afraid of having to answer questions about a poorly designed evaluation. The second reason to document all of these phases is so evaluators, as well as other readers, are able to accurately determine how each presented result was obtained. Irregularities in test results can often be explained via a thorough analysis of the test protocol. Documenting the results along with the test protocol allows others to improve the evaluation protocol for future evaluations.

A final ideal, which was not taken from this article, is that you need to understand the true requirements for your application in order to determine if results from any evaluation show whether or not a technology investment is warranted. There is a difference between true requirements and desired requirements. In the majority of requirement definitions that are not performed by a combination of technologists that are knowledgeable in the subject area and practitioners that are familiar with the activity, these are incorrectly mixed together. You may have a desired goal of 90% on some measurable. If you are currently only obtaining 15 to 20% using existing methods, then an improvement to only 30 to 40% could easily make up the costs incurred because of the addition of the technology. This would be your true requirement. If your evaluations show the technology can hit at a 70% rate and you are measuring against your desired requirement, it would appear as if the technology would fail miserably. However, if you use your true requirement, the evaluations would show that the addition of technology would be a resounding success!

In addition to the ideals presented above, the article "An Introduction to Evaluating Biometric Systems" provides a structured approach to a complete evaluation that moves from the general to the specific through three major steps: a Technology Evaluation, a Scenario Evaluation, and an Operational Evaluation.

The most general type of evaluation is a Technology Evaluation. The goal of the Technology Evaluation is to determine the underlying technical capabilities of the systems for a particular technology, in this case facial recognition. The testing is performed in laboratories using a standard set of data that was collected by a universal sensor. In the vast majority of technologies, the same data can and should be used as input for each system. Technology Evaluations are always completely repeatable. Technology Evaluations typically take a short time to complete, depending on the type of technology being evaluated

The next step in the structured evaluation process is a Scenario Evaluation. Scenario evaluations aim to evaluate the overall capabilities of the entire system in a specific scenario, rather than a subset of the system which was evaluated in the Technology Evaluation. For example, in evaluating facial recognition systems the Technology Evaluation would study the face recognition algorithms only but the Scenario Evaluations studies the entire system – including camera and camera-algorithm interface, in a given scenario. Each tested system would have its own acquisition sensor and would thus receive slightly different data. Scenario Evaluations are not always completely repeatable for this reason, but the approach used can always be completely repeatable. Scenario Evaluations typically take a few weeks to complete because multiple trials, and for some Scenario Evaluations, multiple trials of multiple subjects/areas, must be completed.

The most specific step in the structured evaluation process is an Operational Evaluation, which is very similar to a Scenario Evaluation except that it is performed at the actual site, using the actual subjects/areas. Operational Evaluations are not very repeatable unless the actual operational environment naturally creates repeatable data. Operational Evaluations typically last from several weeks to several months.

The three steps described in this structured evaluation process not only flow from the general to the specific, but also flow from one to another. Technology Evaluations are performed on all applicable technologies that could conceivably meet your requirements. Results from the Technology Evaluations will be of immediate interest to the vendors as well as the evaluators. The Technology Evaluation results will provide the vendors a direction towards what new developments they will need to undertake to improve their product. The evaluator can then select a subset of these technologies for a Scenario Evaluation. Once the Scenario Evaluation has been completed, the evaluator can select one, or possibly two, systems for an extended Operational Evaluation at the actual site. If the Operational Evaluation is successful, the evaluator can then make the decision to implement the technology permanently on site(s).

There are multiple reasons to not skip the Technology Evaluation and go straight for a Scenario Evaluation, or even worse, an Operational Evaluation. The first reason is that you would be selecting technology based on a whim rather than scientific analysis. The second reason is that if a Scenario Evaluation is successful, or fails, we will not truly understand why unless we have the technical information from the Technology Evaluations to analyze with the Scenario Evaluation results. The FRVT 2000 evaluation provided an example of this. There were two different, yet similar, scenarios tested and the results from each evaluation varied wildly. There were two main variances in the setup of the Scenario Evaluations. Data from the Technical Evaluation showed that one of the setup variances did not contribute to the result variances. The other setup variance, therefore, must have been the culprit. Without having the data from the Technology Evaluation, no one would have been able to determine where the limitation in the systems occurred.

This three-step evaluation plan has also been adopted by Great Britain's Best Practices in Testing and Reporting Performance of Biometric Devices. This report can be found at http://www.afb.org.uk/bwg/bestprac10.pdf.

2.2 The FERET Program

The DoD Counterdrug Technology Development Program Office began the FacE REcognition Technology (FERET) program in 1993. The program consists of three important parts:

- Sponsoring research.
- Collecting the FERET database.
- The FERET evaluations

FERET-sponsored research was instrumental in moving facial recognition algorithms from concept to reality. Many commercial systems still use concepts that were involved in the FERET program.

The FERET database was designed to advance the state of the art in facial recognition, with the images collected directly supporting algorithm development and the FERET evaluations. The database is divided into a development set, which was provided to researchers, and a set of images that was sequestered. The sequestering was necessary so that additional FERET evaluations and future evaluations such as the FRVT 2000 could be administered using images that researchers have not previously used with their systems. If previously used images are used in an evaluation, it is possible that researchers may tune their algorithms to handle that specific set of images. The FERET database contains 14,126 facial images of 1,199 individuals. Before the FRVT 2000, only one-third of the FERET database had ever been used by anyone outside the government.

The final and most recognized part of the FERET program was the FERET evaluations that compared the abilities of facial recognition algorithms using the FERET database. Three sets of evaluations were performed in August 1994, March 1995, and September 1996.

A portion of the FRVT 2000 was based very heavily on the FERET evaluation. Numerous images from the unreleased portion of the FERET database, the scoring software, and baseline facial recognition algorithms for comparison purposes were used in FRVT 2000. The FERET program also provided insight into what the author should expect from participants and outside entities before, during, and after the evaluations.

2.3 A Previous Scenario Evaluation for a COTS Facial Recognition System

In 1998, the DoD Counterdrug Technology Development Program Office was asked to study the feasibility of using facial recognition at an access control point in a federal building. The technical agents assigned from NAVSEA Crane Division studied the layout and arranged a scenario evaluation for a facial recognition vendor at their facilities. The selected vendor brought a demonstration system to NAVSEA Crane, set it up, and taught the technical agents how to use the system.

A subject was enrolled into the system according to the procedures outlined by the vendor. During the evaluation, the technical agent entered the subject's ID number into the system, which was configured for access control (verification) mode. A stopwatch was used to measure the recognition time starting with the moment the ID number was entered and ending when the subject was correctly identified by the system. The resulting time, measured in seconds, was recorded in a table. This timed test was repeated at several distances with the subject being cooperative and indifferent. System parameters were also varied incrementally from one extreme to the other. The methodology of the evaluation was never explained to the vendor.

When the system was returned to the vendor, they looked at the system settings for the final iteration of the timed test and immediately complained that NAVSEA Crane had not tested the system at an optimal point. They offered to return to NAVSEA Crane with another system so they could retest using the vendor's own test data and test plan and then write a report that the sponsors could use instead of the sponsor-written evaluation report. The invitation was not accepted because the proposed effort had been canceled for other reasons.

The author learned several lessons from this simple evaluation. The first was how to develop a scenario evaluation and improve on it for future evaluations such as the FRVT 2000. The second lesson was the importance of being completely candid about the evaluation plan so the vendor is less inclined to dispute its validity after the evaluation. The final and most important lesson was to continue to let a non-biased sponsor run the evaluations, but allow a vendor representative to run their own machines and set the system parameters under the sponsor's supervision. Because the sponsor, rather than the vendor representative, ran the system during this previous evaluation, it gave the vendor an opportunity to blame poor results on operator error rather than the system. All three lessons were used to develop the evaluation methodology for the FRVT 2000.

3.0 Evaluation Methodology Overview

The Facial Recognition Vendor Test 2000 was divided into two evaluation steps: the Recognition Performance Test and the Product Usability Test. The FRVT 2000 Recognition Performance Test is a technology evaluation of commercially available facial recognition systems. The FRVT 2000 Product Usability Test is an example of a scenario evaluation, albeit a limited one.

3.1 Technology Evaluation: Recognition Performance Test

Each vendor was given a set of 13,872 images to process. They were instructed to compare each image with itself and with all other images, and return a matching score for each comparison. The matching scores were stored in similarity files that were returned to the test agent along with the original images. Each vendor was given 72 continuous hours to process the images. Some vendors were able to process the entire set of images, while others were only able to process a subset of the images in the allotted time. At the conclusion of the test, each vendor's hard disk was wiped to eliminate the images, similarity files, and any intermediate files.

After all testing activities were complete, the similarity files were processed using the scoring software. The images were divided into different probe and gallery sets to test performance for various parameters such as lighting, pose, expression and temporal variation.

3.2 Scenario Evaluation: Product Usability Test

The scenario chosen for the Product Usability Test was access control with live subjects. The Product Usability Test was administered in two parts: the Old Image Database Timed Test and the Enrollment Timed Test. For the Old Image Database Timed Test, vendors were given a set of 165 images captured with a standard access control badge system, including one image of each of three test subjects. The set contained two images for five people, and one image for each of the other 155 people. Vendors enrolled these images into their system for comparison with the live subjects. The operational scenario was that of a low-security access control point into the lobby of a building. The building's security officers did not want to mandate that the employees take the time to enroll into the new facial recognition system so they used their existing digital image database taken from the employee's picture ID badges.

For the Enrollment Timed Test, the images of the three test subjects were removed from the system while the other images were retained. Vendors were then allowed to enroll the three subjects using their standard procedures, including the use of multiple images. The purpose of the test was to measure system performance using vendor enrollment procedures. The operational scenario was that of an access control door for a medium-to-high security area within the building previously described. In this case, employees were enrolled in the facial recognition system using the standard procedures recommended by the vendor.

During the Product Usability Test, several parameters were varied including start distance, behavior mode, and backlighting. Tests were performed for each subject at distances of 12, 8, and 4 feet for all trials except for the variability test. Test subjects performed each test using both a cooperative mode and a simulated, and thus repeatable, indifferent behavior mode. For the cooperative mode, subjects looked directly at the camera for the duration of the trial. For the indifferent mode, subjects instead moved their focus along a triangular path made up of three visual targets surrounding the camera. Each trial was performed with and without backlighting provided by a custom light box.

3.3 Information Sharing Plan

Previous evaluations of biometric technology were flawed in one or more ways because of poor evaluation protocols, poor adherence to these protocols, poor documentation of the evaluations, and inherent bias by the entities performing the evaluations. This revelation provided the second purpose for performing this evaluation - to design, implement, and document a proper evaluation that would become the standard that all future evaluations will be compared against. This goal can be met by, for the first time, developing a thorough evaluation protocol, following the protocol, and then documenting the entire evaluation process rather than the industry standard approach of showing only the results.

An overview of the evaluation protocol for the FRVT 2000 was described in the previous two sections. In many prior biometric evaluations, this level of detail was all that was provided either before or after the evaluation. This practice is flawed as potential vendors would like to understand more details about an evaluation before deciding to participate in it, there is no test plan for use during the evaluations, and there is not enough documentation for other individuals to fully understand what any published results mean.

The author decided that the FRVT 2000 evaluations would set a precedent of documentation for before, during, and after the evaluations. A web site for the Facial Recognition Vendor Test 2000 was created as the primary method for sharing information among vendors, sponsors and the public about the evaluation. A copy of the web site is available in Appendix B. The web site was divided into two areas - public and restricted. The public area contained the following pages.

Home Page. Menu for subsequent pages.

Overview. Provided the main description of the evaluation including an introduction, discussions on participant qualifications, release of the results and test make-up. This page also provided reports from the latest FERET evaluations. The author decided it was important for not only the participants in the FRVT 2000, but also the general biometrics community to understand the purpose and format of the evaluations.

Upcoming Dates. Provided a list of important dates and their significance in the evaluation.

Sponsors. Described the various agencies that either sponsored or provided assistance for the FRVT 2000. POCs for each agency and hyperlinks to the agency's web site were provided. In many prior biometrics evaluations, the individuals and agencies/companies performing the evaluations were for some reason hidden. The author decided to let the entire community know who was performing the evaluations so that they would be able to determine the validity of the evaluation, as well as the results to be presented.

Points of Contact (POCs). Listed for test-specific questions, media inquiries and for all other questions. The author decided to let the entire community know who was performing the evaluations so that they would be able to determine the validity of the evaluation, as well as the results to be presented.

How to Participate. Discussed how a vendor would request to participate in the evaluation. Providing this information up-front and openly showed that these evaluations would not be biased towards any particular system and would be open to anyone wishing to participate.

Frequently Asked Questions (FAQ). Established to submit questions and read the responses from the evaluation sponsors. This section enabled the biometrics community to ask questions about the evaluation while it was being performed. Both the questions and the answers were available to the entire biometrics community.

Forms. Online forms to request participation in the evaluation and for access to portions of the FERET and HumanID databases. The forms were used to show that the participating vendors understood the methodology of the evaluations before the actual tests were performed and to clear up any legal issues. The forms were available to anyone to view, whether participating in the evaluations or not.

Participating Vendors. Provided a list of the vendors that are participating in the evaluation, a hyperlink to their web sites, and point-of-contact information. This enabled the biometrics community to see who was participating in the evaluations and how to contact them.

The restricted area of the FRVT 2000 web site was encrypted using 128-bit SSL encryption. Access was controlled using an ID and password provided to participating vendors and sponsors. The purpose of the restricted area was to provide a forum for discussion between the sponsors and the participating vendors without having to be concerned about the impact of any statements on the community's perception of the participating vendors and the overall facial recognition community. The restricted site also enabled the sponsors to share information with the participating vendors that they didn't want made publicly available until after the evaluation results were released. The author's fear in releasing some of these documents to the public prior to the evaluation report was that someone would use the information in these documents incorrectly before the report was released. The restricted area contained the following pages.

Test Plan. Provided the detailed test plan for the evaluations. This test plan showed step-by-step the procedures that would be used during the evaluations. The test plan was provided to the participating vendors one month before the evaluations took place and was followed for each of the vendor evaluations. A copy of the test plan is provided in Appendix C.

Application Programmer's Interface (API). Provided the application API document that shows how the vendor's similarity files would need to be written so that the vendor similarity files could be computed using the sponsor's scoring software. The API document was made available in both HTML and PDF formats.

Images. Provided the Facial Recognition Vendor Test 2000 Demonstration Data Set, which consisted of 17 facial images in one compressed (zip) file. The purpose of the Demonstration Data Set was to give the vendors an idea of the types of images they would encounter in the evaluations and to provide a data set for them to use to make sample similarity for sponsor review prior to the actual evaluations. The number of images in the Demonstration Data Set was chosen to enable the vendors to make a set of similarity files that was large enough to check for accuracy, but not large enough for the vendors to be able to tune their algorithms for this data set. Samples of the FRVT 2000 Demonstration Data Set is provided in Appendix D.

FAQ. This page was established for the participating vendors to submit questions and to read the responses from the evaluation sponsors. The restricted area FAQ was more test-specific in nature than the public area FAQ which focused on the overview of the evaluation. The sponsors encouraged all participating vendor questions related to the evaluation to be submitted via the restricted area FAQ page. In many cases, the sponsors refused to answer questions unless they were submitted through the restricted area FAQ pages. This practice maintained an even playing field amongst the vendors as no one would have any information that the others did not also have.

4.0 Results & Discussion

4.1 Technology Evaluation: Recognition Performance Test

4.1.1 Overview

Each vendor was given a set of 13,872 images to process. They were instructed to compare each image with itself and with all other images, and return a matching score for each comparison. The matching scores were stored in similarity files that were returned to the test agent along with the original images. Each vendor was given 72 continuous hours to process the images. Three vendors were able to process the entire set of images, while two others were only able to process a subset of the images in the allotted time.

After all testing activities were complete, the similarity files were processed using the scoring software. The images were divided into different probe and gallery sets to test performance for various parameters such as lighting, pose, expression and temporal variation. The results for each of these probe and gallery sets were reported in the FRVT 2000 Evaluation Reports in quick-look bar charts that highlight key results as well as the full receiver operator characteristic (ROC) and cumulative match characteristic (CMC).

4.1.2 Recognition Performance Test Results Format

Biometric developers and vendors will, in many cases, quote a false acceptance rate (sometimes referred to as the false alarm rate) and a false reject rate. A false acceptance (or alarm) rate (FAR) is the percentage of imposters (an imposter may be trying to defeat the system or may inadvertently be an imposter) wrongly matched. A false rejection rate (FRR) is the percentage of valid users wrongly rejected. In most cases, the numbers quoted are quite extraordinary. They are, however, only telling part of the story.

The false acceptance rate and false rejection rate are not mutually exclusive. Instead, there is a give-take relationship. The system parameters can be changed to receive a lower false acceptance rate, but this also raises the false rejection rate and vice versa. A plot of numerous false acceptance rate-false rejection rate combinations is called a receiver operator characteristic curve. A generic ROC curve is shown in figure 2. The probability of verification on the y-axis ranges from zero to one and is equal to one minus the false reject rate. The false acceptance (or alarm) rate and the false reject rate quoted by the vendors could fall anywhere on this curve and are not necessarily each other's accompanying rate. Some spec sheets also list an equal error rate (EER). This is simply the location on the curve where the false acceptance rate and the false reject rate are equal. A low EER can indicate better performance if one wants to keep the FAR equal to the FRR, but many applications naturally prefer a FAR/FRR combination that is closer to the end points of the ROC curve. Rather than using EER alone to determine the best system for a particular purpose, one should use the entire ROC curve to determine the system that performs best at the desired operating location. The ROC curve shown in figure 2 uses a linear axis to easily show how the equal error rate corresponds to the false acceptance and false reject rate. The ROC curves provided in the FRVT 2000 Evaluation Report use a semi-log axis so that low-false-alarm rate results can be viewed. The equal error rates are listed as text on the graphs.

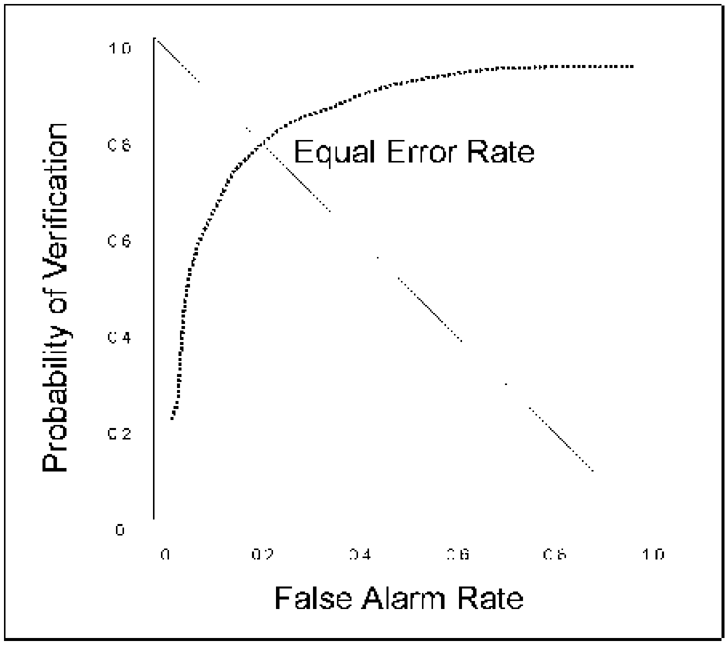


Figure 2 - Generic ROC Curve

Although an ROC curve shows more of the story than a quote of particular rates, it will be difficult to have a good understanding of the system capabilities unless one knows what data was used to make these curves. An ROC curve for a fingerprint system that obtained data from coal miners would be significantly different than one that obtained data from office workers. Facial recognition systems differ in the same way. Lighting, camera types, background information, aging and other factors would each impact a facial recognition system's ROC curve. For the Facial Recognition Vendor Test 2000, participating vendors compared 13,872 images to one another. These images can be subdivided into different experiments to make an ROC curve that shows the results of comparing one type of image to another type of image. Section 4.1.3 describes the different experiments that were reported.

The above description is valid for displaying verification results. In a verification application, a user claims an identity and provides their biometric. The biometric system compares the biometric template (the digital representation of the user's distinct biometric characteristics) with the user's stored (upon previous enrollment) template and gives a match or no-match decision. Biometric systems can also act in an identification mode, where a user does not claim an identity but only provides their biometric. The biometric system then compares this biometric template with all of the stored templates in the database and produces a similarity score for each of the stored templates. The template with the best similarity score is the system's best guess at who this person is. The score for this template is known as the top match.

It is unrealistic to assume that a biometric system can determine the exact identity of an individual out of a large database. The system's chances of returning the correct result increases if it is allowed to return the best two similarity scores, and increased even more if it is allowed to return the best three similarity scores. A plot of probabilities of correct match versus the number of best similarity scores is called a cumulative match characteristics curve. A generic CMC curve is shown in figure 3.

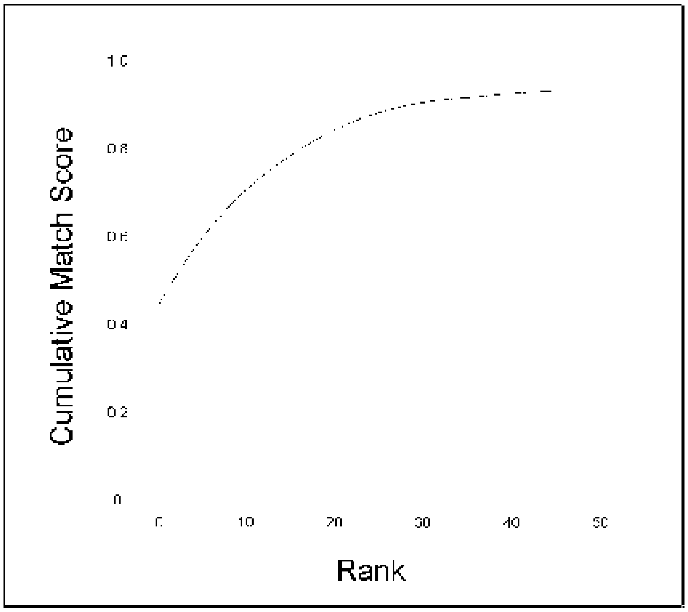


Figure 3 - Generic CMC Curve

Just as with ROC curves, these results can vary wildly based on the data that was used by the biometric system. Results for the same experiments described in Section 4.1.3 for verification results will also be shown for identification results. One other item must be provided to complete the story for CMC results: the number of biometric templates in the system database. This number was also provided alongside the results.

4.1.3 Recognition Performance Test Experiment Descriptions

Numerous experiments can be performed based on the similarity files returned by the participating vendors. The FRVT 2000 Evaluation Report provided results for the following types of experiments:

> ***Compression Experiments***. The compression experiments were designed to estimate the effect of lossy image compression on the performance of face-matching algorithms. Although image compression is widely used to satisfy space and bandwidth constraints, its effect in machine vision applications is often assumed to be deleterious; therefore, compression is avoided. This study mimics a situation in which the gallery images were obtained under favorable, uncompressed circumstances, but the probe sets were obtained in a less favorable environment in which compression has been applied.
>
> ***Distance Experiments***. The distance experiments were designed to evaluate the performance of face matching algorithms on images of subjects at different distances to the

fixed camera. The results of these experiments should be considered for situations where the distance from the subject to the camera for enrollment is different from that used for verification or identification. In all experiments, the probe images were frames taken from relatively low-resolution, lightly compressed, video sequences obtained using a consumer grade tripod-mounted auto-focus camcorder. In these sequences the subjects walked down a hallway toward the camera. Overhead fluorescent lights were spaced at regular intervals in the hallway, so the illumination changed between frames in the video sequence. This may be thought of as mimicking a low-end video surveillance scenario such as that widely deployed in building lobbies and convenience stores.

Expression Experiments. The expression experiments were designed to evaluate the performance of face matching algorithms when comparing images of the same person with different facial expressions. This is an important consideration in almost any situation because it would be rare for a person to have the exact same expression for enrollment as for verification or identification.

Illumination Experiments. The problem of algorithm sensitivity to subject illumination is one of the most studied factors affecting recognition performance. When an image of the subject is taken under different lighting conditions than the condition used at enrollment, recognition performance can be expected to degrade. This is important for systems where the enrollment and the verification or identification are performed using different artificial lights, or when one operation is performed indoors and another outdoors.

Media Experiments. The media experiments were designed to evaluate the performance of face-matching algorithms when comparing images stored on different media. In this case, digital CCD images and 35mm film images were used. This is an important consideration for a scenario such as using an image captured with a video camera to search through a mugshot database created from a film source.

Pose Experiments. The performance of face-matching algorithms applied to images of subjects taken from different viewpoints is of great interest in certain applications, most notably those using indifferent or uncooperative subjects, such as surveillance. Although a subject may look up or down and thereby vary the declination angle, the more frequently occurring and important case is where the subject is looking ahead but is not facing the camera. This variation is quantified by the azimuthal head angle, referred to here as the pose.

Resolution Experiments. Image resolution is critical to face recognition systems. There is always some low resolution at which the face image will be of sufficiently small size that the face is unrecognizable. The resolution experiments described below were designed to evaluate the performance of face matching as resolution is decreased. The metric used to quantify resolution is eye-to-eye distance in pixels.

Temporal Experiments. The temporal experiments address the effect of time delay between first and subsequent captures of facial images. The problem of recognizing subjects during extended periods is intuitively significant and is germane to many applications. Robust testing of this effect is difficult because of a lack of long-term data.

A full description of the experiments performed during the Recognition Performance Test portion of the FRVT 2000 evaluations is provided in Appendix E

4.1.4 Recognition Performance Test Results

Results from the experiments explained in Appendix E were provided via two methods in the FRVT 2000 Evaluation Report. The first used bar charts to highlight key results from each experiment. These are shown in Appendix F. The second provided the full ROC and CMC curves. These are shown in Appendix G. Overall conclusions of each group of experiments are provided below.

Compression Experiments. The compression experiments show that compression of facial images does not necessarily adversely affect performance. Results actually show that performance increased slightly (approximately two percentage points) for 10:1 and 20:1 compression rates versus uncompressed probe images. A compression ratio of 30:1 returned scores nearly identical to uncompressed probe images. It is not until a compression ratio of 40:1 that the performance rate drops below that of the uncompressed probes by approximately two percentage points. Images for the compression experiments used the FERET database, so only the top scores for identification were reported. Top match scores fell in the low to mid 60 percent range except for the 40:1 case. The probability of identification increased to the mid to high 80s for the top 20 matches. Because the results are aggregated and consider only JPEG compression, we recommend that additional studies on the effect of compression on face recognition systems be conducted.

Top Rank Scores for Compression Identification Experiments					
	C0	C1	C2	C3	C4
Top	0.63	0.65	0.66	0.63	0.56

Distance Experiments. The distance experiments across all algorithms and the three sets of distance experiments show that performance decreased as distance between the person and camera increased. There were three sets of distance experiments: experiments D1-D3 (indoor digital gallery images, indoor video probes 2, 3 and 5 meters from the camera), D4 and D5 (indoor video gallery images, indoor video probes 3 and 5 meters from the camera) and, D6 and D7 (outdoor video gallery images, outdoor video probes 3 and 5 meters from the camera). With the exception of the D1 experiment, Lau was the top performer followed by Visionics and then C-Vis. This trend continued for other rank scores as well.

Top Rank Scores for Distance Identification Experiments							
	D1	D2	D3	D4	D5	D6	D7
C-Vis	0.05	0.01	0.00	0.02	0.00	0.02	0.00
Lau	0.24	0.15	0.09	0.43	0.22	0.30	0.10
Visionics	0.33	0.09	0.01	0.29	0.10	0.16	0.07

Expression Experiments. For the identification performance in the expression experiment, all three algorithms performed better on the E1 case; whereas during verification, all three algorithms achieved their best performance on the E2 case. The difference in identification performance between the E1 and E2 cases for the top match score ranged from three to five percentage points, and zero to two percentage points for the verification equal error rate. This shows that for the FRVT 2000, identification is more sensitive to changes in expression than verification. Visionics was the top performer for both experiments, followed by Lau and then C-Vis.

Top Rank Scores for Expression Identification Experiments		
	E1	E2
C-Vis	0.39	0.34
Lau	0.83	0.8
Visionics	0.96	0.93

Illumination Experiments. In the illumination experiment, the I3 case was the most difficult and I2 was the least difficult. Illumination experiments I1 and I3 used the same gallery of digital mugshots taken indoors, the I1 probe set had indoor digital images with overhead lighting, and the I3 probe set's images were taken outdoors. The I1 experiment's performance was significantly better than the I3 experiment, which shows that an area of future investigation is handling lighting changes that occur when one image is taken indoors and the other is taken outdoors. Visionics was by far the top performer, followed by Lau and then C-Vis.

Top Rank Scores for Illumination Identification Experiments			
	I1	I2	I3
C-Vis	0.16	0.43	0.06
Lau	0.8	0.85	0.32
Visionics	0.91	0.97	0.55

Media Experiments. For Lau Technologies and Visionics Corp., switching between media did not significantly affect performance. For case M1, the gallery consisted of 35mm images and the probe set consisted of digital images. For the M2 case, the gallery contained digital images and the probe set 35mm images. Visionics was slightly ahead of Lau and both were significantly ahead of C-Vis.

Top Rank Scores for Media Identification Experiments		
	M1	M2
C-Vis	0.58	0.54
Lau	0.96	0.96
Visionics	0.99	0.97

Pose Experiments. The pose experiments show that performance is stable (top match in the mid to high 0.90 and nearing 0.99 between rank 5 and 10) when the angle between a frontal gallery image and a probe is less than 25 degrees. Performance dramatically falls off when the angle is greater than 40 degrees. The top match score for a pose of 40 degrees falls to 0.68, and to 0.30 for a pose of 60 degrees.

Images from the HumanID database with a pose of 45 degrees and the additional hardship of being in outside lighting returned even worse scores. Visionics was clearly superior with a top rank of 0.24, followed by Lau at 0.09 and C-Vis at 0.02.

Top Rank Scores for Pose Identification Experiments					
	P1	P2	P3	P4	P5
Top	0.98	0.94	0.68	0.3	
C-Vis					0.02
Lau					0.09
Visionics					0.24

Resolution Experiments. In this experiment, the R2 (eye separation of 45 pixels) performance values were better than the R1 (eye separation of 60 pixels). Performance then began to degrade for Lau and Visionics for R3 (eye separation of 30 pixels), but increased slightly for C-Vis. All systems had their worst performance on the R4 experiment (inter-pupil distance of 15). In all experiments, Visionics was clearly superior, followed by Lau and then C-Vis.

Top Rank Scores for Resolution Identification Experiments				
	R1	R2	R3	R4
C-Vis	0.33	0.35	0.37	0.18
Lau	0.79	0.83	0.76	0.74
Visionics	0.94	0.94	0.94	0.92

Temporal Experiments. For the FERET temporal probe sets, the FRVT 2000 performance for the duplicate I (T1) and duplicate II (T2) probes have almost the same top rank score - 0.63 for T1 and 0.64 for T2. (The duplicate I probes are probes taken on different days or under different conditions than the gallery images; the duplicate II probes and gallery images were taken at least 18 months apart.) In the FERET 1996 evaluation, the algorithms evaluated performed better on the duplicate I probe set. In the FERET evaluations, there was approximately a seven percentage point difference in performance between duplicate I and II probes for the best partially automatic algorithm. This shows improvement since the FERET program, but the low scores show that further developments are still necessary for most applications.

The T3, T4 and T5 experiments use the same probe set and vary the type of images in the gallery. The time between the collection of the gallery and probe images was approximately one year. The T3, T4 and T5 experiments are similar to the FERET duplicate II probe set (T2 experiment) because there was at least one year between the time the gallery and probe images were acquired. The gallery in T3 consisted of images taken with best-practice mugshot lighting, the T4 gallery contained FERET-style images and the T5 gallery's images were taken with overhead lighting. Based on the top match score, the hardest experiment was T5; the easiest was T3. The verification scores do not produce such a ranking of the experiments.

Top Rank Scores for Temporal Identification Experiments					
	T1	T2	T3	T4	T5
Top	0.63	0.64			
C-Vis			0.06	0.03	0.03
Lau			0.34	0.34	0.26
Visionics			0.55	0.55	0.35

Overall Conclusions for the Recognition Performance Test. The FERET evaluations identified temporal and pose variations as two key areas for future research in face recognition. The FRVT 2000 shows that progress has been made in temporal changes, but developing algorithms that can handle temporal variations is still a necessary research area. In addition, developing algorithms that can compensate for pose variations, and illumination and distance changes were noted as other areas for future research.

The FRVT 2000 experiments on compression show that moderate levels of compression do not adversely affect performance. The resolution experiments find that moderately decreasing the resolution can slightly improve performance.

In experiments where results are given by vendor, there was a clear order to placement for top rank scores. Visionics, in all cases except distance, outperformed Lau and both clearly outperformed C-VIS. All applications of facial recognition would encounter situations that are defined by multiple FRVT 2000 experiments, so a detailed analysis of all results would need to be performed prior to selecting any system for a scenario evaluation.

4.2 Scenario Evaluation: Product Usability Test

4.2.1 Overview

The scenario chosen for the Product Usability Test was access control with live subjects. The Product Usability Test was administered in two parts: the Old Image Database Timed Test and the Enrollment Timed Test. For the Old Image Database Timed Test, vendors were given a set of 165 images captured with a standard access control badge system, including one image of each of the three test subjects. Images of the three test subjects are shown in Figure 4. The set contained two images for five people, and one image for each of the other 155 people. Vendors enrolled these images into their system for comparison with the live subjects. The operational scenario was that of a low-security access control point into the lobby of a building. The building's security officers did not want to mandate that the employees take the time to enroll into the new facial recognition system so they used their existing digital image database taken from the employee's picture ID badges.

Figure 4 - Product Usability Test's Test Subjects

For the Enrollment Timed Test, the images of the three test subjects were removed from the system while the other images were retained. Vendors were then allowed to enroll the three subjects using their standard procedures, including the use of multiple images. The purpose of the test was to measure system performance using vendor enrollment procedures. The enrollment procedures were not evaluated. The operational scenario was that of an access control door for a medium-to-high security area within the building previously described. In this case, employees were enrolled in the facial recognition system using the standard procedures recommended by the vendor.

During the Product Usability Test, several parameters were varied including start distance, behavior mode, and backlighting. Tests were performed for each subject at distances of 12, 8, and 4 feet for all trials except for the variability test. Test subjects performed each test using cooperative and simulated, repeatable, indifferent behavior modes. For the cooperative mode, subjects looked directly at the camera for the duration of the trial. For the indifferent mode, subjects instead moved their focus along a triangular path made up of three visual targets surrounding the camera. Each trial was performed with and without backlighting provided by a custom light box.

For the Old Image Database Timed Test, subjects began each trial standing at the specified start distance then walked toward the camera when the timer was started. Each subject started at 12, 8 and 4 feet in cooperative mode then repeated in indifferent mode. Subject 1 then performed 8 cooperative trials from a start distance of 12 feet for the variability test, a test to determine the consistency of the subject-system interaction. Subject 1 then performed three more cooperative trials from 12, 8, and 4 feet holding a photograph of his own face to determine if the system could detect liveness. The photograph was an 8" x 10" color glossy print taken in a professional photo studio. This entire sequence was followed four times: once in verification mode without backlighting, once in identification mode without backlighting, once in verification mode with backlighting, and once in identification mode with backlighting.

The Enrollment Timed Test was performed exactly as the Old Image Database Timed Test described above except the subjects stood in place at the specified start distance rather than walking toward the camera.

4.2.2 Product Usability Test Results Format

For the Old Image Database Timed Test, three parameters were recorded:

> ***Final distance*** is the distance in feet between the camera and the test subject at the end of the trial. This was recorded in increments of one foot.
>
> ***Acquire time*** is the time in seconds it took the system to report a match, regardless of whether or not the answer was correct. This was recorded in increments of 1/100 second. An X indicates that a match was not acquired within the 10-second time limit.
>
> ***Correct match*** tells whether or not the system matched the live subject with the correct person in the database. Again, an X indicates that a match was not acquired within the 10 second time limit.

For the Enrollment Timed Test, the parameters were recorded as described; however, the subjects stood in place for each of these trials so it was unnecessary to record the final distance.

For the variability test, subject 1 performed eight cooperative-mode trials for both the verification and identification modes, with and without backlighting. A start distance of 12 feet was used for each trial. Note that it is desirable to have a correct match on all trials except the photo tests, where a photo of subject 1 was used to attempt access.

A complete description of the Product Usability Tests is provided in Appendix C.

4.2.3 Product Usability Test Results

Results from the Product Usability Test were provided via tables in the FRVT 2000 Evaluation Report. These are shown in Appendix H. Overall conclusions are provided below.

In the Product Usability Tests, all vendors performed considerably better in the Enrollment Timed Tests than in the Old Image Database Timed Tests. There are two main differences between the two tests. The first is that the subjects are walking towards the camera in the Old Image Database Timed Test and are stationary for the Enrollment Timed Tests. Results from the Recognition Performance Test show us that performance actually increases as the subjects get closer to the camera, so this would not cause the degradation in performance seen in the Old Image Database Timed Test.

The second difference between the two Product Usability Tests is the enrollment method of gallery images. In the Old Image Database Test, the gallery images were provided to the vendors before testing began. These images were taken with different camera systems and in a different location than where the testing occurred. In the Enrollment Timed Test, the gallery images were enrolled using the vendor system and in the same room where testing took place. By default, this difference in enrollment procedures is the cause of the change in performance by the systems in the Product Usability Tests. This shows that potential users of facial recognition technology should enroll subjects using images gathered by the facial recognition system at the installation location if at all possible. These results also show facial recognition vendors that this is an area for additional research.

In all cases, there was very little difference in performance between cooperative and simulated indifferent results. The lack of a difference is mainly because of the pose angles introduced by the simulated indifferent behavior. The initial pose angle varied between 17 and 24 degrees, depending on the start distance, and decreased as the subject began simulating indifferent behavior. These results are in agreement with the pose experiments in the Recognition Performance Test and show that facial recognition systems will not show significant changes in performance if a subject is cooperative versus indifferent as long as the indifferent subject is facing toward the camera.

Adding moderate, non-varying backlighting generally introduced a small degree of difficulty for the facial recognition vendors, but in most cases it was negligible. Further experimentation with higher intensity backlighting, lighting from various angles, and varying intensity are necessary to fully understand the impact of lighting in this scenario.

In all cases, the facial recognition systems were quicker and more accurate when performing verification experiments than in identification experiments. The gallery size for identification experiments in the Product Usability Test was 165, which is a fairly small number. It is anticipated that performance disparity will increase as the identification gallery increases, but further tests are required to know for sure.

Two of the five companies correctly returned no score for the photo tests in the Enrollment Timed Test. This evaluation was a very quick look at the liveness issue that is important for any form of access control using biometrics, but it may not be an issue for other applications. Additional research on this issue should be carried out for the three systems that attempted to identify the individual and on the two that correctly returned no score to determine their consistency.

4.3 Information Sharing Plan

Besides the results themselves, the other major goal of the FRVT 2000 evaluation was to fully document the evaluation for release to the public. This had never been done before in the field of biometrics, and the FRVT 2000 Evaluation Report should serve as an example for all future evaluations.

4.3.1 Information Available Prior to Evaluation

Several forms of documentation were provided to the public prior to the actual testing of the vendor systems. An overview of the evaluations along with the evaluation methodology, evaluation sponsors, and participating vendors was provided via the FRVT 2000 web page (Appendix B). API documentation for the Recognition Performance Test, a detailed test plan, and a demonstration data set (Appendices B, C, and D)was also provided to the participating vendors well before the evaluations took place.

The FRVT 2000 web site's FAQ pages (Appendix B) provided an excellent method for communication between the sponsors, the participating vendors, and the public. Because of the lessons learned from previous scenario evaluations (described in Section 2.3), the author believed that the vendors would propose modifications to the evaluation protocol because the FERET program participants did also. This issue was successfully settled at the start by addressing this in the FAQ section of the web site as shown below:

> ***25. Can my company propose changes to the planned tests?***
>
> *Absolutely. We are always looking for new ideas on how to compare one system to another. The sponsors, however, spent considerable time developing the test plan for the Facial Recognition Vendor Test 2000 and decided that the method given on this web site is how the tests will be performed. It would be unfair to other test participants to change the tests at this point. We will gladly hold on to all proposed changes and study them if we should do another series of tests.*

This approach proved to be effective as only one vendor voiced objections regarding the evaluation methodology. This vendor eventually withdrew from the evaluation. The letter requesting to withdraw from the evaluations stated that the reason was their disapproval of the evaluation method used in FRVT 2000. When the sponsors received this letter via e-mail, they sent a reply granting the vendor's request and also described the validity of the evaluation method.

The sponsors did not find out until much later that the vendor also sent copies of the withdrawal request e-mail to all of the other vendors participating in the FRVT 2000. The message was sent separately to the other vendors, so the sponsors therefore, did not copy any of the other vendors on their reply letter to this vendor.

When seen from the viewpoint of the other participants, one vendor had questioned the validity of the evaluation method in an apparently open forum without the evaluation sponsors responding whatsoever. In hindsight, the sponsors feel that this may have had a negative effect as two other vendors subsequently withdrew within the next 36 hours. Fortunately, one of these vendors requested to rejoin the evaluation the following week.

The lesson learned from this chain of events is that all discussions with anyone outside those running the evaluation should be completely open to the public. The sponsors had worked to ensure that the participating vendors had a level playing field via the Q&A restrictions but, in this case, a further degree of restrictions on discussion would have been beneficial. More information about this case study can be found in Appendix I.

4.3.2 Information Available During the Evaluation

During the FRVT 2000 Evaluations the author modified the FRVT 2000 Web Site from that shown in Appendix B to the format still in use at http://www.dodcounterdrug.com/facialrecognition. The new format still provided an overview of the FRVT 2000 evaluations, but also provided a comprehensive overview of the FERET program. Links and/or information about the FRVT 2000 was also maintained on the Counterdrug Technology Information Network (http://www.ctin.com), the Biometric Consortium's website (http://www.biometrics.org), and several smaller websites.

The author maintained interest in the FRVT 2000 evaluations by speaking at various conferences. The most significant of these was an invited presentation at the Biometric Consortium's 2000 Annual Conference. This conference attracts the world's foremost experts in biometric technology each year and provided an excellent opportunity to discuss the evaluations and to answer questions from the biometric community. The author also spoke at several other conferences, workshops, and government meetings to maintain the evaluation's visibility in the biometrics and government communities.

4.3.3 Information Available After the Evaluations

The author wrote the FRVT 2000 Evaluation Report and released it to the public on 17 February 2001. The evaluation report is a very extensive document numbering 235 pages that explains in detail the FRVT 2000 evaluations in its entirety. In addition to the information previously described in this thesis, the FRVT 2000 Evaluation Report also provides the following information that makes the FRVT 2000 evaluations the most documented evaluation of biometric technology to date:

- Data Collection (Appendix J)
- Evaluation Preparations (Appendix K)
- Modifications to the Recognition Performance Test Scoring Algorithm (Appendix L)
- Minor Reporting Practice Changes (Appendix M)
- FRVT 2000 Evaluation Announcements (Appendix N)
- Sample Images for Recognition Performance Test Experiments (Appendix O)
- Vendor Product Descriptions (Appendix P)
- Vendor Position Papers on FRVT 2000 Evaluation Report (Appendix Q)

5.0 Conclusions

The Facial Recognition Vendor Test 2000 has been a worthwhile endeavor. It will help numerous readers evaluate facial recognition systems for their own uses and will serve as a benchmark for all future evaluations of biometric technologies.

The FRVT 2000 evaluations were not designed, and the FRVT 2000 Evaluation Report was not written, to be a buyer's guide for facial recognition. No one will be able to open the report to a specific page to determine which facial recognition system is best because there is not one system for all applications. The only way to determine the best facial recognition system for any application is to follow the three-step evaluation methodology described in the FRVT 2000 Evaluation Report and analyze the data as it pertains to each individual application. It is possible that some of the experiments performed in the Recognition Performance and Product Usability portions of this evaluation have no relation to a particular application and should be ignored.

The FRVT 2000 Evaluation Report was downloaded 439 times in the first 10 days of availability. The author plans to generate further interest in the report via upcoming speaking engagements and articles. Upcoming speaking engagements currently include:

- CardTech/SecureTech - the country's largest conference on physical security
- The Office of National Drug Control Policy's 2001 International Technology Symposium - Large technology conference that attracts technical and practitioner attendees from around the world
- American Correctional Association's Annual Conference - the nationwide conference for the corrections industry
- Northeast Technology Product and Assessment Committee Meeting - Quarterly meeting of state government officials from the NE United States
- National Institute of Justice Advisory Council Meeting - The advisory council has influence on the future direction of technology development efforts within the U.S. Department of Justice.

Upcoming authoring engagements currently include:

- Corrections Today Articles - two articles in the July issue of Corrections Today Magazine
- Biometrics Technology Today Articles - Overview article in an upcoming issue of Biometrics Technology Today
- Book Chapter - An untitled book is currently being written about evaluating biometric technologies. The author was invited to write the chapter concerning evaluating facial recognition systems.

Bibliography

"FERET (FacE REcognition Technology) Recognition Algorithm Development and Test Report", P. J. Phillips, P. Rauss, S. Der, ARL-TR-995, U.S. Army Research Laboratory, 1996

"The FERET Database and Evaluation Procedure for Face Recognition Algorithms," P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, Image and Vision Computing Journal, Vol. 16, No. 5, pp. 295-306, 1998.

"The FERET Verification Testing Protocol for Face Recognition Algorithms," S. Rizvi, P. J. Phillips, and H. Moon, Image and Vision Computing Journal

"The FERET Evaluation Methodology for Face-recognition Algorithms," P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, IEEE trans. Pattern Analysis and Machine Intelligence.

"An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki in IEEE Computer, February, pp. 56-63, 2000.

"Facial Recognition Vendor Test 2000 Evaluation Report," by D.M Blackburn, J.M. Bone, and P.J. Phillips. http://www.dodcounterdrug.com/facialrecognition.

Appendix A
Biometric Term Glossary

Biometric signature (biometric template).	A digital representation of the user's distinct characteristics. Exact definition depends on the type of biometric system used.
Cooperative subject.	A subject that actively assists the biometric system.
Cumulative match characteristic (CMC).	A set of data points that describes the identifi cation performance of a biometric system when the system returns the top n matches, where n is a user-defi ned number.
Duplicate.	A probe image of a person whose corresponding gallery image was taken from a different image set. Usually, a duplicate is taken on a different day than the corresponding gallery image.
Duplicate I probes.	Set of duplicate probes for a gallery used in the FERET evaluations.
Duplicate II probes.	Set of duplicate probes in the FERET program where there is at least one year between the acquisition of the corresponding probe and gallery images.
Enrollment.	The process of observing an individual's chosen characteristic, normalizing it and storing it in the biometric system's database.
Equal error rate.	The operating point in a biometric system where the false acceptance rate and false reject rate are equal.
False acceptance (alarm) rate (FAR).	The percentage of imposters whose identity claims are incorrectly accepted.
False reject rate (FRR).	The percentage of valid users wrongly rejected.
FERET.	The FacE REcognition Technology program sponsored by the DoD Counterdrug Technology Development Program Office from 1993 through 1998. Detailed information about the FERET program can be found at http://www.dodcounterdrug.com/facialrecognition.
Fully automatic algorithm.	An algorithm that can locate a face in an image and recognize the face. All algorithms tested in FRVT 2000 were fully automatic.
Gallery set.	The collection of images of individuals known to the algorithm. A gallery set always has only one image per person. See probe set.
Identification mode.	The biometric system compares the given individual's biometric signature to all biometric signatures in its data-base and returns the top n matches.
Indifferent subject.	A subject that does not actively help or hinder the biometric system.

Noncooperative subject. See *uncooperative subject*

Operational Evaluation. The third in a sequence of three evaluations outlined in "An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki in *IEEE Computer*, February 2000, p. 56-63, 2000. The primary goal of an operational evaluation is to determine if a biometric system meets the requirements of a specific application. See *technology evaluation* and *scenario evaluation*.

Partially automatic algorithm. An algorithm that requires that the centers of the eyes be provided prior to recognizing a face. Most of the results cited in the FERET reports are from partially automatic algorithms.

Probe set. A set of images containing the face of an unknown individual that is presented to an algorithm to be recognized. Probe can also refer to the identity of the person in a probe image. A probe set may contain more than one image of an individual. See *gallery set*.

Recognition. A generic term used in the description of biometric systems. Recognition does not inherently mean either identification or verification but is sometimes used as such.

Receiver operating characteristic (ROC). A collection of data points that describe a biometric system's numerous FAR/FRR associations.

Scenario Evaluation. The second in a sequence of three evaluations outlined in "An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki in *IEEE Computer*, February 2000, p. 56-63, 2000. The primary goal of a scenario evaluation is to determine whether a biometric technology is sufficiently mature to meet performance requirements for a class of applications. See *technology evaluation* and *operational evaluation*.

Scoring software. A software package that uses similarity files and experiment definitions as input, and then returns output that can be displayed graphically by a ROC or CMC curve.

Similarity score. A numerical description returned by a facial recognition algorithm that describes that algorithm's confidence that the probe and gallery image were of the same individual.

Similarity file. A predefined file structure used in the FERET program and the FRVT 2000 evaluation that documented the similarity scores from numerous image comparisons.

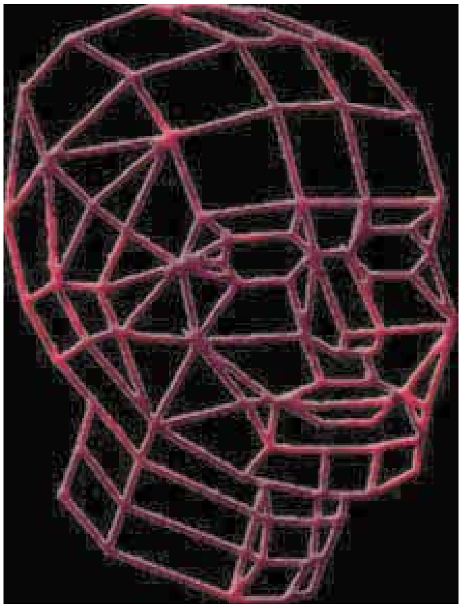
Technology evaluation. The first in a sequence of three evaluations outlined in "An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki in *IEEE Computer*, February 2000, p. 56-63, 2000. The primary goal of a technology evaluation is to measure the state of the art, to determine technological progress, and to identify the most promising approaches. See *scenario evaluation* and *operational evaluation*.

Uncooperative subject. A subject that attempts to actively hinder the biometric system. Sometimes referred to as *noncooperative*.

Verification mode. The biometric system compares the given individual with who that individual says they are and gives a go/no-go decision.

Appendix B
FRVT 2000 Web Site

Facial Recognition Vendor Test 2000



- Overview
- Sponsors
- How to Participate
- Participating Vendors
- FAQ
- Points of Contact
- Upcoming Dates
- Forms
- Restricted Area

NOTE:††Due to participating vendor's requested evaluation dates and the sponsor's schedules, the evaluation of systems for FRVT 2000 extended into June. The sponsors are currently in the process of preparing the results and final report, which will be made available as soon as possible at this location.

Overview

Overview Topics

INTRODUCTION

The DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), the Defense Advanced Research Projects Agency (DARPA), NAVSEA Crane Division and NAVSEA Dahlgren Division are sponsoring an evaluation of commercial off the shelf (COTS) facial recognition products. The purpose of these evaluations is to accurately gauge the capabilities of facial recognition biometric systems that are currently available for purchase. The sponsoring agencies, as well as other government agencies, will use this information as a major factor when determining future procurement or development efforts. Participation in these evaluations is open to all facial recognition systems on the US commercial market. The U.S. Government will not compensate vendors to participate in these evaluations.

Participating vendors will be given access to the restricted area of this web site where they will be able to download important documents - such as the API for the recognition performance tests. The restricted area also has a question/answer forum where all discussion between participating vendors and the Government sponsors will take place. Participating vendors will also be given a set of practice images that are similar to those that they will be tested on.

Two categories of tests will be conducted: Recognition Performance Tests and Product Usability Tests. For each category, multiple tests will be performed to measure system performance in verification mode and in identification mode. The Recognition Performance Tests will use the FERET test methodology with a new database of images. The Product Usability Tests will evaluate performance in both low and medium security access control scenarios. A more detailed description of the test is provided below.

QUALIFICATIONS

To participate in these evaluations, each vendor must provide commercial facial recognition system(s) that will be available on the U.S. open market at the time of testing and up to two technicians to run the system through the tests. Vendors will also need to fill out forms requesting entrance in the test, and submit a four page (maximum) document that:

- Provides an overview of the submitted system
- Provides a component list for the submitted system
- Provide a detailed cost breakdown of the submitted system

Due dates for these requirements vary. Please see the Upcoming Dates.

University and research systems will not be permitted (separate tests for these systems are being developed for the near future). Vendors with multi-biometric types in their systems will be allowed to participate if they have the proper interface to run using only facial recognition. Vendors will be able to pick the components of the system as they see fit keeping in mind that results from these tests, as well as the street price of each system at the time of testing, will be made available to the public.

Each vendor may be allowed to submit up to two systems for testing, but vendors must show that there is a clear difference between the two systems. For example, if a vendor has a high-end/high cost version and a low-end/low cost version they will be allowed to enter both versions. However, if the difference is merely a change of cameras, the second version will not be allowed. Final decision to allow more than one system will be made by the Government. The basis of this decision could be technical or to limit the number of systems to be tested.

TEST RESULTS

Results of the tests will be documented in a final report and, possibly, several international conference papers. The final report will contain each participating vendorís four page document as well as the results of the government test. All reports and papers will be made available to the public. Testing activities will be recorded using video cameras but the footage will not be released to the public. Portions of the video, however, may be used to provide a quick 5-10 minute overview of the tests. General observations of the capabilities within the facial recognition community may be made on this promotional video, but individual vendorís results will not be given.

RESTRICTED AREA INFO

Participating vendors will be given access to the restricted area of this web site where they will be able to download important documents - such as the API for the recognition performance. The restricted area also has a question/answer forum where all discussion between participating vendors and the Government sponsors will take place. A small ìdevelopment setî of images will be available for download that will contain images similar to those that will be used for the actual tests.

RECOGNITION PERFORMANCE TESTS

The Recognition Performance Tests will be very similar to the original FERET tests that were sponsored by the DoD Counterdrug Technology Development Program Office. Since the conclusion of the original FERET program, the data sets and reports have been transferred to the National Institute of Standards and Technology (NIST), who is serving as a technical consultant for these tests. Images used in this test will be a combination of images from the FERET database as well as DARPAís new HumanID database.

For more information on the FERET tests, please see the following papers

The FERET Evaluation Methodology for Face-Recognition Algorithms

The FERET Verification Testing Protocol for Face Recognition Algorithms

On the day of the Recognition Performance Tests, the vendor will be given a set of test images in JPG format. The vendor may convert the images to another format if necessary, but no extra time will be given for this. The vendor will use their algorithm to compare each image to the others and report the similarity scores in a format that will be provided (in an API document located in the Restricted Area of this web site). The final report will show the results in the form of Receiver Operating Characteristic (ROC) curves for verification tests and Cumulative Match Characteristics (CMC) Curves for identification tests. Other forms of displaying information may also be used.

PRODUCT USABILITY TESTS

The Product Usability Tests will consist of two timed tests and an optional access control system interface test. The timed tests will be used to measure the response time of the overall system for two different operational scenario simulations: the Old Image Database Timed Test and the Enrollment Timed Test. These tests are described below. It is not necessary for a vendor to have an access control product to participate in these tests - these are operational scenarios that were developed to give the public a means of comparing the test results with something they would be familiar with. Each of the timed tests will be performed for both verification and identification and will be performed once with overhead fluorescent lighting and again with the addition of simulated back lighting. Results from the timed tests will be given in a format similar to the chart below.

Subject/Distance	4'	8'	12'
Cooperative	6.73	4.62	9.34
Uncooperative	9.67	7.43	X

Each test will show the time for the system to make a decision (if it makes a decision) and whether the decision is correct or not. The "X" in the chart shows that the system did not reach a decision before the time-out limit occurred. The number that is in red and underlined means that the system reached a decision, but it was the wrong decision.

Old Image Database Timed Test

The operational scenario for the Old Image Database Timed Test is that of a low security access control point into the lobby of a building. The buildingís security officers want to improve security into the area but do not want to slow down the flow through the entry area. The security officers also do not want to mandate that the employees take the time to enroll into the new system so they will use their existing digital image database taken from the employeeís picture ID badges. The employees may not be aware that they are being checked using a facial recognition system, so they will not be fully cooperative. The security officers would also like to install a facial recognition system in the lobby to automatically identify known bad guys who may be loitering outside the access points.

On the day of the test, the vendor will be given a database of images in JPG format for entry into their system. There will be only one image per subject and it is expected that the quality of these images will be fairly poor, but uniform. The vendor will be shown the low security target region of a hypothetical ROC curve. They will then adjust the sensitivities of the system to meet this goal according to the lighting conditions. The ROC curve location is merely a suggestion for meeting the desired low security level. Vendors may adjust sensitivities as they see fit before beginning the test, but once the test begins, they will not be allowed to adjust the settings. Once the system has been tuned, recognition will be attempted with several live test subjects. The test database will contain one image for each of the test subjects, in addition to other images.

For the verification portion of the test, each subject will stand at a maximum distance from the camera, and an ID number will be entered into the system to simulate a magnetic card swipe. A timer will then be started, and the subject will walk toward the camera. The subject will be looking in the general direction of the camera, but will not be looking at it directly. The timer will be stopped and the subject will stop walking if the system gives a verification results. If the subject reaches the minimum camera distance before the system gives a verification result, he will stop until the time limit is reached. The time required for the system to give a verification result, if any, will then be recorded as well as the distance at which the verification took place. If a verification result is not reached within the time limit or the system gives an incorrect verification result, these facts will be recorded. Each system will be tested using three different subjects. Someone will then attempt verification by holding an 8x10 color photograph of one test subject in front of his face.

For the identification portion of the test, each subject will stand at the maximum camera distance facing away from the camera. The timer will be started and the subject will turn and walk toward the camera as in the verification test. Identification time and distance will be recorded along with the correctness of the result and whether or not results were obtained within the time limit. Each system will again have three tests subjects and someone will attempt identification with an 8x10 color photograph.

After both the verification and identification portions of the test have been completed using overhead fluorescent lights, back lighting will be used to simulate conditions in a building with windows. Vendors will be allowed to adjust the system sensitivities for these new lighting conditions, but once the test begins, they will not be allowed to adjust them further. The verification and identification portions of the test will then be repeated using the back lighting.

Enrollment Timed Test

The operational scenario for the Enrollment Timed Test is that of an access control door for a medium security area within the building previously described. In this case, employees will be enrolled in the facial recognition system using the standard procedures recommended by the vendor. The access control system on one door has been setup so that an individual enters his identity with a magnetic stripe card and the system must verify if this is indeed the correct individual. On another door, the system has been setup so that an individual simply walks up to the camera and the door opens if the identity of the individual matches an individual in the database with valid credentials. The employees will be aware that they are being checked using a facial recognition system, but may or may not be cooperative.

This test will be performed after the Old Image Database Timed Test. Vendors will remove the images of the live test subjects from the database and enroll those subjects using their standard procedures. The vendor will then be shown a medium security target region of a hypothetical ROC curve so that sensitivities can be adjusted to meet this desired security level based on the lighting conditions. Recognition attempts will then be made using the same live test subjects of the previous test.

For the verification portion of the test, each subject will first stand at a maximum camera distance and simulate uncooperative behavior by using head movement without looking directly at the camera. An ID number will be entered into the system to simulate a magnetic card swipe and a timer will be started. If the system makes a verification result, the timer will be stopped and the time will be recorded. If verification results are not reached within the time limit or the system gives an incorrect verification result, these facts will be recorded. The subject will then repeat the test at the same location in a cooperative manner by looking directly at the camera. Each subject will repeat both the uncooperative and cooperative modes at several different camera distances. An 8x10 color photograph of one test subject will then be used for verification attempts at each distance.

For the identification portion of the test, each subject will again stand at several distinct camera distances and behave in both uncooperative and cooperative modes. However, this time they will begin by facing away from the camera and turn toward the camera when the timer is started. Identification times will be recorded in the same manner as the verification times. An 8x10 color photograph of one test subject will then be used for identification attempts at each distance.

After both the verification and identification portions of the test have been completed using overhead fluorescent lights, back lighting will be added and vendors may adjust system sensitivities. The verification and identification portions of the test will then be repeated using the back lighting.

Access Control System Interface Test

This is an optional test meant to determine if the facial recognition system can interface successfully with an access control system. To participate in this optional test, the facial recognition system must have a WIEGAND interface.

The goal is to test the interface rather than the facial recognition algorithm. The vendor will connect the facial recognition system to an access control system via a standard WIEGAND interface and enroll a single test subject that was successfully verified in the Enrollment Timed Test. The subject will present their identity to the access control system and then make three attempts to use the facial recognition system for verification. If the access control system receives a valid signal through the WIEGAND interface upon successful verification, the system will have passed this portion of the test. Vendors must indicate their intent to participate in this optional test. Vendors should also indicate any particular form of the WIEGAND standard required by their system, i.e. number of wires or number of data bits.

Sponsors

The DoD Counterdrug Technology Development Program Office sponsors the research, development, testing, evaluation, demonstration and integration of prototype systems to satisfy shortfalls in current capabilities to detect, identify, monitor, locate, track, analyze, and disseminate information regarding illegal drug related activities. The projects are intended to have dual mission applications, supporting both general purpose and counterdrug military requirements. In addition, individual projects may also support the counterdrug needs of Domestic Law Enforcement Agencies (DLEAs). The DoD Counterdrug Technology Development Program Office has been actively involved in facial recognition research and application through its sponsorship of the FERET program.

POC: Mr. Duane Blackburn, BlackburnDM@nswc.navy.mil

The National Institute of Justice is the research agency of the U.S. Department of Justice. Created by the Omnibus Crime Control and Safe Streets Act of 1968, as amended, NIJ is authorized to support research, evaluation, and demonstration programs, development of technology, and both national and international information dissemination.

POC: Mr. Tom Coty, cotyt@ojp.usdoj.gov

The Defense Advanced Research Projects Agency is the central research and development organization for the Department of Defense (DoD). It manages and directs selected basic and applied research and development projects for DoD, and pursues research and technology where risk and payoff are both very high and where success may provide dramatic advances for traditional military roles and missions and dual-use applications.

POC: Dr. Jonathon Phillips, jphillips@darpa.mil

NAVSEA-Crane, Defense Security Systems is the Navy's Center of Expertise and Acquisition Agent for the procurement and installation of all badging and access control systems for the entire Department of Defense.

POC: Mr. Mike Bone (Bone_Mike@crane.navy.mil)

NAVSEA - Dahlgren Division - Our mission is to be the Navy's principal research, development, and test and evaluation activity for surface ship combat systems, ordnance, mines, strategic systems, amphibious warfare, mine countermeasures, and special warfare systems.

POC: Mr. Duane Blackburn, BlackburnDM@nswc.navy.mil

The National Institute of Standards and Technology was established by Congress "to assist industry in the development of technology ... needed to improve product quality, to modernize manufacturing processes, to ensure product reliability ... and to facilitate rapid commercialization ... of products based on new scientific discoveries." An agency of the U.S. Department of Commerce's Technology Administration, NIST strengthens the U.S. economy and improves the quality of life by working with industry to develop and apply technology, measurements, and standards.

POC: Mr. Patrick Grother, pgrother@nist.gov

How to Participate

If you are interested in participating in these evaluations you must complete the two forms listed in the Forms section of this web site and have a company official witness and sign them. Upon receipt of the original, signed forms, we will send you a confirmation note that has your ID and password for access to the restricted area of this web site.

On the day that the vendor is being tested, you will need to provide a four page (maximum) document that:

- Provides an overview of the submitted system
- Provides a component list for the submitted system
- Provide a detailed cost breakdown of the submitted system

Please contact Mr. Mike Bone for additional information.† We prefer all communication to be via E-mail when possible so that we can log this activity on the FAQ and Discussion (restricted area) pages of this web site.† Some questions will ONLY be answered via the FAQ and Discussion pages.

Mr. Mike Bone

Bone_Mike@crane.navy.mil
Crane Division, Naval Surface Warfare Center
C4041, B180
300 Hwy. 361
Crane, IN 47522-5001
Phone: (812) 854-1141
Fax: (812) 854-2655

Participating Vendors

Company Name: Miros, Inc.
Product Name: TrueFace Engine SDK
Point of Contact: Jim Kottas
Address: Miros, Inc.
572 Washington St., Suite 18
Wellesley, MA 2482-6418
Phone/Fax: 781-235-0330 x225†/†781-235-0720
Email: jkottas@miros.com

Company Name: Visionics Corporation
Product Name: FaceIt(R)
Point of Contact: Kirsten Nobel
Address: One Exchange Place
Suite 800
Jersey City, NJ 07302
Phone/Fax: 201-332-9213, #207†/†201-332-9313
Email: kirsten@visionics.com

Company Name: Banque-Tec International Pty. Ltd.
Product Name: Eidolon
Point of Contact: Geoff Poulton
Address: CSIRO Division of Radiophysics,
PO Box 76, Epping, NSW 1710, AUSTRALIA
Phone/Fax: 61 2 9372 4287†/†61 2 9372 4411
Email: Geoff.Poulton@tip.csiro.au

Company Name: C-VIS Computer Vision und Automation GmbH
Product Name: FaceSnap Recorder (R)
Point of Contact: Dr.-Ing. Volker Vetter
Address: Universitaetsstrasse. 142
Bochum, Germany, 44799
Phone/Fax: +49 (0)234/97066-0†/†+49 (0)234/97066-30
Email: V.Vetter@c-vis.com

Company Name: LAU Technologies
Product Name: Face in the Crowd
Point of Contact: Hyeonjoon Moon
Address: 30 Porter Road
Littleton, MA 14600
Phone/Fax: 978-952-2055†/†978-952-2001
Email: hm@lautechnologies.com

Confirmed Participants

FAQ

Overview | Sponsors | How to Participate | Participating Vendors | Points of Contact | Upcoming Dates | Forms | Restricted Area | HOME

Submit an unanswered question

Frequently Asked Questions

1. Who is sponsoring/running these evaluations?
2. Is this evaluation part of the FERET program?
3. So, this could be the first step of future government activities with facial recognition?
4. Who is eligible to participate in these evaluations?
5. How do I sign up to participate in these evaluations?
6. Where do I send the forms?
7. What is the purpose of the required forms?
8. How will the four page writeup be used by the government?
9. When will I receive my companyís ID and PIN to access the restricted area of this site?
10. When will the Image Development Test Set be available?
11. When will the API documentation be available & what is its purpose?
12. If the sample images & API documentation are not available until March 8, why should I sign up for these tests before then?
13. What if my COTS product cannot produce a similarity file?
14. Why do I have to send a similarity file to the government based on the Image Development evaluation Set?
15. What happens if I do not send the sample similarity file by the required date?
16. What happens if there is something wrong with my sample similarity file?
17. When will I receive the actual images that will be used for the evaluation?
18. Can I keep a copy of these images once I have finished the evaluation?
19. When will I need to provide the government the similarity scores for the Recognition Performance Test?
20. What media can I use to provide the similarity scores?
21. How long will it take me to perform these tests?
22. Where and when will the tests take place?
23. Can my company request preferred test dates?
24. Will I be compensated for participating in these evaluation?
25. Can my company propose changes to the planned tests?
26. Can my company enter a facial recognition system based on thermal imaging?

27. When will the results from the Facial Recognition Vendor Test 2000 be released & where can we get them?

1. Who is sponsoring/running these evaluations?
The DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), the Defense Advanced Research Projects Agency (DARPA), and NAVSEA Crane Division are sponsoring these evaluations. Numerous individuals from NAVSEA-Dahlgren Division volunteered their time to help us make part of the image database. The National Institute of Standards and Technology has been very busy advising us throughout the evaluation development and were instrumental in assembling the various picture databases that will be used for this evaluation (and others). Technical Agents from the DoD Counterdrug Technology Development Office will be administering the evaluation, scoring the results, and writing the final report.

2. Is this evaluation part of the FERET program?
Not really. Although the DoD Counterdrug Technology Development Program also sponsored FERET, Dr. P. Jonathon Phillips (who was the Technical Agent for the FERET program) is actively involved with these tests, and this test uses part of the FERET database and scoring algorithms, this is not considered a part of the FERET program. The FERET program, which began in 1993, consists of three parts:

- Sponsoring research
- Collecting the FERET database
- The FERET evaluations

The goal of the FERET program is to advance the state of the art in facial recognition. The purpose of these evaluations is to measure the current capabilities of facial recognition to determine if it is ready for application or if further development work is still needed.

3. So, this could be the first step of future government activities with facial recognition?
Absolutely, although no specific plans are currently in place. The sponsors, as well as other government agencies that are unable to help sponsor these evaluations, have been studying facial recognition technology for several years and see numerous potential applications. The results of these evaluations will form the basis of our future efforts over the next few years.

4. Who is eligible to participate in these evaluations?
Anyone that has a commercially available system that is available on the United States market is eligible to participate. The government, or a private company that reads the results of the evaluation, must be able to call the vendor and purchase the system that was tested without any development efforts.

5. How do I sign up to participate in these evaluations?
You need to fill out, sign, and mail (original copy) two forms to the government to participate in these evaluations.

6. Where do I send the forms?
Send them to:

Mr. Duane Blackburn
DoD Counterdrug Technology Development Program Office
NSWCDD Code T43

17320 Dahlgren Road
Dahlgren, VA 22448

7. What is the purpose of the required forms?
The government needs a record that the vendor has volunteered to participate in these evaluations and understands how the evaluations will be performed and how the results will be released. Additionally, we need a separate record that each vendor has requested and promised to use NIST databases according to the limitations listed on the form.

8. How will the four page writeup be used by the government?
The four page (maximum) writeup will be a section in the final report that will be released to the public. This is your opportunity to describe your system to everyone that reads the report. Please, no salesman language!

9. When will I receive my company's ID and PIN to access the restricted area of this site?
You will be E-mailed your company's ID and PIN to access the restricted area of this site as soon as the government receives the original copies of the two required forms.

10. When will the Image Development Test Set be available?
The Image Development Test Set will be available on the restricted area of this site on March 8.

11. When will the API documentation be available & what is its purpose?
The API documentation will be available on the restricted area of this site on March 8. The Recognition Performance portion of these tests uses the FERET scoring code. In order to use that code, we must have your results in a standard format.

12. If the sample images & API documentation are not available until March 8, why should I sign up for these tests before then?
The restricted area of this web site has a second FAQ page. The FAQ on the public side will answer all general questions. The restricted FAQ page will answer more in-depth questions and allow a forum for Q&A that is not in the public eye.

13. What if my COTS product cannot produce a similarity file?
You will be allowed to modify your COTS system so that it will produce a similarity file. However, this ìpatchî must also be available to the general public. A key of any test is the ability for someone else to run the same tests using their images. NIST is currently writing a version of the scoring code that will eventually be made available to the general public.

14. Why do I have to send a similarity file to the government based on the Image Development Test

Set?
We need to verify that we can properly read your similarity file before you arrive for testing.

15. What happens if I do not send the sample similarity file by the required date?
You will not be allowed to participate in these evaluations.

16. What happens if there is something wrong with my sample similarity file?
You will be allowed to fix the problem and resend the sample similarity file as many times as needed until the day before you are scheduled to take the tests.

17. When will I receive the actual images that will be used for the evaluation?
You will receive the actual evaluation images on the same day that you take the Recognition Performance portion of these evaluations.

18. Can I keep a copy of these images once I have finished the evaluation?
No. In fact, you will not be allowed to even temporarily copy these images onto your hard drive. You will need to access the images directly off the CD during the evaluation.

19. When will I need to provide the government the similarity scores for the Recognition Performance Test?
As soon as you have finished taking the Recognition Performance Test.

20. What media can I use to provide the similarity scores?
Jaz disk.

21. How long will it take me to perform these tests?
It will take two days for each vendor to complete these tests. You will perform the Recognition Performance test on one day and the Product Usability test on the other.

22. Where and when will the tests take place?
The tests will begin in April. You will be notified by March 24 the two days that you will be taking the tests. The tests will take place at NAVSEA-Crane Division in Crane, Indiana (approximately 60 miles south of Indianapolis).

23. Can my company request preferred test dates?
Yes. If you've already registered for the tests, just send your preferred dates to Mike Bone, bone_mike@crane.navy.mil. We can't guarantee that your dates will be available, but we'll try to work out any conflicts between you and other vendors that request the same dates. It would be very helpful if you send

alternate dates in case your first choice is unavailable.

24. Will I be compensated for participating in these evaluations?
Unfortunately, no. The government will not be able to fund participating vendors. The added exposure to your product and the benefits of healthy competition with your peers is all we are able to provide.

25. Can my company propose changes to the planned tests?
Absolutely, we are always looking for new ideas on how to compare one system to another. The sponsors, however, have spent considerable time developing the test plan for the Facial Recognition Vendor Test 2000, and have decided that the method given on this web site is how we will be performing these tests. It would be unfair to other test participants to change the tests at this point. We will gladly hold on to all proposed changes and will study them if we should do another series of tests in the future.

26. Can my company enter a facial recognition system based on thermal imaging?
Unfortunately, no. All the images collected for this test were captured using cameras sensitive in the visible spectrum. Testing thermal imaging systems would require us to collect images using thermal sensors. If government agencies show increased interest in these systems, we may consider doing future tests with a separate thermal imaging category.

27. When will the results from the Facial Recognition Vendor Test 2000 be released & where can we get them?
Due to the preferred evaluation dates of the participating vendors, as well as lead-time requirements for some of the non-US based vendors, the evaluation schedule was pushed into the early part of June. Once this is complete it will take a few weeks to analyze the data to ensure its accuracy and then a couple of weeks to finalize the report. We plan to make the results available as soon as possible, but we will make sure that the results are correct before releasing them. The results will be made available at this web site (http://www.dodcounterdrug.com/facialrecognition) which will be undergoing a major renovation. Look for these changes soon!

Points of Contact

Overview | Sponsors | How to Participate | Participating Vendors | FAQ | Upcoming Dates | Forms | Restricted Area | HOME

If you are interested in participating in these evaluations, or if you are participating and have specific questions about these tests, please contact:

Mr. Mike Bone

Bone_Mike@crane.navy.mil
Crane Division, Naval Surface Warfare Center
C4041, B180
300 Hwy. 361
Crane, IN 47522-5001
Phone: (812) 854-1141
Fax: (812) 854-2655

For media inquiries, please contact:

Mrs. Debra O. Eubanks

eubanksdo@nswc.navy.mil
NSWC Dahlgren Laboratory Public Affairs Officer
Code CD06
17320 Dahlgren Road
Dahlgren, VA 22448
Phone: (540) 653-8152
Fax: (540) 653-4679

For all other inquiries, please contact:

Mr. Duane Blackburn

BlackburnDM@nswc.navy.mil
DoD Counterdrug Technology Development Program Office
NSWCDD Code T43
17320 Dahlgren Road
Dahlgren, VA 22448
Phone: (540) 653-6062
Fax: (540) 653-7471

NOTE: We prefer all communication to be via E-mail when possible so that we can log this activity on the FAQ and Discussion (restricted area) pages of this web site. Some questions will ONLY be answered via the FAQ and Discussion pages.

February 11, 2000	On-line announcement of Facial Recognition Vendor Tests 2000
March 8, 2000	Image Development Set Available in Restricted Area
	API available in Restricted Area
March 17, 2000	Last day for vendors to sign up to participate in the evaluation
March 24, 2000	Test schedule announced
March 27, 2000	Vendors must provide readable (correct media) and valid (syntax) similarity files to the Government based on initial runs of the Image Development Set. This is to ensure that data coming from the vendors is readable and in the correct format for scoring of the actual Image Test Set
April 10, 2000	Formal testing begins
June 2000	Final report made available to the public

To sign up for these tests, complete the two forms via the links on this page and have a company official witness and sign them. Upon receipt of the original, signed forms, we will send you a confirmation note that has your ID and password for access to the restricted area of this web site.

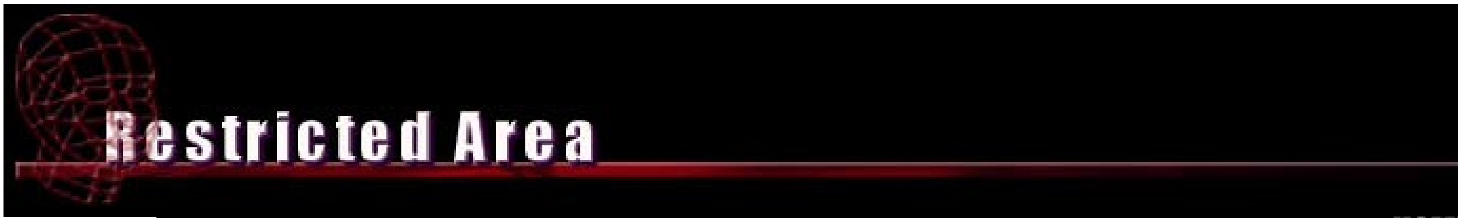
Send the forms with original signatures to:

Mr. Duane Blackburn
DoD Counterdrug Technology Development Program Office
NSWCDD Code T43
17320 Dahlgren Road
Dahlgren, VA 22448
Phone: (540) 653-6062

Forms

Application for Participating in Facial Recognition Vendor Test 2000

Application for Access to a Portion of the Development HumanID Data Set and FERET Database



Overview | Sponsors | How to Participate | Participating Vendors | FAQ | Points of Contact | Upcoming Dates | Forms | HOME

Login Lpoindexte

Password

Change Password

Login

Application for Participating in Facial Recognition Vendor Test 2000

1. Overview

The DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), the Defense Advanced Research Projects Agency (DARPA), NAVSEA Crane Division and NAVSEA Dahlgren Division are sponsoring an evaluation of commercial off the shelf (COTS) facial recognition products. The purpose of these evaluations is to accurately gauge the capabilities of facial recognition biometric systems that are currently available for purchase. The sponsoring agencies, as well as other government agencies, will use this information as a major factor when determining future procurement or development efforts. Participation in these tests is open to all facial recognition systems on the US commercial market. The U.S. Government will not compensate vendors to participate in these tests.

2. Test Description

Two categories of tests will be conducted: Recognition Performance Tests and Product Usability Tests. For each category, multiple tests will be performed to measure system performance in verification mode and in identification mode. The Recognition Performance Tests will use the FERET test methodology with a new database of images. The Product Usability Tests will evaluate performance in both low and medium security access control scenarios.

2.1 Recognition Performance Test

The Recognition Performance Tests will be very similar to the original FERET tests that were sponsored by the DoD Counterdrug Technology Development Program Office. Since the conclusion of the original FERET program, the data sets and reports have been transferred to the National Institute of Standards and Technology (NIST), who is serving as a technical consultant for these tests. Images used in this test will be a combination of images from the FERET database as well as DARPA's new HumanID database.

2.2 Product Usability Test

The Product Usability Tests will consist of two timed tests and an optional access control system interface test. Each of the timed tests will show the time for the system to make a decision (if it makes a decision) and whether the decision is correct or not. These tests will be performed for both identification and verification, as well as for different lighting conditions. The optional access control system interface tests will test the facial recognition systems' ability to communicate with an access control system using the WIEGAND standard.

3. Required System Description

On the first day of testing, participating vendors will be required to submit a four page (maximum) document that:

- Provides an overview of the submitted system
- Provides a component list for the submitted system
- Provide a detailed cost breakdown of the submitted system

4. Release of Evaluation Results

Results of the evaluations will be documented in a final report and, possibly, several international conference papers. The final report will contain each participating vendor's four page document as well as the results of the government evaluation. All reports and papers will be made available to the public. Testing activities will be recorded using video cameras but the footage will not be released to the public. Portions of the video, however, may be used to provide a quick 5-10 minute overview of the tests.

5. The Facial Recognition Vendor Test 2000 is being conducted for the sole purpose of determining the capability of facial recognition systems and not for fulfilling immediate or long-term mission requirements. The examination and test of these systems will in no way, expressed or implied, obligate the DoD Counterdrug Technology Development Program Office, or any of the test co-sponsors (hereinafter referred to as the "sponsors"), to purchase, rent, or otherwise acquire the systems tested. Manufacture, transportation, maintenance, and company test representatives shall be accomplished without cost to the sponsors. Tests will be conducted by an authorized representative of the vendor furnishing the system, but will be proctored by government personnel. Test sponsors will not endorse the vendor's products after the test conclusion. The test sponsors assumes no cost or obligations, expressed or implied, for damage to, destruction of, or loss of such equipment, or for damages or injuries resulting from the submission to the sponsors of defective items for test.

6. The vendor understands that any data obtained during these evaluations, as well as the four page system description, becomes the property of the DoD Counterdrug Technology Development Program Office and the vendor does not possess a proprietary interest in neither the data nor the system description.

7. The vendor will not file any claim against the sponsors or otherwise seek compensation for any equipment, materials, supplies, information, or vendor services provided.

8. The sponsors are not bound, or obligated, to follow any recommendations of the vendor. The United States Government is not bound, nor is it obligated, in any way to give any special consideration to the vendor on future contracts.

9.If the vendor decides to use results of these evaluations in any form of product literature, it must be accompanied by the following phrase. "Results shown from the Facial Recognition Vendor Test 2000 do not constitute endorsement of any particular system by the Government." It must also be accompanied by a link to the final report that will be generated by the Government.

10. Participating Vendor Information
NOTE:All information is required

Company Name: ______

Product Name: ______

Point of Contact: ______ ______ (First Name / Last Name)

Mailing Address: ______
______

City, State, & Zip ______ ______ ______ - ______

Phone: ______

Fax: ______

Email: ______

Web Site Address: ______

Number of systems to submit for testing:
○ One (1) ○ Two (2)

Number of systems to submit for optional Access Control System Interface Test:
○ Zero (0) ○ One (1)

If submitting system(s) for optional access control system interface test, please list any weigand interface requirements such as number of wires or number of data bits below:

Please proof the information you entered above to ensure it is correct before submitting.

NOTE: When you press 'Submit Form' (below), a completed form will be displayed. **Your submission is not complete until you print, obtain appropriate signatures, and mail the completed form to Duane Blackburn (address is provided on the form).**

Submit Form

11. Request for Vendor Participation

"With my signature I authorize my company to participate in the Facial Recognition Vendor Test 2000. I have read, and agree to be tested according to, the test description on this form and on the Facial Recognition Vendor Test 2000 website at http://www.dodcounterdrug.com/facial recognition. I understand how the facial recognition systems will be tested and how the results will be used. I understand that only commercially available facial recognition systems will be allowed to participate in these tests. Biometric systems that include facial recognition in cooperation with another biometric type will be allowed to participate, but only the facial recognition algorithm portion of that product will be tested.

"I understand that I must send original signed copies of this form and the Application for Access to NIST Special Database for Facial Recognition form to be allowed to participate in these tests. I must also provide a four page (maximum) document that explains the submitted system. I understand that I must provide a sample similarity file based on the development set of images that is available on the website. Results from the Recognition Performance Test must be written onto a Jaz disk and given to the government immediately following completion of the test. If I am requesting to have two of my systems tested, I understand that I must provide with this application a written description that shows the difference between the systems so that the government will be able to decide if both will be allowed to participate

"I understand that test activities will be videotaped and that portions of the video may be used for promotional purposes. Any questions that I have had were answered on the FAQ page of the website. I understand that further test details and sample images will be provided in the future on the Facial Recognition Vendor Test 2000 web site. I understand that test details and modifications that are listed on the website supersede any details in the test overview. I understand that the exact testing schedule at NAVSEA Crane will be will be released in the future."

FAQ

HOME

If you have a question not currently addressed in the Public or Restricted (for those who have access) FAQ areas, please ask it here by completing the form below. Your question will automatically be Emailed to the appropriate individuals. Answers to questions will be made available in the appropriate FAQ area (Public or Restricted). Due to volume, all questions may not receive individual attention. Please check this site for responses.

Your Company:

Your Name:

Your Email:

Your Question:

Submit Question

Forms

HOME

Application for Access to a Portion of the Development HumanID Data Set and FERET Database

1. Overview
The National Institute of Standards and Technology collects and maintains facial image databases for use by the Government for evaluating human identification technology. The Facial Recognition Vendor Test 2000 is one such evaluation.

2. Database Subsets to be Used
The Facial Recognition Vendor Test 2000 will use portions of the FERET database that was collected as part of the FERET program and the HumanID Data Set.

3. Vendor Access to Facial Recognition Vendor Test 2000 Demonstration Data Set
A small subset (~30 JPG images) will be placed on the restricted portion of the Facial Recognition Vendor Test 2000 website (http://www.dodcounterdrug.com/facialrecognition) on March 1. These images are given out as an example of the pictures in the databases and will give the vendor an opportunity to write sample similarity files to verify that they are in the correct format and are readable by the Government's scoring code.

The remainder of the test images, also in JPG format, will be given to the vendor on the day they arrive at NAVSEA Crane to take the test. The images will be given to the vendor on a CD-ROM. Vendors will not be allowed to copy these images onto their hard drive; they must be read directly off the CD-ROM. The vendor must return the CD-ROM to the government at test completion, assure the Government that none of the images are still resident on the test computer, and allow the Government to inspect all disks on the system to verify compliance.

4. Participating Vendor Information
NOTE:All information is required

Company Name: [ ]

Product Name: [ ]

Point of Contact: [ ] [ ] (First Name / Last Name)

Mailing Address: [ ]
[ ]

City, State, & Zip [ ] [ ] [ ] - [ ]

Phone: [ ]

Fax: [ ]

Email: [ ]

Web Site Address: [ ]

Number of systems to submit for testing:
○ One (1) ○ Two (2)

Number of systems to submit for optional Access Control System Interface Test:
○ Zero (0) ○ One (1)

Web Site Address: [ ]

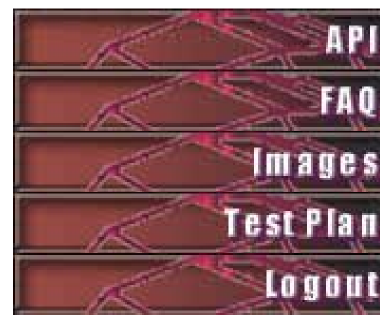
Please proof the information you entered above to ensure it is correct before submitting.

NOTE: When you press 'Submit Form' (below), a completed form will be displayed. **Your submission is not complete until you print, obtain appropriate signatures, and mail the completed form to Duane Blackburn (address is provided on the form).**

[Submit Form]

5. Request for Access to the Facial Recognition Vendor Test 2000 Demonstration Data Test Set
"With my signature I authorize my company to use the NIST Special Database for Facial Recognition, in association with the Facial Recognition Vendor Test 2000, and promise to do so according to the rules and limitations listed on this form."


Facial Recognition
Vendor Test
2000
Restricted Area


API
FAQ
Images
Test Plan
Logout

Download API (.pdf)

1. Overview

This document describes the image data supplied to vendors and the similarity files expected from vendors. The text refers to two data sets: the first, for release in early March 2000 is termed the Development Set; the second is the much larger Test Collection that will be used at the tests in April.

2. Images to Vendors

Development Set

The images provided in the development set are intended to ensure that vendors are capable of handling the images that will be provided in the Recognition Test. There are 17 images in the development set. They are a small subset of those that will be used for the Vendor Recognition Test. They are representative of the images in the larger set. The Test Collection will contain on the order of 10000 images.

Image Formats

The images are all in standard JPEG/JFIF format. They are readable by most image processing and image viewing utilities including the major web browsers. All images may be read using source code from the Independent JPEG Group (http://www.ijg.org/) available for download at ftp://ftp.uu.net/graphics/jpeg/jpegsrc.v6b.tar.gz.

The images come from various sources and were generally obtained from multiple different devices on different dates. The images do not all have the same width, height nor precision; some are color, some are grayscale; the amount of compression varies.

Naming convention

The files have arbitrary filenames of this form: i00000.jpg ... i00016.jpg. For the larger set containing N images the integer part of the filename will range from 0 to N-1. All filenames contain precisely 10 characters (1+5+1+3).

3. Results from Vendors

Similarity Files

The vendors are required to submit their results as "similarity files", which are described below. This implements the FERET protocol described in ìThe FERET Evaluation Methodology for Face Recognition Algorithmsî published as NIST IR 6264 available here: http://www.itl.nist.gov/iaui/894.03/pubs.html#face.

A similarity file contains a numerical similarity match between an image ìxî and all other images including ìxî itself. For each image provided, vendors must generate and submit a similarity file. The format is described below. For a test collection containing N images, vendors must submit N similarity files, each file containing N lines. Each line must contain an image name and a scalar value indicating the similarity between two images; a large value indicates that two images are closely matched whereas a small value indicates dissimilarity.

Naming convention

For an image, named for example i00016.jpg, vendors must submit a corresponding similarity file named i00016.sim. Any other file name is illegal.

Format

Similarity files are plain ASCII text files. For a data collection containing N images, a similarity file contains exactly N lines. Each line contains two fields separated by white space ("space Hex 20, or "\t" tab Hex 09). Blank lines are illegal.

First Field

The first field on each line contains the name of one of the images provided to the vendors in the test, e.g. i00016.jpg. Pathnames (for example: /data/results/ or d:\data\results) that precede the file name are illegal. Each file name occurs exactly once in the file.

Second Field

The second item is a floating point value that typically would be generated using the ISO C idioms for printing floating point numbers, namely: "%f", "%e" or "%E" as supplied to fprintf(). While a value may also be a decimal integer it must be readable as a floating point value. Legal examples of similarity values are 3.14159 6.626e-34 2.998e+08 42 Negative numbers are permissible. Large negative numbers indicate very dissimilar images; smaller negative numbers indicate less dissimilarity. Large positive numbers indicated very similar images. Undefined numbers (e.g. NaN and Inf) are illegal.

Files must be sorted

The file must be sorted in numerically decreasing order of the value in the second field; i.e. any given value must be less than or equal to the value on the previous line.

Submission

Vendors will place all similarity files on 2GB Iomega Jaz disks (http://www.iomega.com) using as many as are necessary to hold all the files. Partial submissions are ineligible ñ the total number of files submitted must be identically N. All files must end with the ".sim" extension. Files must not be compressed, merged or archived in any way. Zip drives and smaller Jaz drives are not permitted.

The collection of files submitted by a vendor will be screened before scoring to affirm their validity. A vendor will be notified if their submission does not conform to the specifications above.

Submission of the Development Set

The development set must be submitted according to the rules given above. This paragraph is a specific reiteration thereof and is a description of the steps that Vendors must take. There are seventeen images in the development set, so N = 17. Therefore the images are named i00000.jpg through i00016.jpg. Vendors must supply exactly 17 similarity files named i00000.sim through i00016.sim. Further each similarity file must contain exactly 17 lines; each line will contain the name of one of the images, and the similarity of that image to the one for which the file is named. An image name can occur only once in each similarity file. The image for which the file is named will also occur in its own similarity file. The files are sorted.

Caution

As stated above, for a collection containing N images, a vendor must submit N similarity files and each file has N lines. The total volume of similarity data therefore increases quadratically with N. Vendors should be aware that when the number of images is on the order of 10000 the total disk space required for the similarity files may well be in excess of 2 gigabytes. Vendors are advised that this volume of data may fill common hard drive partitions, and should plan proactively.

Example Similarity File: i00005.sim

```
i00006.jpg 96.3
i00005.jpg 96.0
i00000.jpg 92.2
i00002.jpg 90.9
i00001.jpg 88.5
i00008.jpg 8.81e+01
i00007.jpg 84.0
i00003.jpg 80.04001
i00009.jpg 77
i00010.jpg 63.1
i00004.jpg 50.1111
i00011.jpg 33.2
i00012.jpg 33.2
i00013.jpg 33.0
i00015.jpg 20
i00014.jpg ñ11.03
i00016.jpg ñ200.00
```

Note that there are 17 entries and each image is referenced exactly once.

Submit an unanswered question

Frequently Asked Questions (Restricted Area)

1. How many images will be used for the Recognition Performance test?
2. What if I have one system that does only live video recognition and one system that does only offline recognition?
3. For the Recognition Performance Test, will we be allowed to read the images and make/store a template of the image on our hard disk so that we only have to make a template once? Will I be able to write the similarity files to the hard disk during processing & then copy the files to the Jaz disk?
4. You have stated that we will be allowed one day to perform the Recognition Performance Test, but our studies have indicated that it will take XX hours to do these. What will we do?
5. Will we be able to have the computer make comparisons & write similarity files at night?
6. In some of the images in the Facial Recognition Vendor Test 2000 Demonstration Data Set my algorithm was not able to automatically find the eyes. How will this affect my test results?
7. My company does not have a copy of the FERET development image set (or recently acquired these images) and we are worried that other vendors, who have these images, may have an advantage. How is this being taken care of?
8. How long will we be allowed to take the Product Usability Test?
9. My commercial product does not have a verification mode. How will this affect my participation?
10. The use of live subjects in the Product Usability Test throws some uncertainties in the evaluation. How will this be handled?
11. Will the test subjects always be moving?
12. Will there be individuals other than the test subjects in the field of view?
13. My product wasn't developed for access control so we don't have a live subject measure. Won't the test using the picture paint an unfair portrait of my product?
14. Approximately how many images will be in the database for the Product Usability Test?
15. Asking us to set our systems so that they fall in an approximate range on a hypothetical ROC curve seems a bit too vague. Could you narrow this down for us?
16. Will the vendors need to bring their own equipment (computer, camera, etc.) or will the government be providing these?
17. Who do I send the development set similarity files to?
18. Can we bring two separate systems to the test in order to run the Recognition Performance Test and the Product Usability Tests in parallel?
19. Why separate performance and usability tests? How can the results be combined and assessed? If I was choosing a system to buy, I'd want to know the recognition performance on real-world images - i.e. those from the usability test. Performance and usability are not separable so it seems scientifically dangerous to test them separately - surely the answers you need are to the question 'Which system works best on real-world images?'

20. Two questions in one really: What proportion of the recognition test set are of very small faces, such as 'i00011'? Our system will not return a similarity score for such images. It is a perfectly reasonable response for a system to 'abstain' when it does not consider the input data to be reliable enough to give an accurate similarity score. How is 'abstention' dealt with in the test?
21. Are the images 24 bit color and 8 bit grayscale only or are some color mapped (i.e. 8 bit color)?
22. What format will the images be in for the Product Usability Test? Where will the images be located?
23. What would happen if I was unable to generate an application in time that would satisfy the Product Usability Test but I do have one that would do the Recognition Performance Test? Could I still take the Recognition Performance Test but not the Product Usability Test?
24. When doing the Enrollment Timed Test, how many images will we be able to enroll per person? Our normal procedure is to enroll more than one.
25. I'd like to ship my system(s) to you in advance. Can you supply monitors for us to use so we can save shipping costs and avoid the risk of damage?
26. Can you provide an uninterruptible power supply for the Recognition Performance Test in case of a power outage during the overnight processing?
27. Can we still submit questions about the test once testing begins?
28. Can we e-mail our sample similarity files instead of mailing a JAZ disk?
29. Can we setup our systems on the day we arrive so we're ready to begin the next day?
30. Where are the targets located relative to the camera?

1. How many images will be used for the Recognition Performance test?
The exact number of images has not been determined since the image are still being processed. However, we expect the database to contain about 10,000 images.

2. What if I have one system that does only live video recognition and one system that does only offline recognition?
As long as both systems use the same recognition engine, both may be tested. The offline system will undergo the Recognition Performance test. The live video system will undergo the Product Usability tests. The product descriptions that you provide should state the intended use of each system as well as the fact that each system is built around the same recognition algorithm.

3. For the Recognition Performance Test, will we be allowed to read the images and make/store a template of the image on our hard disk so that we only have to make a template once? Will I be able to write the similarity files to the hard disk during processing & then copy the files to the Jaz disk?
Yes, this is the preferred method since it is significantly faster. All data and files derived from the images (i.e. templates, similarity files) must be removed at the conclusion of the test. The government must be allowed to ensure this has been completed.

4. You have stated that we will be allowed one day to perform the Recognition Performance Test, but our studies have indicated that it will take XX hours to do these. What will we do?
This question surprised the sponsors of the test since our baseline algorithm was able to perform this test with 10000 images in 6.5 hours (on a standard 1998 400MHz pc) and the longest period required in the FERET program, with significantly slower computers, was two days. We will have to modify our testing procedures slightly. Vendors will be given up to three days to perform the Recognition Performance Test - this includes template generation as well as matching, sorting results, and writing the similarity files. Matching and template

generation must begin with i00000.jpg, continue with i00001.jpg, then i00002.jpg and so forth. Vendors will be allowed to continue testing until the time limit is reached. Test results will be released based only on those images that all vendors provided results. However, if one vendor finishes a small number of the images & all the other vendors complete the entire set, the Government will be forced to make a decision on how to release the results.

5. Will we be able to have the computer make comparisons & write similarity files at night?
Yes. Hours for the test area(s) are 0900-1700 every day. This is the only time that anyone will be allowed in the test area(s). You may setup your computer to run scores during off-hours, but this will be unsupervised by vendor representatives or the Government. The Government suggests that you provide some technique so that if your system crashes during the analysis you will not have to start the process again from the beginning.

6. In some of the images in the Facial Recognition Vendor Test 2000 Demonstration Data Set my algorithm was not able to automatically find the eyes. How will this affect my test results?
This is a problem that was first encountered in the FERET evaluations and is accounted for in the scoring code. This is a system level test, so a failure to acquire on a particular image (which is scored the same as a failure to identify) is a valid measurement of system performance. For a real world example, consider a facial recognition product that is comparing a picture database to images from a surveillance camera. If the facial recognition engine cannot find the eyes on a subject in the surveillance camera it could turn into a significant problem!

7. My company does not have a copy of the FERET development image set (or recently acquired these images) and we are worried that other vendors, who have these images, may have an advantage. How is this being taken care of?
The FERET development image set has been available to anyone that would be eligible to participate in these evaluations. However, only 1/3 of the FERET database has been seen by those outside the government. The other 2/3 has been sequestered for tests such as these. The FERET images used in these evaluations, which will only be a portion of the total images, will be from the previously sequestered images.

8. How long will we be allowed to take the Product Usability Test?
You will be given one full test day (0900-1700)

9. My commercial product does not have a verification mode. How will this affect my participation?
You will be allowed to play with templates & directories so that you can do an "identification" analysis with only the verification test subject in the database as long as you are able to complete the entire test during the test period.

10. The use of live subjects in the Product Usability Test throws some uncertainties in the evaluation. How will this be handled?
The Product Usability Test is more of an operability measure than a performance analysis (which is the Recognition Performance Test). The goal is to show that time is a concern when choosing these systems and that certain parameters such as distance and lighting can change this time. Consistent test subjects are still necessary, however. Two control methods will be in place to help with this. The first is that the test subjects will be required to walk a certain path and look at specific locations. The second is that the test subjects will practice this several times before the first vendor is tested so that they are in a routine before the tests begin.

11. Will the test subjects always be moving?
No. At times they will be standing in a stationary location.

12. Will there be individuals other than the test subjects in the field of view?
No.

13. My product wasn't developed for access control so we don't have a live subject measure. Won't the test using the picture paint an unfair portrait of my product?
The sponsors fully understand this concern and have planned accordingly. Access control was chosen as the scenario for the product usability test. Other scenarios do exist, such as booking stations, where a "live" test would not be needed but would limit the amount of realistic data that we could collect. The Government plans to have text in the test report stating that failure of the "live" test is not a concern for non-access control products (it is for access control products!).

We also invite participating vendors to comment on this in their 4 page system write-up.

14. Approximately how many images will be in the database for the Product Usability Test?
On the order of 150 images.

15. Asking us to set our systems so that they fall in an approximate range on a hypothetical ROC curve seems a bit too vague. Could you narrow this down for us?
Sure, we're not completely inflexible! Instead of a hypothetical ROC curve, we will provide you with a target false acceptance rate to shoot for. These rates will be included in the test plan that will be available on the restricted portion of the Facial Recognition Vendor Test 2000 web site.

16. Will the vendors need to bring their own equipment (computer, camera, etc.) or will the government be providing these?
This is a system test & the Government wants each vendor to provide the components they would normally recommend to someone that has these requirements. The makeup of the system (including computer, camera, etc) and the cost breakdown should be provided in the vendor's four page product description.

17. Who do I send the development set similarity files to?
You need to send these to Mike Bone so that he receives them by March 27.

18. Can we bring two separate systems to the test in order to run the Recognition Performance Test and the Product Usability Tests in parallel?
Absolutely. In fact we encourage you to do so. This will help speed up the testing process and eliminate some of the "down time" while the Recognition Performance Test is processing. The systems must be the same (including recognition algorithm, computer, and miscellaneous components such as processor and memory). Otherwise you are attempting to enter two different systems.

19. Why separate performance and usability tests? How can the results be combined and assessed? If I was choosing a system to buy, I'd want to know the recognition performance on real-world images - i.e. those from the usability test. Performance and usability are not separable so it seems scientifically dangerous to test them separately - surely the answers you need are to the question 'Which system works best on real-world images?'
There are a number of different reasons for conducting an evaluation, and the design of the evaluation is based on the desired purpose. The purposes of the Facial Recognition Vendor Test 2000 are to advance the state-of-the-art, to measure the current state-of-the-art, and to measure the performance of system X at control access to building Y. A more detailed discussion on the subject of evaluating biometric systems can be found in "Introduction to Evaluating Biometrics Systems" in the February 2000 issue of IEEE Computer.

The purpose of the recognition performance part of the Facial Recognition Vendor Test 2000 is assessing the state-of-the-art of commercially available face recognition systems. In terms of the above article, it is a technology test and is designed to assess general ability. This includes the ability to perform identification and verification. This part of the test is designed to evaluate performance on a large dataset.

The purpose of the product usability part is to assess and determine how the systems would function from an operational point of view. Performing two separate, but complimentary, tests allows for a much more detailed understanding of the state-of-the-art. If one were to field or consider a face recognition product for a specific application, we recommend testing candidate face recognition products in that specific application.

20. Two questions in one really: What proportion of the recognition test set are of very small faces, such as 'i00011'? Our system will not return a similarity score for such images. It is a perfectly reasonable response for a system to 'abstain' when it does not consider the input data to be reliable enough to give an accurate similarity score. How is 'abstention' dealt with in the test?
The performance test is intended to assess the state of the art of face recognition "at a distance" - the distance may vary (between 1 meter and maybe 12 meters) as is evident in the Development Set.

The FERET protocol of Sep 96 allows for the scoring of subsets of the images; ie those that belong to certain categories. There will not be one aggregated performance number over all images so if a vendor abstains from certain subsets it will only show in the scoring of those subsets.

One such result of the test will include how state of the art algorithms degrade as the distance increases (or equivalently as resolution decreases). If a vendor abstains from this category of images, it will not affect their performance figures on other subsets (lighting, pose, facial expression etc.)

The vendor should supply a small similarity value in the cases where it abstains. "Small" in this case should be smaller than any value they supplied for the images that they did not abstain on.

21. Are the images 24 bit color and 8 bit grayscale only or are some color mapped (i.e. 8 bit color)?
All images are either 24 bit color or 8 bit grayscale. There are no color mapped images.

22. What format will the images be in for the Product Usability Test? Where will the images be located?
The images will be similar to image i00012.jpg in the Image Development Set. They will be located in a subdirectory of a JAZ disk.

23. What would happen if I was unable to generate an application in time that would satisfy the Product Usability Test but I do have one that would do the Recognition Performance Test? Could I still take the Recognition Performance Test but not the Product Usability Test?
The test report would simply state that your company declined to submit a product for the Product Usability Test. It would be up to you to explain why in your 4-page system description.

24. When doing the Enrollment Timed Test, how many images will we be able to enroll per person? Our normal procedure is to enroll more than one.
You may enroll the live subjects with as many images as your procedure requires. The idea is that you do the things that you would recommend to your customers to achieve the best results. This distinguishes the results of this test from the Old Image Database Timed Test where you must enroll just the images made available to you.

25. I'd like to ship my system(s) to you in advance. Can you supply monitors for us to use so we can save shipping costs and avoid the risk of damage?
We have a few monitors that we can make available to you during the tests. Just send e-mail to Mike Bone stating your requirements and we'll see if any of our monitors fit your needs. Also, be sure to include the monitor you would normally use in your system description and price.

26. Can you provide an uninterruptible power supply for the Recognition Performance Test in case of a power outage during the overnight processing?
Yes, we will provide a small UPS. It won't last through an extended outage, but should handle the unlikely event of a small power glitch.

27. Can we still submit questions about the test once testing begins?
It wouldn't be fair to the vendors who have already been tested if the other vendors are able to get additional information about how the test will be run. We will continue to freely answer questions and post them to the restricted area FAQ until COB April 7. After that, you can still send questions, but we will use our judgement in deciding whether or not to answer them. If we decide that answering your question would give your company an unfair advantage, we will decline to do so.

28. Can we e-mail our sample similarity files instead of mailing a JAZ disk?
Yes. You can e-mail them to Mike Bone at bone_mike@crane.navy.mil, but you must ensure that you are able to read/write to a Jaz disk prior to testing. We will not give you additional time to setup a Jaz drive on your system at the test site.

29. Can we setup our systems on the day we arrive so we're ready to begin the next day?
You can setup your system for the Recognition Performance Test the day you arrive then begin the test the next day if you wish. You can also begin the test on the day you arrive. The current test plan states that the 72 hour time limit includes setup time, but this will be changed in the next revision.

The Product Usability Tests will be run differently, however. The 0900 - 1700 time limit will include setup time. You will not be allowed to setup on your arrival date and start the test the next day. The reason for this is that the amount of time available on the arrival date may vary between vendors. This means that some would gain an advantage by having more time to tune their system to the environment. To be fair to all vendors, we must limit setup activities to the 0900 - 1700 time limit imposed on the test.

30. Where are the targets located relative to the camera?
The left and right targets are spaced 18 feet apart on a wall located 16.5 feet behind the camera mark. The camera mark is located in the center of these 2 targets. The center target is located at the camera mark and the center of the target is 6 inches above the floor.

Logout

Images

Restricted Area

API | FAQ | Test Plan

The images available for download constitute the Facial Recognition Vendor Test 2000 Demonstration Data Set, which is a subset of the HumanID Data Set. As such, access to these images are controlled. You have been given permission to use these images since you have signed the form "Application for Access to a Portion of the Development HumanID Data Set and FERET Database" on this web site. Do not share these images with anyone outside of your organizational control.

For your convenience, the Data Set is provided in one compressed (.zip) file.

Download Data Set, FRVTImages.zip (approximately 1MB)

Logout

Test Plan Document

The following document contains the test plan that will be followed for FRVT 2000. We're still doing trial runs with the test subjects to give them some practice and fine tune the procedures, so there may be some minor modifications to the plan in the following week. However, the basic testing structure will remain the same and no new tests will be added.

Note: The conversion of the test plan to PDF format didn't seem to work very well on the data recording tables in the appendix. Although they don't look right when viewed on screen, they seem to look OK when printed to a PostScript printer.

Test Plan Document (March 24, 2000)

Test Plan Document (March 31, 2000)

UPDATE March 31, 2000: The test plan has been modified slightly to clear up a few points of confusion and to incorporate some improvements in the live subject procedures that were identified in the practice sessions. Changes are as follows:

- Test Overview -> GeneralFixed typing error in first paragraph. Updated start time requirements for tests.
- Test Overview -> Test Space -> Room Layout.Reworded Station 2 description to clarify.
- Test Overview -> Test Space -> Floor Marks.Reworded to clarify.
- Test Overview -> Test Space -> Visual Targets.Added more details on the location of targets.
- Test Overview -> Testing Conventions -> Start Location.Changed the way subjects begin identification trials. Rather than beginning with their backs to the camera then turning 180 degrees when the timer is started, we found it was easier for the subjects to begin facing perpendicular to the camera path, but with their heads turned to face away from the camera. This requires subjects to merely turn their bodies 90 degrees when the timer is started.
- Recognition Performance Test -> Test Description -> Time Limits.Updated to allow setup of vendor system before starting 72 hour timer.
- Recognition Performance Test -> Test Procedure -> Test Procedure.Added steps to record start and end times and amendeded test procedures to state that system settings may not be changed once a particular scenario has started.

Schedule

May 1 - 5	Visionics Corporation
May 8 - 12	Lau Technologies
May 15 - 19	Miros, Inc.
May 22 - 26	C-VIS Computer Vision und Automation GmbH
June 5 - 9	Banque-Tec International Pty. Ltd.

If you haven't already done so, please send e-mail to bone_mike@crane.navy.mil with the names of the representative(s) that will be coming along with arrival and departure dates. This will depend on how long you expect your system will need to process all images in the Recognition Performance Test and whether or not you bring two systems to run the Product Usability Tests in parallel. Remember that up to 72 hours will be permitted for the Recognition Performance Test and one business day from 0900 - 1700 will be permitted for the Product Usability Tests. All testing must be completed by 1700 on the Friday of your test week.

Travel Information

The test will be held at NAVSEA Crane. Crane is located approximately 70 miles southwest of Indianapolis, IN. The nearest terminal is Indianapolis International Airport, and lodging can be found in Bloomington, IN. See http://www.crane.navy.mil/General/visit_info.htm for directions to Bloomington and NAVSEA Crane from the terminal and a list of Bloomington hotels. The hotels that are closest to the driving route are the Fairfield Inn, Comfort Inn, Day's Inn, Hampton Inn, Holiday Inn, Quality Inn, and Ramada Inn. Directions from the NAVSEA Crane gate to the test space will be provided before testing begins.

Crane, IN is in the GMT -5:00 time zone.

Logout

Appendix C
FRVT 2000 Test Plan

Facial Recognition Vendor Test 2000

Test Plan

March 31, 2000

Introduction

This document describes the procedures that will be followed for Facial Recognition Vendor Test 2000. The result of each vendor test will be a set of similarity files that will be processed to generate graphs of recognition performance, data recorded for timed tests, a vendor supplied document describing the tested system and associated costs, and video records of all testing activities. At the conclusion of all tests, a final report will be written describing the results.

Test Overview

General

The test is made up of 2 parts: the Recognition Performance Test and the Product Usability Test. Both are described in detail below. Vendors will supply and operate their own equipment. Government personnel will direct test activities and record data. Vendors may bring 2 separate but identical systems to the test so the Recognition Performance Test and Product Usability Tests can be run in parallel. The Recognition Performance Test will be initiated first, either on the day the vendor arrives or the following morning. The Product Usability Tests will begin the morning after the vendor arrival date if 2 systems are available. Otherwise it will begin the morning following completion of the Recognition Performance Test.

Personnel

Test Agent

Test agent refers to the government representative administering the test and recording results, and possibly and assistant.

Vendor

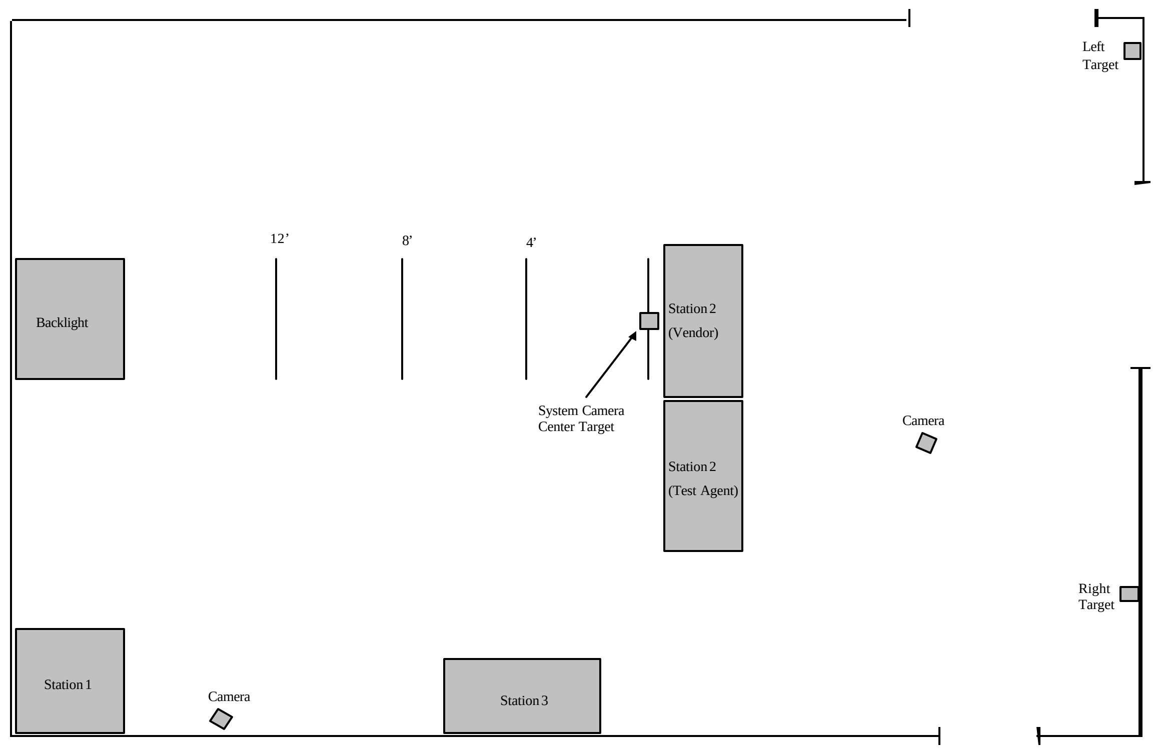
Vendor refers to the representative(s) of the company whose product is being tested.

Subjects

There will be 3 live test subjects that will take part in the timed tests. They will be referred to specifically as subject 1, subject 2, and subject 3, or generically as subjects. There will be one female subject and two male subjects. One male subject will be wearing glasses. An 8” x 10” color photograph of one subject will be used in some trials. It will be held in front of one subject’s face during these trials. For clothing uniformity, all subjects will be wearing lab coats of the same color. All subjects will practice the trials before the first vendor test in order to establish a consistent routine.

Test Space

The following diagram shows the approximate layout of the testing area.



<u>Room Layout</u>

All tests will be held in a single room. There will be 3 stations setup in the testing room, each near a 120 volt standard US power outlet. The stations are assigned as follows:

Station 1 will be used for the Recognition Performance Test. It consists of a table to hold the vendor system, and chairs.

Station 2 will be used for the timed tests. It consists of 2 tables with chairs, markings on the floor to direct subject position, and visual targets to focus subjects' attention during non-cooperative behavior modes. One table with adjustable height will be dedicated to the vendor system. If the vendor system includes camera and lighting supports, they may be placed on the table or on the floor in front of the table. The other table will be used by the test agent when recording data and to hold a timer.

Station 3 will be used for the Access Control System Interface Test. It consists of a table for the vendor system and the access control system to which it will be interfaced.

Floor Marks

Marks have been taped to the floor in front of station 2 as follows:

A camera mark designates the location at which the front of the system camera lens should be aligned. All other floor marks are assigned distances relative to the camera mark in the direction of the subject. The camera mark is labelled “CAMERA”.

Marks have been placed at one foot intervals starting one foot in front of the camera mark and ending 12 feet from the camera mark. These marks are used to determine where subjects start the trials, where they stop if the system has not achieved identification or verification, and to determine the final distance between the subject and the camera. Each mark is labelled with its distance from the camera mark and is referred to as a distance mark. The mark placed one foot from the camera mark is also labelled “STOP” and is referred to as the stop mark. The marks at 4 feet, 8 feet, and 12 feet are also labelled “START” and are referred to as start marks.

Visual Targets

Two visual targets printed on 8-1/2” x 11” paper have been posted on the walls behind the vendor system table for subjects to focus their attention during the non-cooperative behavior modes of the timed tests. The center of each target is 6’ above the floor, spaced 18’ apart on a wall located 16.5’ behind the camera mark. These targets are referred to as the left target and right target. A third target, referred to as the center target, has been placed in a vertical position with its center 6” above the floor at the camera mark. The system camera will serve as the visual target for cooperative behavior modes.

Room Lighting

The test room is illuminated with overhead fluorescent lights. The room also has outside windows that will be completely covered by opaque material for the duration of the tests.

Back Lighting

Back lighting will be used for some trials in the timed tests. This is meant to simulate the presence of an outside window behind the subject in a controlled manner. To accomplish this, a custom lighting device has been built consisting of a track lighting system with fixtures arranged in a 4 x 4 grid.

The lights are mounted inside a box facing toward the camera. The front side of the box, which faces the camera, has approximate dimensions of 4’ x 4’ and is covered by a translucent diffusing material.

Testing Conventions

Start Location

When the subject is to begin a timed test by facing the camera, (s)he will stand with toes behind the designated start mark. When the subject is to begin by facing away from the camera, (s)he will stand with body facing perpendicular to the camera path, head turned away from the camera, and feet behind the designated start mark.

Time Measurement

For each trial of the timed tests, the time will be recorded to the nearest 1/10 second.

Behavior Mode

Two behavior modes will be employed by subjects during the timed tests: cooperative and non-cooperative.

In the cooperative mode, subjects will look directly at the camera with very little head movement while walking or standing.

In the non-cooperative mode, subjects will begin a trial looking at the right target. Once the timer is started, the subject will turn his/her head slowly, moving visual focus on a triangular path from the right target to the left target, down to the center target, then back up to the right target. This will be done using a cadence that allows 2 complete cycles in 10 seconds. This will be done for both standing and walking trials.

Distance Measurement

The final distance from subject to camera in the timed tests will be recorded as the label of the last distance mark that the subject's toes have reached.

Video Recording

All testing activities will be recorded with video cameras to ensure accurate records. One camera will be used to record the activities at station 1 during the Recognition Performance Test. The camera will be placed so that the screen of the vendor system and any operator activity will be in the field of view. The camera will be recording for the duration of the test but only when the testing room is occupied.

Another camera will be placed behind station 2 so that the vendor system and the subjects will be in the field of view. This camera will be recording during system setup and during all trials of the timed tests. This camera will also be used to record all activities of the Access Control System Interface Test at station 3.

Recognition Performance Test

Test Description

Overview

The Recognition Performance Test will be very similar to the original FERET tests that were sponsored by the DoD Counterdrug Technology Development Program Office. Since the conclusion of the original FERET program, the data sets and reports have been transferred to the National Institute of Standards and Technology (NIST), who is serving as a technical consultant for these tests. Images used in this test will be a combination of images from the FERET database as well as DARPA's new HumanID database.

On the day of the Recognition Performance Test, the vendor will be given a set of test images in JPG format. The vendor may convert the images to another format if necessary, but no extra time will be given for this. The vendor will use their algorithm to compare each image to the others and report the similarity scores in the format defined in the API document.

Time Limits

Once the system is setup, vendors will be allowed 72 continuous hours to process the test images. Vendors should process test images in filename order so that if time runs out before processing is complete, a common set of similarity scores can be identified among all vendors. Vendors will only be allowed access to the test space between the hours of 0900 and 1700. The system may continue to process test images outside these hours during the 72 hour time period, but in the event of an overnight system crash, vendors will not be allowed to restart the system until 0900 the following day. Vendors are encouraged to implement their system in a manner that allows restarting from the point where a system crash occurred rather than restarting from the beginning.

Data Recording

Vendors will generate a similarity file for each test image. All similarity files will be stored on one or more 2GB Jaz disks and submitted to the test agent at test completion. The final report will show the results in the form of Receiver Operating Characteristic (ROC) curves for verification tests and Cumulative Match Characteristics (CMC) Curves for identification tests. Other forms of displaying information may also be used.

Test Procedure

Preparation

1. Test agent records available space on system hard disk.
2. Test agent releases Jaz disk containing image database to vendor.

Test Procedure

1. Test agent records start time.
2. Vendor inserts Jaz disk into system and initiates test sequence.
3. Test ends when all images have been processed or 72 hours has elapsed.
4. Test Agent records end time.
5. Test agent collects Jaz disk containing similarity files from vendor.
6. Vendor deletes any remaining templates and similarity files from system hard disk.
7. Test agent records available space on hard disk.
8. Test agent wipes free space on hard disk.

Product Usability Tests

Test Description

Overview

The Product Usability Tests will consist of two timed tests and an optional interface test. The timed tests will be used to measure the response time of the overall system for two different operational scenario simulations: the Old Image Database Timed Test and the Enrollment Timed Test. Optionally, the Access Control System Interface Test will be used to determine if the system can communicate with an access control system. It is not necessary for a vendor to have an access control product to participate in these tests - these are operational scenarios that were developed to give the public a means of comparing the test results with something they would be familiar with.

The operational scenario for the Old Image Database Timed Test is that of a low security access control point into the lobby of a building. The building's security officers want to improve security into the area but do not want to slow down the flow through the entry area. The security officers also do not want to mandate that the employees take the time to enroll into the new system so they will use their existing digital image database taken from the employee's picture ID badges. Some employees may not be aware that they are being checked using a facial recognition system, so they will not always be fully cooperative.

The operational scenario for the Enrollment Timed Test is that of an access control door for a medium/high security area within the building previously described. In this case, employees will be enrolled in the facial recognition system using the standard procedures recommended by the vendor. The access control system on one door has been setup so that an individual enters his identity and the system must verify if this is indeed the correct individual. On another door, the system has been setup so that an individual simply walks up to the camera and the door opens if the identity of the individual matches an individual in the database with valid credentials. The employees will be aware that they are being checked using a facial recognition system, but may or may not be cooperative.

Each of the timed tests will be performed for both verification and identification and will be performed once with overhead fluorescent lighting and again with the addition of simulated back lighting.

The Access Control System Interface Test is an optional test meant to determine if the facial recognition system can interface successfully with an access control system. To participate in this optional test, the facial recognition system must have a WIEGAND interface. The goal is to test the interface rather than the facial recognition algorithm.

Time Limits

One full business day, between the hours of 0900 and 1700, will be allowed for running all Product Usability Tests, including setup. Each trial for the timed tests will be limited to 10 seconds. Based on our trials, we expect the test to take about 6 hours, not including setup time or the optional Access Control System Interface Test.

Data Recording

Results of the timed tests will be recorded on the tables in Appendix A. The distance between the subject and the camera at the end of each trial will be recorded in the Final Distance column to the nearest 1 foot increment. This column will not be used for trials where the subject stands in place. If the system acquires a match for the subject, the time to make the match will be recorded in the Acquire Time column to the nearest 1/10 second. If the correct match was acquired, the word ‘yes’ will be recorded in the Correct Match column. If a match was not acquired or was incorrect, the word ‘no’ will be recorded in the Correct Match column. These tables will be published in the final report as recorded without any analysis.

Old Image Database Timed Test Procedure

<u>Preparation</u>

1. Vendor sets up system at station 2. Front of camera lens is aligned with camera mark.
2. Test agent releases access control image database to vendor on Jaz disk.
3. Vendor enrolls database images from Jaz disk (~150 images, one image per subject).
4. Vendor adjusts system for the suggested low security false alarm (false positive) rate of 0.4%. Settings may not be changed until backlighting is added. This includes camera zoom, unless the system controls the camera automatically while attempting recognition.

<u>Verification Test</u>

Steps 1 - 6 are repeated with <start distance> = 12', 8', and 4'

1. Subject 1 stands facing camera at <start distance> mark.
2. Vendor enters subject 1 ID into system.
3. Test agent vocally counts to 3. On "3", test agent starts timer while vendor simultaneously presses key to begin verification.
4. Subject 1 walks toward camera using cooperative behavior mode.
5. Subject 1 stops walking if vendor and test agent acknowledge match, time expires, or stop mark is reached. If stop mark is reached before time expires, subject 1 stands at stop mark until vendor and test agent acknowledge match or time expires.
6. If vendor and test agent acknowledge match, test agent records time, distance, and correctness of match. If match does not occur before time expires, test agent records that fact.

Steps 7 - 12 are repeated with <start distance> = 12', 8', and 4'

7. Subject 1 stands facing camera at <start distance> mark.
8. Vendor enters subject 1 ID into system.
9. Test agent vocally counts to 3. On "3", test agent starts timer while vendor simultaneously presses key to begin verification.
10. Subject 1 walks toward camera using non-cooperative behavior mode.
11. Subject 1 stops walking if vendor and test agent acknowledge match, time expires, or stop mark is reached. If stop mark is reached before time expires, subject 1 stands at stop mark until vendor and test agent acknowledge match or time expires.
12. If vendor and test agent acknowledge match, test agent records time, distance, and correctness of match. If match does not occur before time expires, test agent records that fact.

13. Repeat steps 1 – 12 with subject 2 for each value of <start distance>.
14. Repeat steps 1 – 12 with subject 3 for each value of <start distance>.
15. Repeat steps 1 – 6 (8) times with subject 3 for <start distance> = 12' to test variability.
16. Repeat steps 1 – 6 with subject holding photograph for each value of <start distance>.

Identification Test

Steps 1 - 5 are repeated with <start distance> = 12', 8', and 4'.

1. Subject 1 stands facing away from camera at <start distance> mark.
2. Test agent starts timer.
3. Subject 1 turns then walks toward camera using cooperative behavior mode.
4. Subject 1 stops walking when vendor and test agent acknowledge match, time expires, or stop mark is reached. If stop mark is reached before time expires, subject 1 stands at stop mark until vendor and test agent acknowledge match or time expires.
5. If vendor and test agent acknowledge match, test agent records time, distance, and correctness of match. If match does not occur before time expires, test agent records that fact.

Steps 6 - 10 are repeated with <start distance> = 12', 8', and 4'.

6. Subject 1 stands facing away from camera at <start distance> mark.
7. Test agent starts timer.
8. Subject 1 turns then walks toward camera using non-cooperative behavior mode.
9. Subject 1 stops walking when vendor and test agent acknowledge match, time expires, or stop mark is reached. If stop mark is reached before time expires, subject 1 stands at stop mark until vendor and test agent acknowledge match or time expires.
10. If vendor and test agent acknowledge match, test agent records time, distance, and correctness of match. If match does not occur before time expires, test agent records that fact

11. Repeat steps 1 – 10 with subject 2 for each value of <start distance>.
12. Repeat steps 1 – 10 with subject 3 for each value of <start distance>.
13. Repeat steps 1 – 5 (8) times with subject 3 for <start distance> = 12' to test variability.
14. Repeat steps 1 – 5 with subject holding photograph for each value of <start distance>.

Backlighting Test

1. Test agent adds backlighting behind subject.
2. Vendor may adjust settings if necessary. Settings may not be changed until backlighting is removed. This includes camera zoom, unless the system controls the camera automatically while attempting recognition.
3. Repeat steps 1 – 16 of the Verification Test with Subjects 1, 2, 3, variability test, and photograph.

4. Repeat steps 1 – 14 of the Identification Test with Subjects 1, 2, 3, variability test, and photograph.
5. Test agent removes backlighting.

Enrollment Timed Test Procedure

Preparation

1. Vendor deletes enrolled templates of 3 test subjects.
2. Vendor adjusts system for the medium/high security false alarm (false positive) rate of 0.1%. Settings may not be changed until backlighting is added. This includes camera zoom, unless the system controls the camera automatically while attempting recognition.
3. Vendor enrolls 3 test subjects using their standard enrollment procedure.

Verification Test

Steps 1 – 6 are repeated with <start distance> = 12', 8', and 4'.

1. Subject 1 stands facing camera at <start distance> mark.
2. Vendor enters subject 1 ID into system.
3. Test agent vocally counts to 3. On "3", test agent starts timer while vendor simultaneously presses key to begin verification.
4. Subject 1 continues standing using cooperative behavior mode during recognition attempt.
5. Trial ends when vendor and test agent acknowledge match or time expires.
6. If vendor and test agent acknowledge match, test agent records time and correctness of match. If match does not occur before time expires, test agent records that fact.

Steps 7 – 12 are repeated with <start distance> = 12', 8', and 4'.

7. Subject 1 stands facing camera at <start distance> mark.
8. Vendor enters subject 1 ID into system.
9. Test agent vocally counts to 3. On "3", test agent starts timer while vendor simultaneously presses key to begin verification.
10. Subject 1 continues standing using non-cooperative behavior mode during recognition attempt.
11. Trial ends when vendor and test agent acknowledge match or time expires.
12. If vendor and test agent acknowledge match, test agent records time and correctness of match. If match does not occur before time expires, test agent records that fact.

13. Repeat steps 1 – 12 with subject 2 for each value of <start distance>.
14. Repeat steps 1 – 12 with subject 3 for each value of <start distance>.
15. Repeat steps 1 – 6 (8) times with subject 3 for <start distance> = 12' to test variability.
16. Repeat steps 1 – 6 with subject holding photograph for each value of <start distance>.

Identification Test

Steps 1 – 5 are repeated with <start distance> = 12', 8', and 4'.

1. Subject 1 stands facing away from camera at <start distance> mark.
2. Test agent starts timer.
3. Subject 1 turns to face camera then stands using cooperative behavior mode during recognition attempt.
4. Trial ends when vendor and test agent acknowledge match or time expires.
5. If vendor and test agent acknowledge match, test agent records time and correctness of match. If match does not occur before time expires, test agent records that fact.

Steps 6 – 10 are repeated with <start distance> = 12', 8', and 4'.

6. Subject 1 stands facing away from camera at <start distance> mark.
7. Test agent starts timer.
8. Subject 1 turns to face camera then stands using non-cooperative behavior mode during recognition attempt.
9. Trial ends when vendor and test agent acknowledge match or time expires.
10. If vendor and test agent acknowledge match, test agent records time and correctness of match. If match does not occur before time expires, test agent records that fact.

11. Repeat steps 1 – 10 with subject 2 for each value of <start distance>.
12. Repeat steps 1 – 10 with subject 3 for each value of <start distance>.
13. Repeat steps 1 – 5 (8) times with subject 3 for <start distance> = 12' to test variability.
14. Repeat steps 1 – 5 with subject holding photograph for each value of <start distance>.

Backlighting Test

1. Test agent adds backlighting behind subject.
2. Vendor may adjust settings if necessary. Settings may not be changed until backlighting is removed. This includes camera zoom, unless the system controls the camera automatically while attempting recognition.
3. Repeat steps 1 – 16 of the Verification Test with Subjects 1, 2, 3, variability test, and photograph.
4. Repeat steps 1 – 14 of the Identification Test with Subjects 1, 2, 3, variability test, and photograph.
5. Test agent removes backlighting.

Access Control System Interface Test Procedure

Preparation

1. Vendor sets up system at Station 3.
2. Vendor connects system to WIEGAND interface of supplied access control system.
3. Vendor enrolls subject 1 in recognition system and assigns WIEGAND format ID.
4. Test agent enrolls subject 1 in access control system using the same WIEGAND format ID.

Test Procedure

1. Vendor enters subject 1 ID into recognition system.
2. Subject 1 stands in front of camera and cooperates to try to achieve successful verification.
3. If successful verification occurs and access control system receives the correct WIEGAND format ID, test agent records successful completion of test. Otherwise, vendor will be given 2 more attempts at successful completion. If successful completion does not occur in 3 total attempts, test agent will record that fact.

The following information is to be completed for each vendor. By signing below, the vendor agrees that the data recorded in the following tables is accurate. Signatures are for the records of the FRVT 2000 sponsors only and will not appear in the test report.

Vendor Name: ______________________________

Product Usability Test Date: ______________________________

Vendor Representative (print): ______________________________

(sign): ______________________________

(date): ______________________________

Test Agent (print): ______________________________

(sign): ______________________________

(date): ______________________________

Old Image Database Timed Test

Back Light ?	Recognition Mode	Subject Instructions	Subject ID	Behavior Mode	Start Distance	Final Distance	Acquire Time	Correct Match ?
No	Verification	Subject faces camera. ID entered. Timer starts. Subject walks toward camera.	1	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			2	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3 Variability Test	Cooperative	12			
					12			
					12			
					12			
					12			
					12			
					12			
					12			
			Photo	Cooperative	12			
					8			
					4			
	Identification	Subject faces away from camera. Timer starts. Subject turns then walks toward camera.	1	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			2	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3 Variability Test	Cooperative	12			
					12			
					12			
					12			
					12			
					12			
					12			
					12			
			Photo	Cooperative	12			
					8			
					4			

Old Image Database Timed Test

Back Light ?	Recognition Mode	Subject Instructions	Subject ID	Behavior Mode	Start Distance	Final Distance	Acquire Time	Correct Match ?
Yes	Verification	Subject faces camera. ID entered. Timer starts. Subject walks toward camera.	1	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			2	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3 Variability Test	Cooperative	12			
					12			
					12			
					12			
					12			
					12			
					12			
					12			
			Photo	Cooperative	12			
					8			
					4			
	Identification	Subject faces away from camera. Timer starts. Subject turns then walks toward camera.	1	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			2	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3	Cooperative	12			
					8			
					4			
				Non Cooperative	12			
					8			
					4			
			3 Variability Test	Cooperative	12			
					12			
					12			
					12			
					12			
					12			
					12			
					12			
			Photo	Cooperative	12			
					8			
					4			

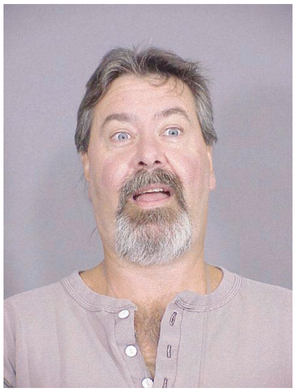
Enrollment Timed Test								
Back Light ?	**Recognition Mode**	**Subject Instructions**	**Subject ID**	**Behavior Mode**	**Start Distance**	**Final Distance**	**Acquire Time**	**Correct Match ?**
No	Verification	Subject faces camera. ID entered. Timer starts. Subject stands in place.	1	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			2	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3 Variability Test	Cooperative	12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
			Photo	Cooperative	12	12		
					8	8		
					4	4		
	Identification	Subject faces away from camera. Timer starts. Subject turns toward camera then stands in place.	1	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			2	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3 Variability Test	Cooperative	12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
			Photo	Cooperative	12	12		
					8	8		
					4	4		

Enrollment Timed Test								
Back Light ?	**Recognition Mode**	**Subject Instructions**	**Subject ID**	**Behavior Mode**	**Start Distance**	**Final Distance**	**Acquire Time**	**Correct Match ?**
Yes	Verification	Subject faces camera. ID entered. Timer starts. Subject stands in place.	1	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			2	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3 Variability Test	Cooperative	12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
			Photo	Cooperative	12	12		
					8	8		
					4	4		
	Identification	Subject faces away from camera. Timer starts. Subject turns toward camera then stands in place.	1	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			2	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3	Cooperative	12	12		
					8	8		
					4	4		
				Non Cooperative	12	12		
					8	8		
					4	4		
			3 Variability Test	Cooperative	12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
					12	12		
			Photo	Cooperative	12	12		
					8	8		
					4	4		

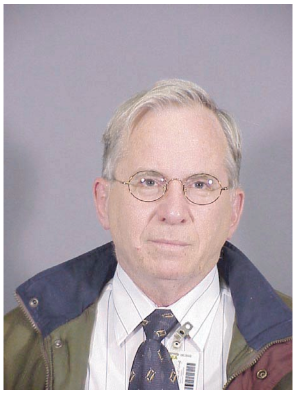
Appendix D
FRVT 2000 Demonstration Data Set

The following images make up a development set made available to participating vendors before the test to ensure that their systems could produce similarity files in the proper format for the scoring software.

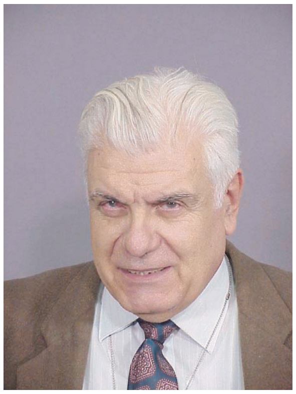
i00000.jpg



i00001.jpg

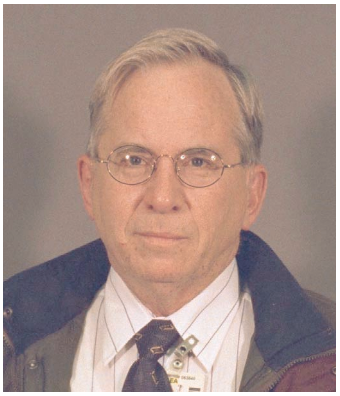


i00002.jpg



i00003.jpg

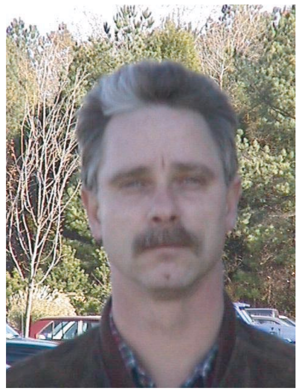
i00004.jpg

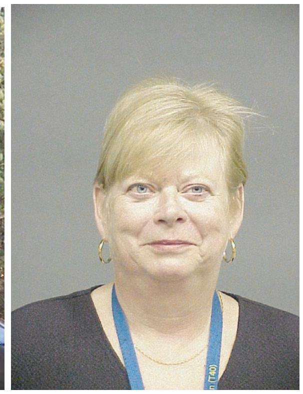


i00005.jpg

i00006.jpg

i00007.jpg





i00008.jpg

i00009.jpg

i00010.jpg

i00011.jpg

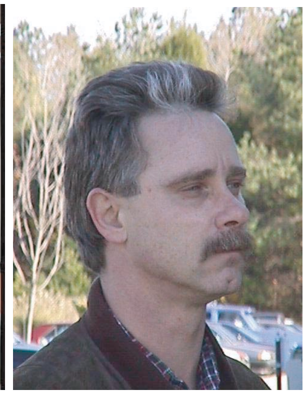
i00012.jpg

i00013.jpg

i00014.jpg

i00015.jpg

i00016.jpg



i00017.jpg

Appendix E
FRVT 2000 Recognition Performance Test
Experiment Descriptions

Recognition Performance Test Experiment Descriptions

Numerous experiments can be performed based on the similarity files returned by the participating vendors. The following subsections, along with tables E1-E9, describe the experiments performed for this thesis. The rows with a white background are designated as FRVT 2000 experiments, while the rows with a gray background are designated as FERET experiments as they used imagery originally collected for the FERET program..

E-1 Compression Experiments

The compression experiments were designed to estimate the effect of lossy image compression on the performance of face-matching algorithms. Although image compression is widely used to satisfy space and bandwidth constraints, its effect in machine vision applications is often assumed to be deleterious; therefore, compression is avoided. This study mimics a situation in which the gallery images were obtained under favorable, uncompressed circumstances, but the probe sets were obtained in a less favorable environment in which compression has been applied. The amount of compression is specified by the compression ratio. The probe sets contain images that were obtained by setting an appropriate quality value on the JPEG compressor such that the output is smaller than the uncompressed input by a factor equal to the compression ratio.

The imagery used in these experiments is part of the FERET corpus; the native source format is uncompressed. The gallery used for the compression experiments is the standard 1,196-image FERET gallery. The probe set used is the 722 images from the FERET duplicate I study.

Table E1 - Compression Experiments

Experiment Name	Compression Ratio	Gallery Size	Probe Size
C0	1:1 (none)	1196	722
C1	10:1	1196	722
C2	20:1	1196	722
C3	30:1	1196	722
C4	40:1	1196	722

E-2 Distance Experiments

The distance experiments were designed to evaluate the performance of face matching algorithms on images of subjects at different distances to the fixed camera. The results of these experiments should be considered for situations where the distance from the subject to the camera for enrollment is different from that used for verification or identification.

In all experiments, the probe images were frames taken from relatively low-resolution, lightly compressed, video sequences obtained using a consumer grade tripod-mounted auto-focus camcorder. In these sequences the subjects walked down a hallway toward the camera. Overhead fluorescent lights were spaced at regular intervals in the hallway, so the illumination changed between frames in the video sequence. This may be thought of as mimicking a low-end video surveillance scenario such as that widely deployed in building lobbies and convenience stores. Two kinds of galleries were used: In experiments D1-D3 the gallery contains images of individuals with normal facial expressions that were acquired indoors using a digital camera under overhead room lights. In experiments D4-D7, however, the gallery itself contains frames extracted from the same video sequences used in the probe sets. Experiments D1-D3, therefore, represent a mugshot vs. subsequent video surveillance scenario in which high-quality imagery is used to populate a database and recognition is performed on images of individuals acquired on video. Experiments D4-D7 test only the effect of distance and avoid the variation due to the camera change.

Note that although the study examines the effect of increasing distance (quoted approximately in meters) the variable often considered relevant to face recognition algorithms is the number of pixels on the face. The distance and this resolution parameter are inversely related. The resolution studies described later also address this effect.

The D4-D5 and D6-D7 studies may be compared to provide a qualitative estimate to the effect of indoor and outdoor lighting. This aspect is covered more fully in the illumination experiments that follow.

Table E2 - Distance Experiments

	Gallery Images		Probe Images			
Experiment Name	Description	Camera Distance	Description	Camera Distanc e	Gallery Size	Probe Size
D1	Indoor, digital, ambient lighting	1.5m	Indoor, video	2m	185	189
D2	Indoor, digital, ambient lighting	1.5m	Indoor, video	3m	185	189
D3	Indoor, digital, ambient lighting	1.5m	Indoor, video	5m	185	189
D4	Indoor, video	2m	Indoor, vido	3m	182	190
D5	Indoor, video	2m	Indoor, video	5m	185	190
D6	Outdoor, video	2m	Outdoor, video	3m	186	195
D7	Outdoor, video	2m	Outdoor, video	5m	186	195

E-3 Expression Experiments

The expression experiments were designed to evaluate the performance of face matching algorithms when comparing images of the same person with different facial expressions. This is an important consideration in almost any situation because it would be rare for a person to have the exact same expression for enrollment as for verification or identification.

The galleries and probe sets contain images of individuals captured at NIST in January 2000 and at Dahlgren in November 1999 using a digital CCD camera and two-lamp, FERET-style lighting. In this and other experiments, fa denotes a normal frontal facial expression, and fb denotes some other frontal expression.

E3 - Expression Experiments				
Experiment Name	Gallery Images	Probe Images	Gallery Size	Probe Size
E1	Regular expression (fa image)	Alternate expression (fb image)	225	228
E2	Alternate expression (fb image)	Regular expression (fa image)	224	228

E-4 Illumination Experiments

The problem of algorithm sensitivity to subject illumination is one of the most studied factors affecting recognition performance. When an image of the subject is taken under different lighting conditions than the condition used at enrollment, recognition performance can be expected to degrade. This is important for systems where the enrollment and the verification or identification are performed using different artificial lights, or when one operation is performed indoors and another outdoors.

The experiments described below use a single gallery containing high-quality, frontal digital stills of individuals taken indoors under mugshot lighting. The variation between experiments is through the probe sets, which are images taken shortly before or after their gallery matches using different lighting arrangements. In all cases, the individuals have normal facial expressions.

E4 - Illumination Experiments				
Experiment Name	Gallery Images	Probe Images	Gallery Size	Probe Size
I1	Mugshot lighting	Overhead lighting	227	189

I2	Mugsho t lighting	Badge system lighting	129	130
I3	Mugsho t lighting	Outdoor lighting	227	190

E-5 Media Experiments

The media experiments were designed to evaluate the performance of face-matching algorithms when comparing images stored on different media. In this case, digital CCD images and 35mm film images are used. This is an important consideration for a scenario such as using an image captured with a video camera to search through a mugshot database created from a film source.

The galleries for the media experiments are made up of images taken at Dahlgren in November 1999 and NIST in December 2000 of individuals wearing normal (fa) facial expressions indoors. The galleries contain either film images or digital CCD images; the probe contains the other. Usually the images were taken simultaneously within a few tenths of a second of each other.

E5 - Media Experiments				
Experiment Name	Gallery Camera	Probe Camera	Gallery Size	Probe Size
M1	35mm	Digital	96	102
M2	Digital	35mm	227	99

E-6 Pose Experiments

The performance of face-matching algorithms applied to images of subjects taken from different viewpoints is of great interest in certain applications, most notably those using indifferent or uncooperative subjects, such as surveillance. Although a subject may look up or down and thereby vary the declination angle, the more frequently occurring and important case is where the subject is looking ahead but is not facing the camera. This variation is quantified by the azimuthal head angle, referred to here as the pose. The experiments described below study the effect of pose variation. These experiments do not address angle of declination or a third variation, side-to-side head tilt.

The imagery used in the pose experiments were taken from two sources. For studies P1-P4, the b15 subset of the FERET collection was used. These images were obtained from 200 individuals who were asked to face in nine different directions under tightly controlled conditions. The P1-P4 gallery contains only frontal images. Each probe set contains images from one of the four different, non-frontal orientations. No distinction was made between left- and right-facing subjects on the assumption that many algorithms behave symmetrically.

The P5 study is distinct because its imagery is not from the FERET collection. Its gallery holds frontal outdoor images, while the probe set contains a corresponding image of the subject facing left or right at about 45 degrees to the camera.

E6 - Pose Experiments					
Experiment Name	Image Type	Gallery Pose	Probe Pose	Gallery Size	Probe Size
P1	FERET	0	15	250	400
P2	FERET	0	25	250	400
P3	FERET	0	40	250	400
P4	FERET	0	60	250	400
P5	HumanID, digital, outdoors	0	45	180	186

E-7 Resolution Experiments

Image resolution is critical to face recognition systems. There is always some low resolution at which the face image will be of sufficiently small size that the face is unrecognizable. The resolution experiments described below were designed to evaluate the performance of face matching as resolution is decreased. The metric used to quantify resolution is eye-to-eye distance in pixels. The imagery used is homogenous in the sense that it was all taken at a fixed distance to a camera, and the resolution is decreased off-line using a standard reduction algorithm. This procedure is driven by the manually keyed pupil coordinates present in the original imagery. The fractional reduction in size is determined simply as the ratio of the original and sought eye-to-eye distances. The resulting eye-to-eye distances are as low as 15 pixels.

A single, high-resolution gallery is used for all the resolution tests. It contains full-resolution, digital CCD images taken indoors under mugshot lighting. The gallery eye separation varies according to the subject with a mean of 138.7 pixels and a range of 88 to 163. In all cases, the probe sets are derived from those same gallery images. The aspect ratio is preserved in the reduction. Note that subjects with large faces are reduced by a greater factor than those with small heads.

E7 - Resolution Experiments			
Experiment Name	Probe Eye Separation (pixels)	Gallery Size	Probe Size
R1	60	101	102

R2	45	101	102
R3	30	101	102
R4	15	101	102

E-8 Temporal Experiments

The temporal experiments address the effect of time delay between first and subsequent captures of facial images. The problem of recognizing subjects during extended periods is intuitively significant and is germane to many applications. Robust testing of this effect is difficult because of a lack of long-term data. Given the absence of meaningful data sets, these experiments rely on imagery gathered during a period of less than two years.

The T1 and T2 studies exactly reproduce the widely reported FERET duplicate I and II tests. They use the standard frontal 1,196-image FERET gallery.

The T2 probe set contains 234 images from subjects whose gallery match was taken between 540 and 1,031 days before (median = 569, mean = 627 days). The T1 probe set is a superset of the T2 probe set with additional images taken closer in time to their gallery matches. The T1 probe set holds 722 images whose matches were taken between 0 and 1031 days after the match (median = 72, mean = 251 days). The difference set (T1-T2 has 488 images) has time delays between 0 and 445 days (median = 4, mean = 70 days). Thus T2 is a set where at least 18 months has elapsed between capturing the gallery match and the probe itself. T1 and T2 also represent an access control situation in which a gallery is rebuilt every year or so.

Experiments T3-T5 are based on the more recent HumanID image collections. The galleries contain about 227 images that were obtained between 11 and 13 months after the probe images. The probe set is fixed and contains 467 images obtained using overhead room lighting. The three studies differ only in the lighting used for the gallery images.

E8(a) - Temporal Experiments			
Experiment Name	Experiment Description	Gallery Size	Probe Size
T1	FERET Duplicate I	1196	722
T2	FERET Duplicate II	1196	234

E8(b) - Temporal Experiments				
Experiment Name	Gallery Lighting	Probe Lighting	Gallery Size	Probe Size

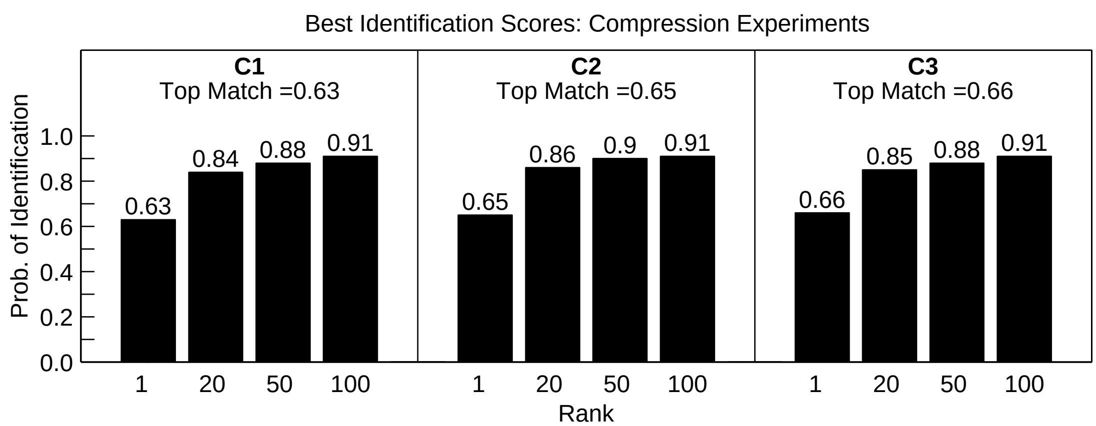
T3	Mugshot	Ambient	227	467
T4	FERET	Ambient	227	467
T5	Overhead	Ambient	226	467

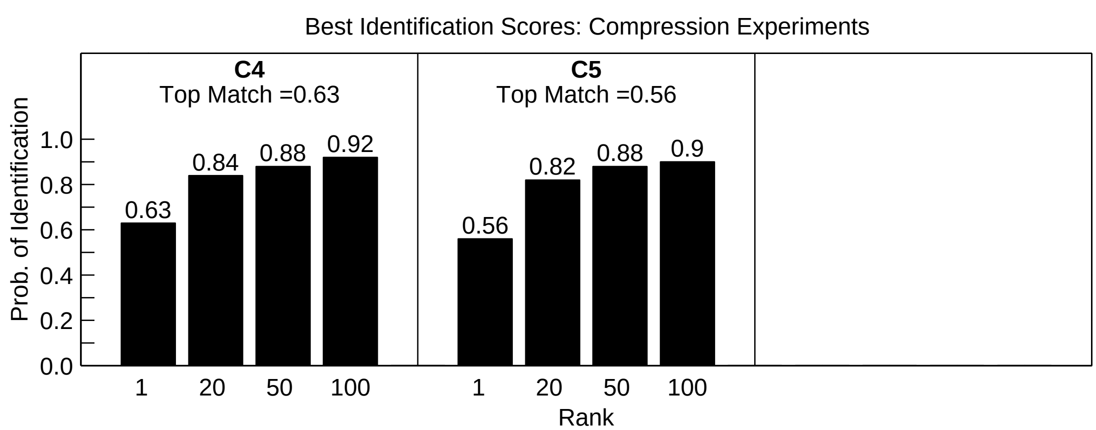
Appendix F
FRVT 2000 Recognition Performance Test Results
(Highlights)

7. 1.4.1 FERET Experiments

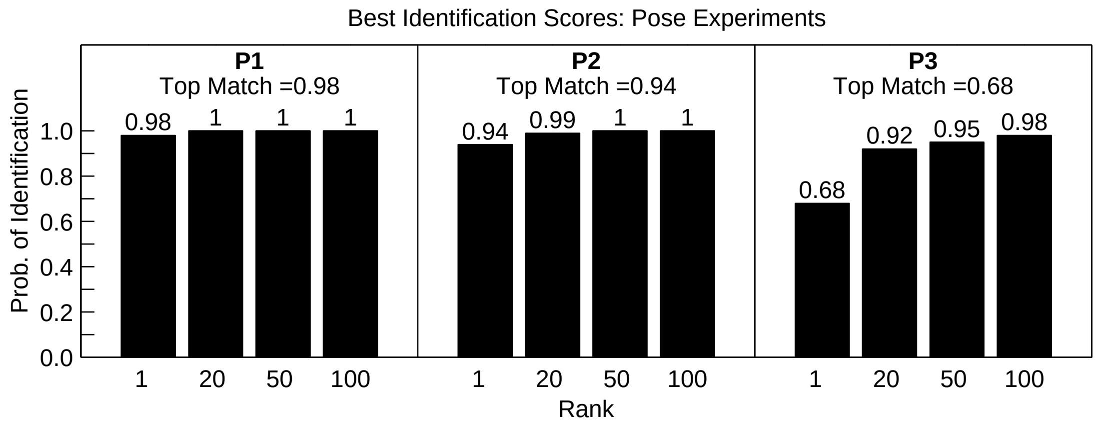
Results shown in this section are from experiments that use images from the FERET database. The purpose of these experiments is to assess the improvement made in the facial recognition community since the conclusion of the FERET program. Results for individual vendors are not given for the FERET experiments. Rather, the sponsors developed a "top match" CMC curve by choosing the top score at each rank from the results obtained from C-Vis, Lau, and Visionics.

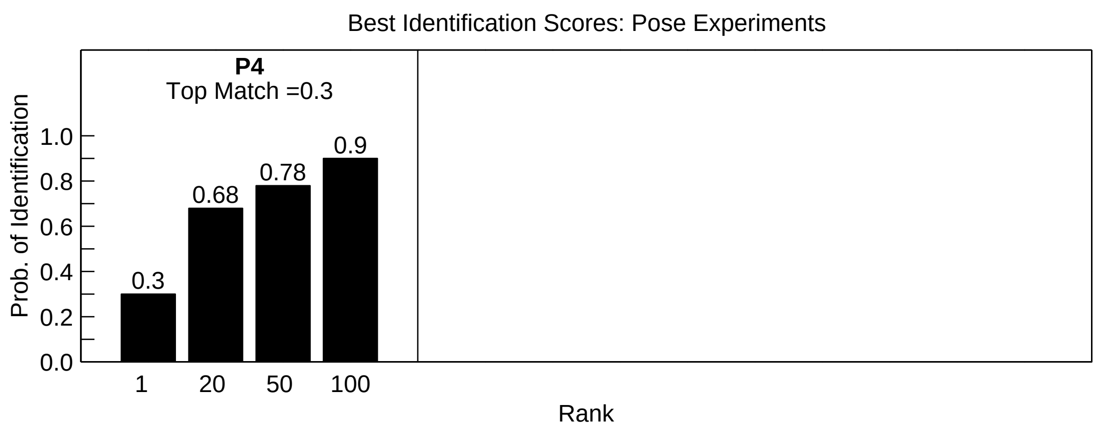
7.1.4.1.1 FERET Results – Compression Experiments



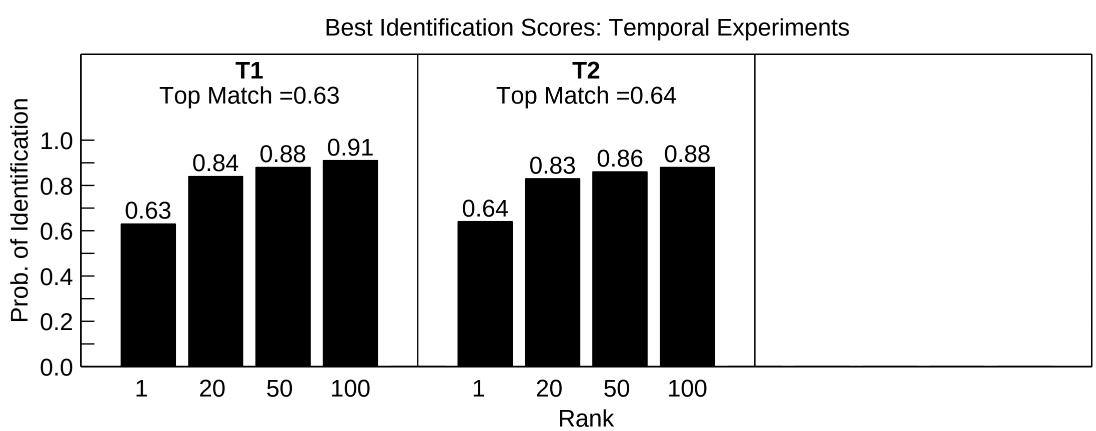


7.1.4.1.2 FERET Results – Pose Experiments





7.1.4.1.3 FERET Results – Temporal Experiments



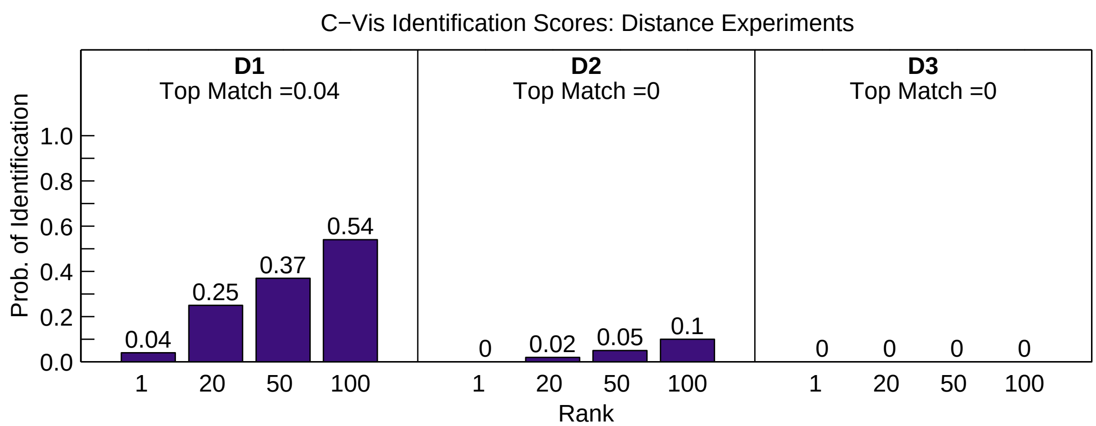
7.1.4.1.4 FERET Results – Comparison with Sep96 FERET Evaluation

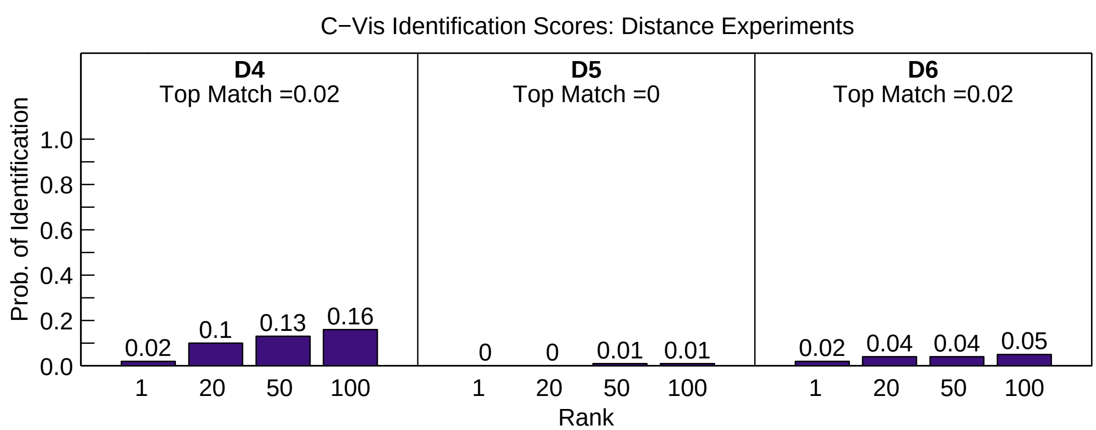
Still to come.

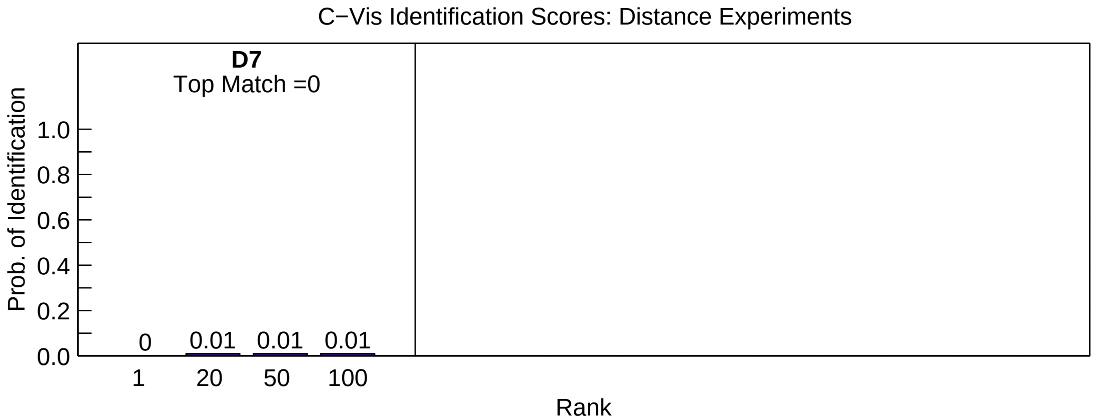
7.1.4.2 FRVT 2000 Experiments

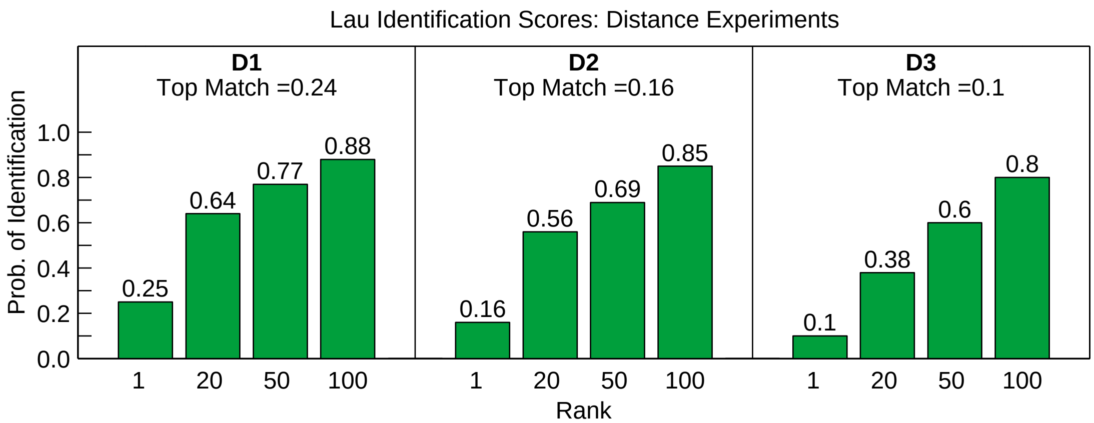
Results shown in this section are from experiments that use images that have been obtained since the conclusion of the FERET program. The purpose of the FRVT 2000 experiments is to measure the capabilities of the facial recognition community overall as well as individual vendors. Results for C-VIS, Lau, and Visionics are given for the FRVT 2000 experiments.

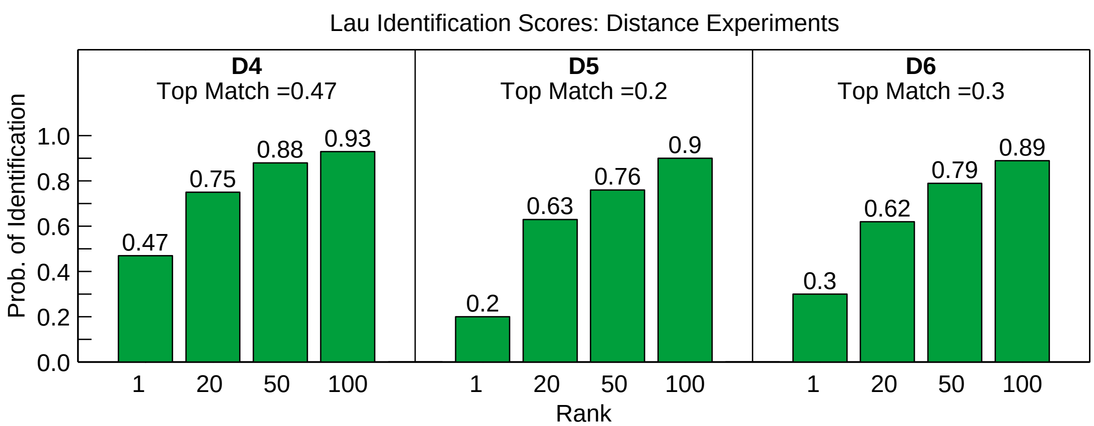
7.1.4.2.1 FRVT 2000 Experiments – Distance

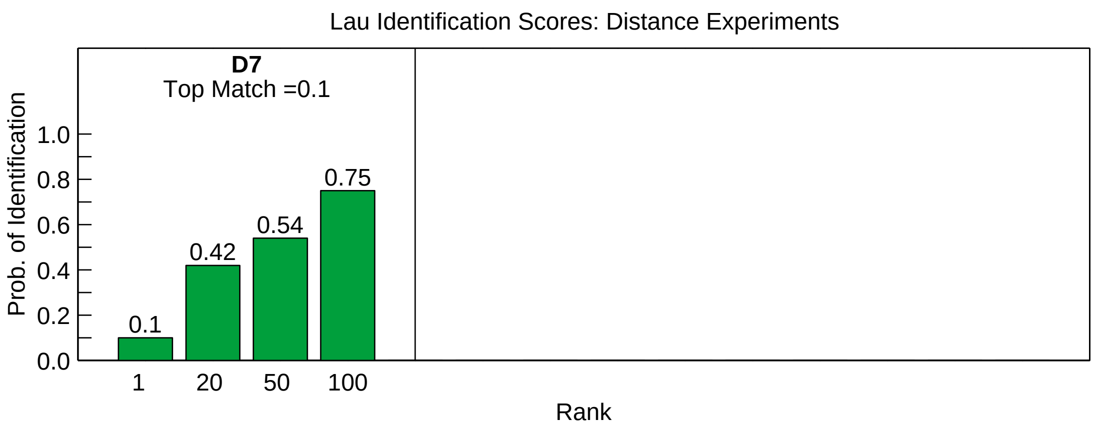


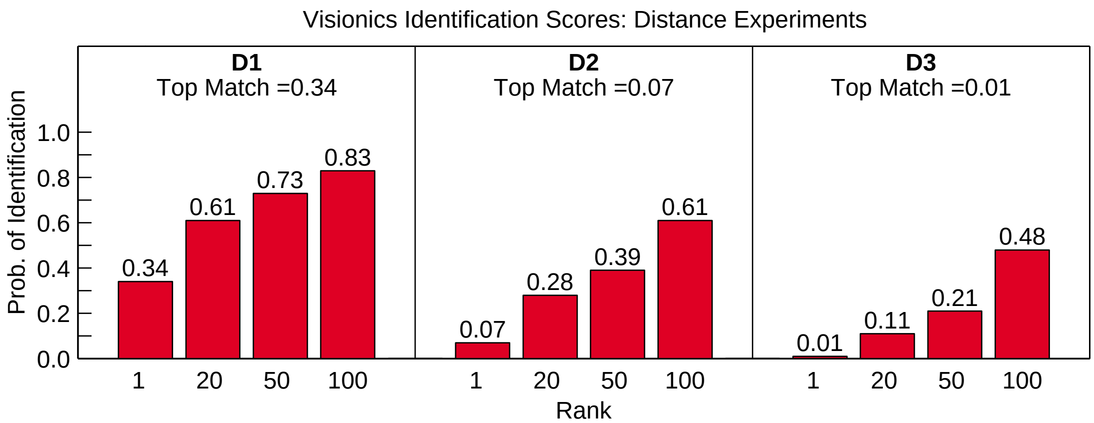


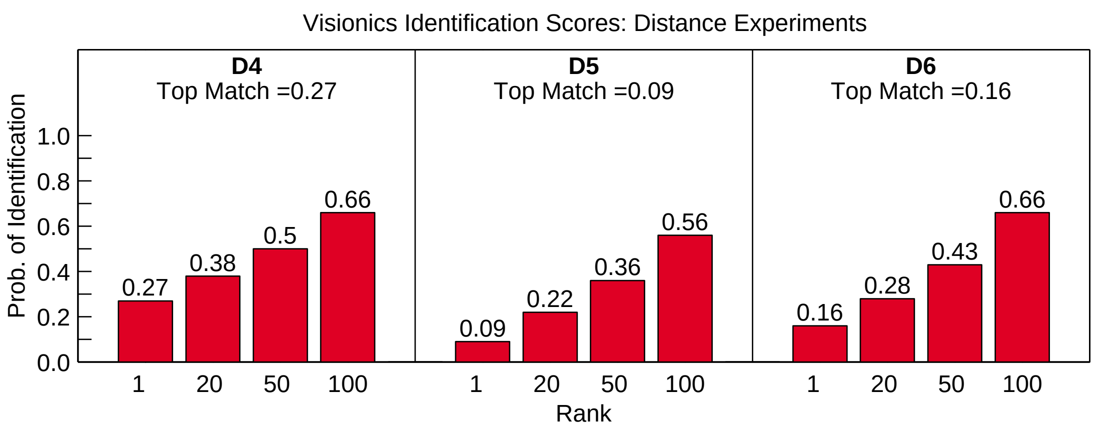

C−Vis Identification Scores: Distance Experiments
D7
Top Match =0
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0
0.01
0.01
0.01
1
20
50
100
Rank

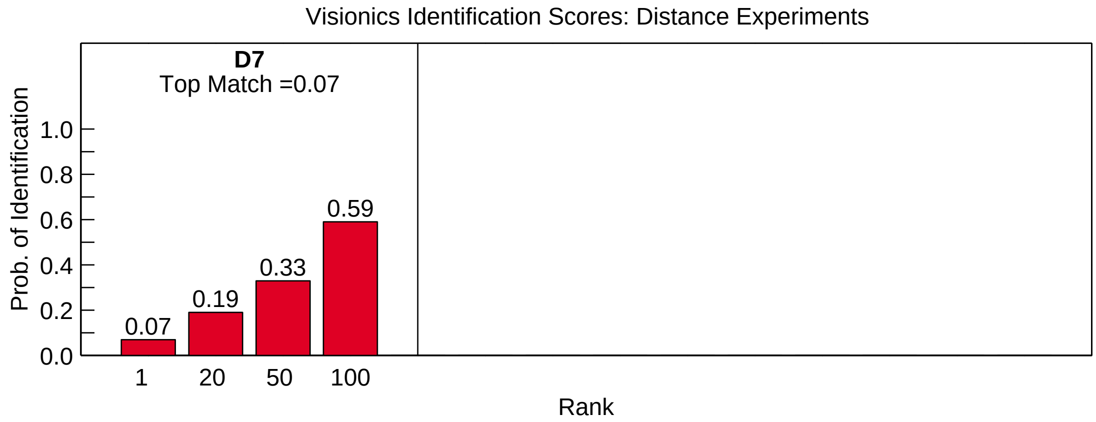

Lau Identification Scores: Distance Experiments
D1
Top Match =0.24
D2
Top Match =0.16
D3
Top Match =0.1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.25
0.64
0.77
0.88
0.16
0.56
0.69
0.85
0.1
0.38
0.6
0.8
1
20
50
100
Rank

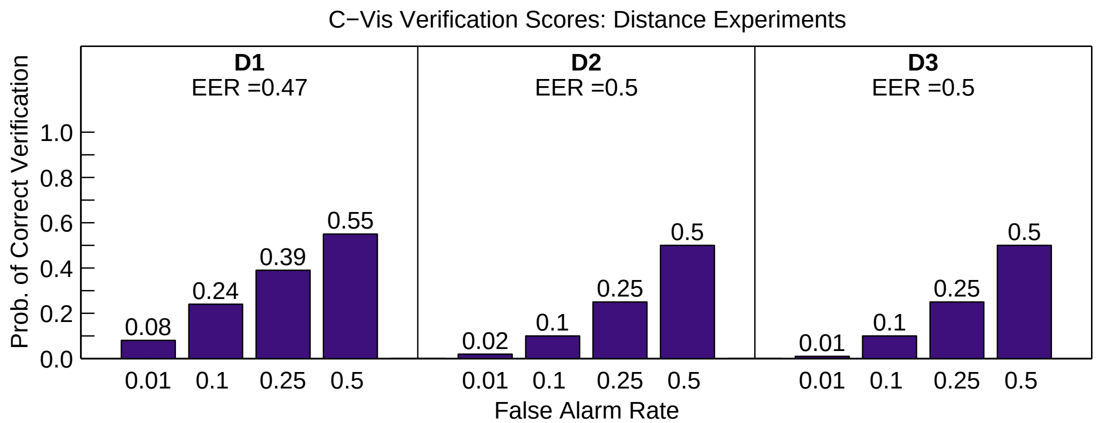

Lau Identification Scores: Distance Experiments
D4
Top Match =0.47
D5
Top Match =0.2
D6
Top Match =0.3
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.47
0.75
0.88
0.93
0.2
0.63
0.76
0.9
0.3
0.62
0.79
0.89
1
20
50
100
Rank

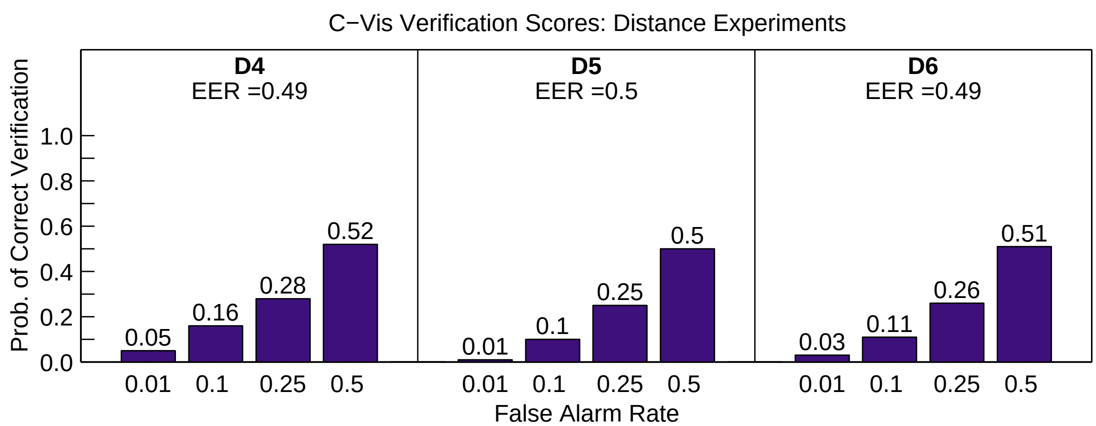

Lau Identification Scores: Distance Experiments
D7
Top Match =0.1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.42
0.54
0.75
1
20
50
100
Rank

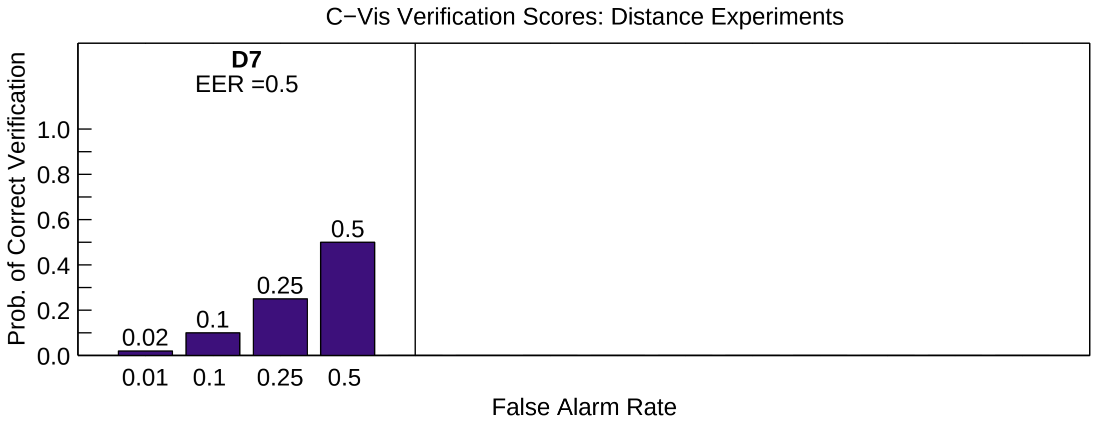

Visionics Identification Scores: Distance Experiments
D1
Top Match =0.34
D2
Top Match =0.07
D3
Top Match =0.01
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.34
0.61
0.73
0.83
0.07
0.28
0.39
0.61
0.01
0.11
0.21
0.48
1
20
50
100
Rank

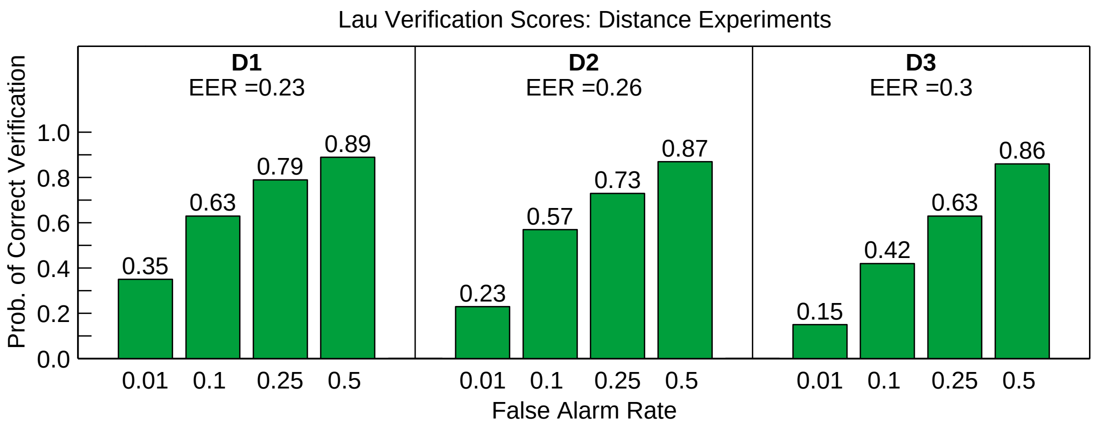

Visionics Identification Scores: Distance Experiments
D4
Top Match =0.27
D5
Top Match =0.09
D6
Top Match =0.16
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.27
0.38
0.5
0.66
0.09
0.22
0.36
0.56
0.16
0.28
0.43
0.66
1
20
50
100
Rank

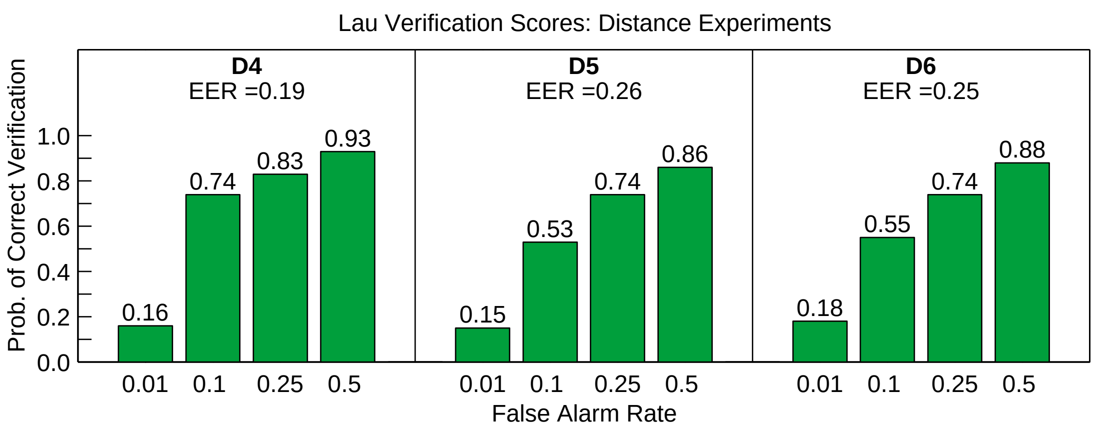

Visionics Identification Scores: Distance Experiments
D7
Top Match =0.07
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.07
0.19
0.33
0.59
1
20
50
100
Rank

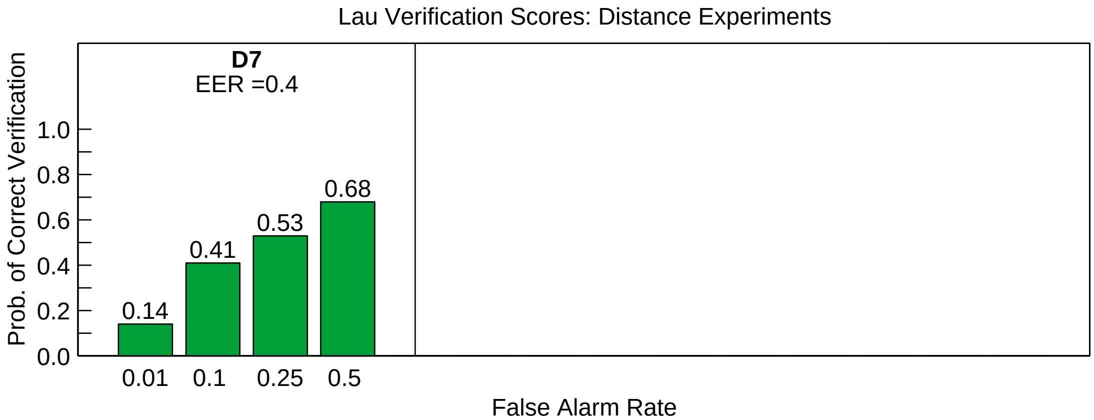

C−Vis Verification Scores: Distance Experiments
D1
EER =0.47
D2
EER =0.5
D3
EER =0.5
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.08
0.24
0.39
0.55
0.02
0.1
0.25
0.5
0.01
0.1
0.25
0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

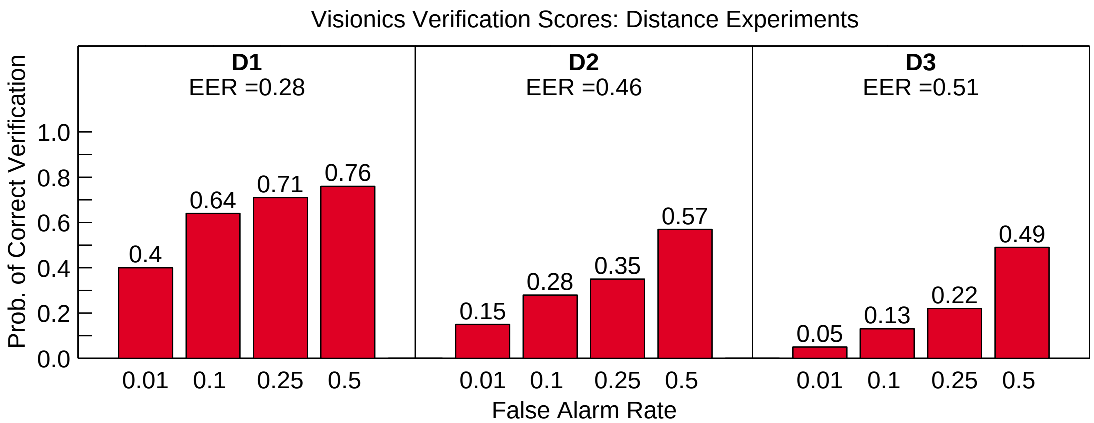

C−Vis Verification Scores: Distance Experiments
D4
EER =0.49
D5
EER =0.5
D6
EER =0.49
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.05
0.16
0.28
0.52
0.01
0.1
0.25
0.5
0.03
0.11
0.26
0.51
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

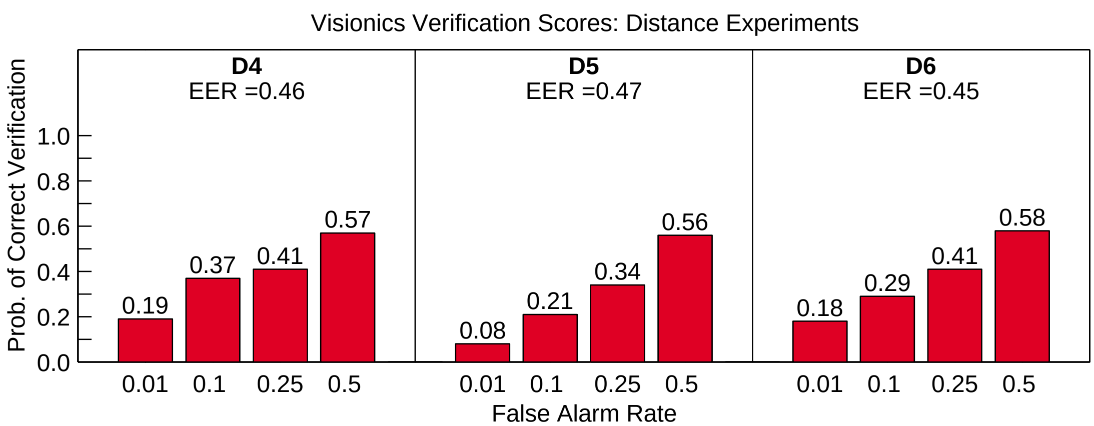

C−Vis Verification Scores: Distance Experiments
D7
EER =0.5
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.02
0.1
0.25
0.5
0.01
0.1
0.25
0.5
False Alarm Rate

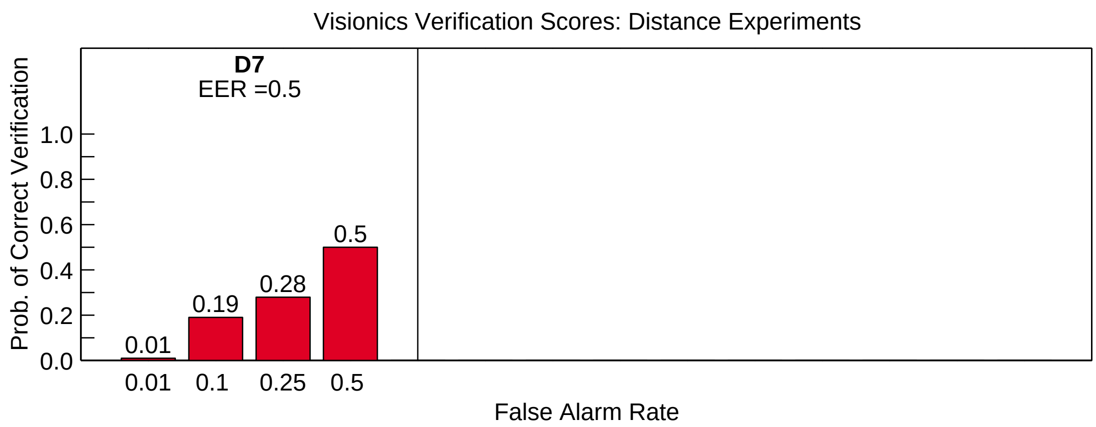

Lau Verification Scores: Distance Experiments
D1
EER =0.23
D2
EER =0.26
D3
EER =0.3
Prob. of Correct Verification
0.35
0.63
0.79
0.89
0.23
0.57
0.73
0.87
0.15
0.42
0.63
0.86
0.01
0.1
0.25
0.5
False Alarm Rate


Lau Verification Scores: Distance Experiments
D4
EER =0.19
D5
EER =0.26
D6
EER =0.25
Prob. of Correct Verification
0.16
0.74
0.83
0.93
0.15
0.53
0.74
0.86
0.18
0.55
0.74
0.88
0.01
0.1
0.25
0.5
False Alarm Rate

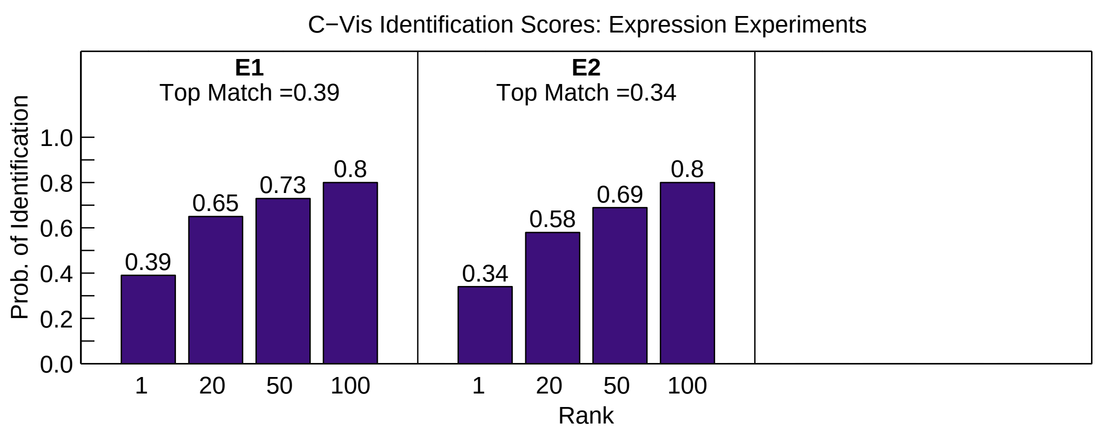

Lau Verification Scores: Distance Experiments
D7
EER =0.4
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.14
0.41
0.53
0.68
0.01
0.1
0.25
0.5
False Alarm Rate

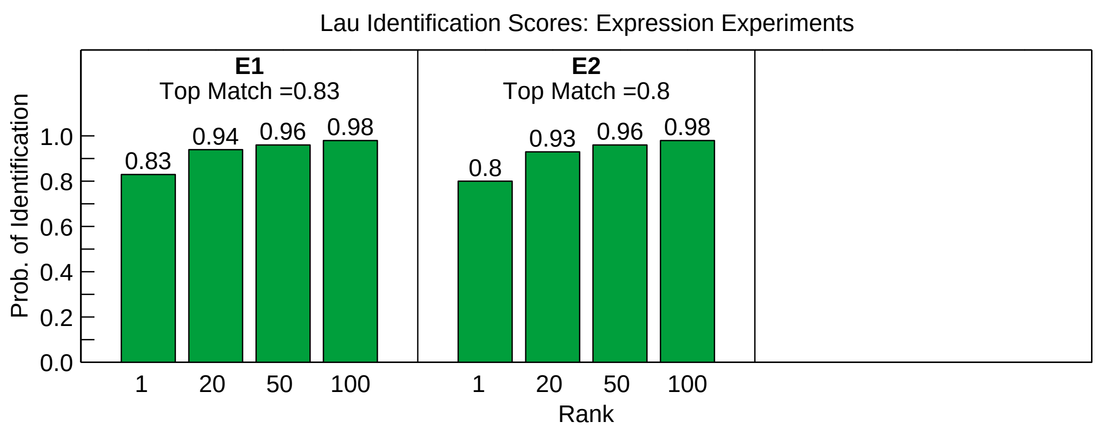

Visionics Verification Scores: Distance Experiments
D1
EER =0.28
D2
EER =0.46
D3
EER =0.51
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.4
0.64
0.71
0.76
0.15
0.28
0.35
0.57
0.05
0.13
0.22
0.49
0.01
0.1
0.25
0.5
False Alarm Rate


Visionics Verification Scores: Distance Experiments
D4
EER =0.46
D5
EER =0.47
D6
EER =0.45
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.19
0.37
0.41
0.57
0.08
0.21
0.34
0.56
0.18
0.29
0.41
0.58
0.01
0.1
0.25
0.5
False Alarm Rate

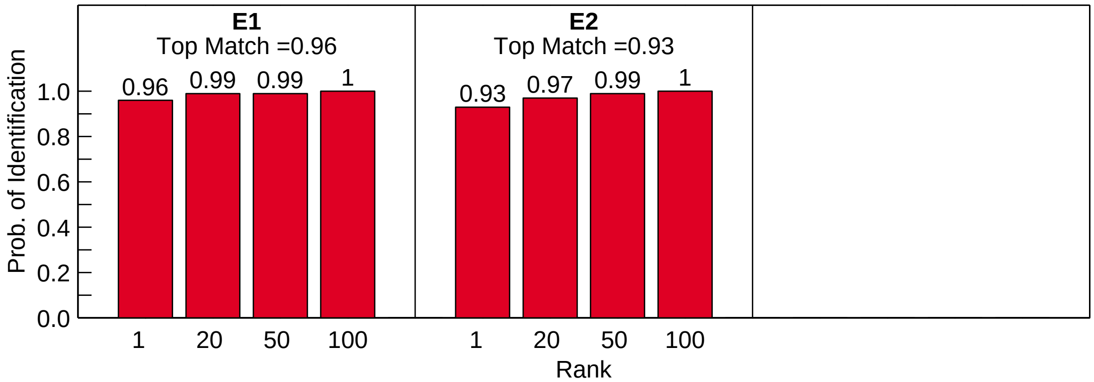


7.1.4.2.2 FRVT 2000 Experiments – Expression

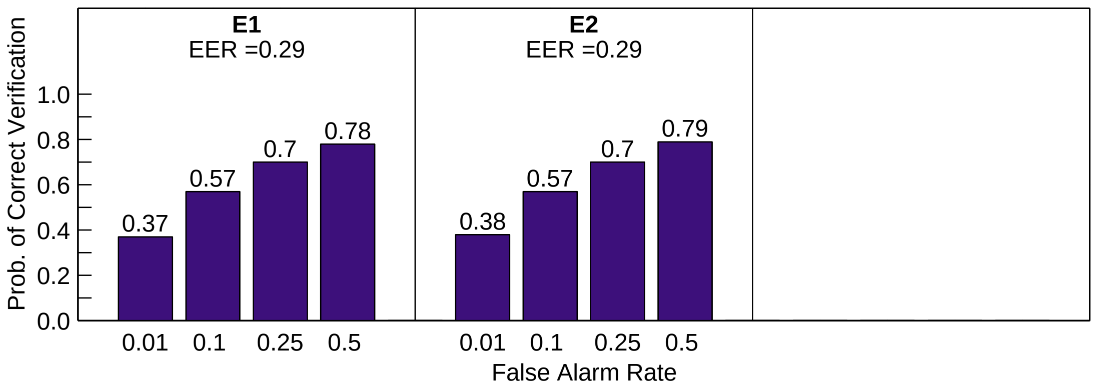




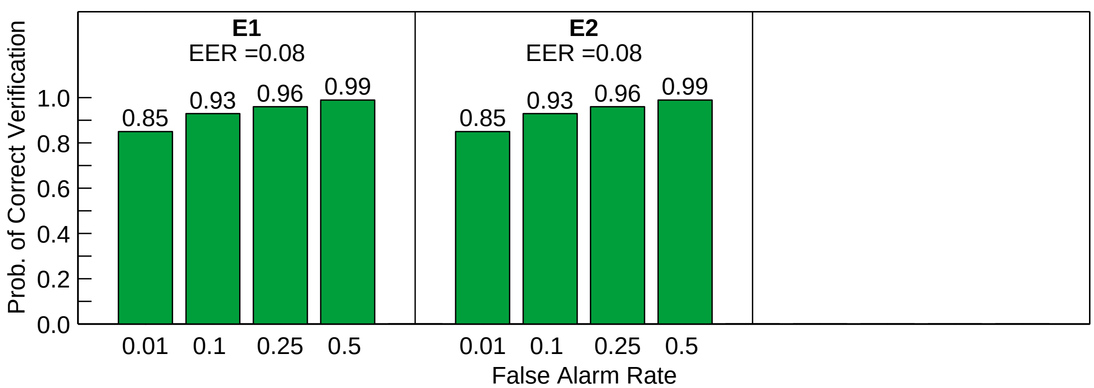
Visionics Identification Scores: Expression Experiments

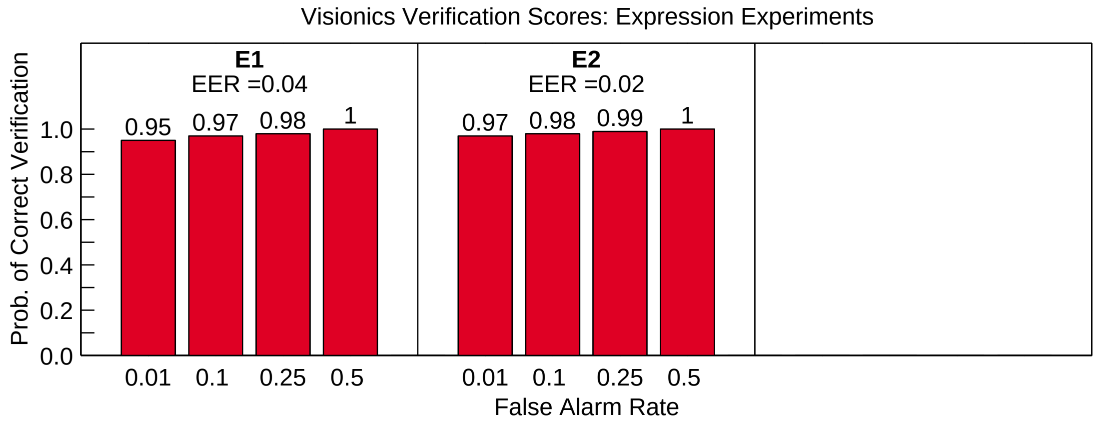


C−Vis Verification Scores: Expression Experiments

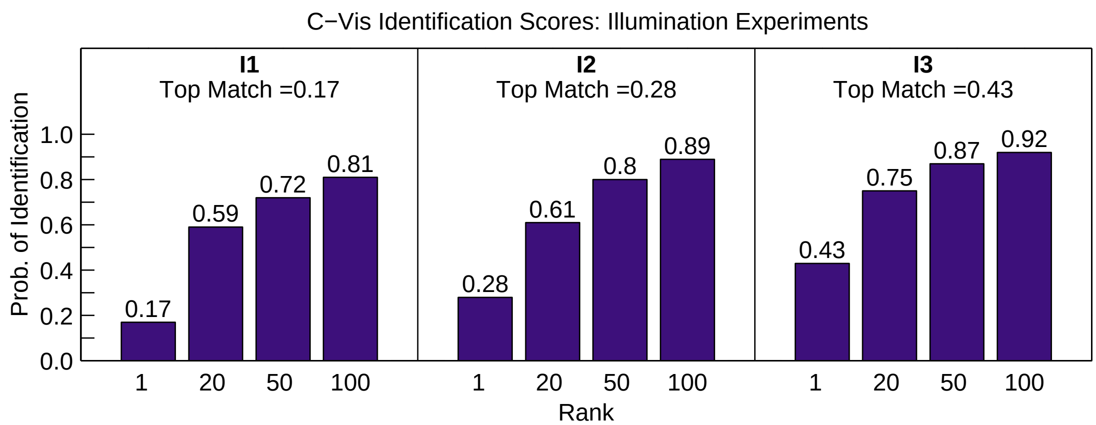


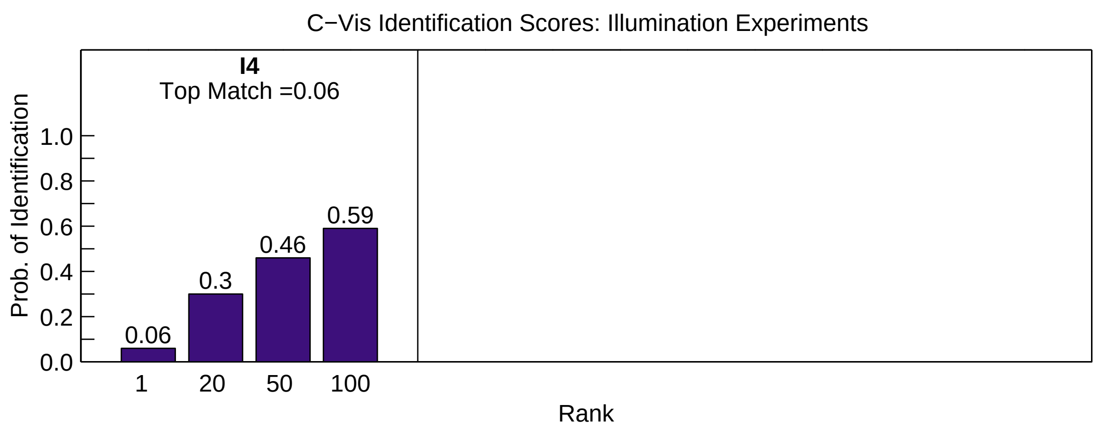
Lau Verification Scores: Expression Experiments

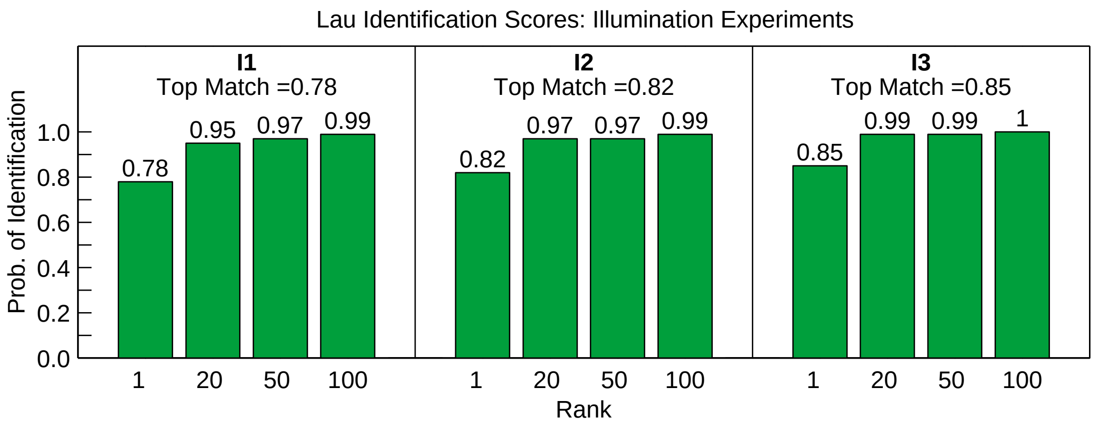


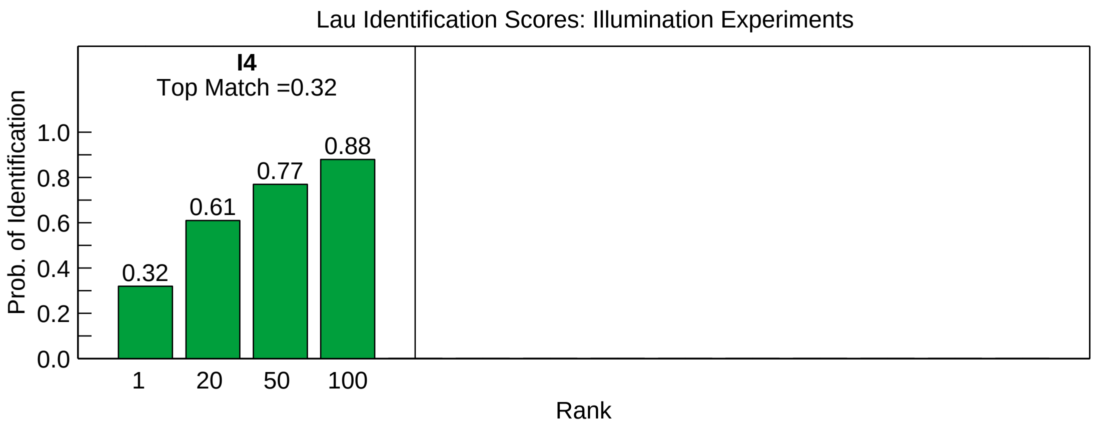


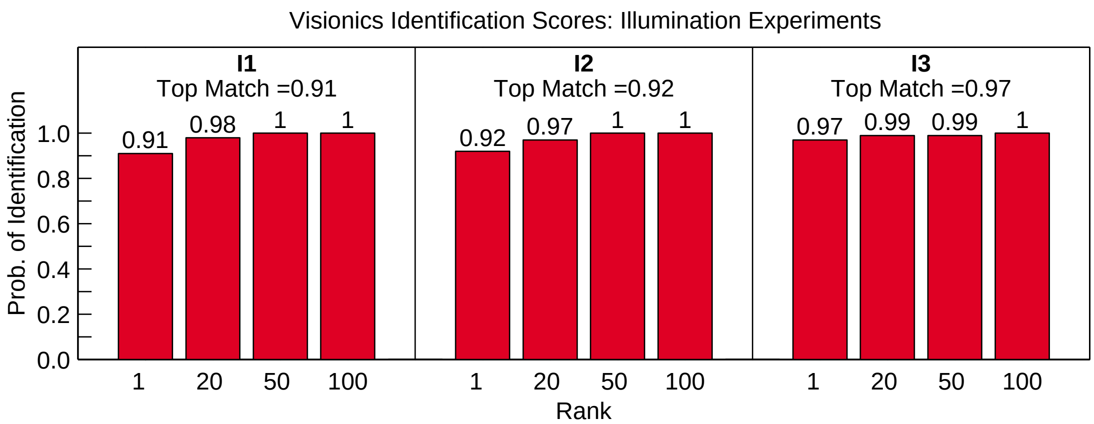
7.1.4.2.3 FRVT 2000 Experiments – Illumination

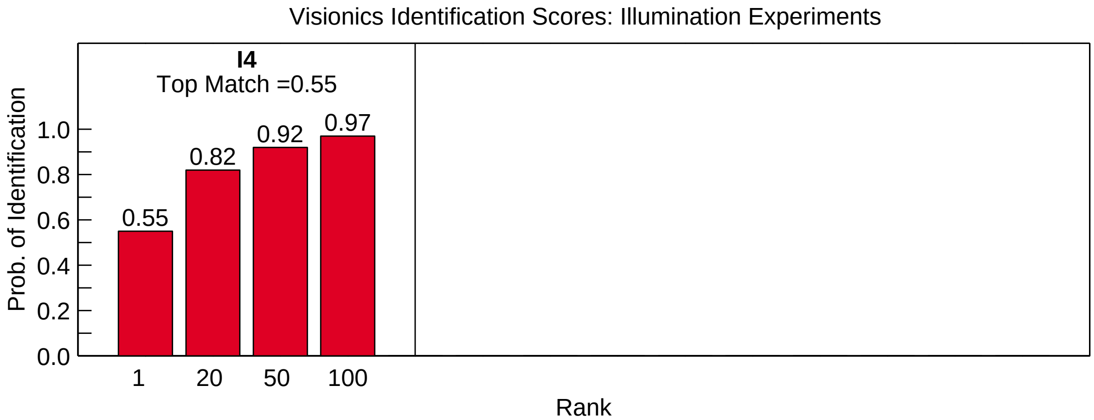


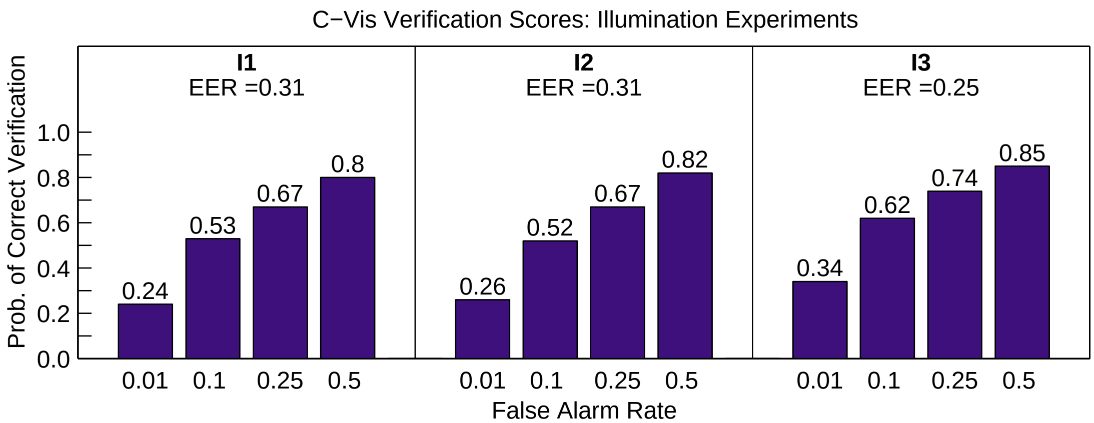


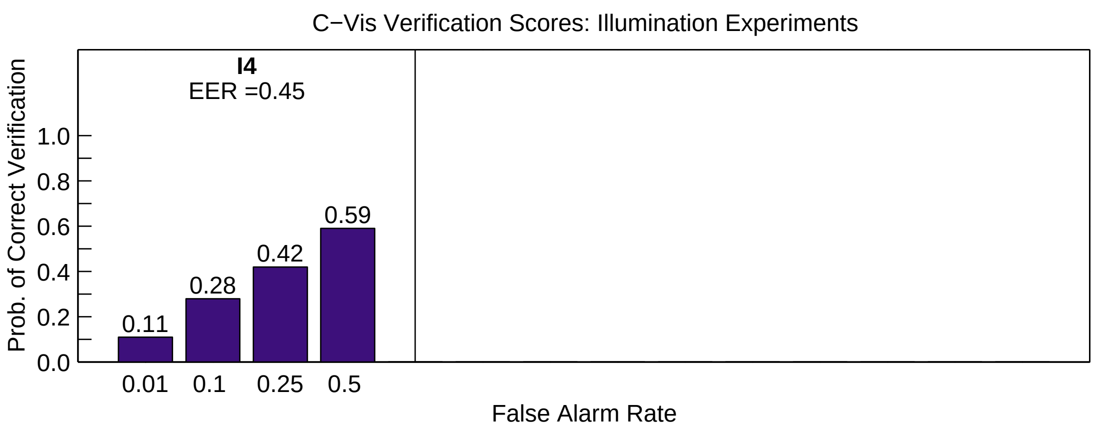

Lau Identification Scores: Illumination Experiments
I1
Top Match =0.78
0.78
0.95
0.97
0.99
I2
Top Match =0.82
0.82
0.97
0.97
0.99
I3
Top Match =0.85
0.85
0.99
0.99
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1 20 50 100
Rank

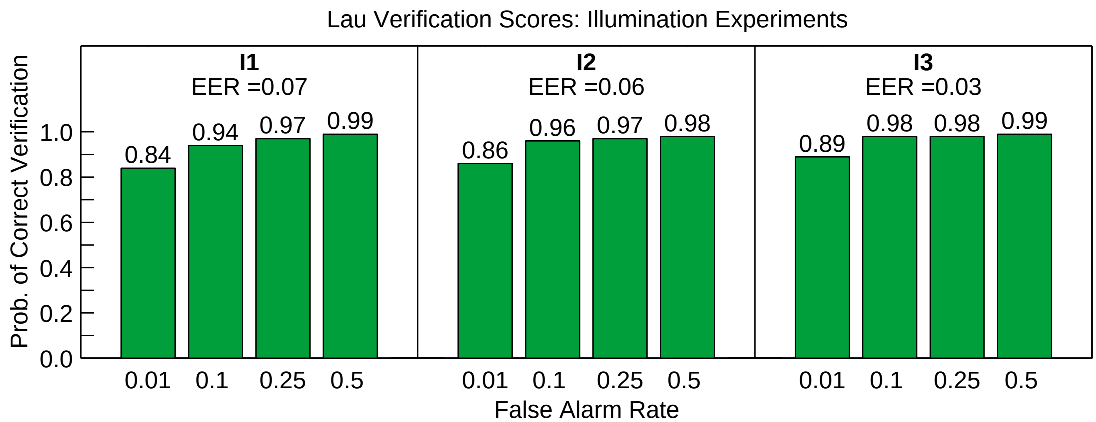

Lau Identification Scores: Illumination Experiments
I4
Top Match =0.32
0.32
0.61
0.77
0.88
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1 20 50 100
Rank

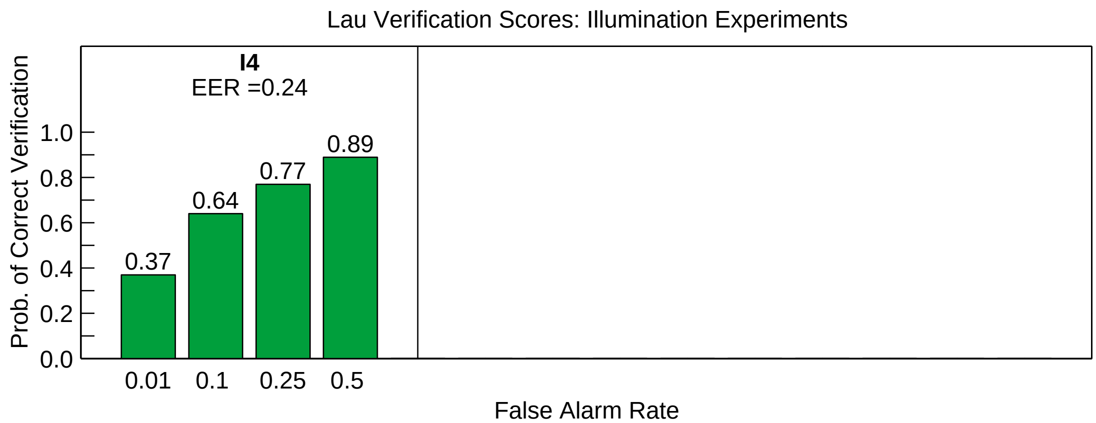

Visionics Identification Scores: Illumination Experiments
I1
Top Match =0.91
0.91
0.98
1
1
I2
Top Match =0.92
0.92
0.97
1
1
I3
Top Match =0.97
0.97
0.99
0.99
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1 20 50 100
Rank

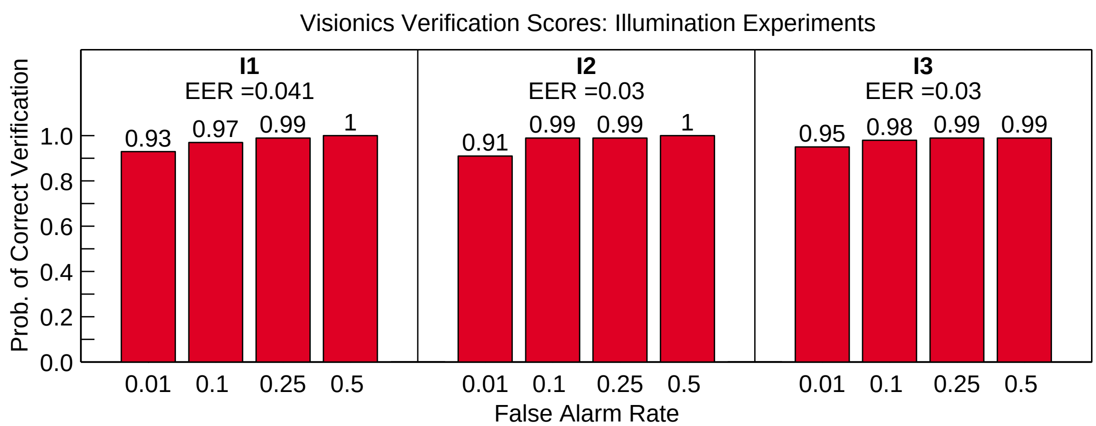

Visionics Identification Scores: Illumination Experiments
I4
Top Match =0.55
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.55
0.82
0.92
0.97
1
20
50
100
Rank

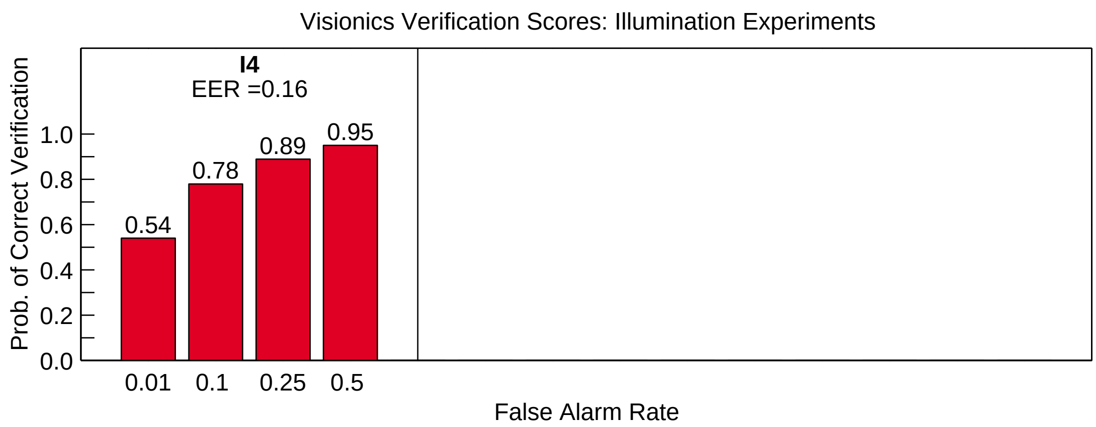

C−Vis Verification Scores: Illumination Experiments
I1
EER =0.31
I2
EER =0.31
I3
EER =0.25
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.24
0.53
0.67
0.8
0.26
0.52
0.67
0.82
0.34
0.62
0.74
0.85
0.01
0.1
0.25
0.5
False Alarm Rate


C−Vis Verification Scores: Illumination Experiments
I4
EER =0.45
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.11
0.28
0.42
0.59
0.01
0.1
0.25
0.5
False Alarm Rate

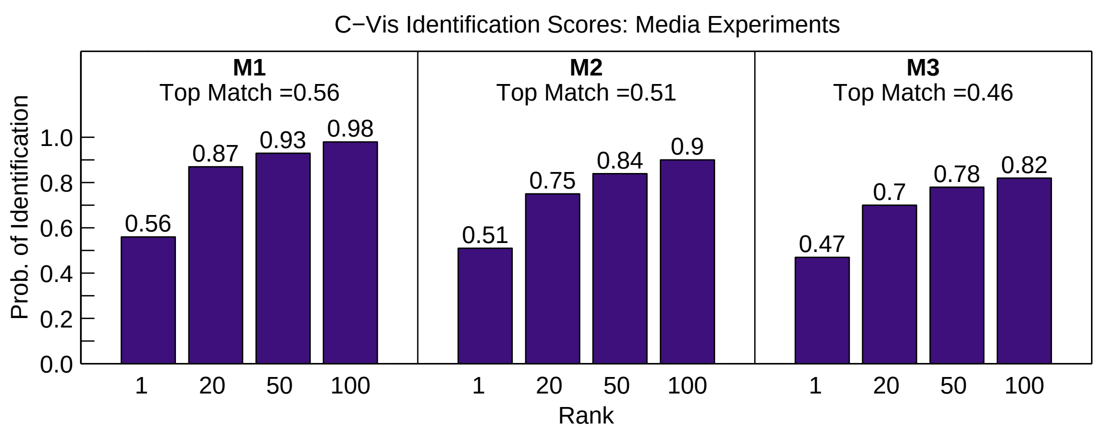

Lau Verification Scores: Illumination Experiments
I1
EER =0.07
0.84
0.94
0.97
0.99
I2
EER =0.06
0.86
0.96
0.97
0.98
I3
EER =0.03
0.89
0.98
0.98
0.99
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

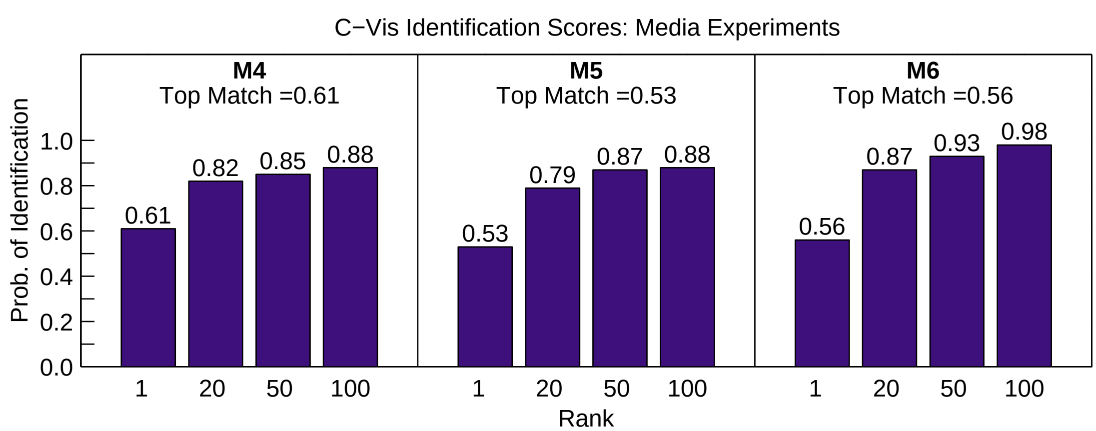

Lau Verification Scores: Illumination Experiments
I4
EER =0.24
0.37
0.64
0.77
0.89
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

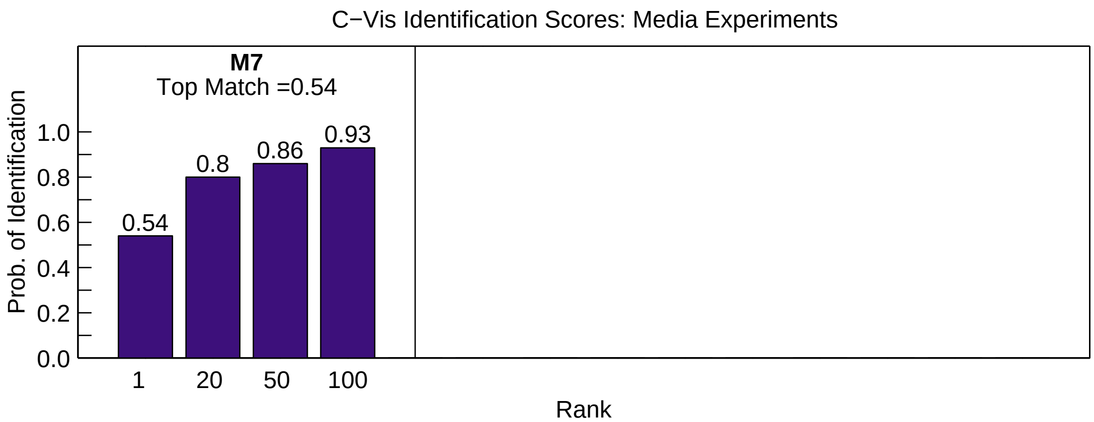

Visionics Verification Scores: Illumination Experiments
I1
EER =0.041
0.93
0.97
0.99
1
I2
EER =0.03
0.91
0.99
0.99
1
I3
EER =0.03
0.95
0.98
0.99
0.99
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

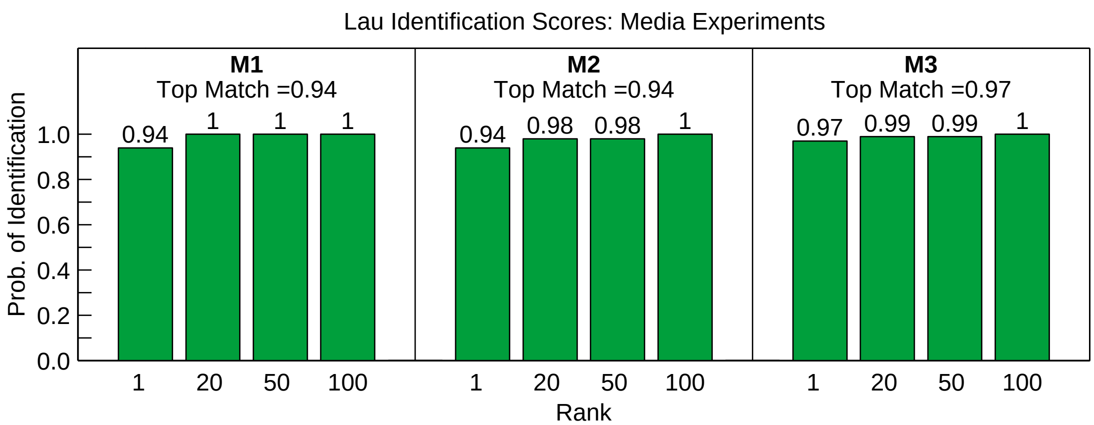


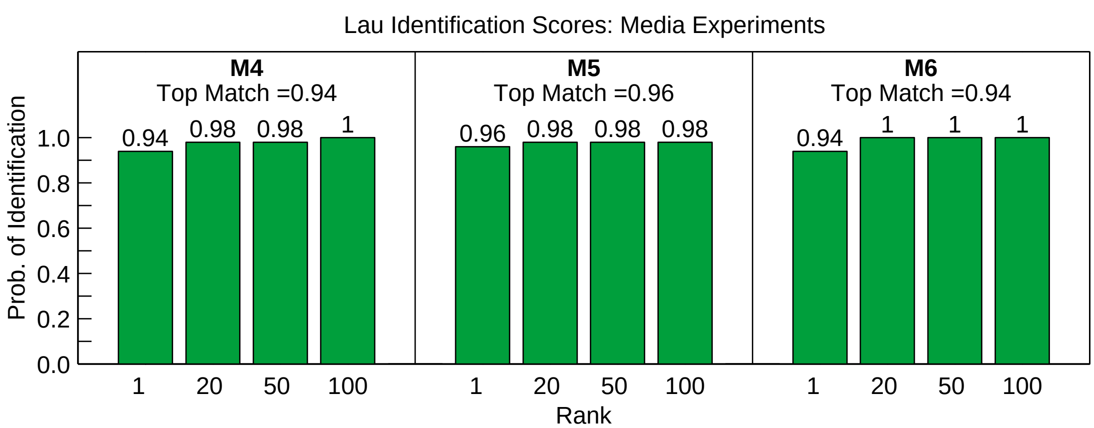
7.1.4.2.4 FRVT 2000 Experiments – Media

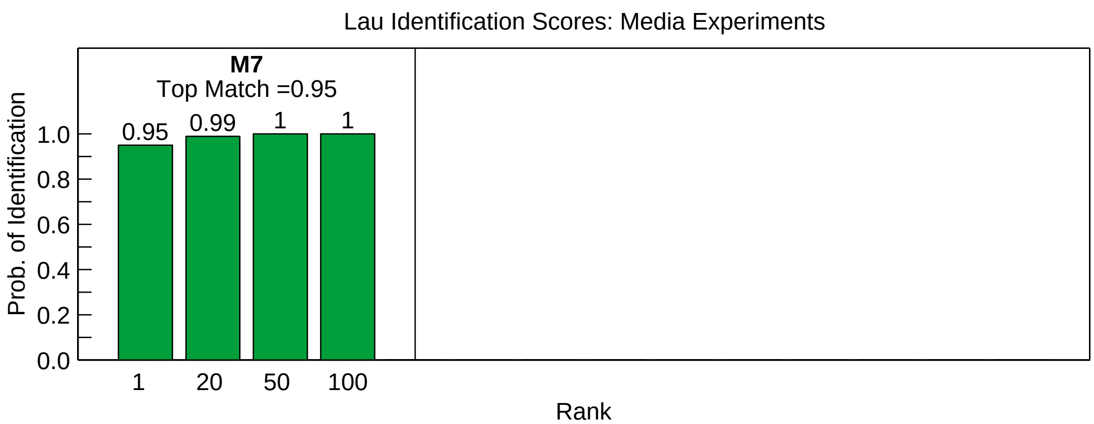


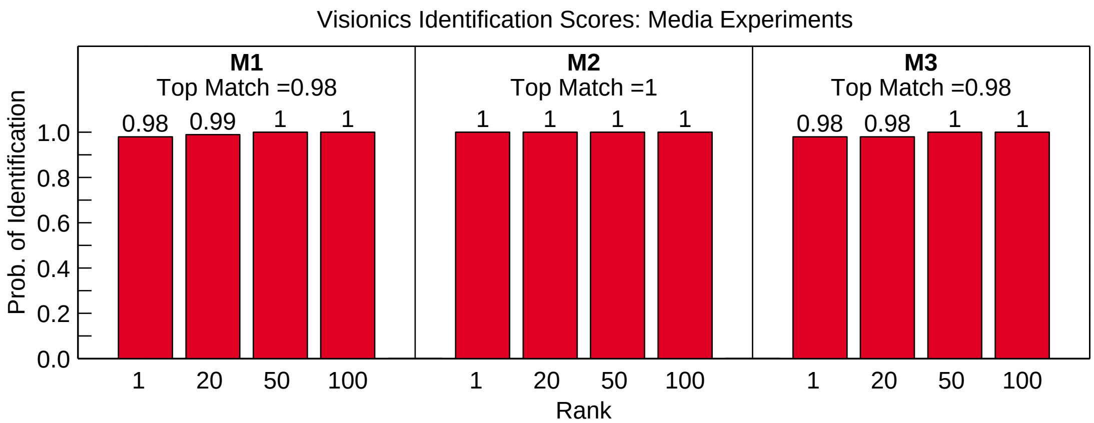


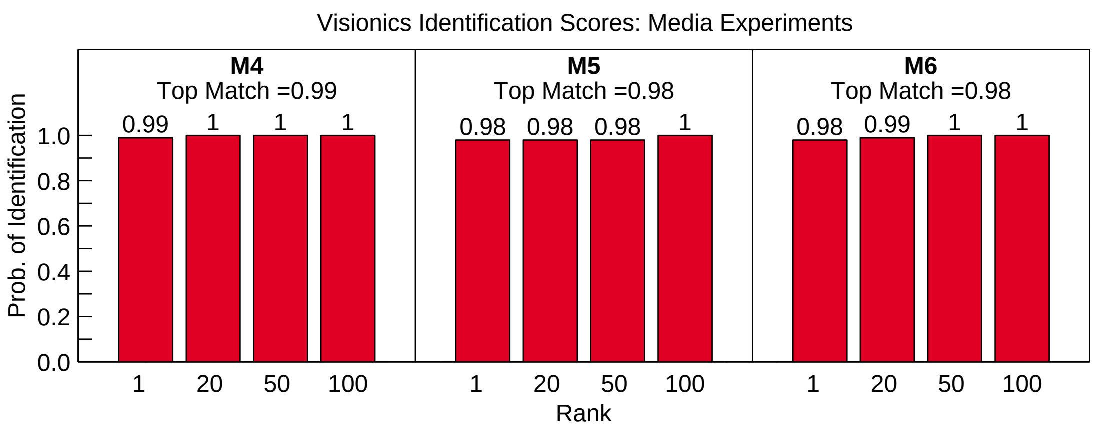

C−Vis Identification Scores: Media Experiments
M7
Top Match =0.54
0.54
0.8
0.86
0.93
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

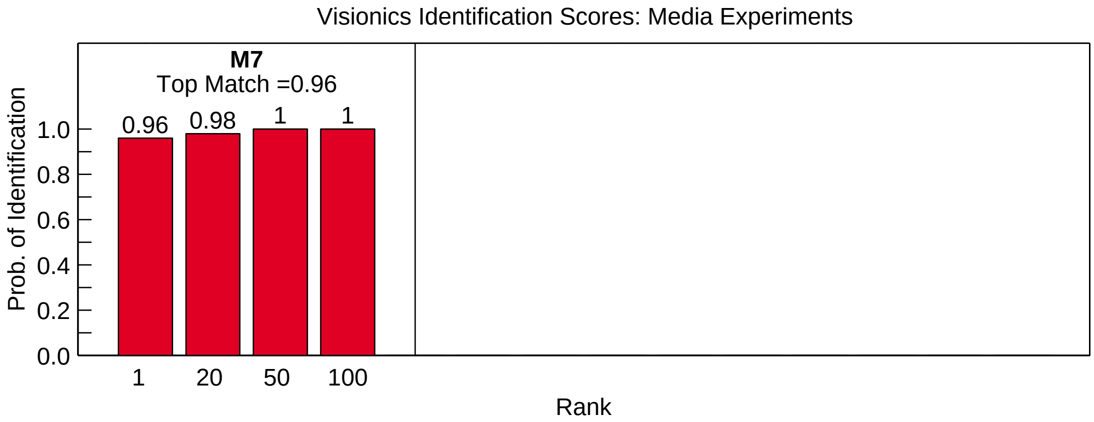

Lau Identification Scores: Media Experiments
M1
Top Match =0.94
0.94
1
1
1
M2
Top Match =0.94
0.94
0.98
0.98
1
M3
Top Match =0.97
0.97
0.99
0.99
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

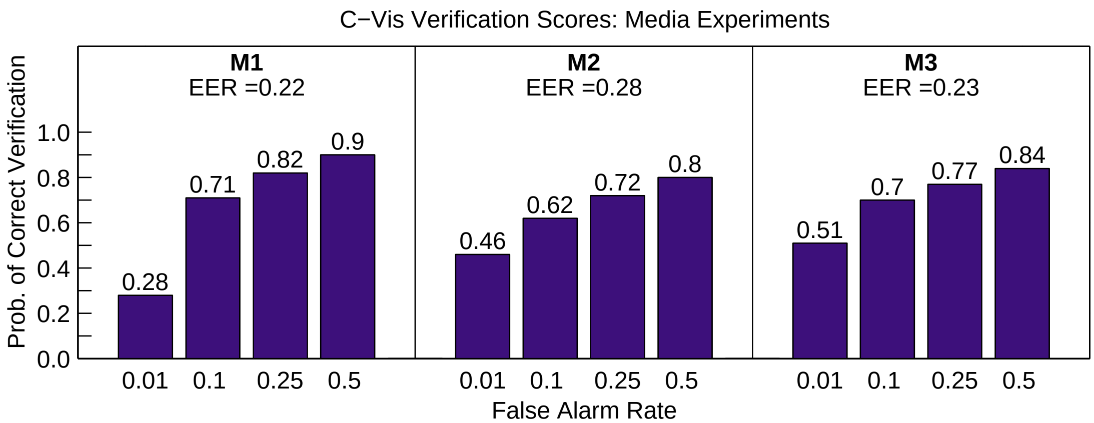

Lau Identification Scores: Media Experiments
M4
Top Match =0.94
0.94
0.98
0.98
1
M5
Top Match =0.96
0.96
0.98
0.98
0.98
M6
Top Match =0.94
0.94
1
1
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

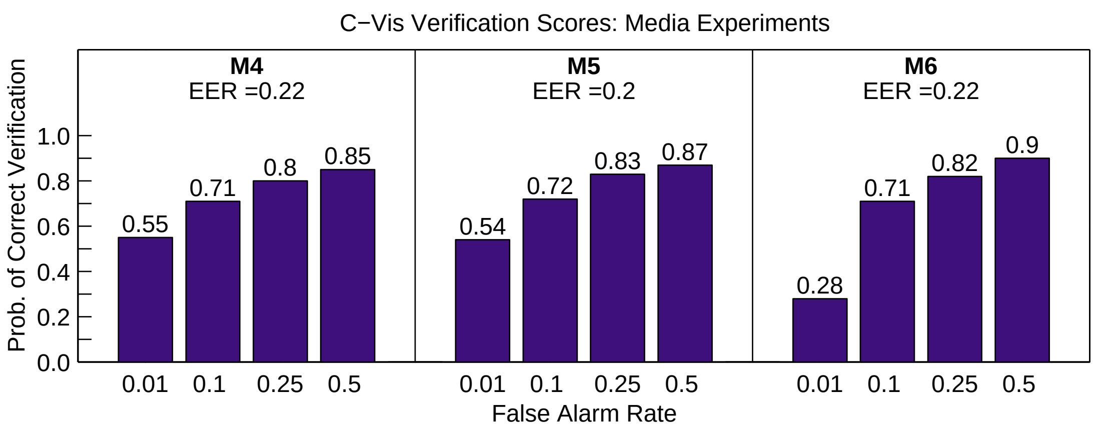

Lau Identification Scores: Media Experiments
M7
Top Match =0.95
0.95 0.99 1 1
Prob. of Identification
1.0 0.8 0.6 0.4 0.2 0.0
1 20 50 100
Rank

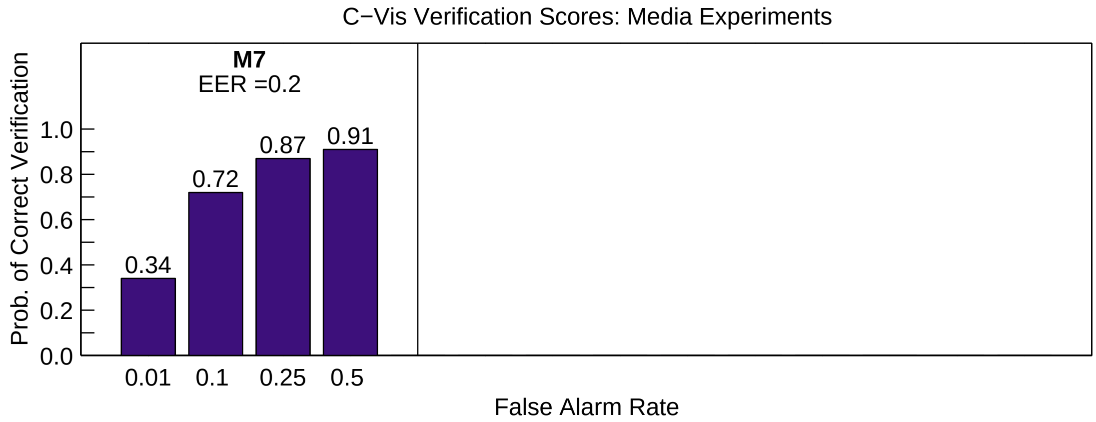

Visionics Identification Scores: Media Experiments
M1
Top Match =0.98
0.98 0.99 1 1
M2
Top Match =1
1 1 1 1
M3
Top Match =0.98
0.98 0.98 1 1
Prob. of Identification
1.0 0.8 0.6 0.4 0.2 0.0
1 20 50 100
Rank

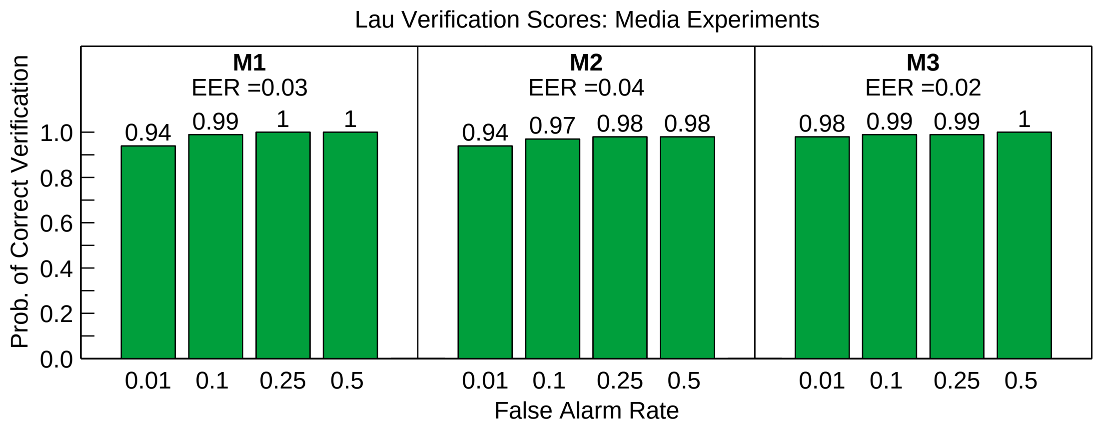

Visionics Identification Scores: Media Experiments
M4
Top Match =0.99
0.99 1 1 1
M5
Top Match =0.98
0.98 0.98 0.98 1
M6
Top Match =0.98
0.98 0.99 1 1
Prob. of Identification
1.0 0.8 0.6 0.4 0.2 0.0
1 20 50 100
Rank

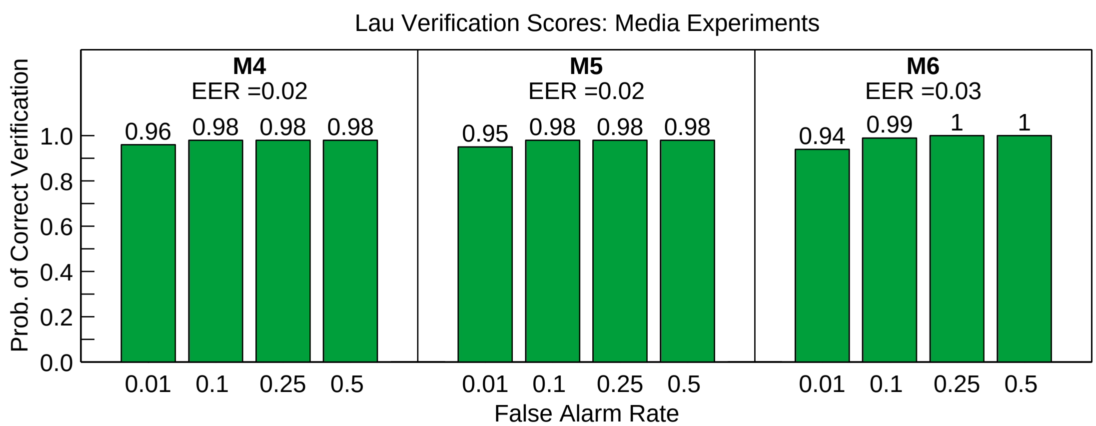

Visionics Identification Scores: Media Experiments
M7
Top Match =0.96
0.96
0.98
1
1
Prob. of Identification
0.0
0.2
0.4
0.6
0.8
1.0
1
20
50
100
Rank

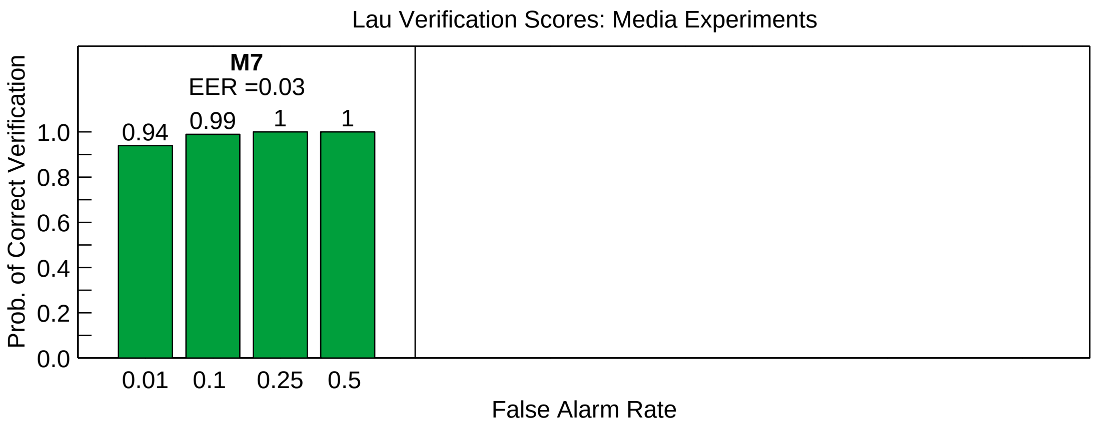

C−Vis Verification Scores: Media Experiments
M1
EER =0.22
0.28
0.71
0.82
0.9
M2
EER =0.28
0.46
0.62
0.72
0.8
M3
EER =0.23
0.51
0.7
0.77
0.84
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.1
0.25
0.5
False Alarm Rate

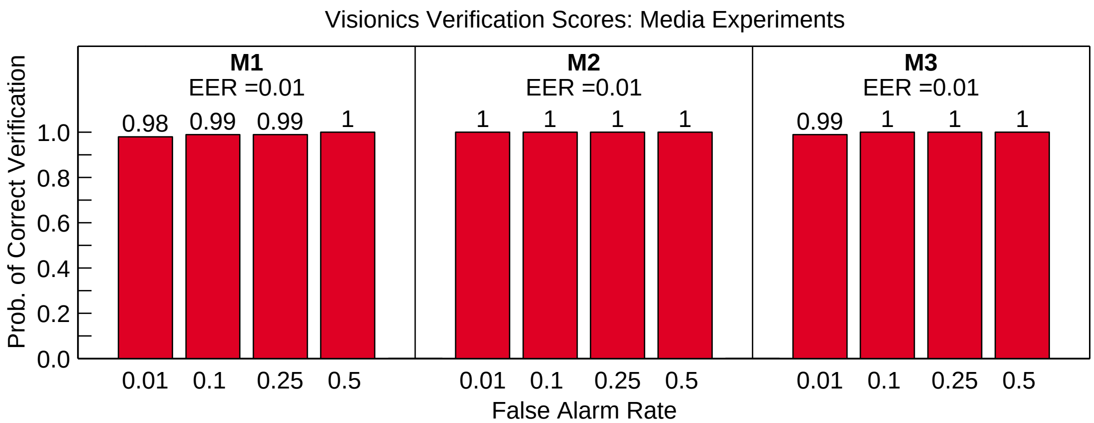

C−Vis Verification Scores: Media Experiments
M4
EER =0.22
0.55
0.71
0.8
0.85
M5
EER =0.2
0.54
0.72
0.83
0.87
M6
EER =0.22
0.28
0.71
0.82
0.9
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.1
0.25
0.5
False Alarm Rate

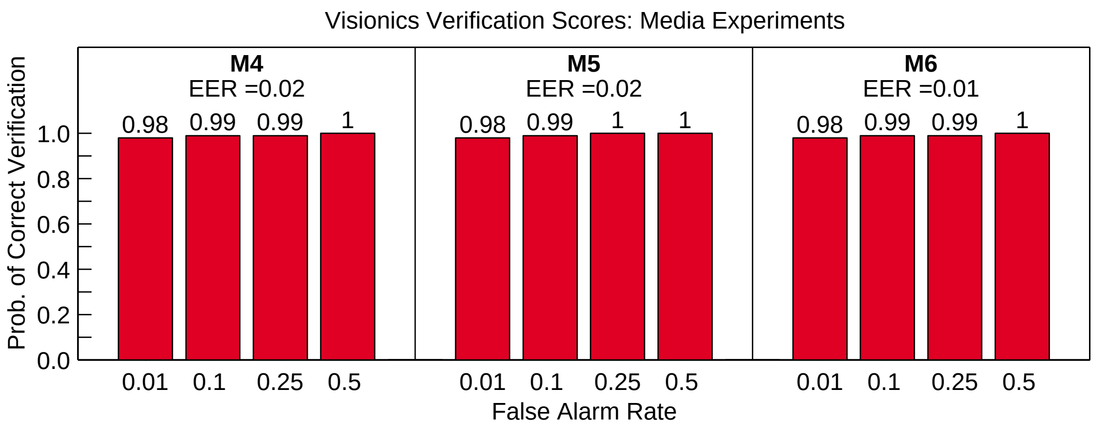

C−Vis Verification Scores: Media Experiments
M7
EER =0.2
0.34
0.72
0.87
0.91
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

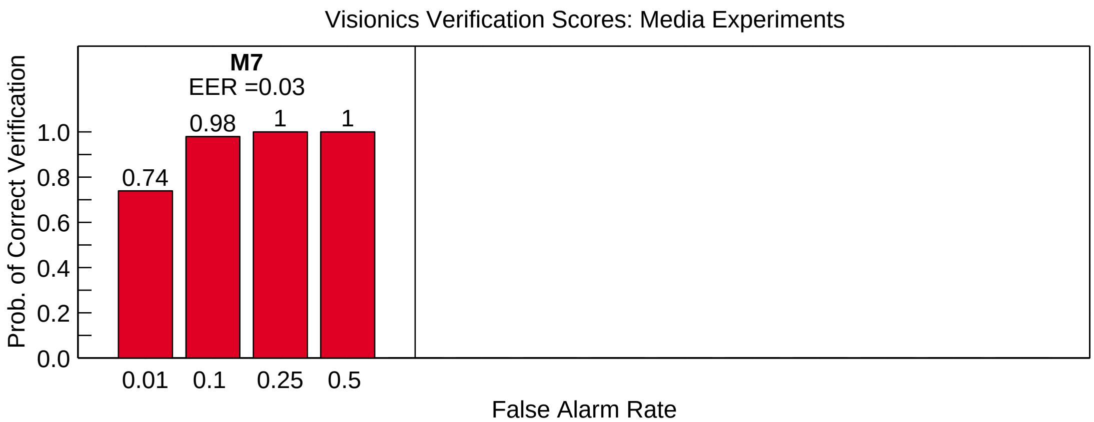

Lau Verification Scores: Media Experiments
M1
EER =0.03
0.94
0.99
1
1
M2
EER =0.04
0.94
0.97
0.98
0.98
M3
EER =0.02
0.98
0.99
0.99
1
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate


Lau Verification Scores: Media Experiments
M4
EER =0.02
0.96
0.98
0.98
0.98
M5
EER =0.02
0.95
0.98
0.98
0.98
M6
EER =0.03
0.94
0.99
1
1
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

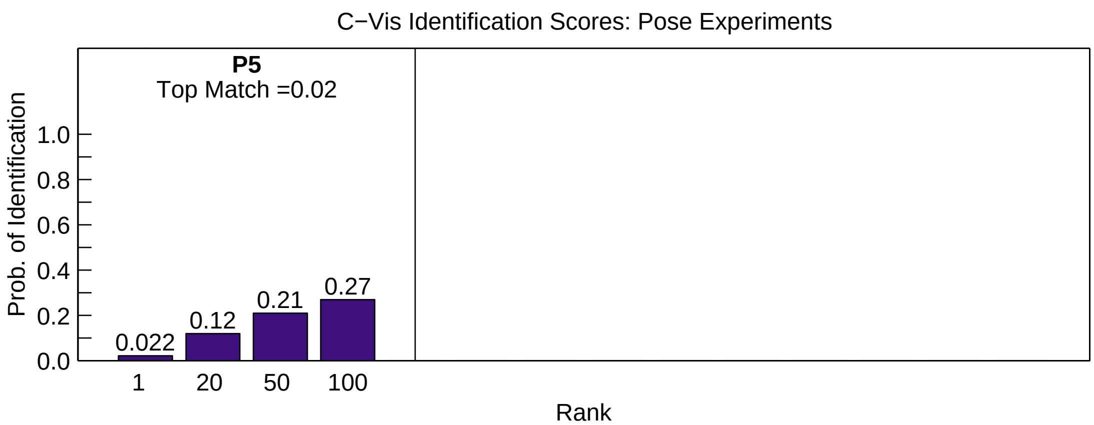

Lau Verification Scores: Media Experiments
M7
EER =0.03
0.94
0.99
1
1
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.25
0.5
False Alarm Rate

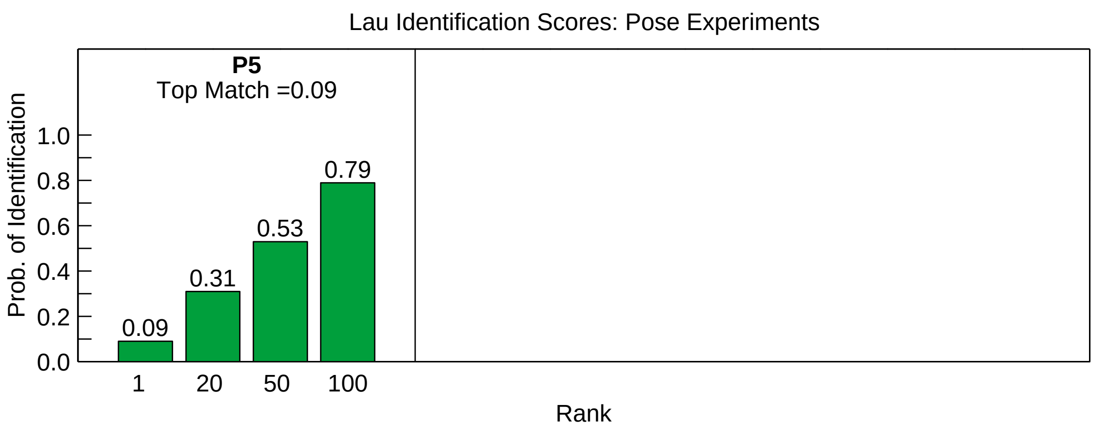

Visionics Verification Scores: Media Experiments
M1
EER =0.01
0.98 0.99 0.99 1
M2
EER =0.01
1 1 1 1
M3
EER =0.01
0.99 1 1 1
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

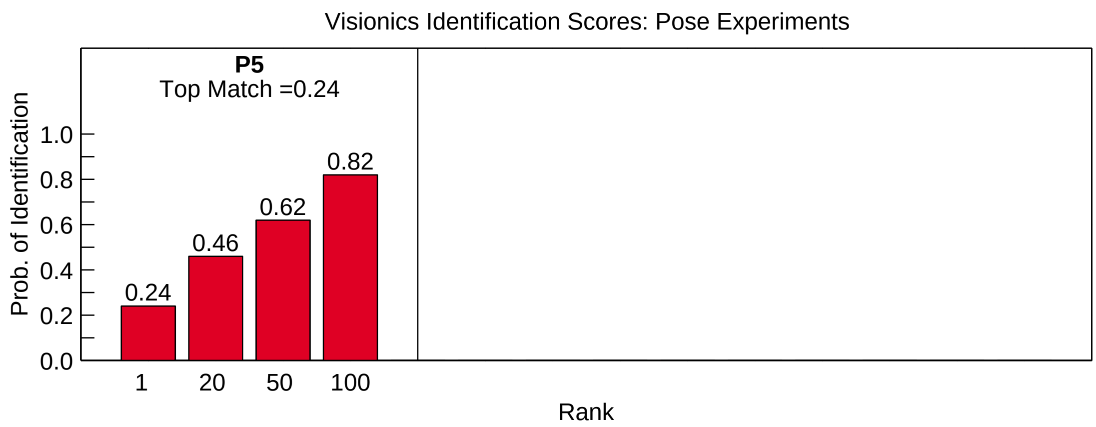

Visionics Verification Scores: Media Experiments
M4
EER =0.02
0.98 0.99 0.99 1
M5
EER =0.02
0.98 0.99 1 1
M6
EER =0.01
0.98 0.99 0.99 1
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

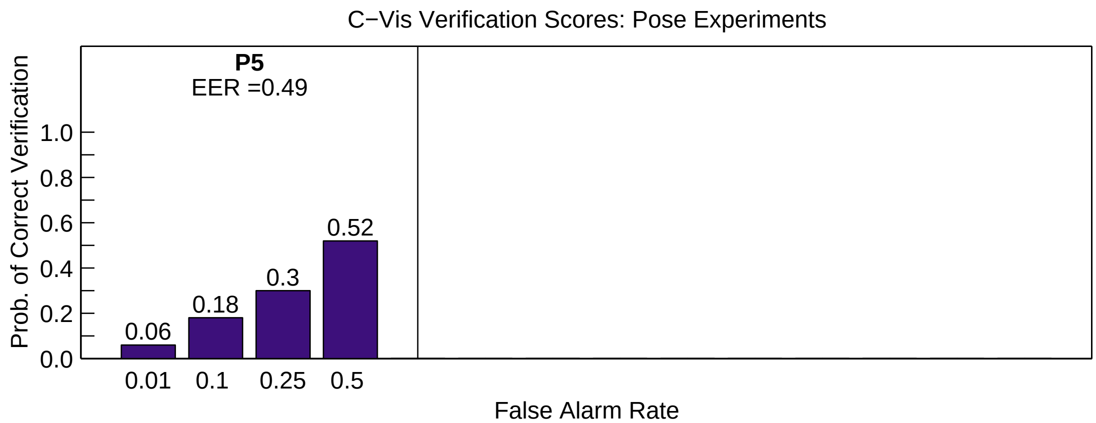


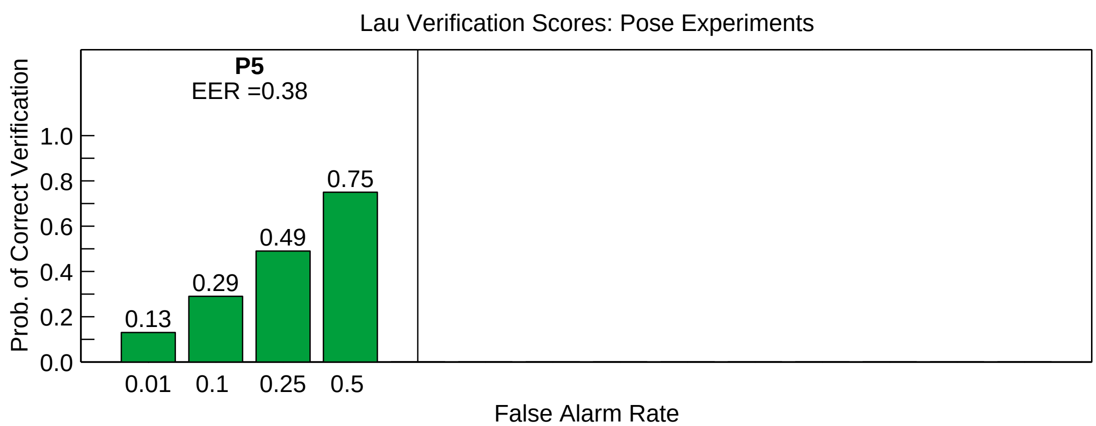
7.1.4.2.5 *FRVT 2000 Experiments – Pose*

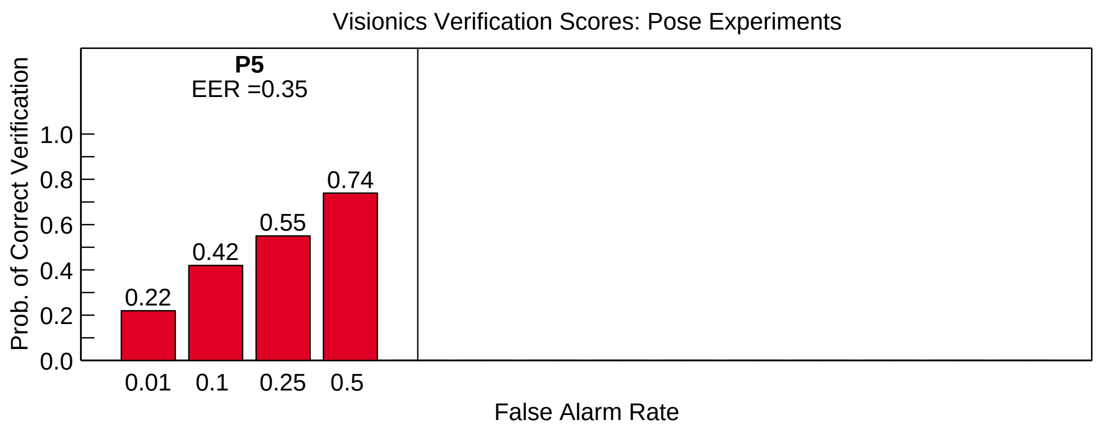




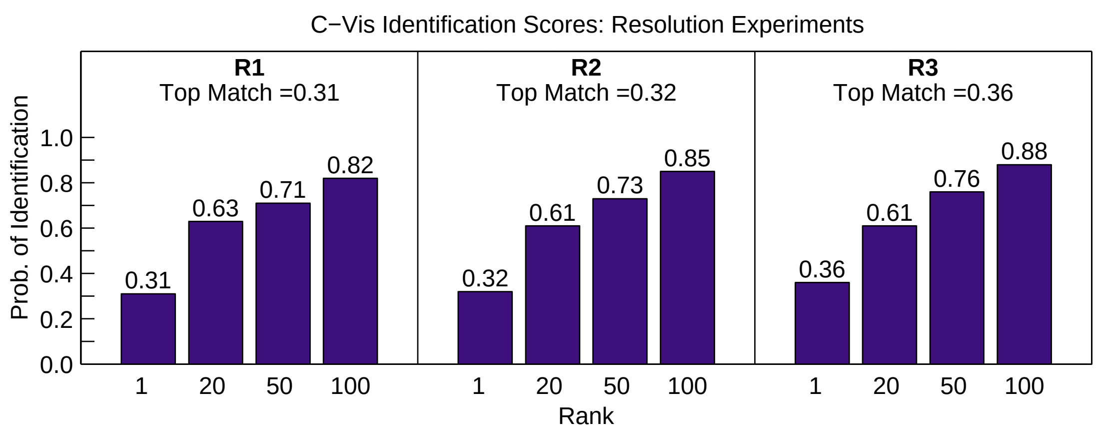

Visionics Identification Scores: Pose Experiments
P5
Top Match =0.24
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
0.24
0.46
0.62
0.82
1
20
50
100
Rank

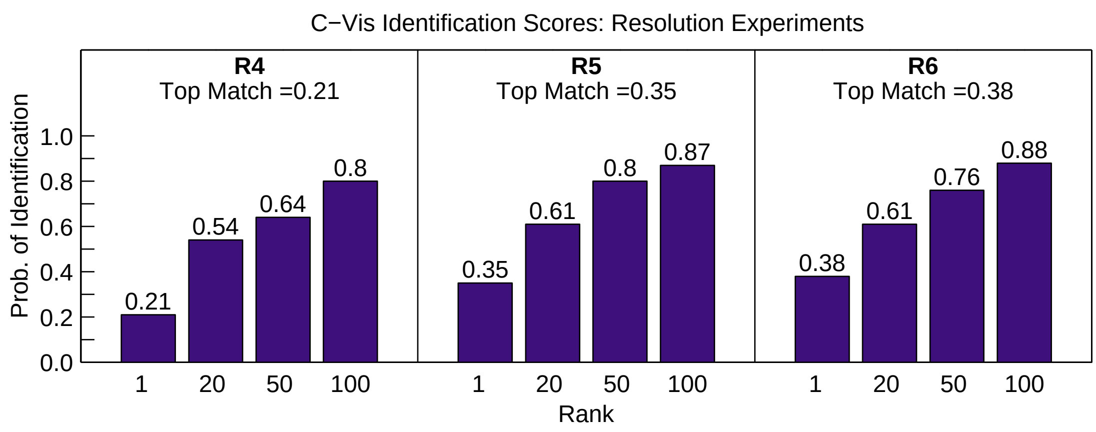

C−Vis Verification Scores: Pose Experiments
P5
EER =0.49
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.06
0.18
0.3
0.52
0.01
0.1
0.25
0.5
False Alarm Rate

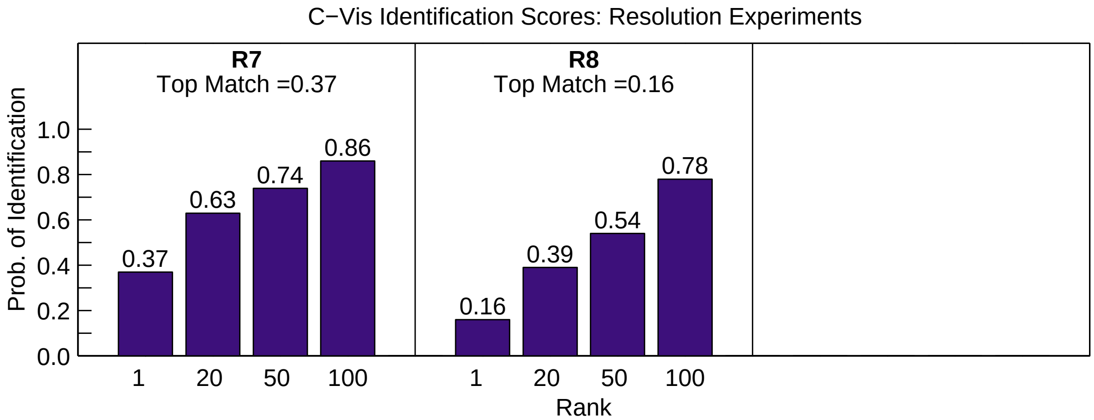

Lau Verification Scores: Pose Experiments
P5
EER =0.38
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.13
0.29
0.49
0.75
0.01
0.1
0.25
0.5
False Alarm Rate

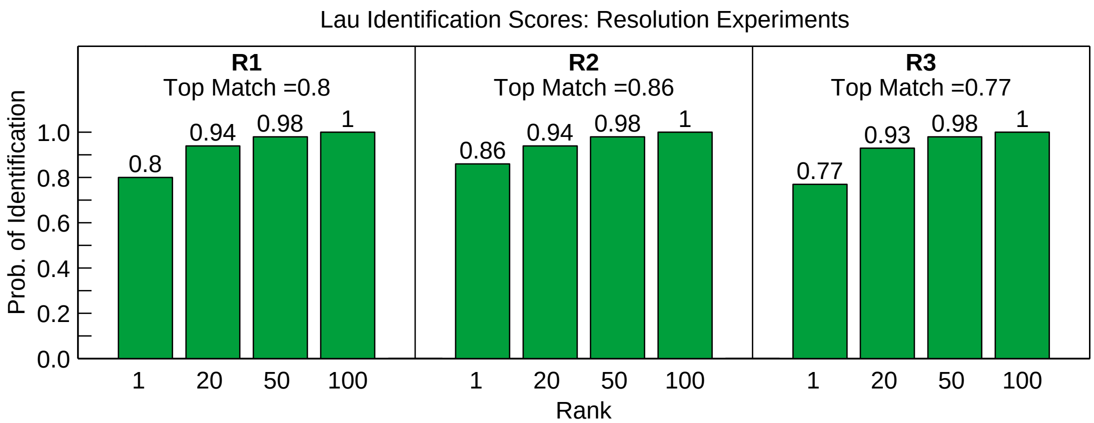


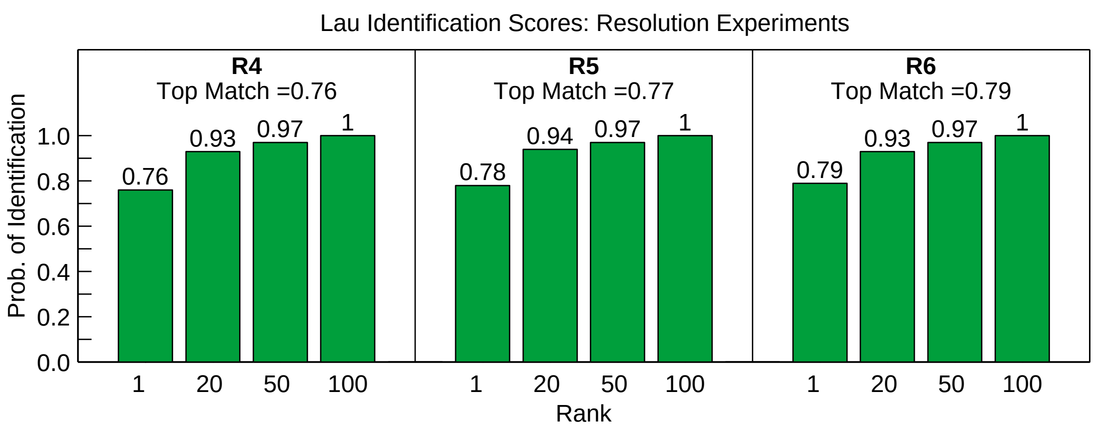
7.1.4.2.6 FRVT 2000 Experiments – Resolution

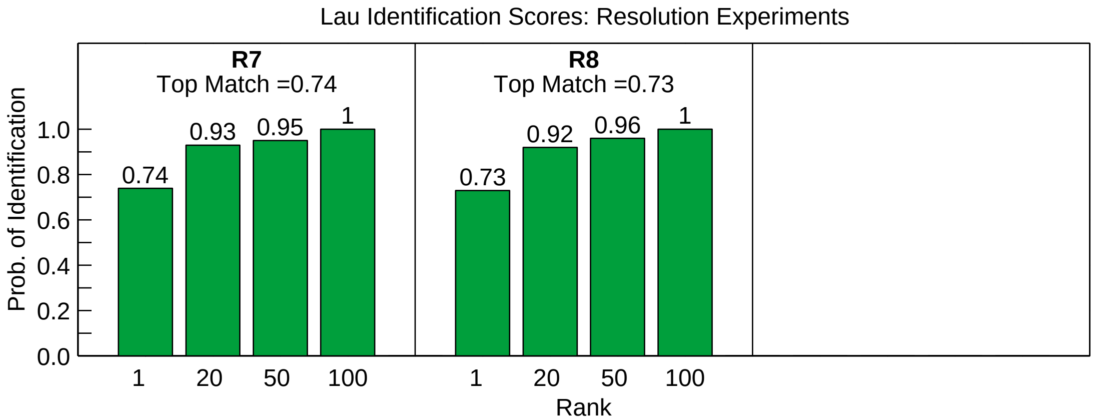


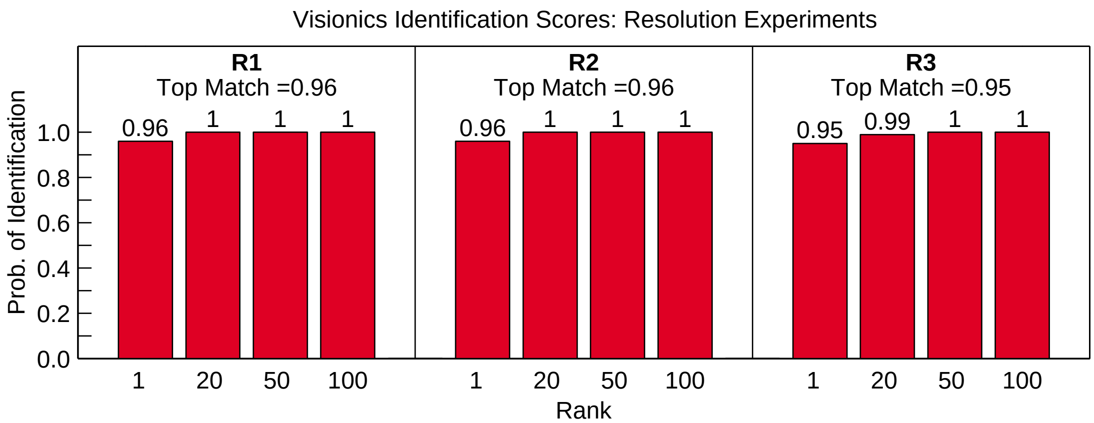


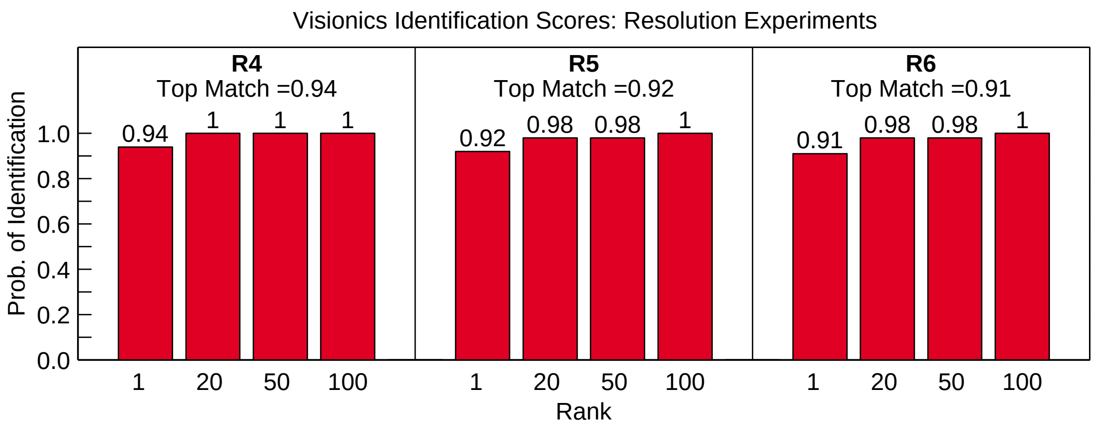

C−Vis Identification Scores: Resolution Experiments
R7
Top Match =0.37
R8
Top Match =0.16
Prob. of Identification
0.0
0.2
0.4
0.6
0.8
1.0
0.37
0.63
0.74
0.86
0.16
0.39
0.54
0.78
1
20
50
100
Rank

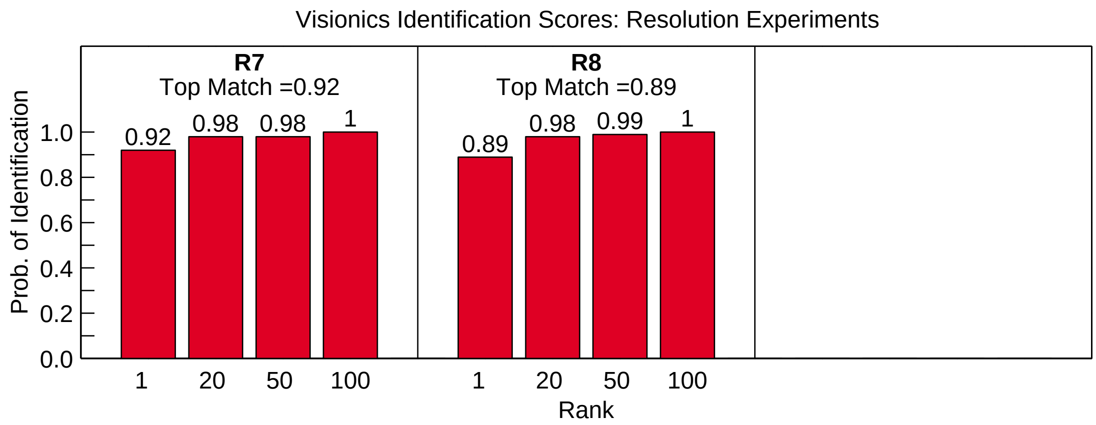

Lau Identification Scores: Resolution Experiments
R1
Top Match =0.8
R2
Top Match =0.86
R3
Top Match =0.77
Prob. of Identification
0.0
0.2
0.4
0.6
0.8
1.0
0.8
0.94
0.98
1
0.86
0.94
0.98
1
0.77
0.93
0.98
1
1
20
50
100
Rank

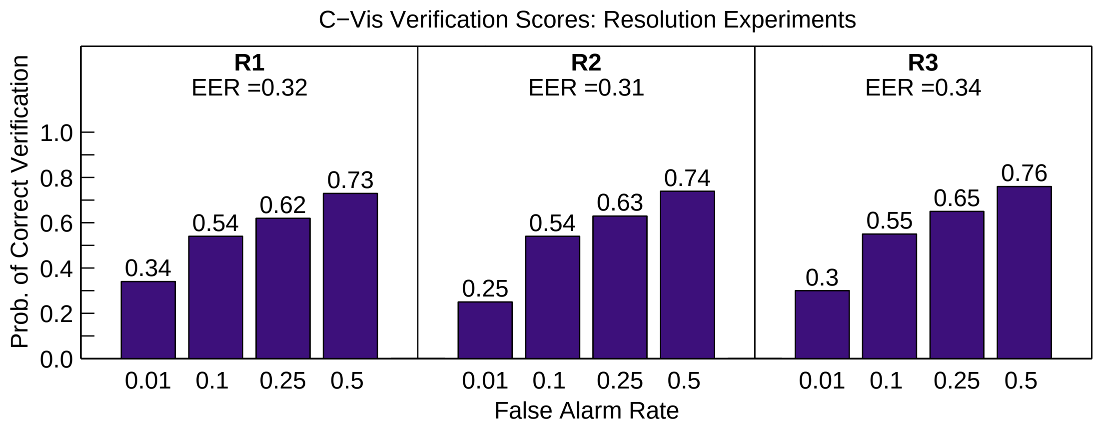

Lau Identification Scores: Resolution Experiments
R4
Top Match =0.76
R5
Top Match =0.77
R6
Top Match =0.79
Prob. of Identification
0.0
0.2
0.4
0.6
0.8
1.0
0.76
0.93
0.97
1
0.78
0.94
0.97
1
0.79
0.93
0.97
1
1
20
50
100
Rank

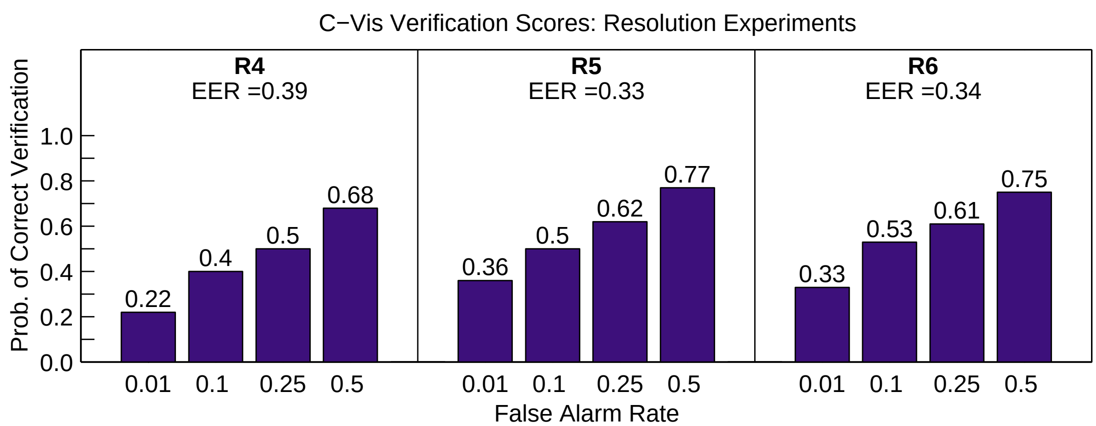

Lau Identification Scores: Resolution Experiments
R7
Top Match =0.74
0.74
0.93
0.95
1
R8
Top Match =0.73
0.73
0.92
0.96
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

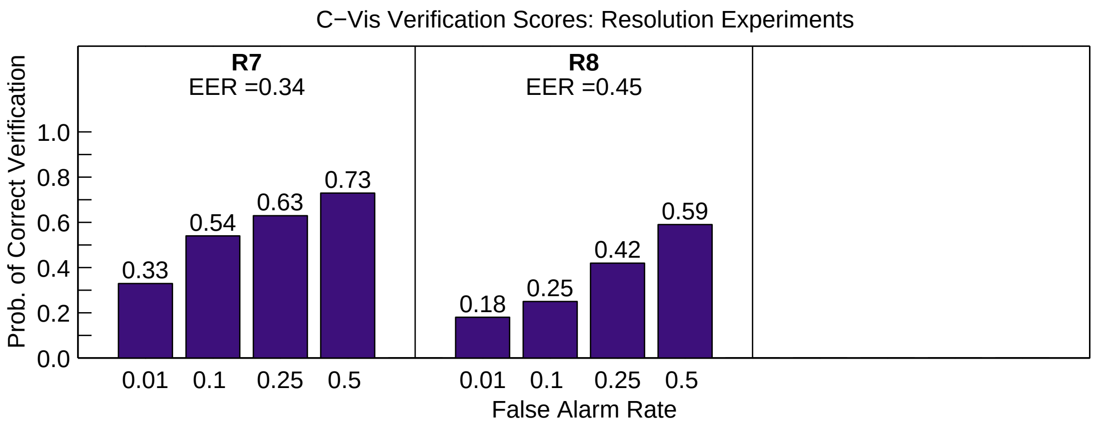

Visionics Identification Scores: Resolution Experiments
R1
Top Match =0.96
0.96
1
1
1
R2
Top Match =0.96
0.96
1
1
1
R3
Top Match =0.95
0.95
0.99
1
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

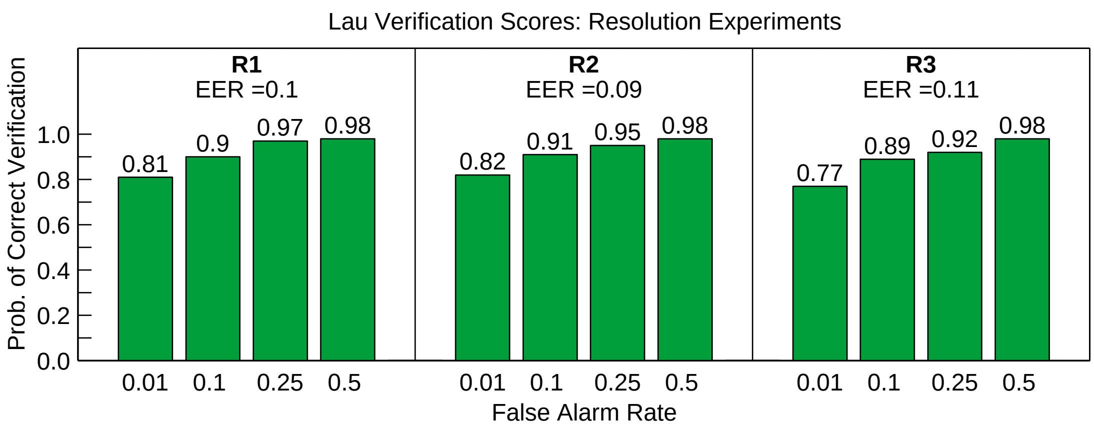

Visionics Identification Scores: Resolution Experiments
R4
Top Match =0.94
0.94
1
1
1
R5
Top Match =0.92
0.92
0.98
0.98
1
R6
Top Match =0.91
0.91
0.98
0.98
1
Prob. of Identification
1.0
0.8
0.6
0.4
0.2
0.0
1
20
50
100
Rank

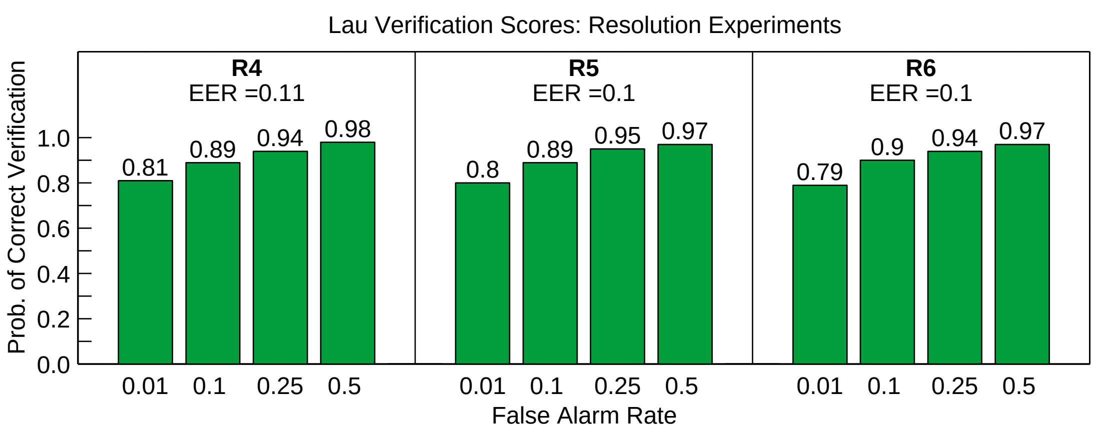

Visionics Identification Scores: Resolution Experiments
R7
Top Match =0.92
0.92 0.98 0.98 1
R8
Top Match =0.89
0.89 0.98 0.99 1
Prob. of Identification
1.0 0.8 0.6 0.4 0.2 0.0
1 20 50 100
1 20 50 100
Rank

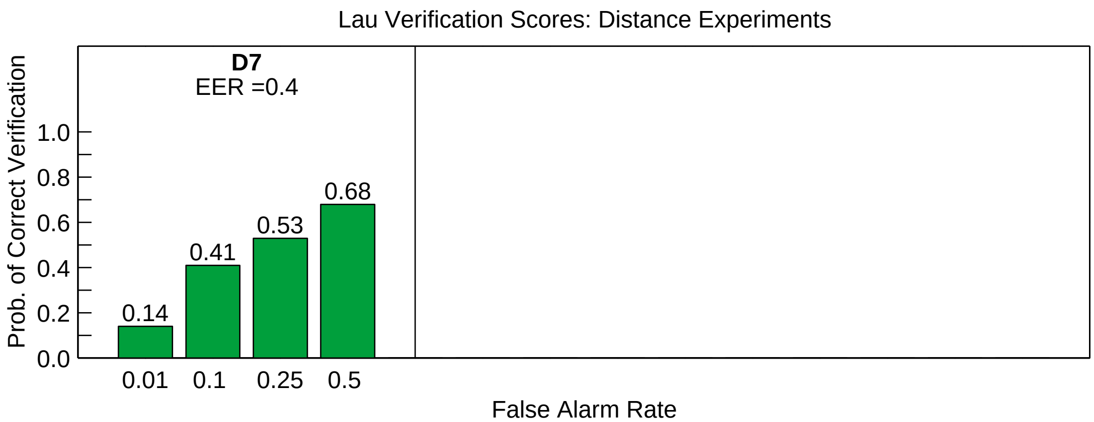

C−Vis Verification Scores: Resolution Experiments
R1
EER =0.32
0.34 0.54 0.62 0.73
R2
EER =0.31
0.25 0.54 0.63 0.74
R3
EER =0.34
0.3 0.55 0.65 0.76
Prob. of Correct Verification
1.0 0.8 0.6 0.4 0.2 0.0
0.01 0.1 0.25 0.5
False Alarm Rate

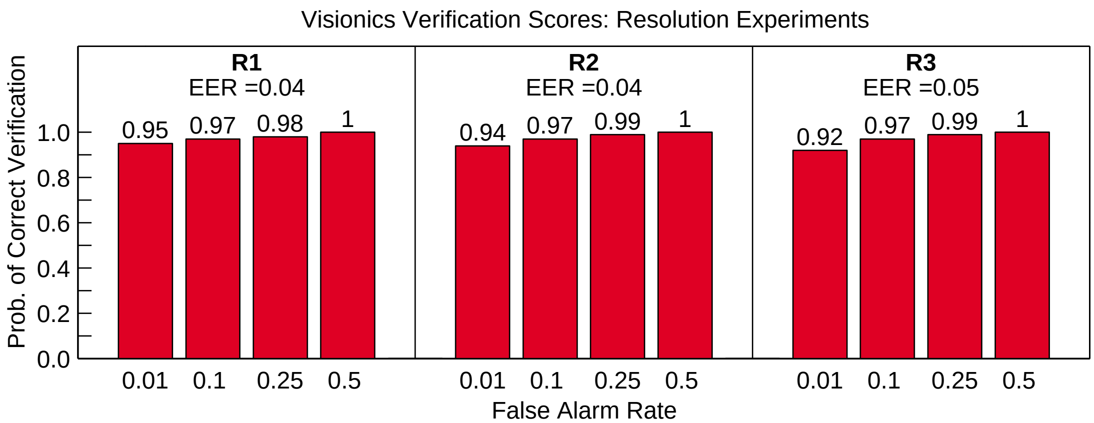

C−Vis Verification Scores: Resolution Experiments
R4
EER =0.39
0.22 0.4 0.5 0.68
R5
EER =0.33
0.36 0.5 0.62 0.77
R6
EER =0.34
0.33 0.53 0.61 0.75
Prob. of Correct Verification
1.0 0.8 0.6 0.4 0.2 0.0
0.01 0.1 0.25 0.5
False Alarm Rate

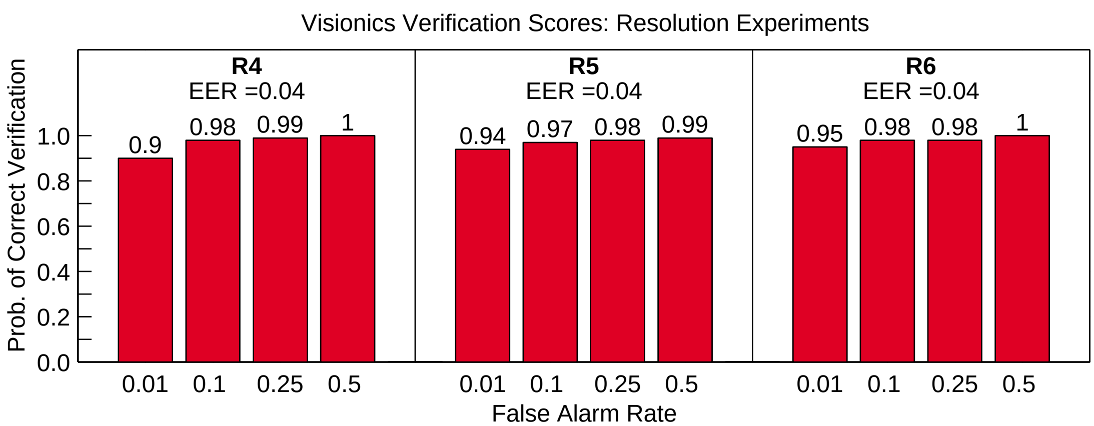

C−Vis Verification Scores: Resolution Experiments
R7
EER =0.34
R8
EER =0.45
0.33
0.54
0.63
0.73
0.18
0.25
0.42
0.59
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

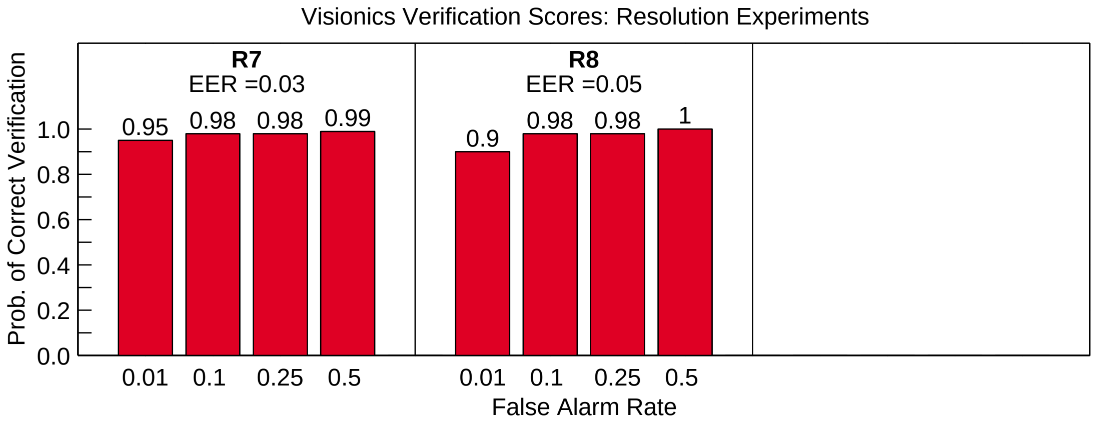

Lau Verification Scores: Resolution Experiments
R1
EER =0.1
R2
EER =0.09
R3
EER =0.11
0.81
0.9
0.97
0.98
0.82
0.91
0.95
0.98
0.77
0.89
0.92
0.98
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate


Lau Verification Scores: Resolution Experiments
R4
EER =0.11
R5
EER =0.1
R6
EER =0.1
0.81
0.89
0.94
0.98
0.8
0.89
0.95
0.97
0.79
0.9
0.94
0.97
Prob. of Correct Verification
1.0
0.8
0.6
0.4
0.2
0.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

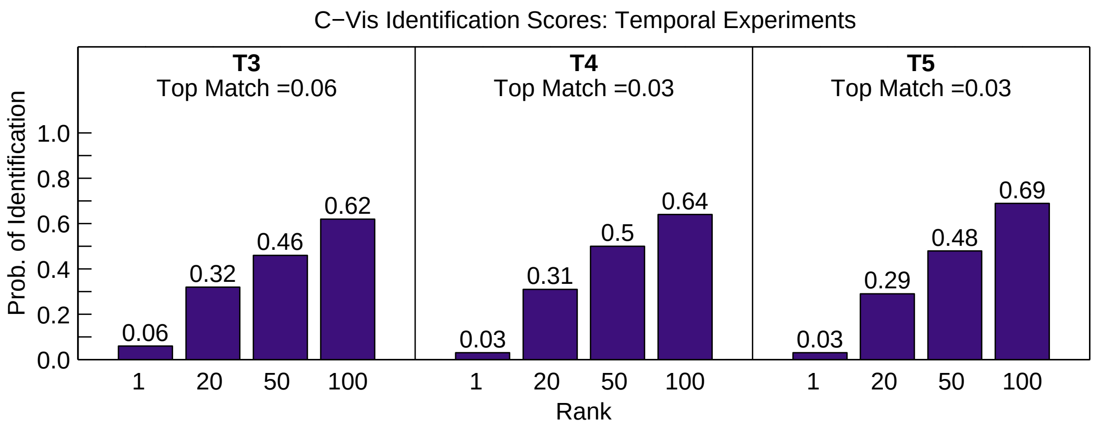

Lau Verification Scores: Distance Experiments
D7
EER =0.4
0.14
0.41
0.53
0.68
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.1
0.25
0.5
False Alarm Rate

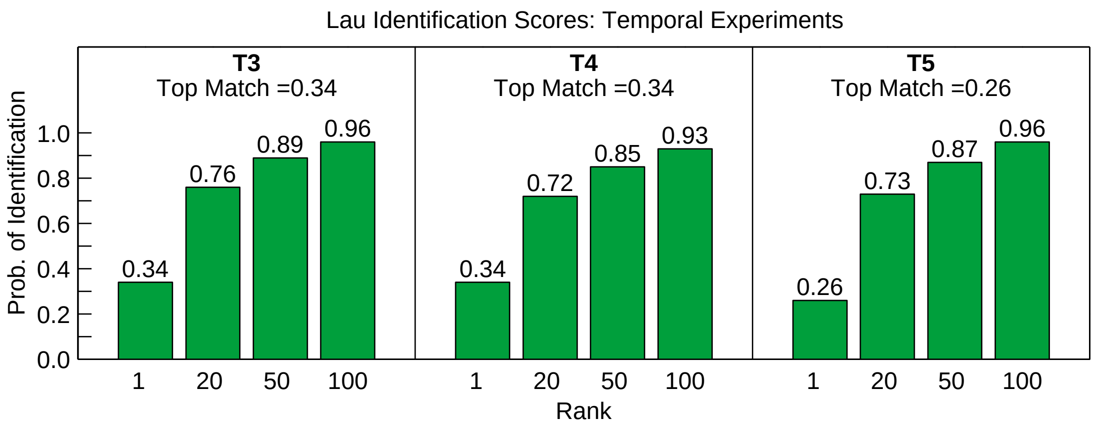

Visionics Verification Scores: Resolution Experiments
R1
EER =0.04
0.95 0.97 0.98 1
R2
EER =0.04
0.94 0.97 0.99 1
R3
EER =0.05
0.92 0.97 0.99 1
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

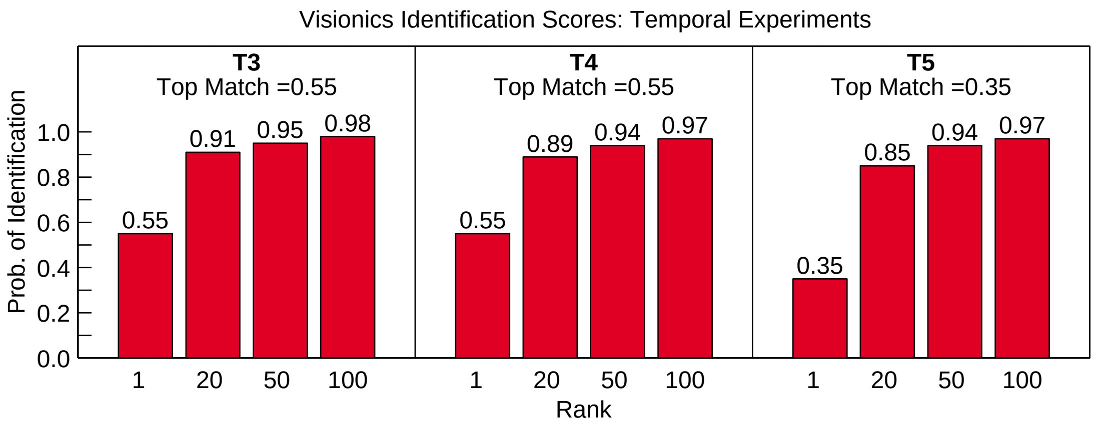

Visionics Verification Scores: Resolution Experiments
R4
EER =0.04
0.9 0.98 0.99 1
R5
EER =0.04
0.94 0.97 0.98 0.99
R6
EER =0.04
0.95 0.98 0.98 1
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
0.01 0.1 0.25 0.5
False Alarm Rate

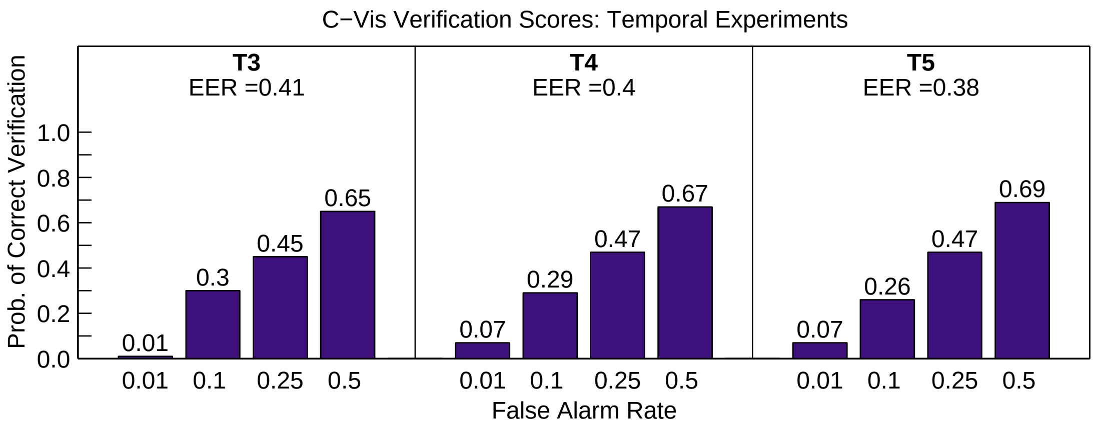


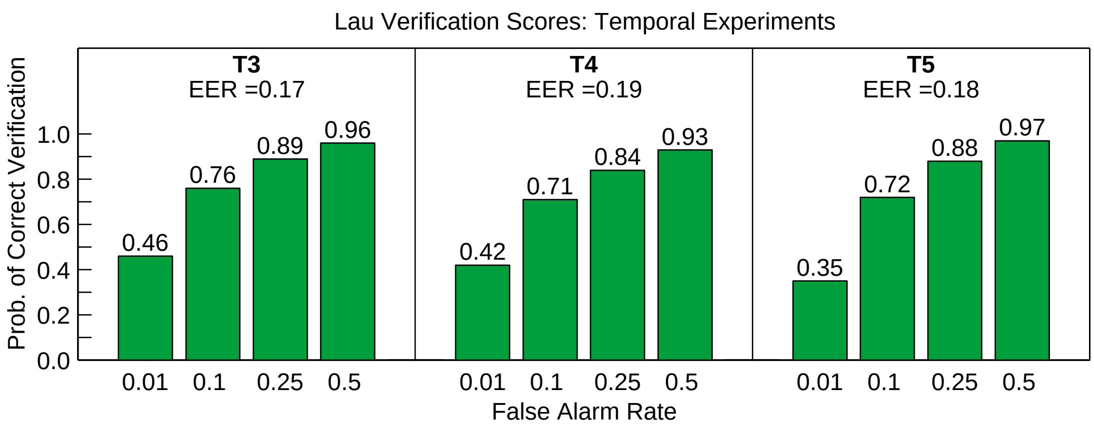
7.1.4.2.7 FRVT 2000 Experiments – Temporal

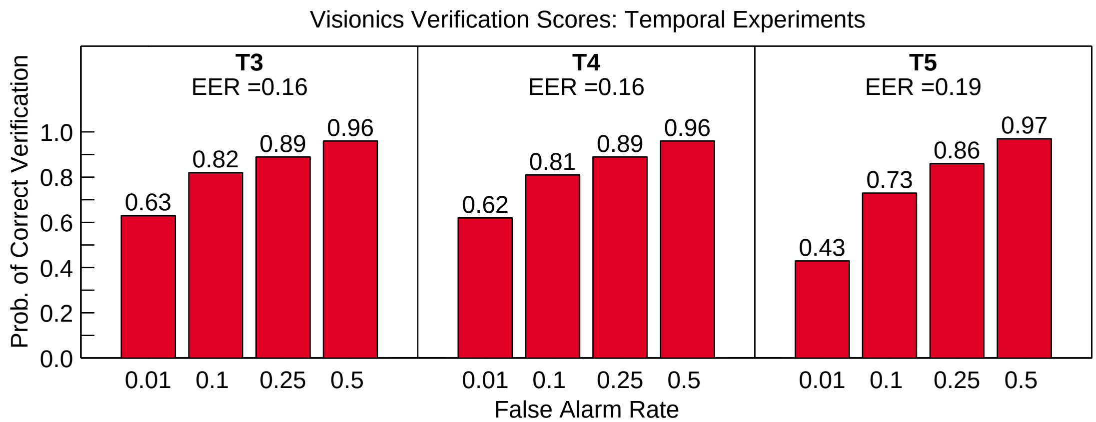





Visionics Identification Scores: Temporal Experiments
T3
Top Match =0.55
T4
Top Match =0.55
T5
Top Match =0.35
Prob. of Identification
0.0
0.2
0.4
0.6
0.8
1.0
0.55
0.91
0.95
0.98
0.55
0.89
0.94
0.97
0.35
0.85
0.94
0.97
1
20
50
100
Rank


C−Vis Verification Scores: Temporal Experiments
T3
EER =0.41
T4
EER =0.4
T5
EER =0.38
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.01
0.3
0.45
0.65
0.07
0.29
0.47
0.67
0.07
0.26
0.47
0.69
0.01
0.1
0.25
0.5
False Alarm Rate


Lau Verification Scores: Temporal Experiments
T3
EER =0.17
T4
EER =0.19
T5
EER =0.18
Prob. of Correct Verification
0.0
0.2
0.4
0.6
0.8
1.0
0.46
0.76
0.89
0.96
0.42
0.71
0.84
0.93
0.35
0.72
0.88
0.97
0.01
0.1
0.25
0.5
False Alarm Rate



7.2 Product Usability Test

7.2.1 Overview

The scenario chosen for the Product Usability Test was access control with live subjects. Some systems tested, however, were not intended for access control applications. The intended application for each system, as shown in Appendix J, should be kept in mind when evaluating the results of the Product Usability Test.

The Product Usability Test was administered in two parts: the Old Image Database Timed Test and the Enrollment Timed Test. For the Old Image Database Timed Test, vendors were given a set of 165 images captured with a standard access control badging system, including images of the three test subjects. Vendors enrolled these images into their system for comparison with the live subjects. The operational scenario was that of a low security access control point into the lobby of a building. The building's security officers did not want to mandate that the employees take the time to enroll into the new facial recognition system so they used their existing digital image database taken from the employee's picture ID badges.

For the Enrollment Timed Test, the images of the three test subjects were removed from the system while the other images were retained. Vendors were then allowed to enroll the three subjects using their standard procedures, including the use of multiple images. The purpose of the test was to measure system performance using vendor enrollment procedures. The enrollment procedures were not evaluated. The operational scenario was that of an access control door for a medium/high security area within the building previously described. In this case, employees were enrolled in the facial recognition system using the standard procedures recommended by the vendor.

During the Product Usability Test, several parameters were varied including start distance, behavior mode, and backlighting. Tests were performed by each subject at start distances of 12, 8, and 4 feet for all trials except for the variability test, which always started at 12 feet. Test subjects performed each test using both a cooperative and non-

Appendix G
FRVT 2000 Recognition Performance Test Results
(Detailed)

Figure M−1: Best Identification Scores: Compression C0

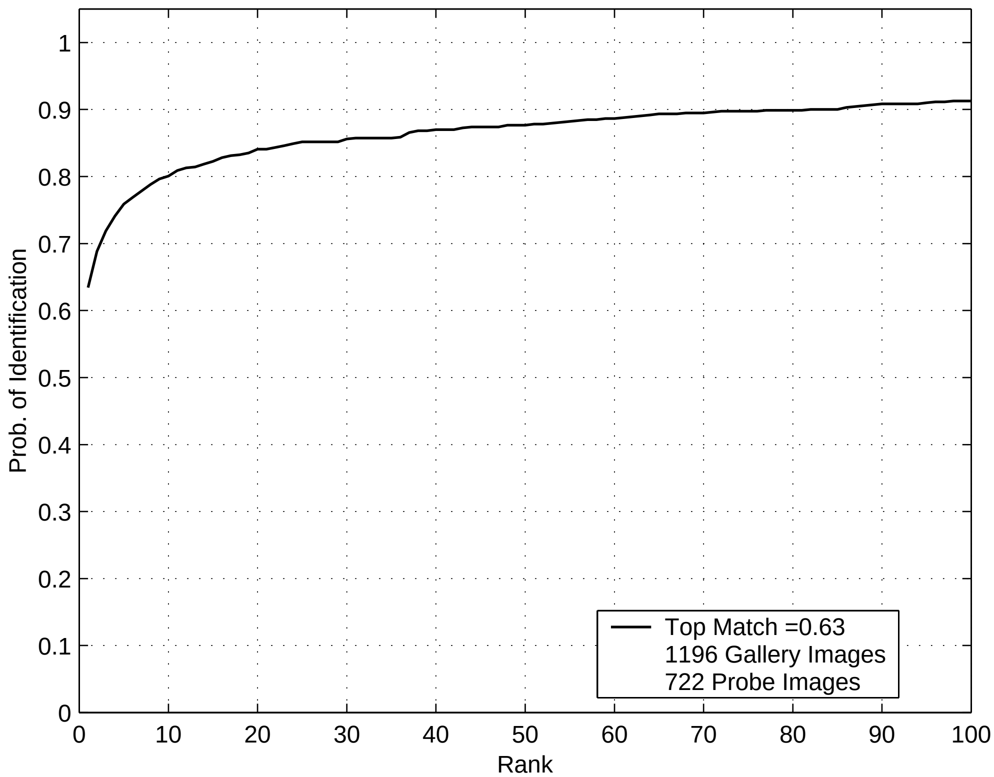


Figure M−2: Best Identification Scores: Compression C1

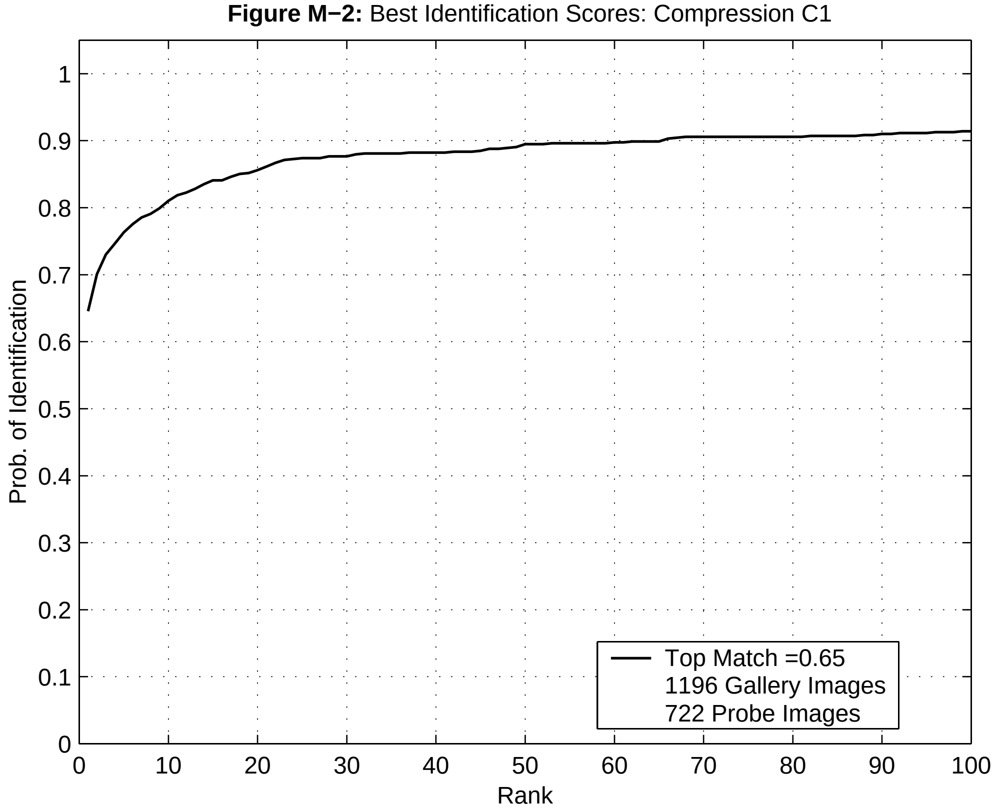


Figure M−3: Best Identification Scores: Compression C2

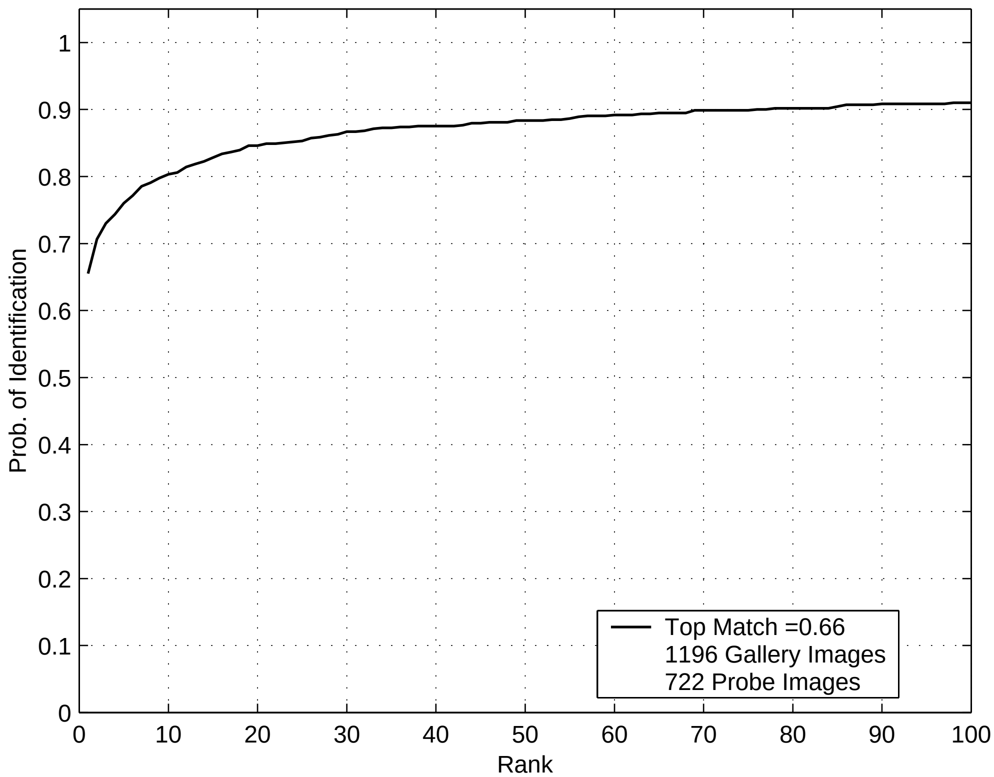


Figure M−4: Best Identification Scores: Compression C3

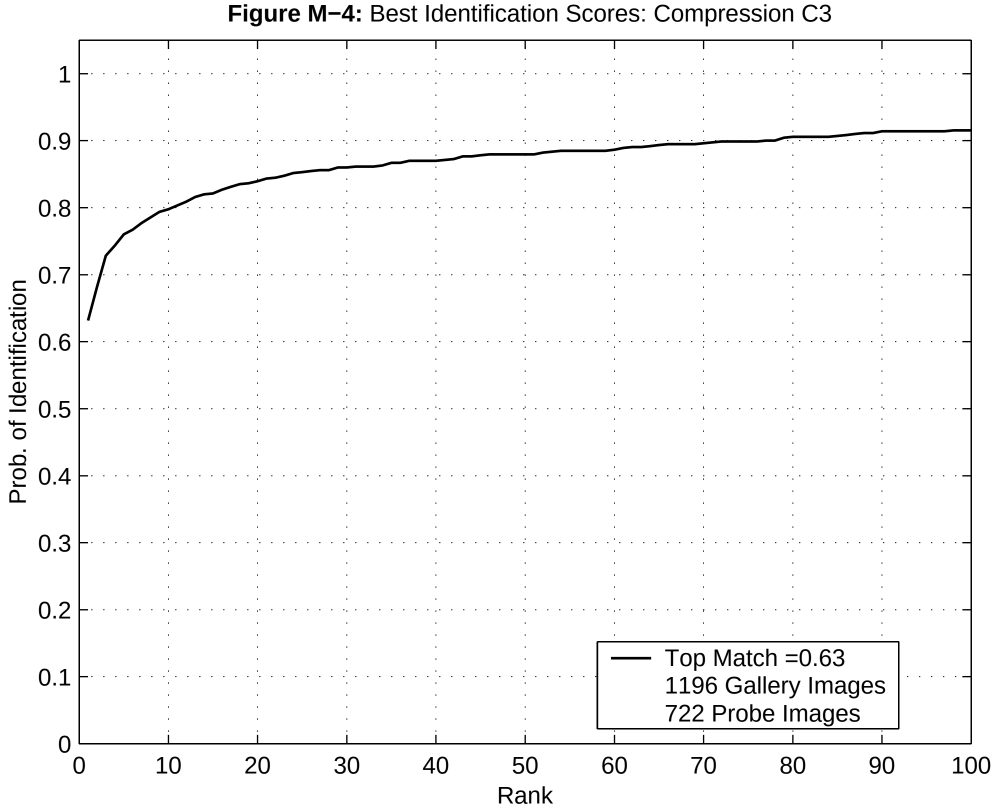


Figure M−5: Best Identification Scores: Compression C4

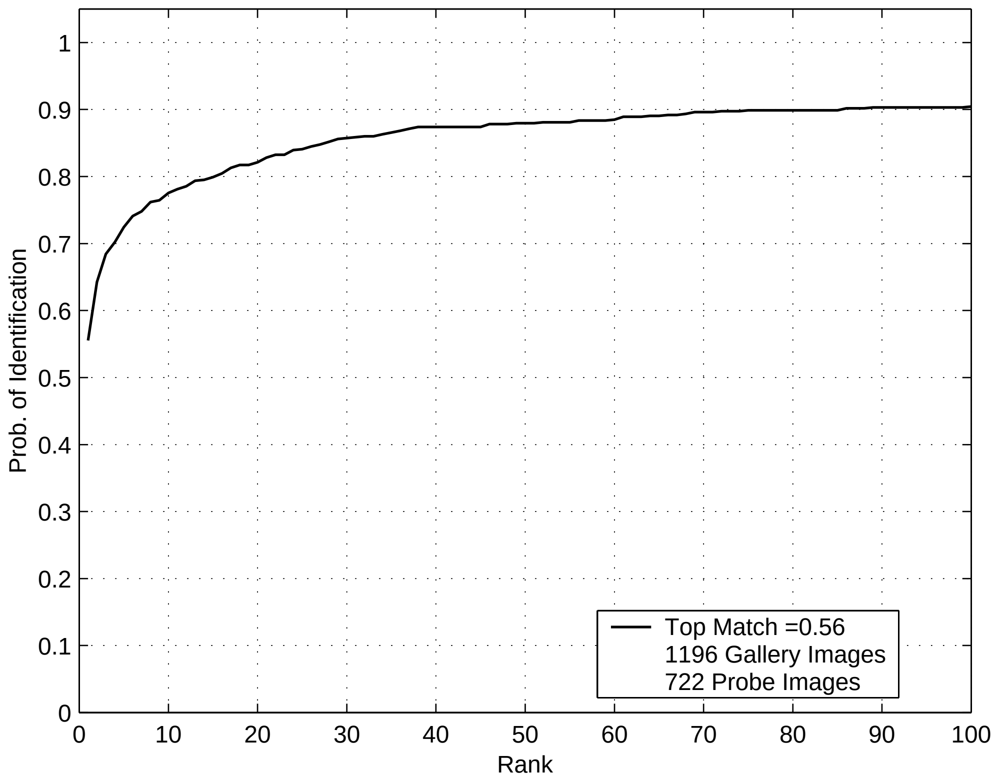


Figure M−6: Best Identification Scores: Pose P1

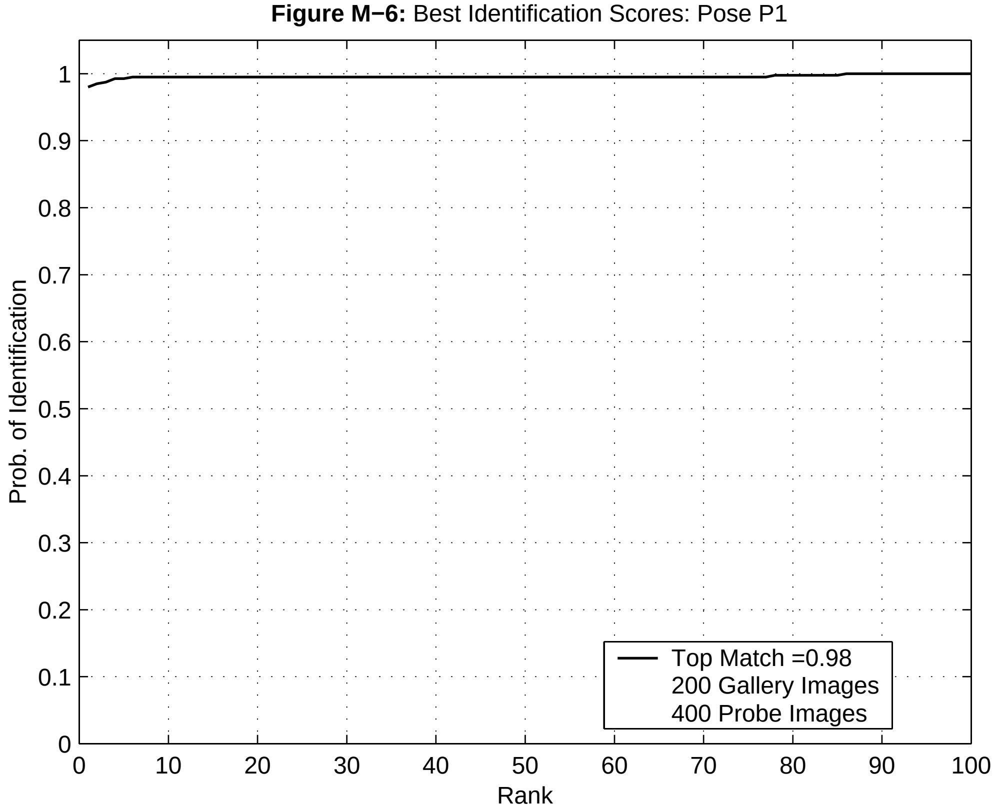


Figure M−7: Best Identification Scores: Pose P2

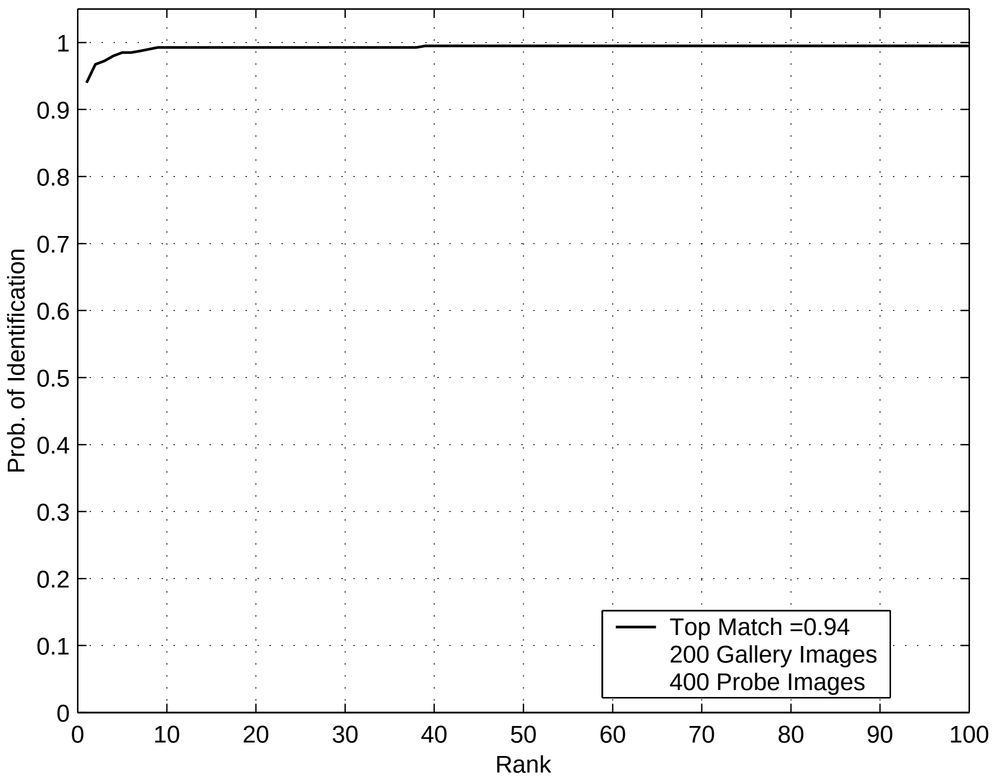


Figure M−8: Best Identification Scores: Pose P3

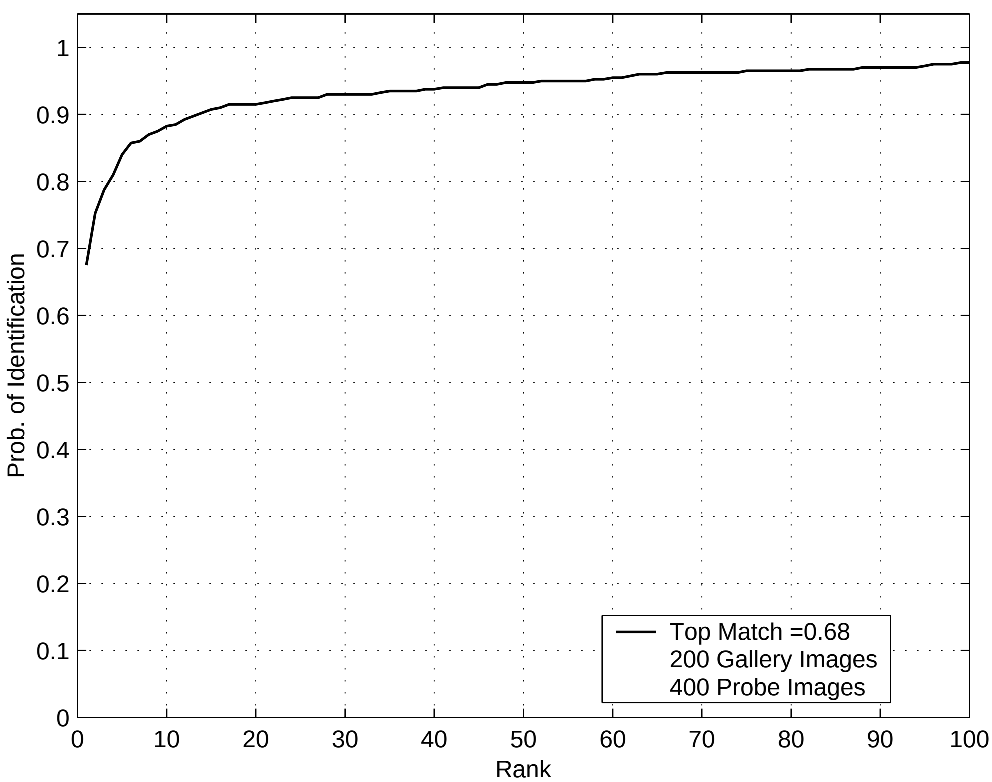


Figure M−9: Best Identification Scores: Pose P4

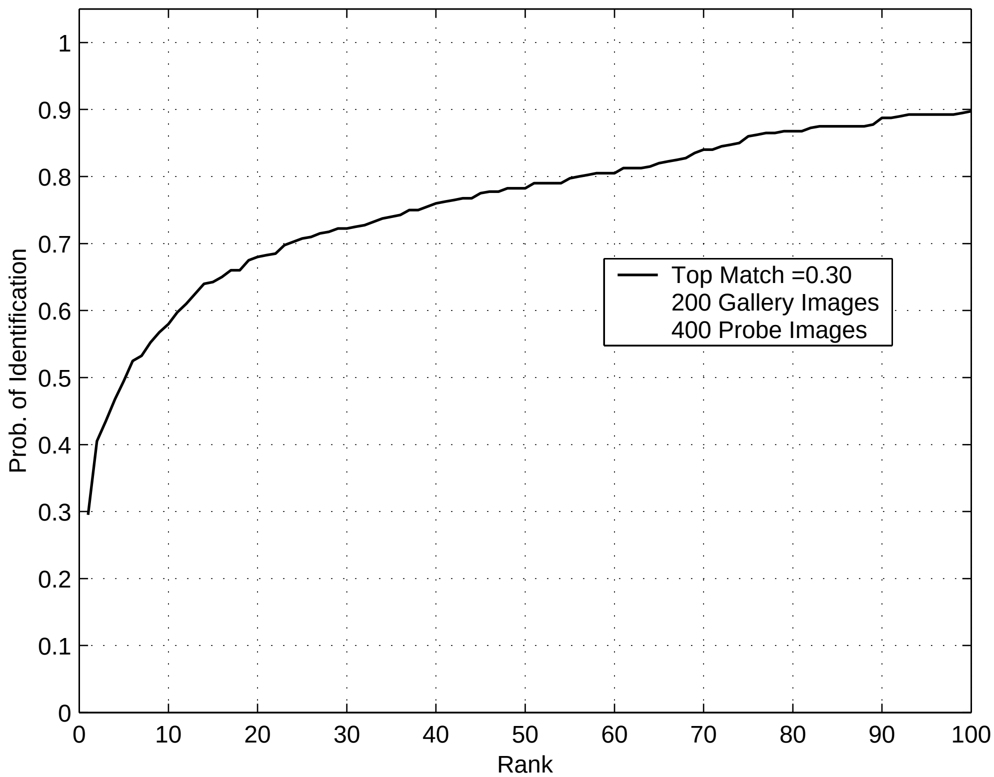
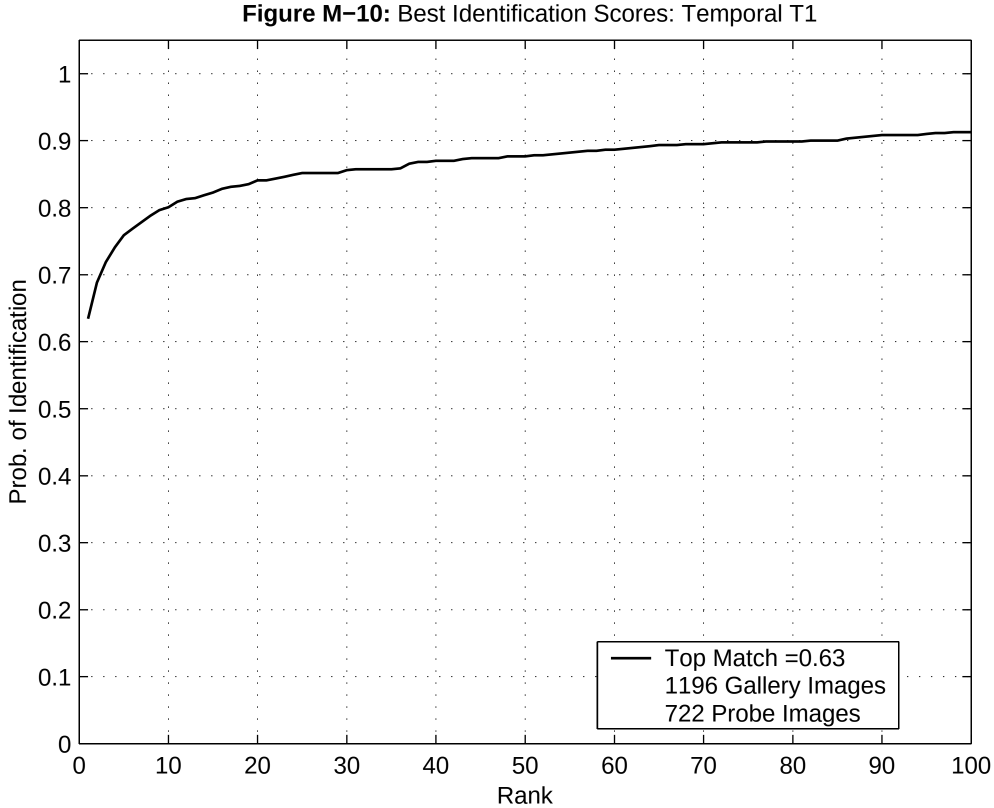


Figure M−10: Best Identification Scores: Temporal T1



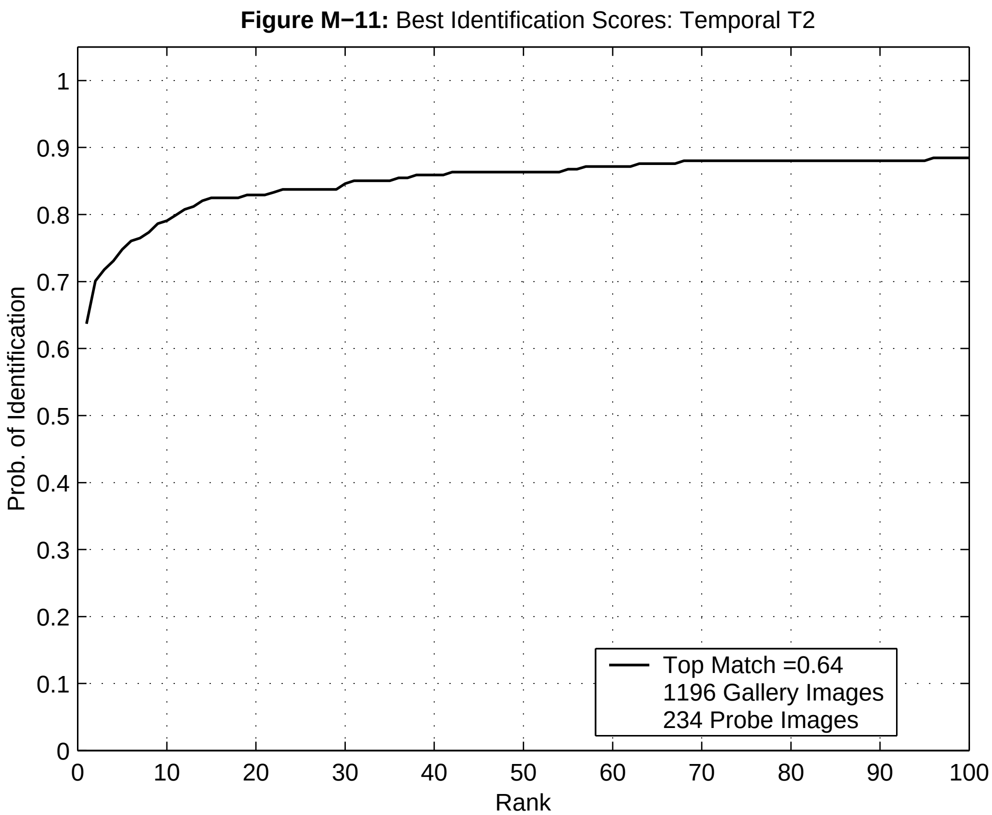


Figure M−11: Best Identification Scores: Temporal T2
Prob. of Identification
Rank
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100
Top Match =0.64
1196 Gallery Images
234 Probe Images

Figure M−12: Identification Scores: Distance D1

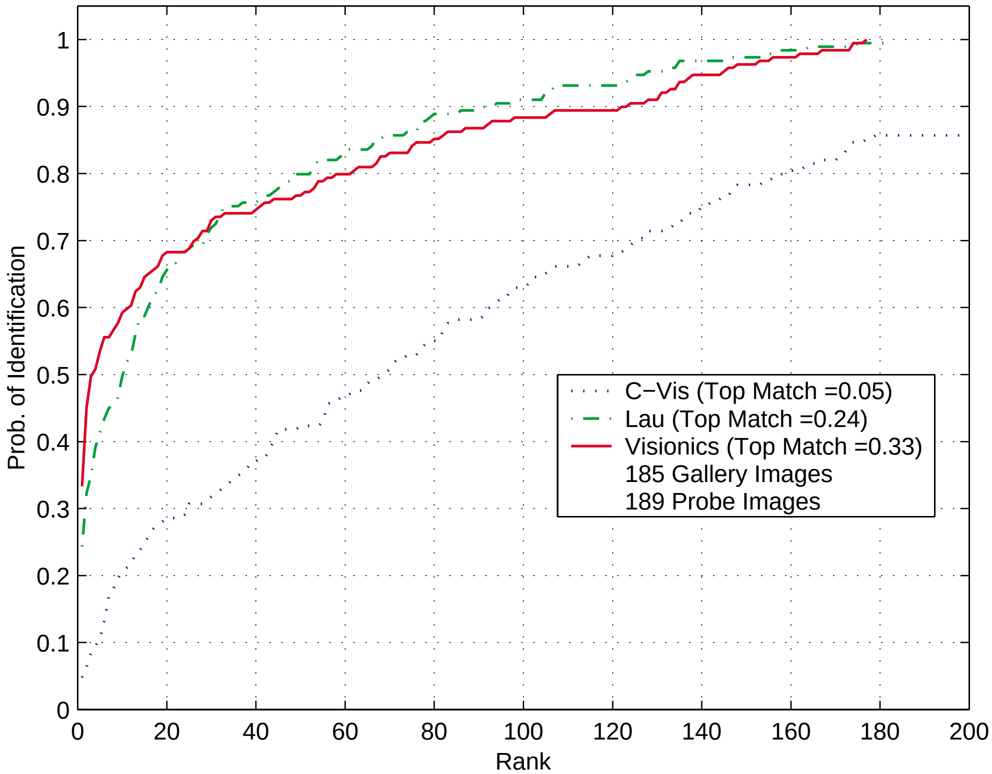


Figure M−13: Identification Scores: Distance D2

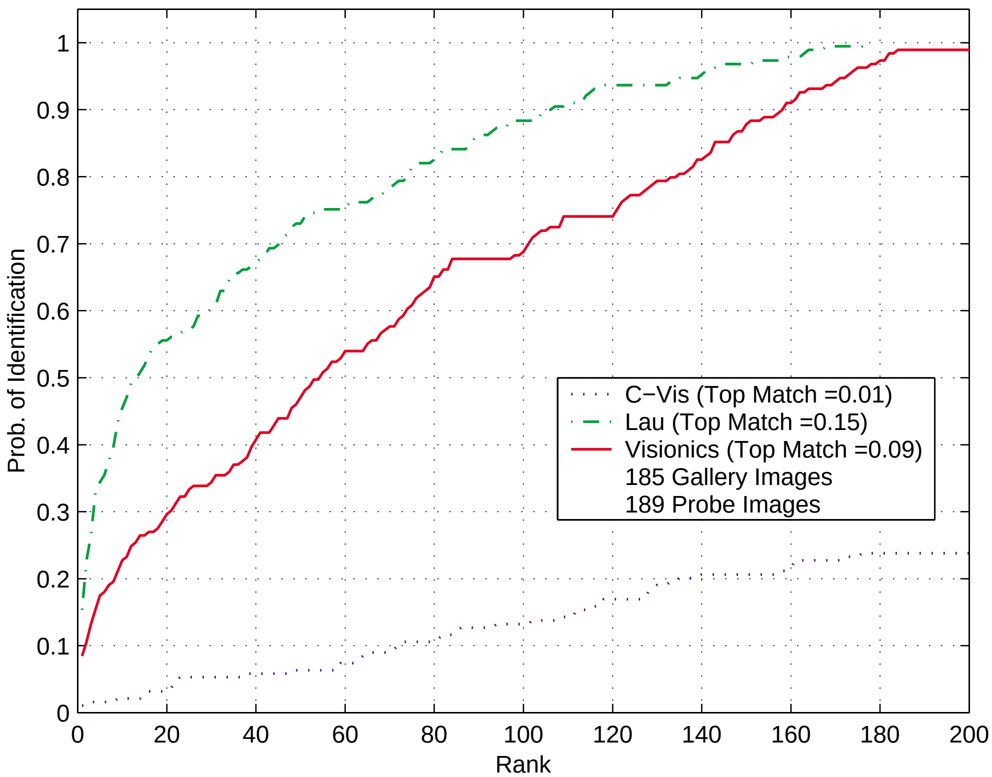


Figure M−14: Identification Scores: Distance D3

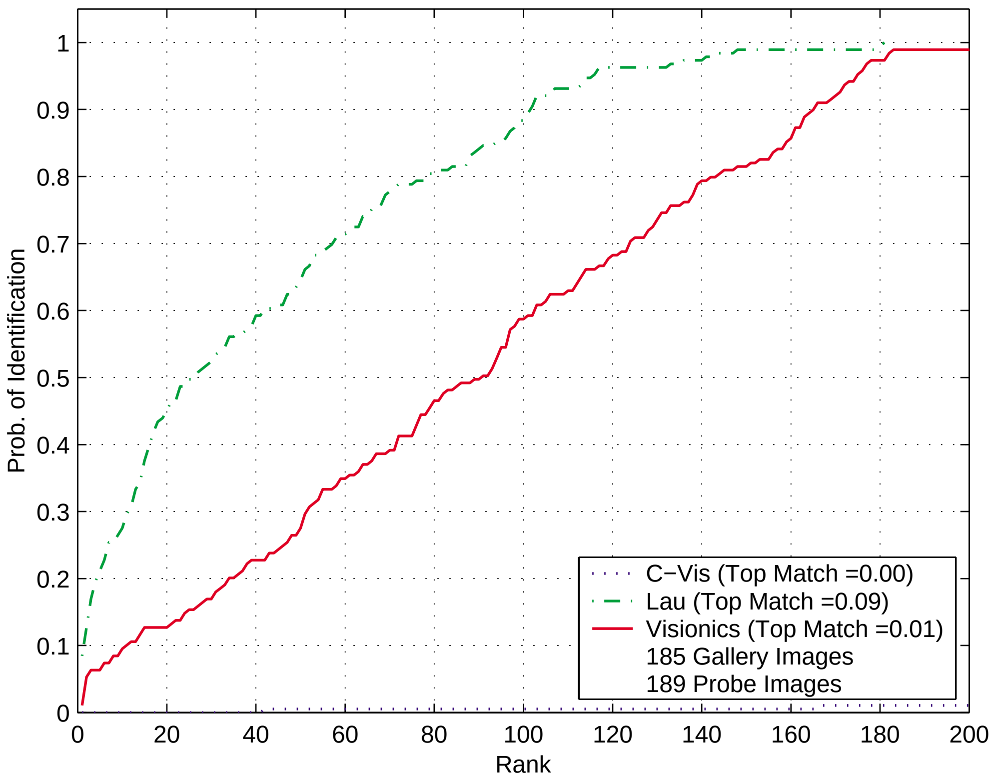


Figure M−15: Identification Scores: Distance D4

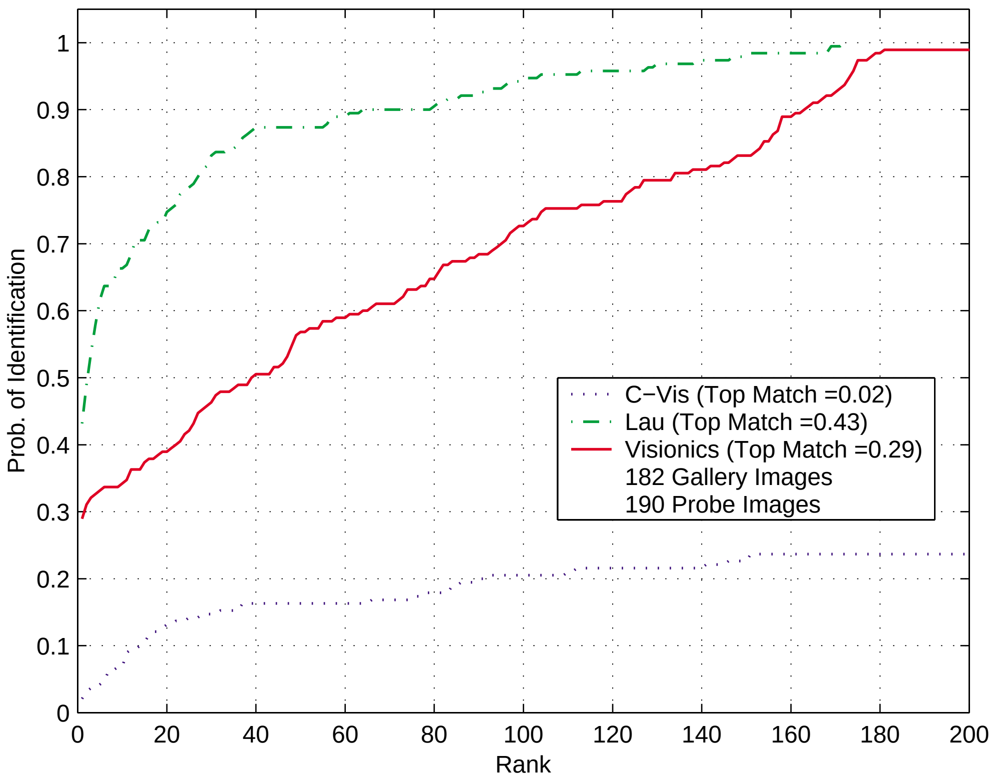


Figure M−16: Identification Scores: Distance D5

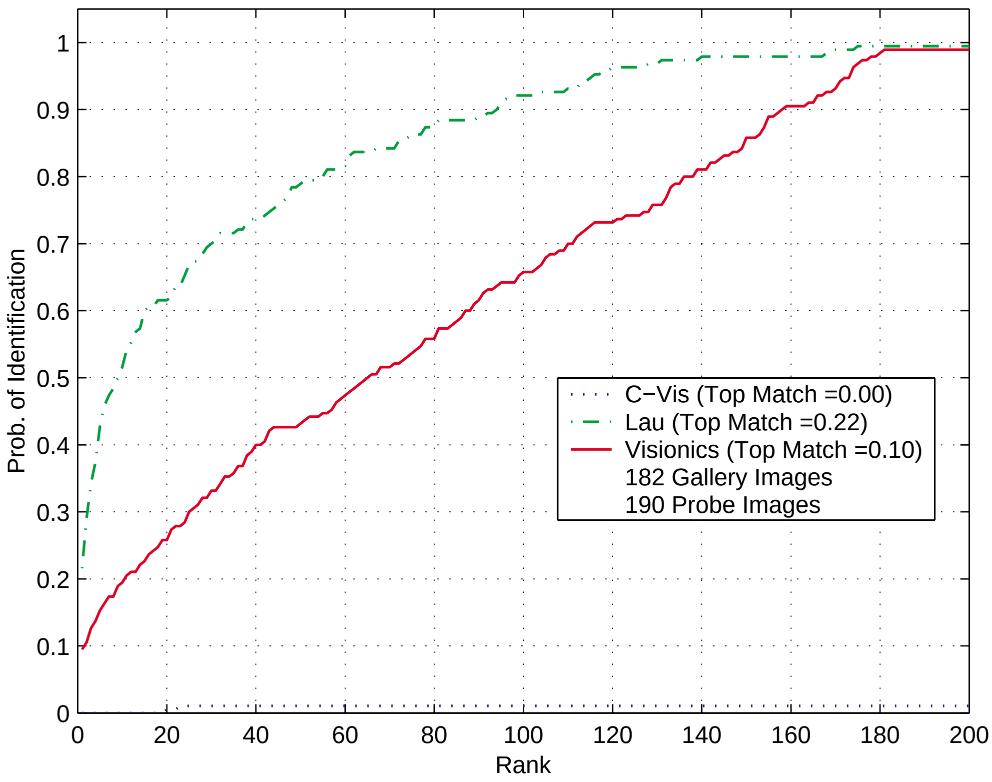


Figure M−17: Identification Scores: Distance D6

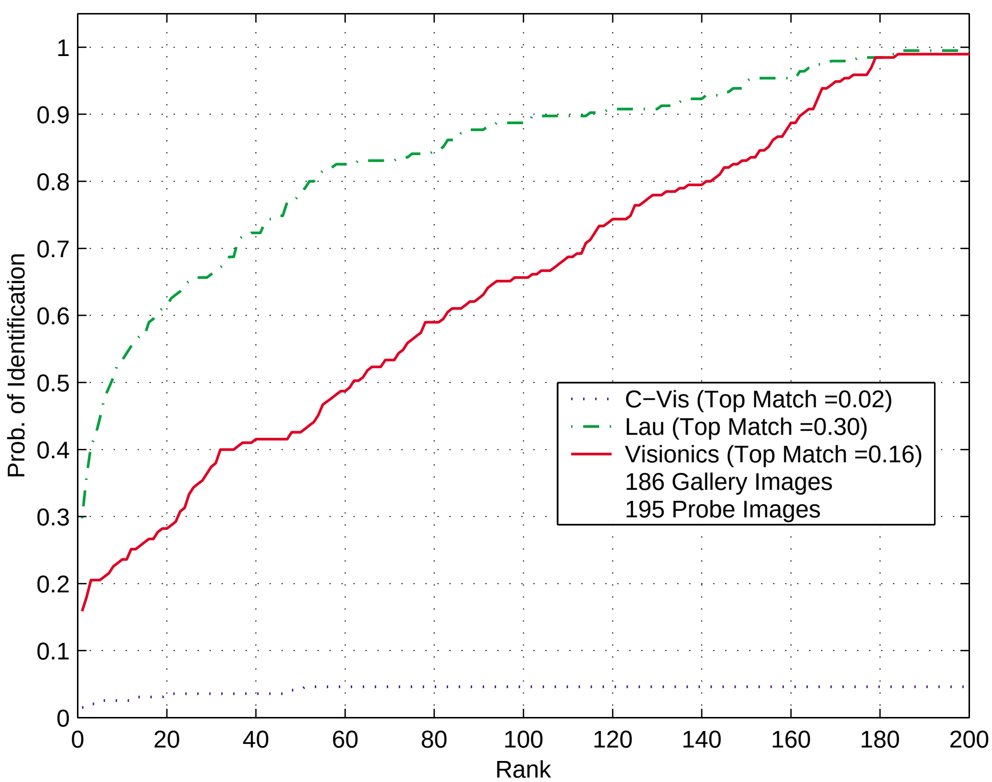


Figure M−18: Identification Scores: Distance D7

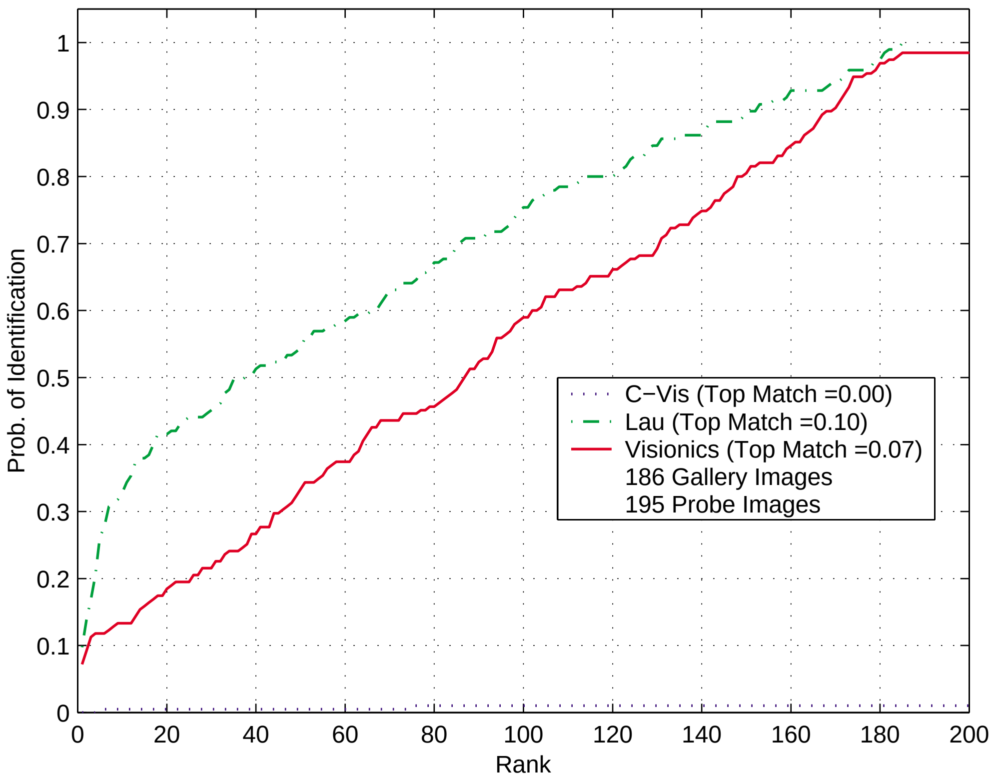


Figure M−19: Identification Scores: Expression E1

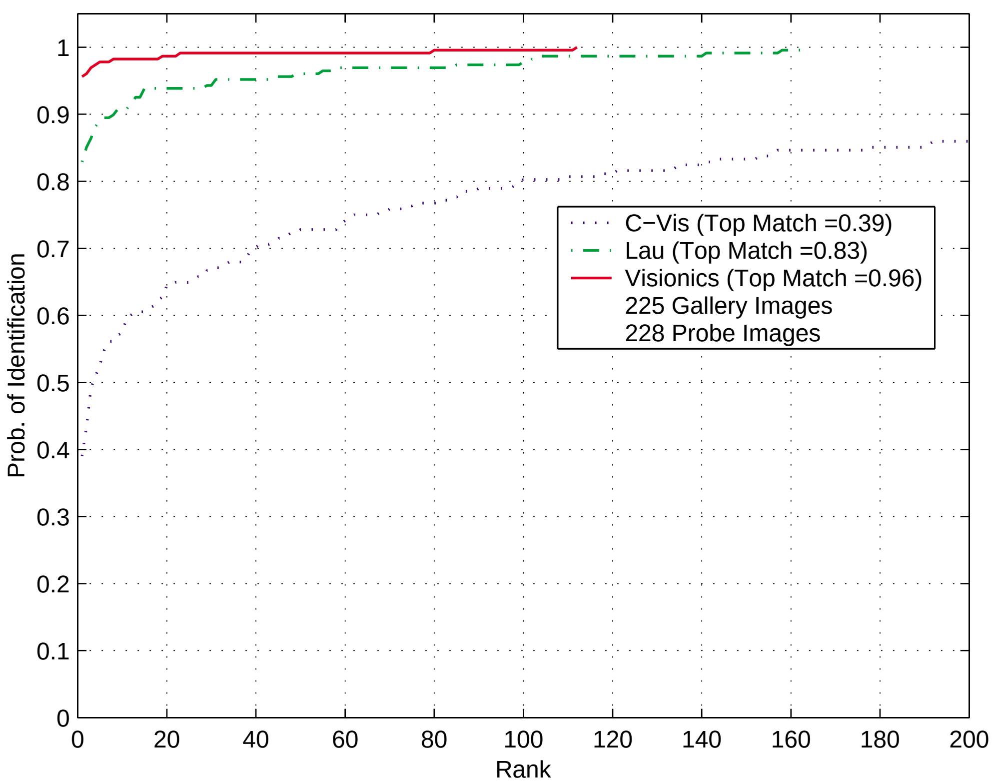


Figure M−20: Identification Scores: Expression E2

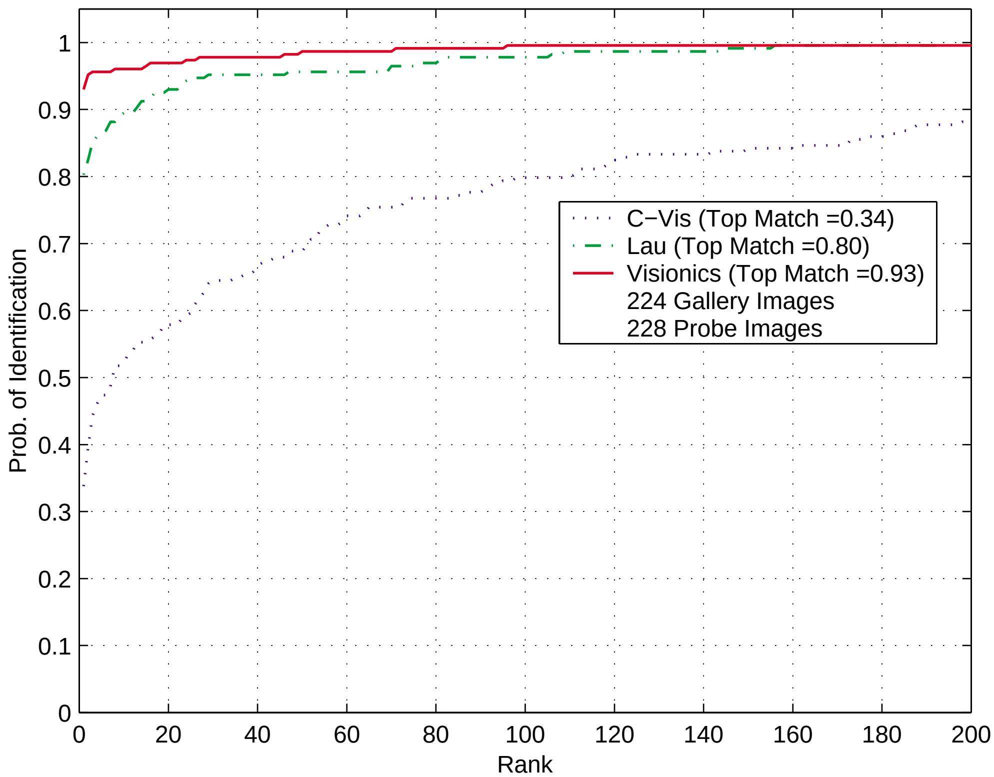


Figure M−21: Identification Scores: Illumination I1

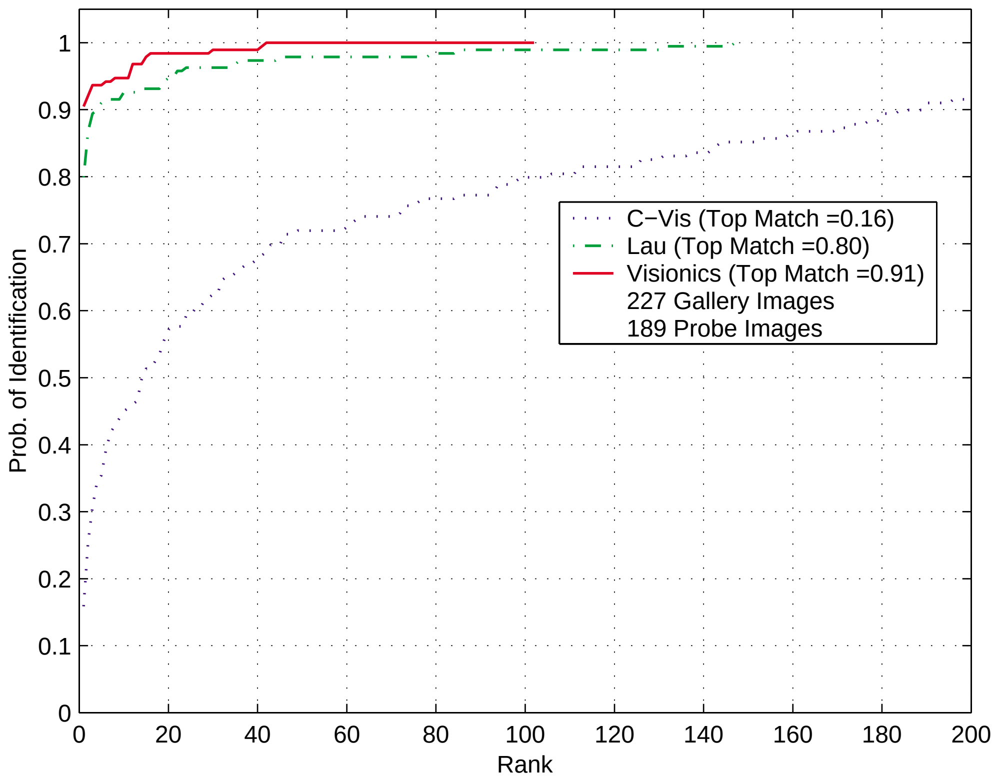


Figure M−22: Identification Scores: Illumination I2

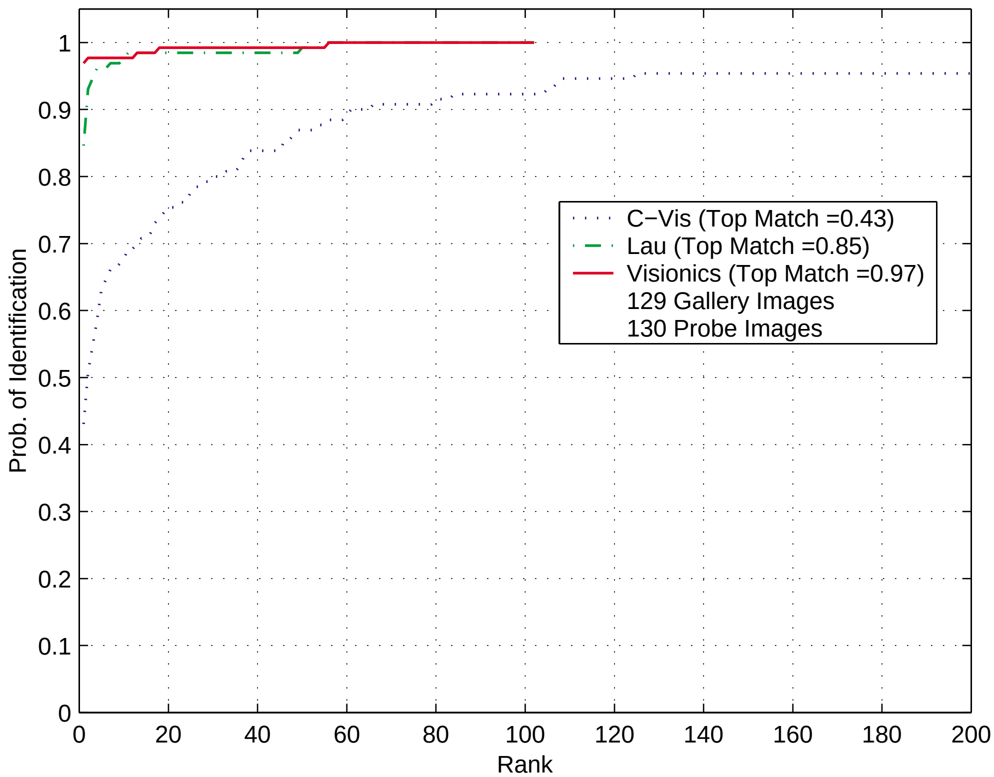


Figure M−23: Identification Scores: Illumination I3

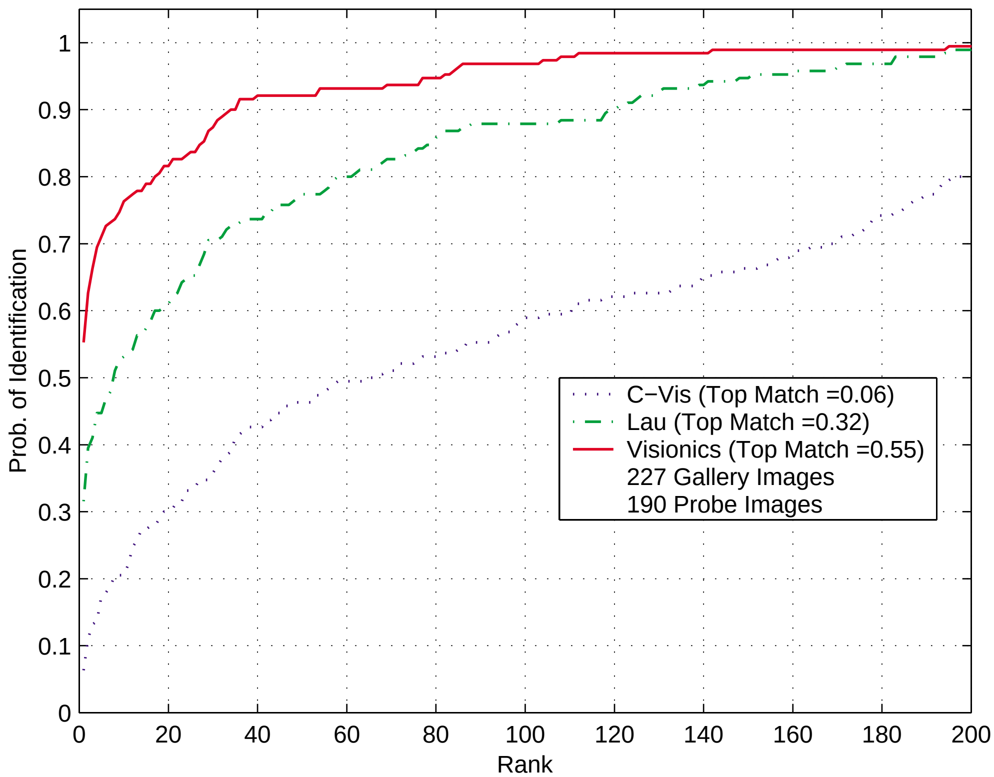


Figure M−24: Identification Scores: Media M1

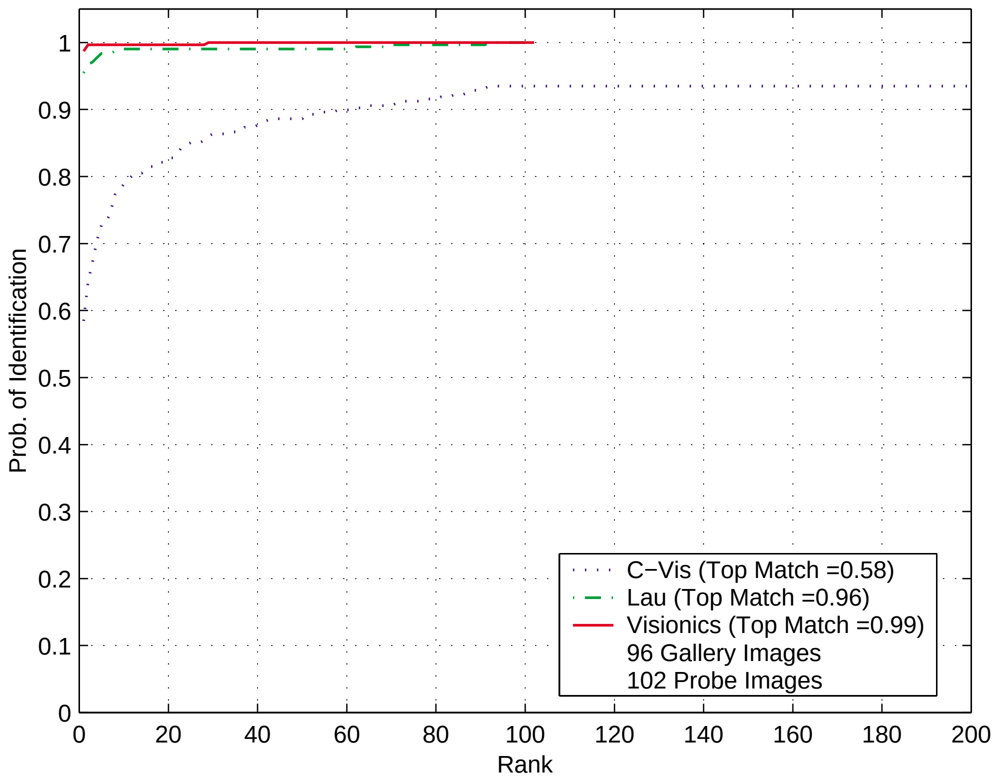


Figure M−25: Identification Scores: Media M2

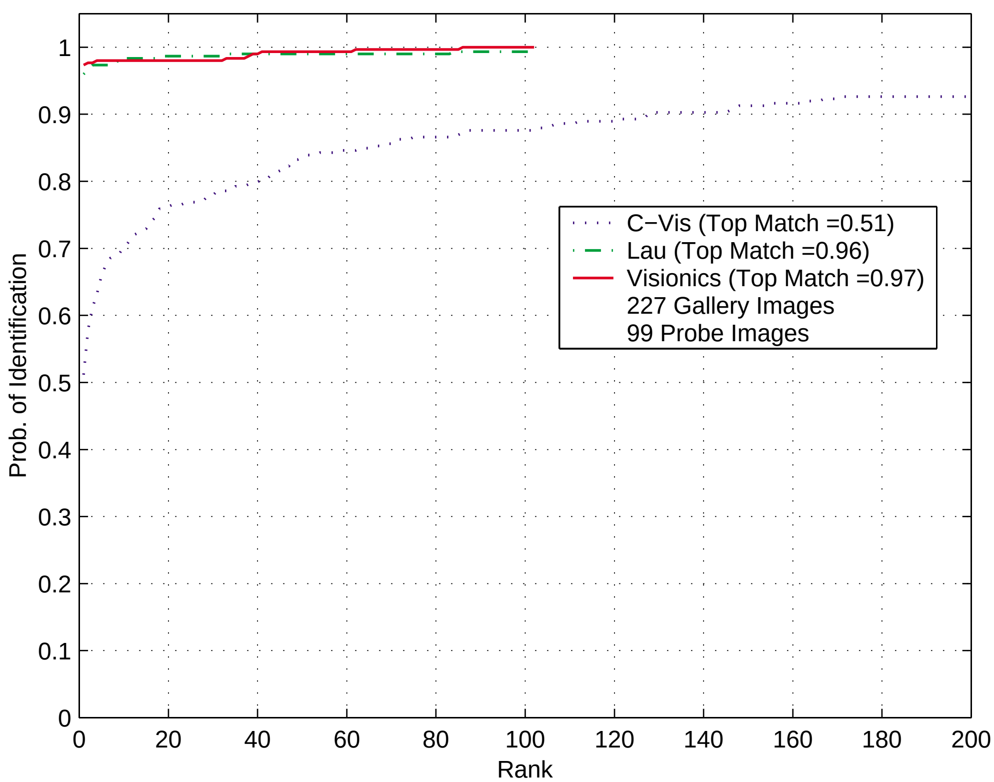


Figure M−26: Identification Scores: Pose P5

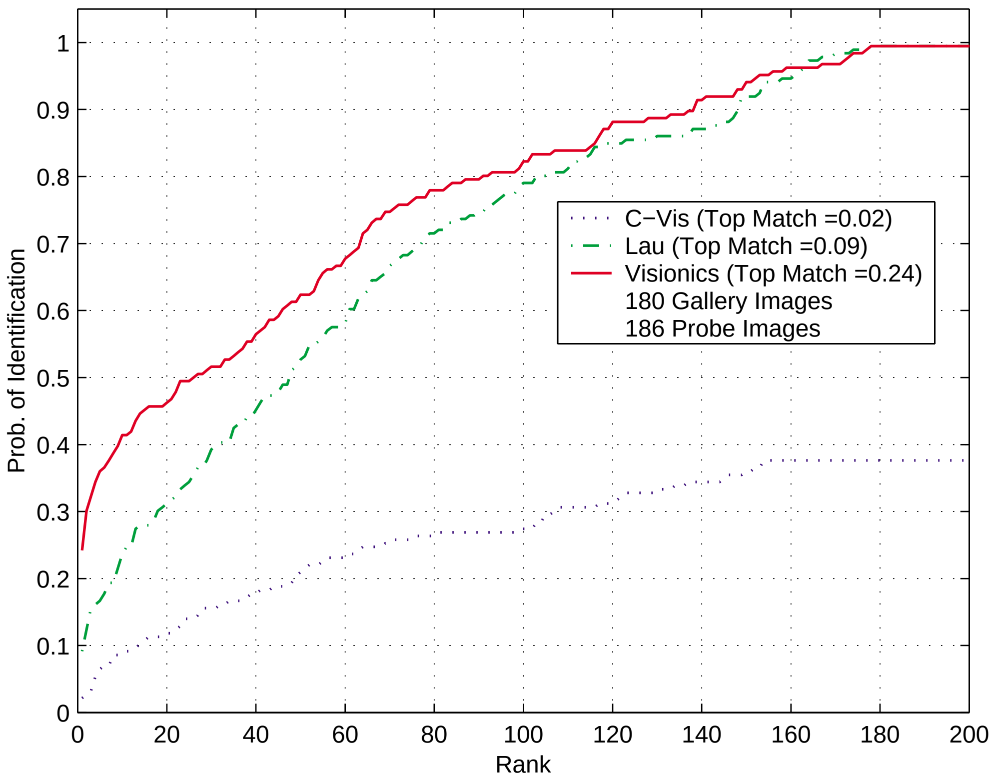


Figure M−27: Identification Scores: Resolution R1

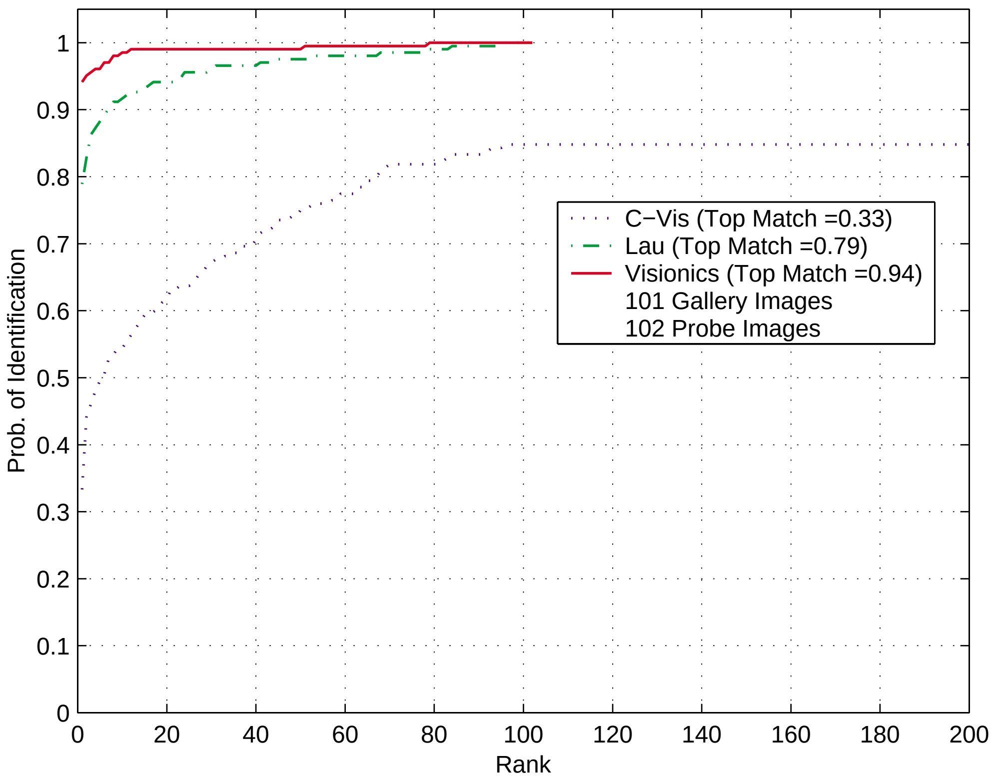


Figure M−28: Identification Scores: Resolution R2

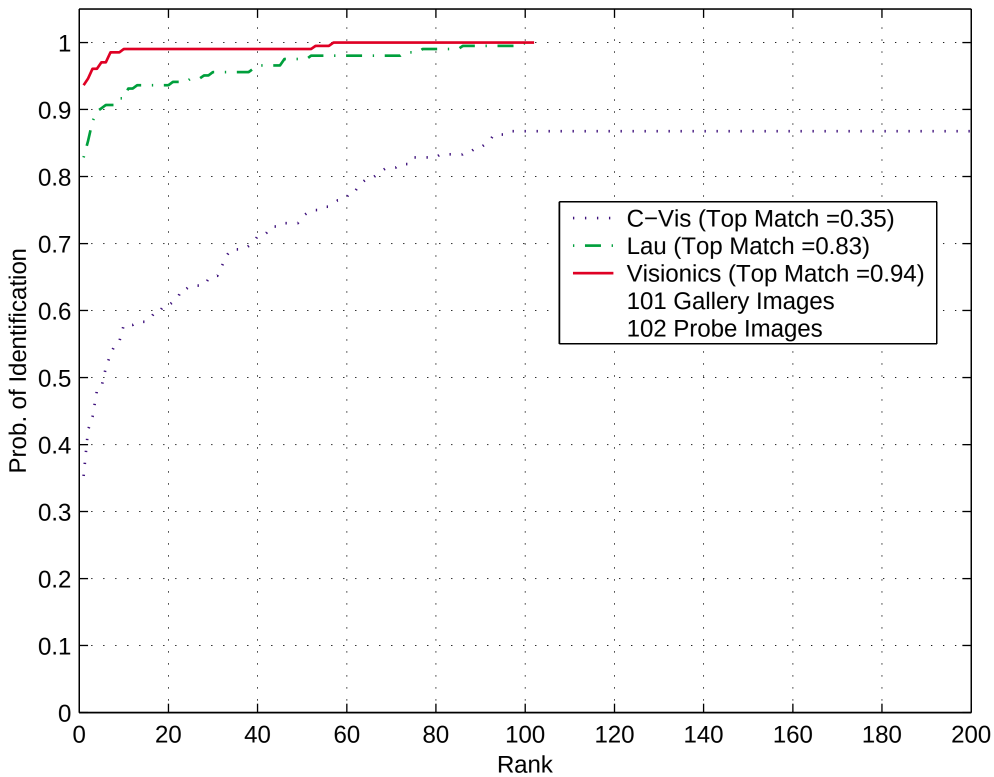


Figure M−29: Identification Scores: Resolution R3

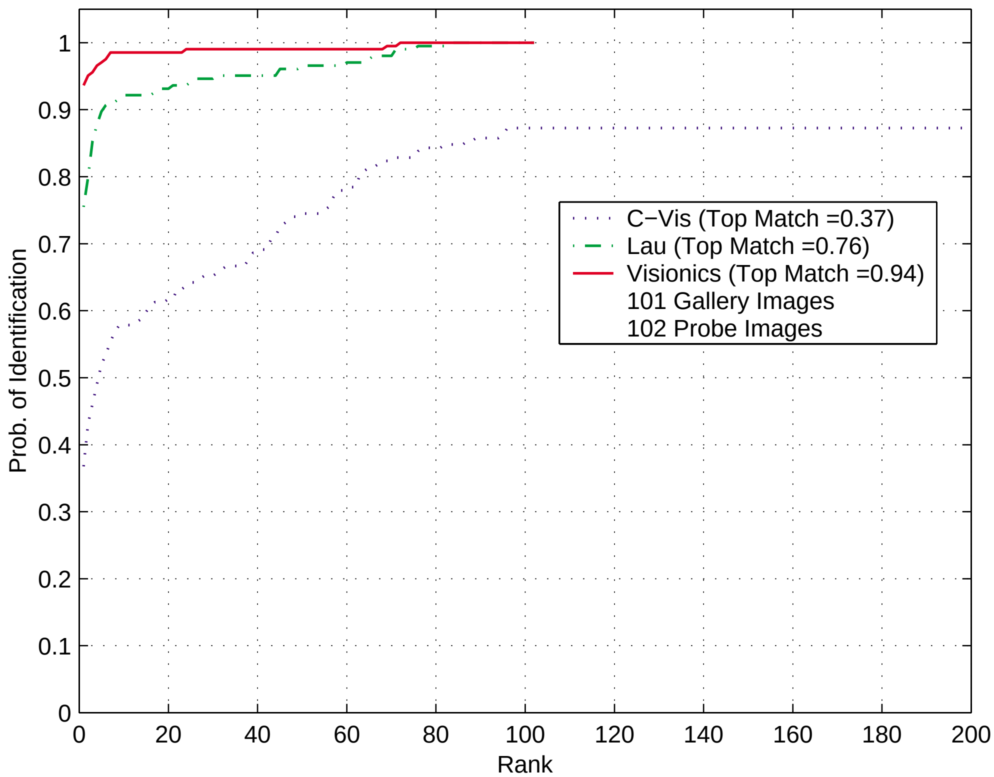


Figure M−30: Identification Scores: Resolution R4

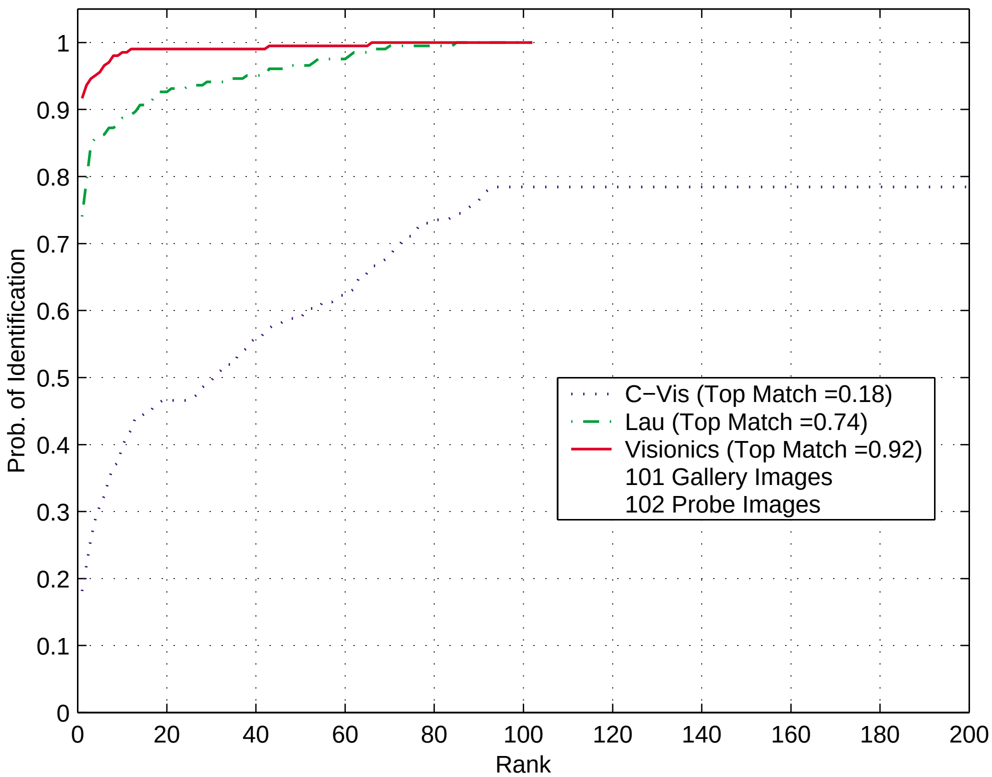


Figure M−31: Identification Scores: Temporal T3

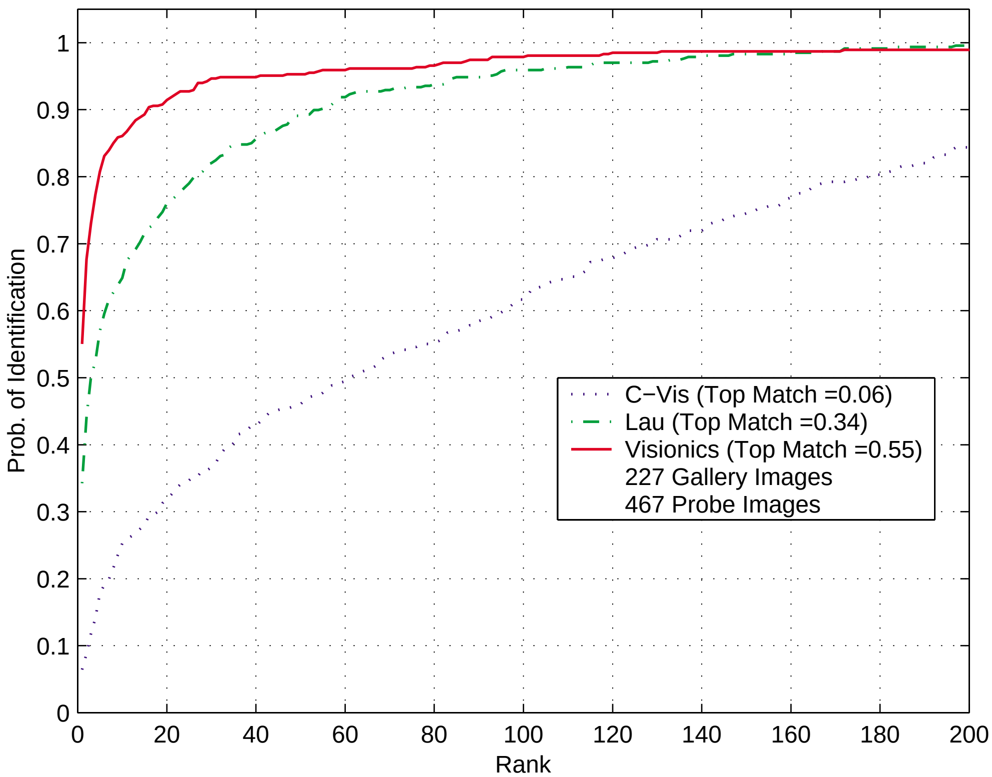


Figure M−32: Identification Scores: Temporal T4

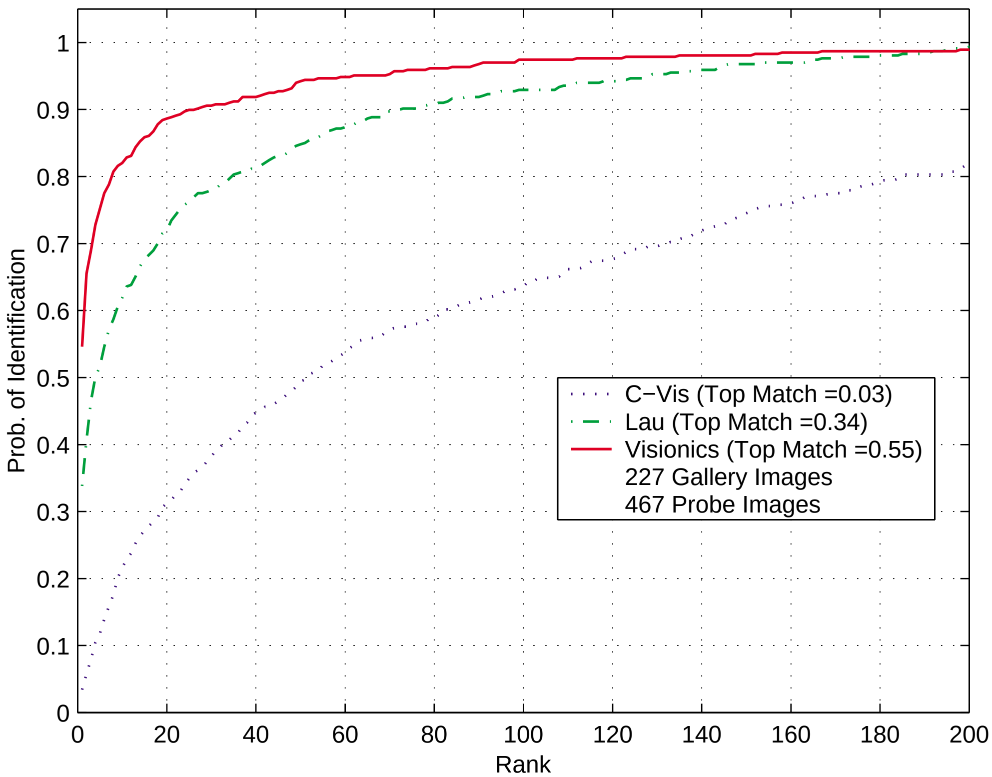


Figure M−33: Identification Scores: Temporal T5

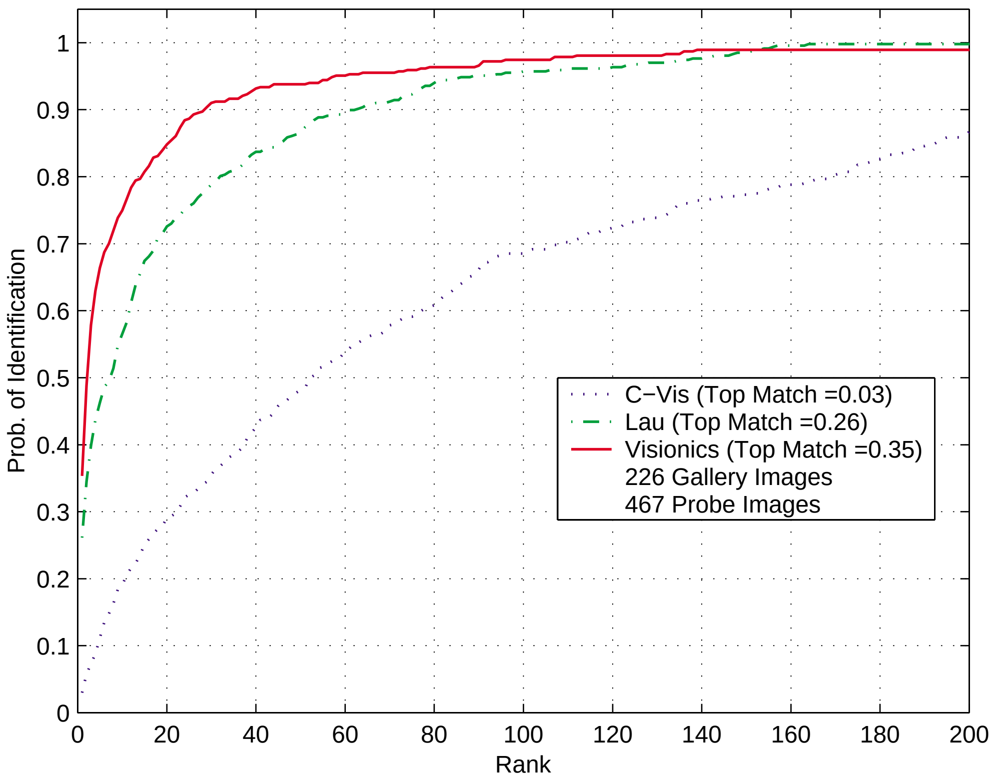


Figure M−34: Verification Scores: Distance D1

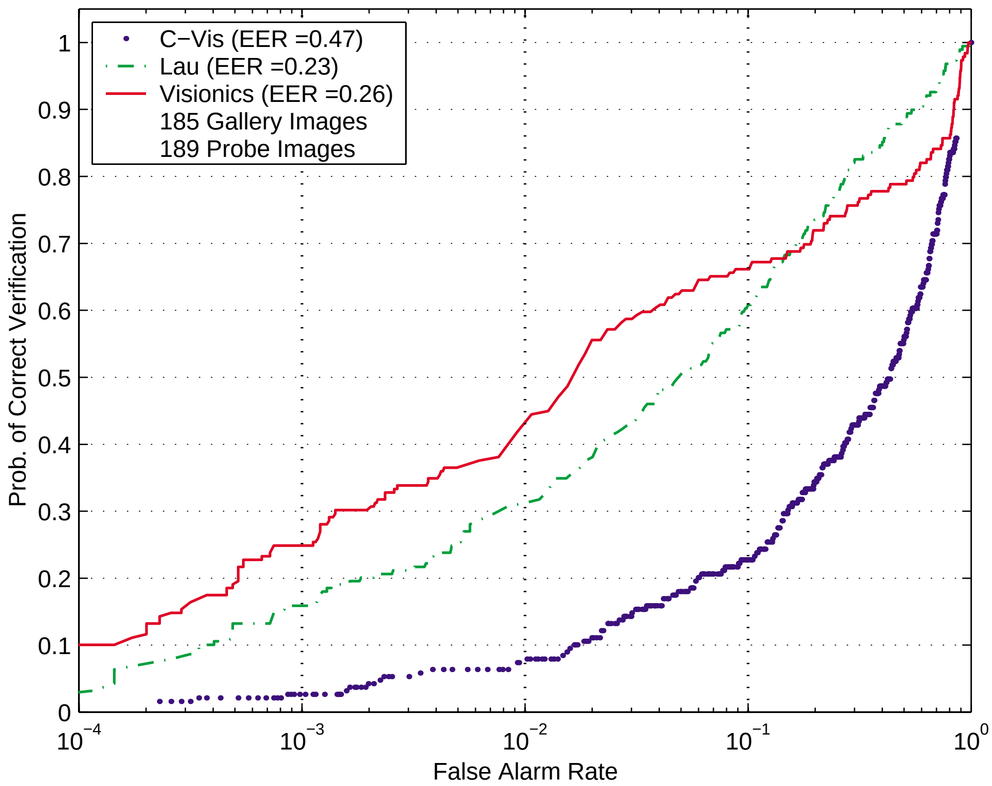


Figure M−35: Verification Scores: Distance D2

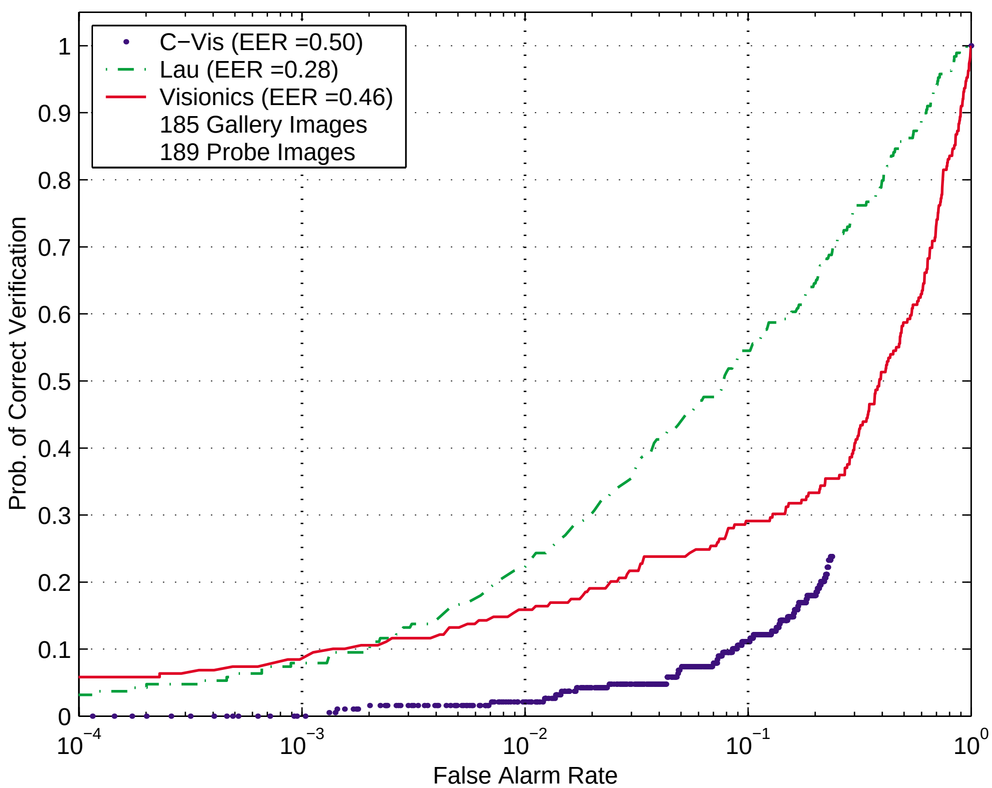


Figure M−36: Verification Scores: Distance D3

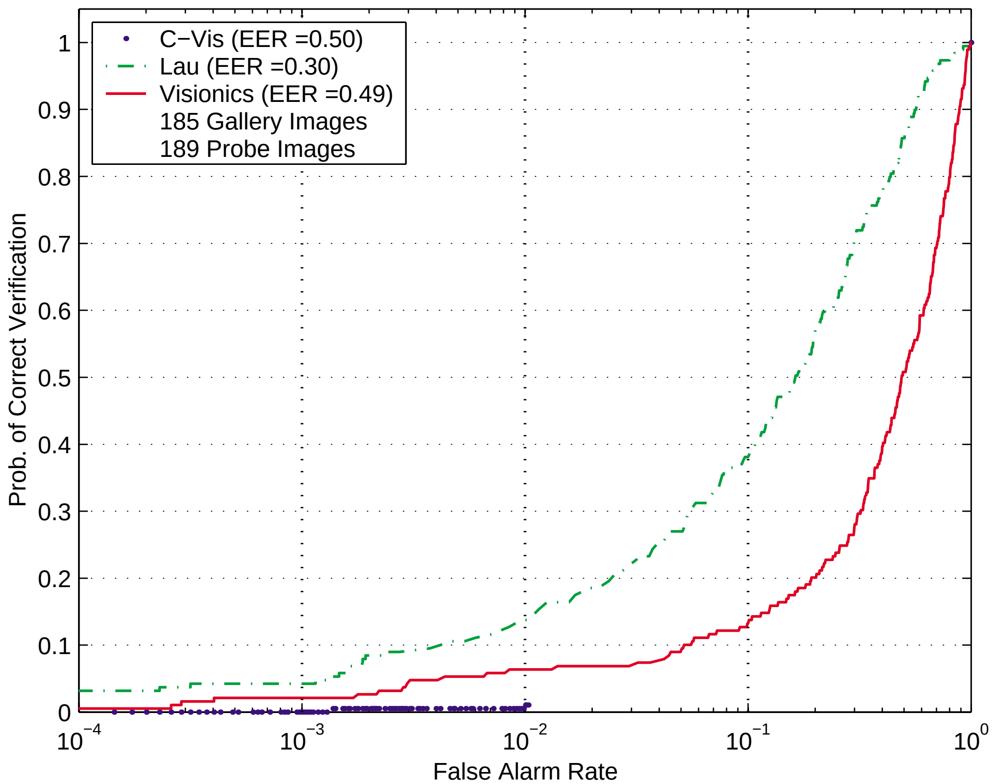


Figure M−37: Verification Scores: Distance D4

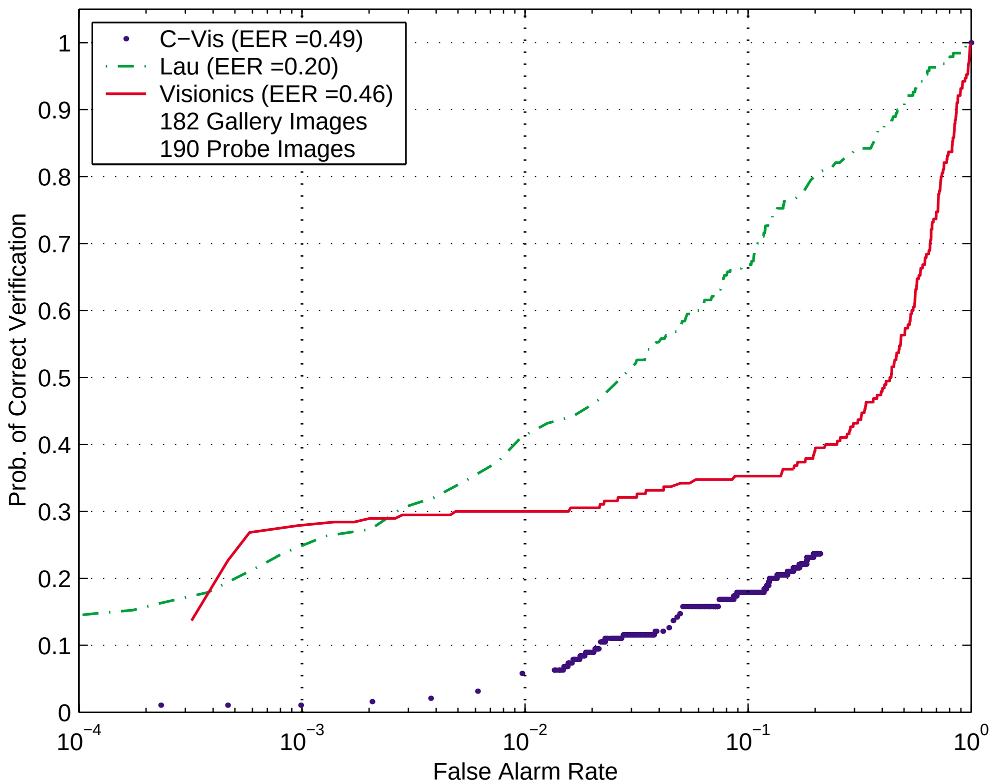


Figure M−38: Verification Scores: Distance D5

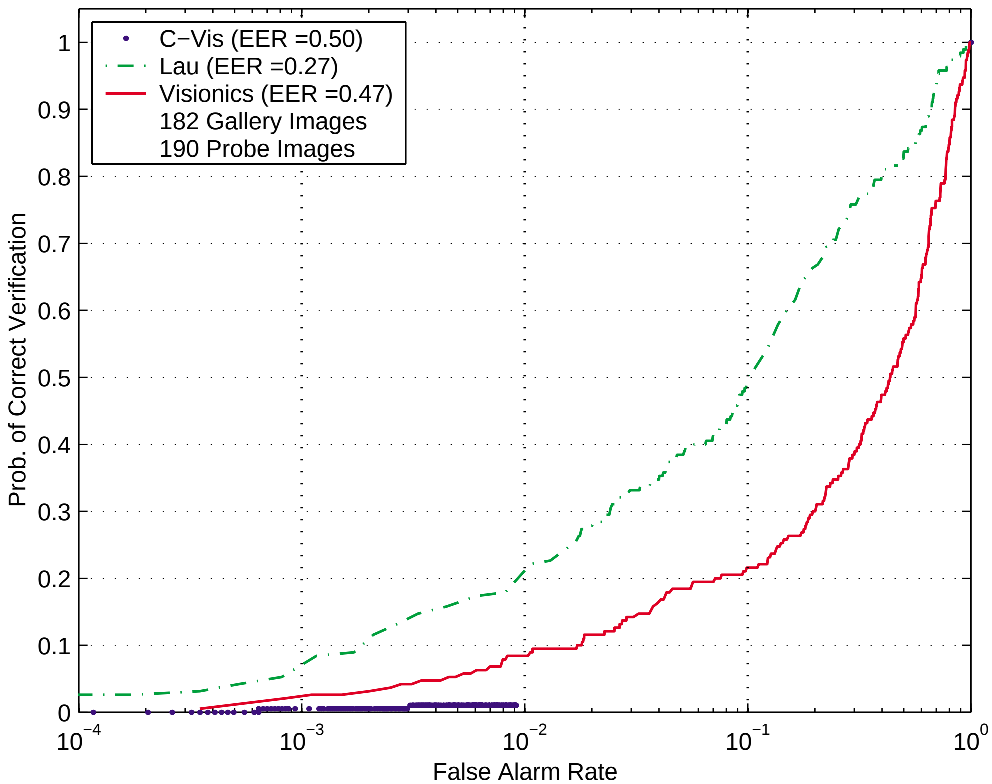


Figure M−39: Verification Scores: Distance D6

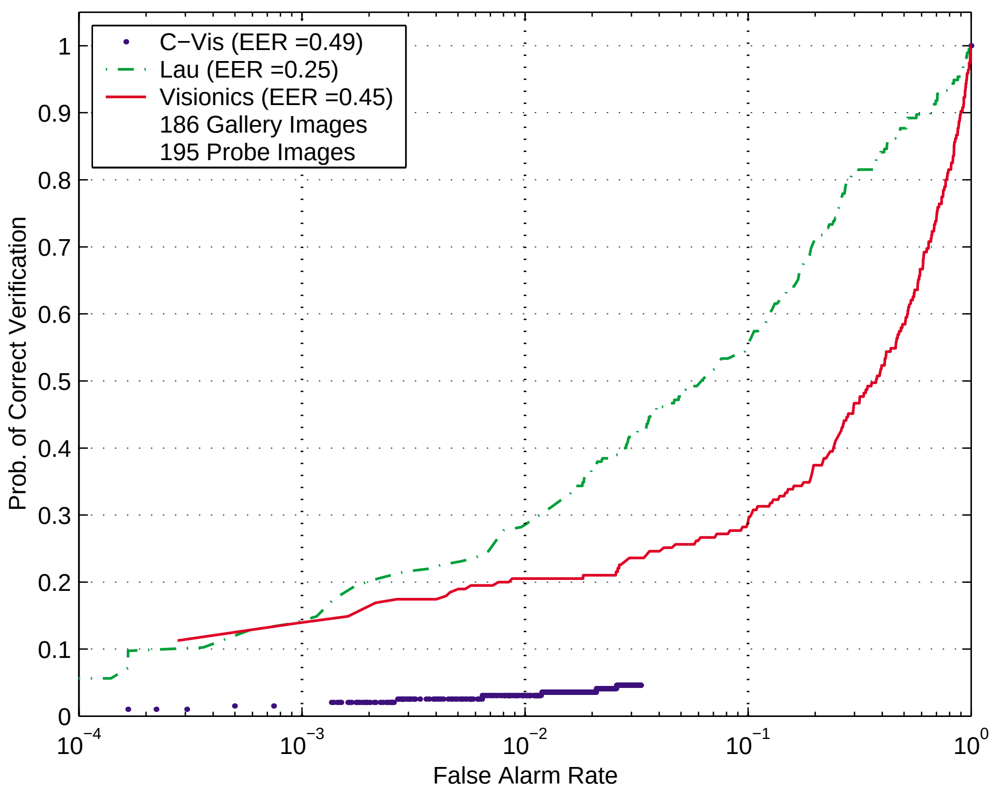


Figure M−40: Verification Scores: Distance D7

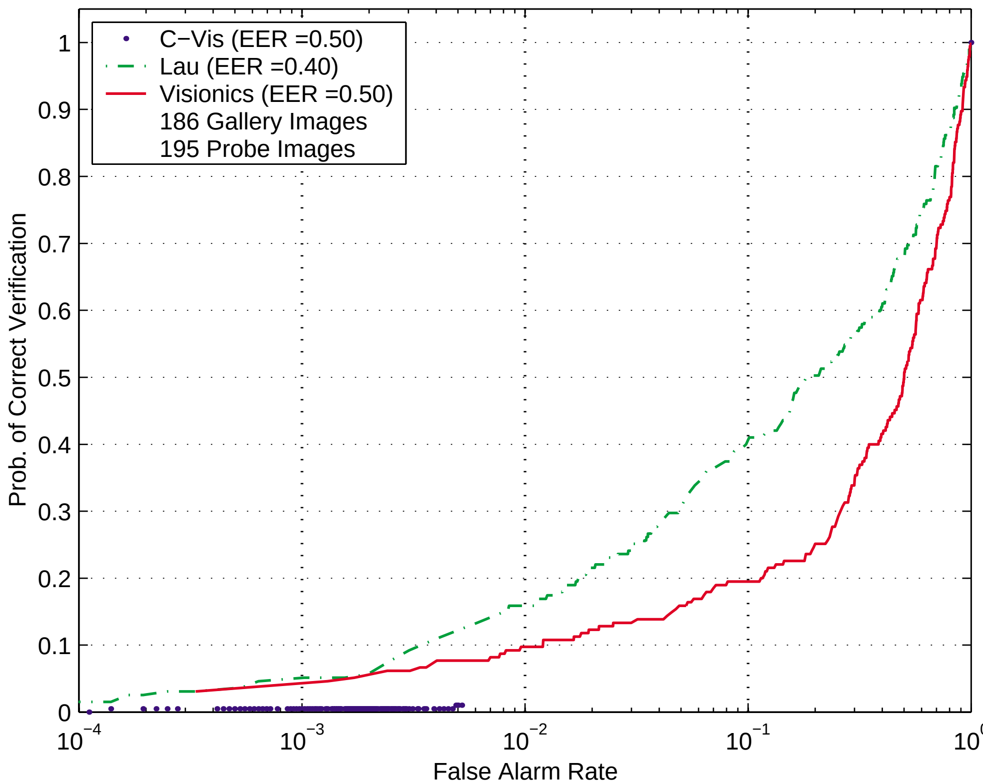


Figure M−41: Verification Scores: Expression E1

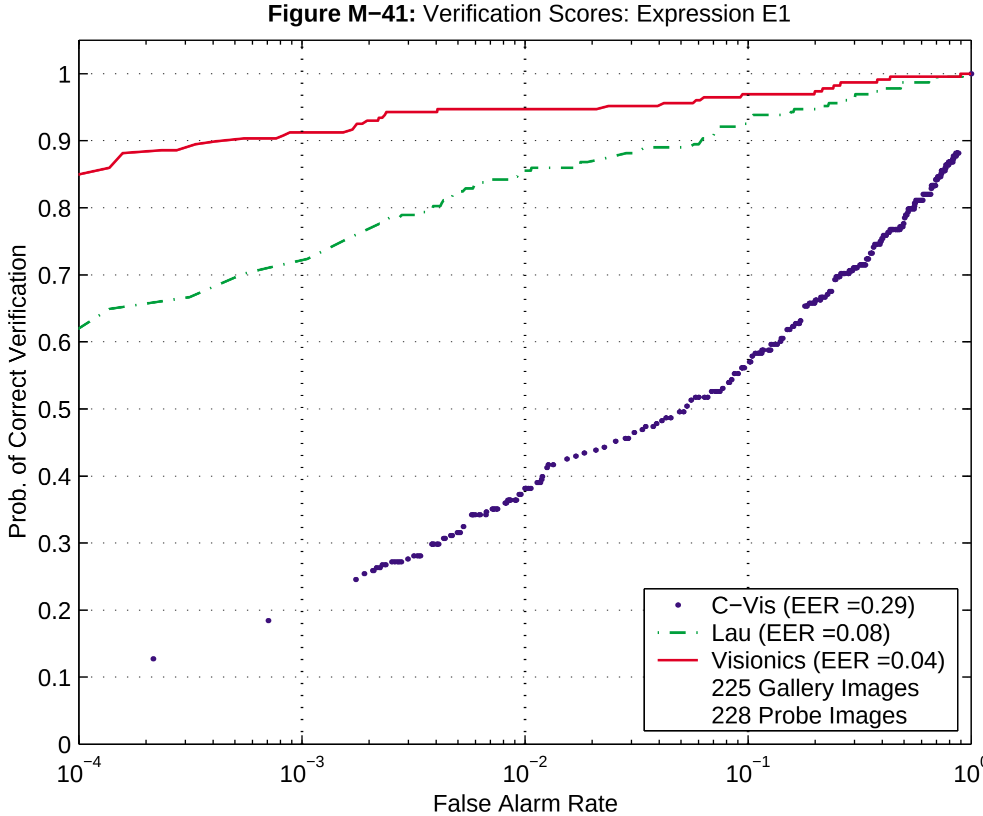


Figure M−42: Verification Scores: Expression E2

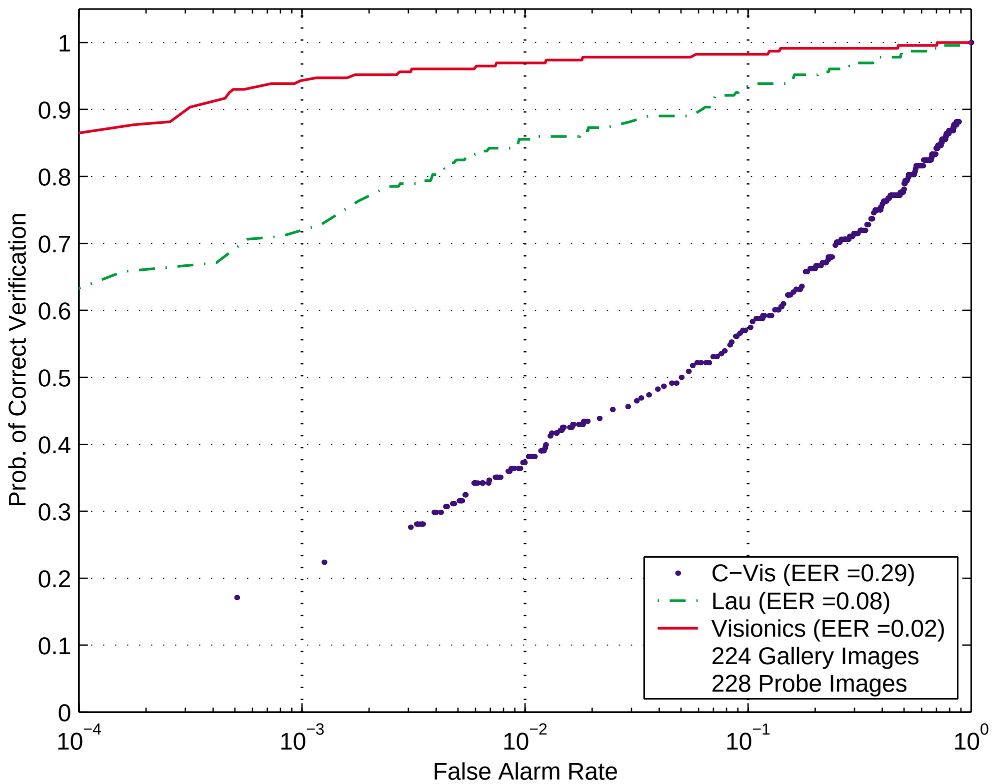


Figure M−43: Verification Scores: Illumination I1

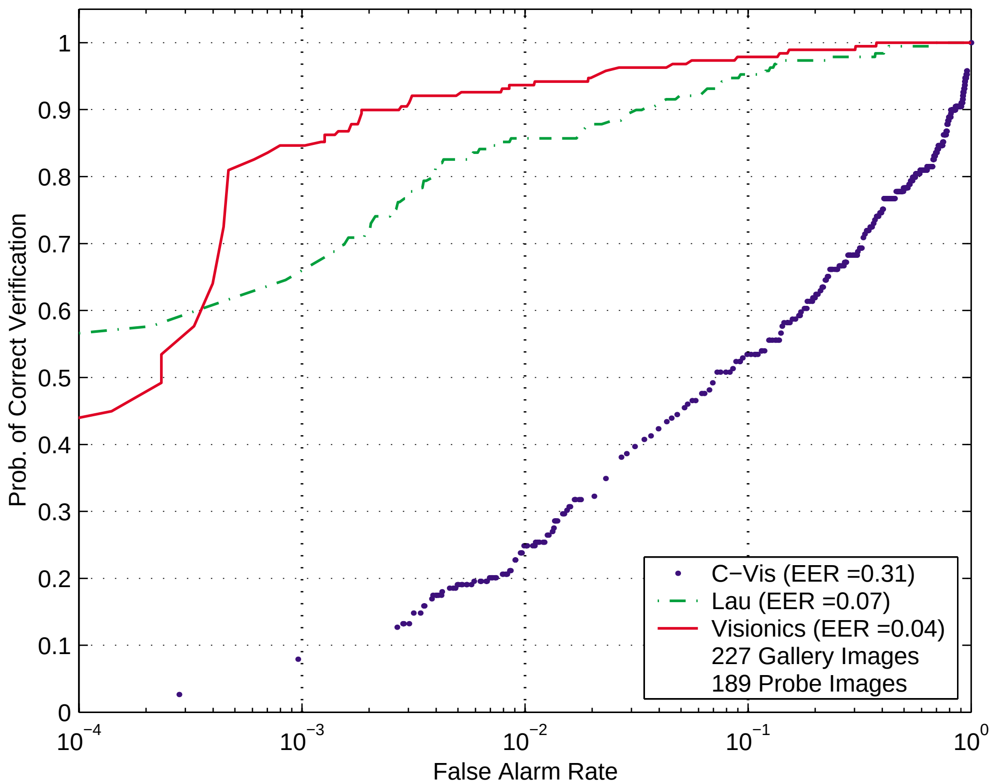


Figure M−44: Verification Scores: Illumination I2

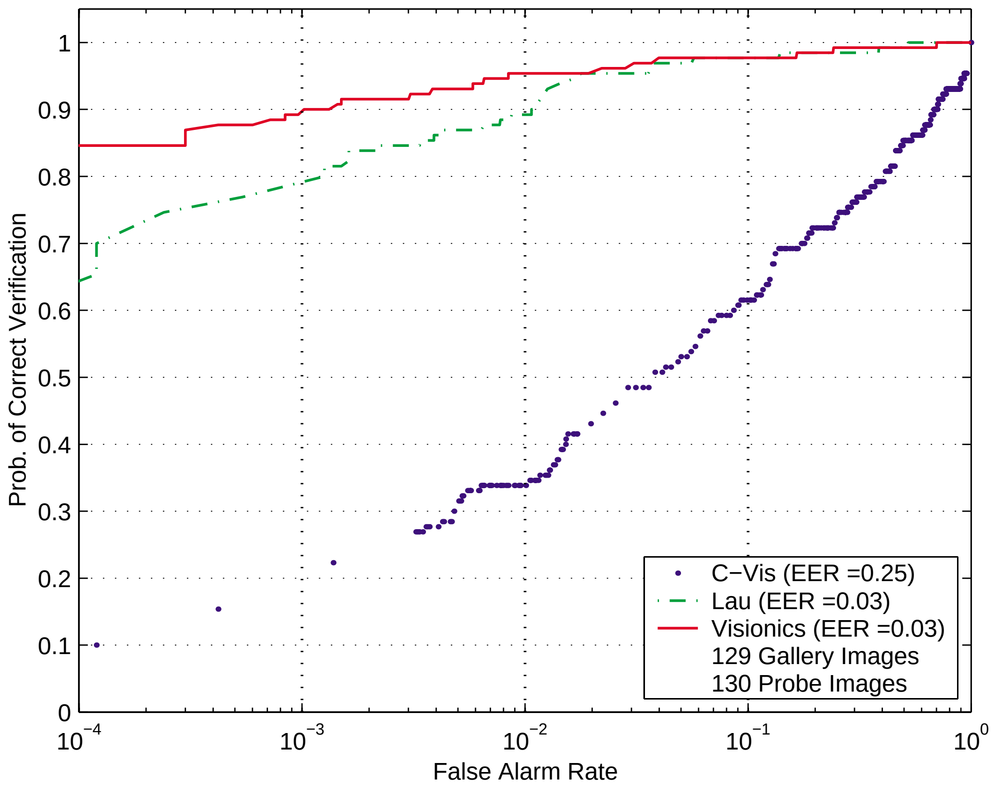


Figure M−45: Verification Scores: Illumination I3

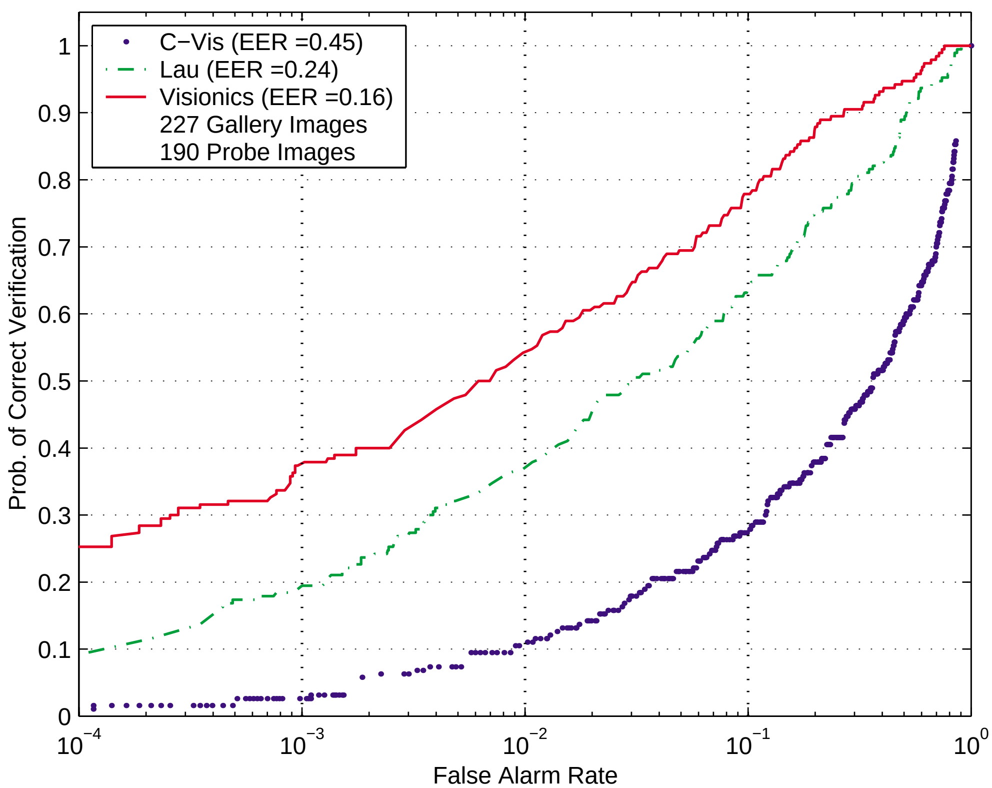


Figure M−46: Verification Scores: Media M1

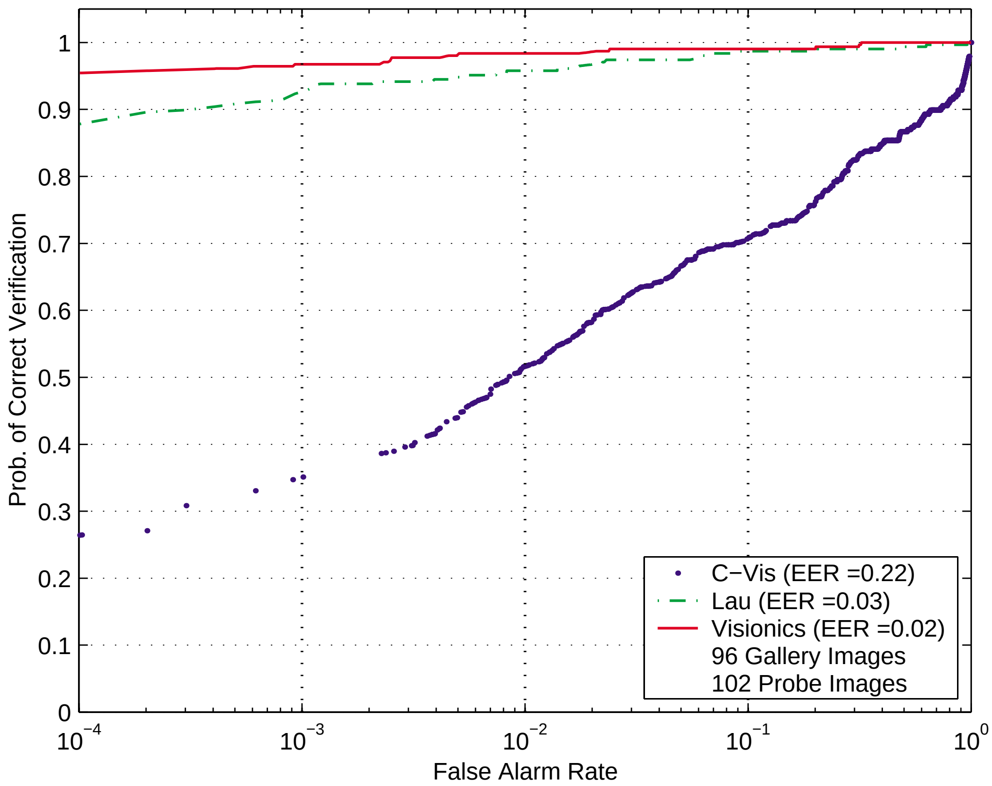
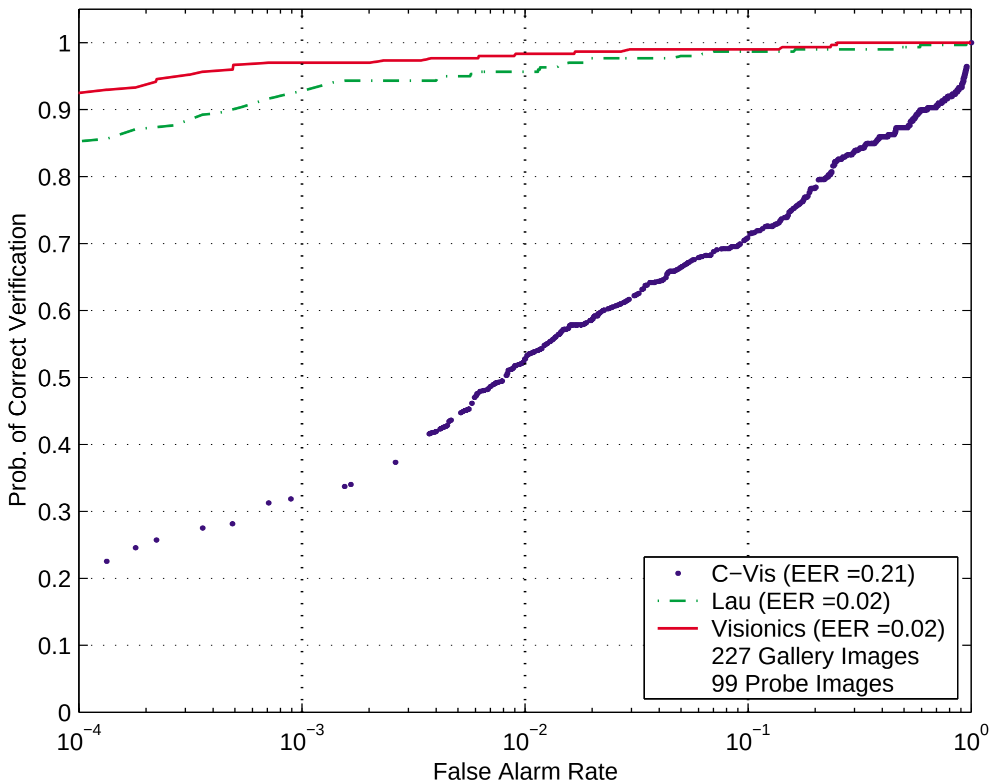


Figure M−47: Verification Scores: Media M2



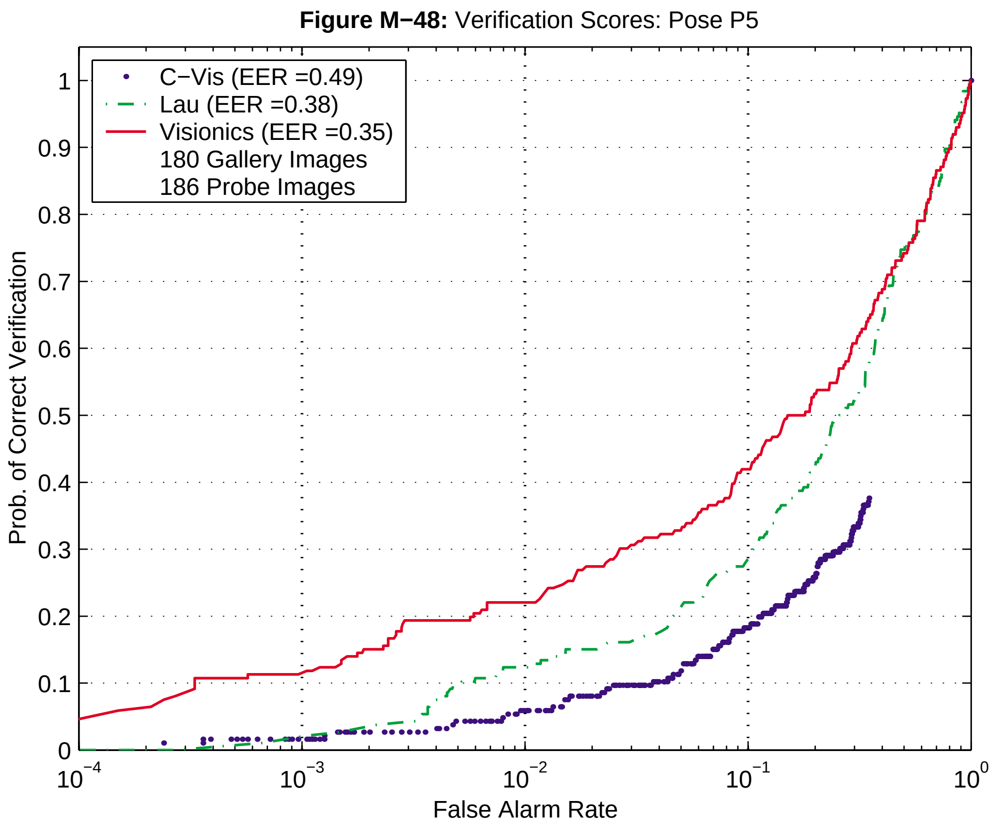


Figure M−48: Verification Scores: Pose P5
C−Vis (EER =0.49)
Lau (EER =0.38)
Visionics (EER =0.35)
180 Gallery Images
186 Probe Images
Prob. of Correct Verification
False Alarm Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10−4
10−3
10−2
10−1
10^0

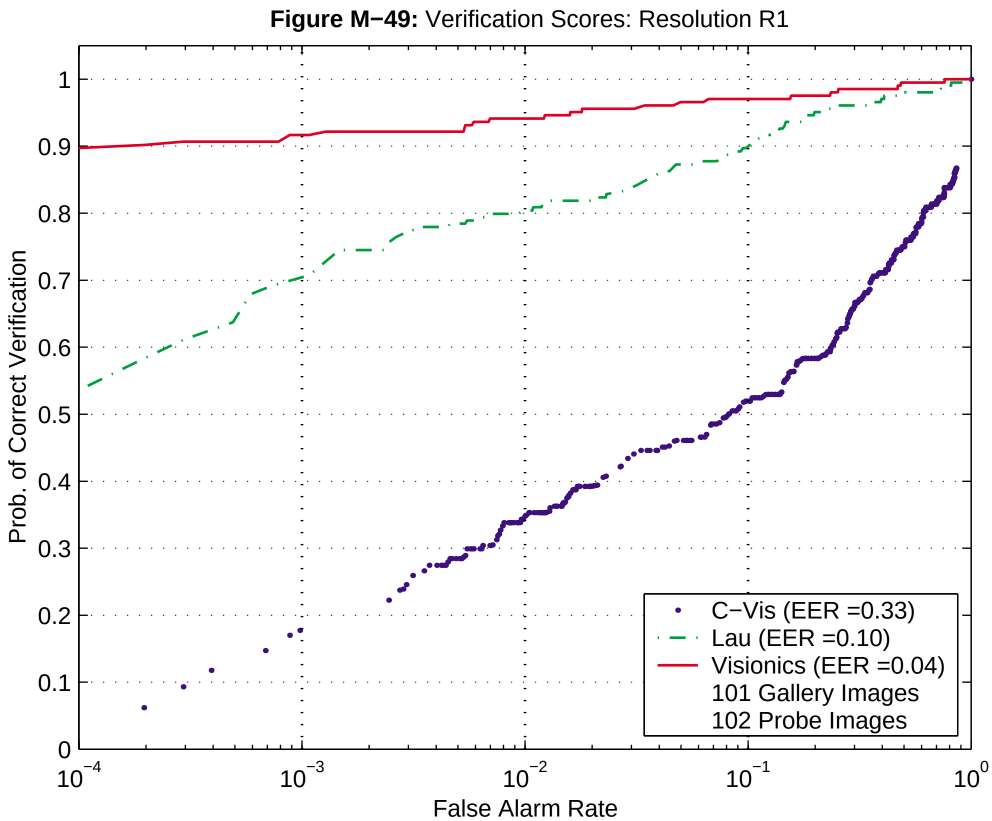


Figure M−49: Verification Scores: Resolution R1
C−Vis (EER =0.33)
Lau (EER =0.10)
Visionics (EER =0.04)
101 Gallery Images
102 Probe Images
Prob. of Correct Verification
False Alarm Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10−4
10−3
10−2
10−1
10^0

Figure M−50: Verification Scores: Resolution R2

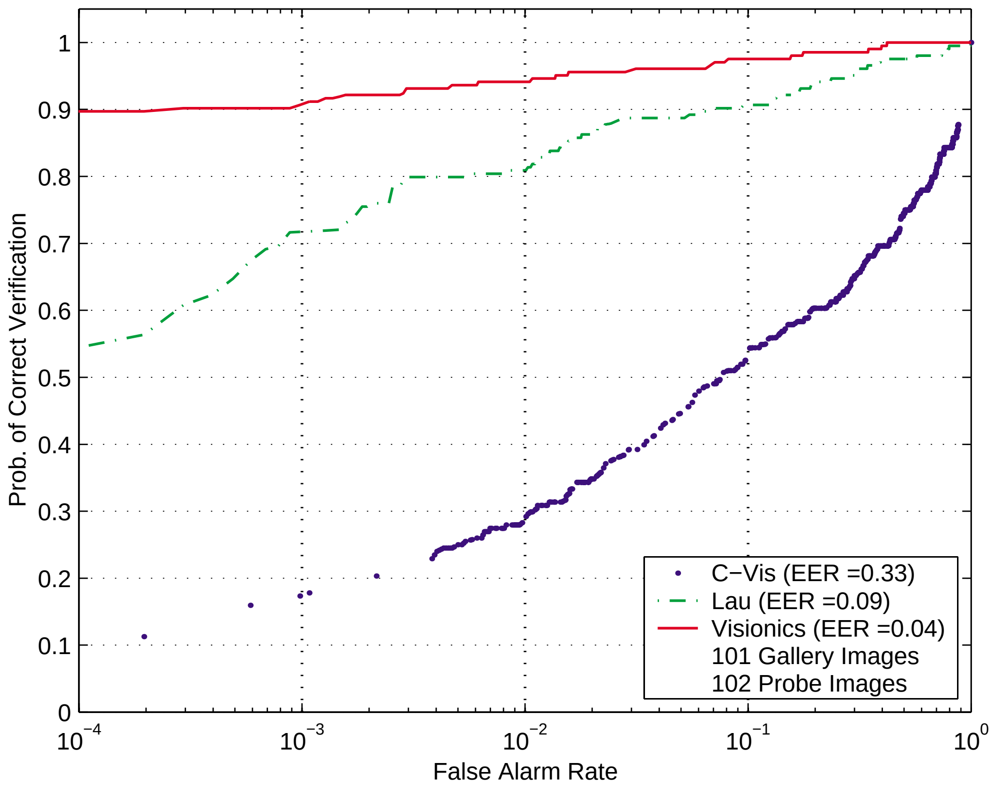


Figure M−51: Verification Scores: Resolution R3

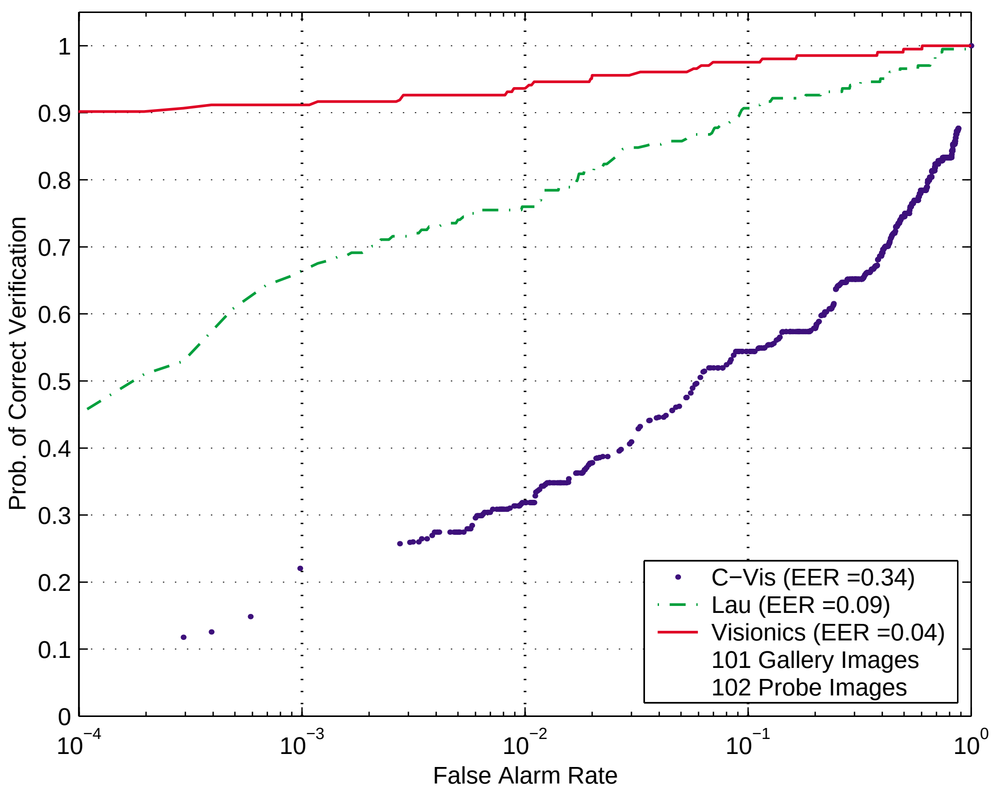


Figure M−52: Verification Scores: Resolution R4

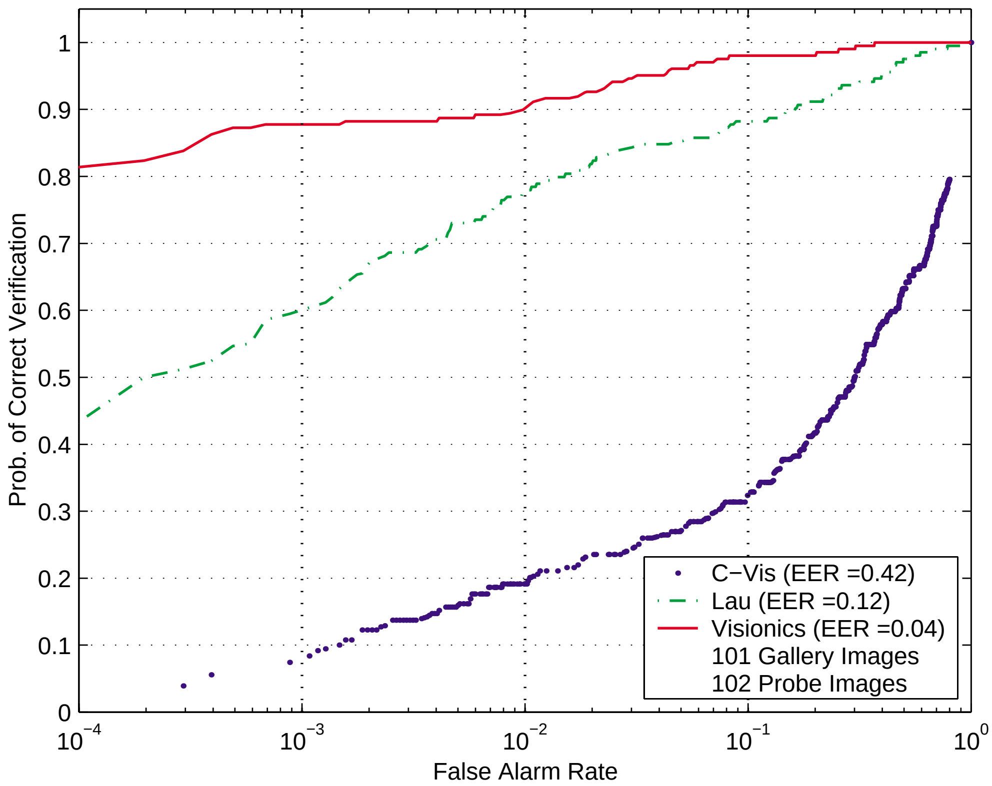
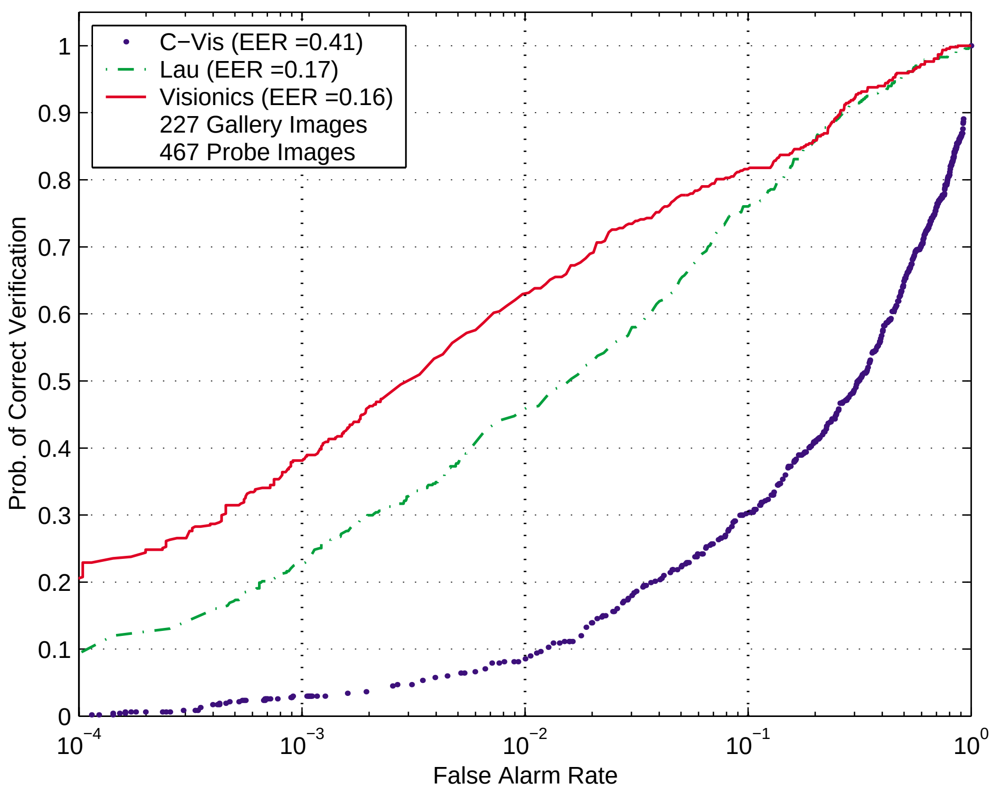


Figure M−53: Verification Scores: Temporal T3



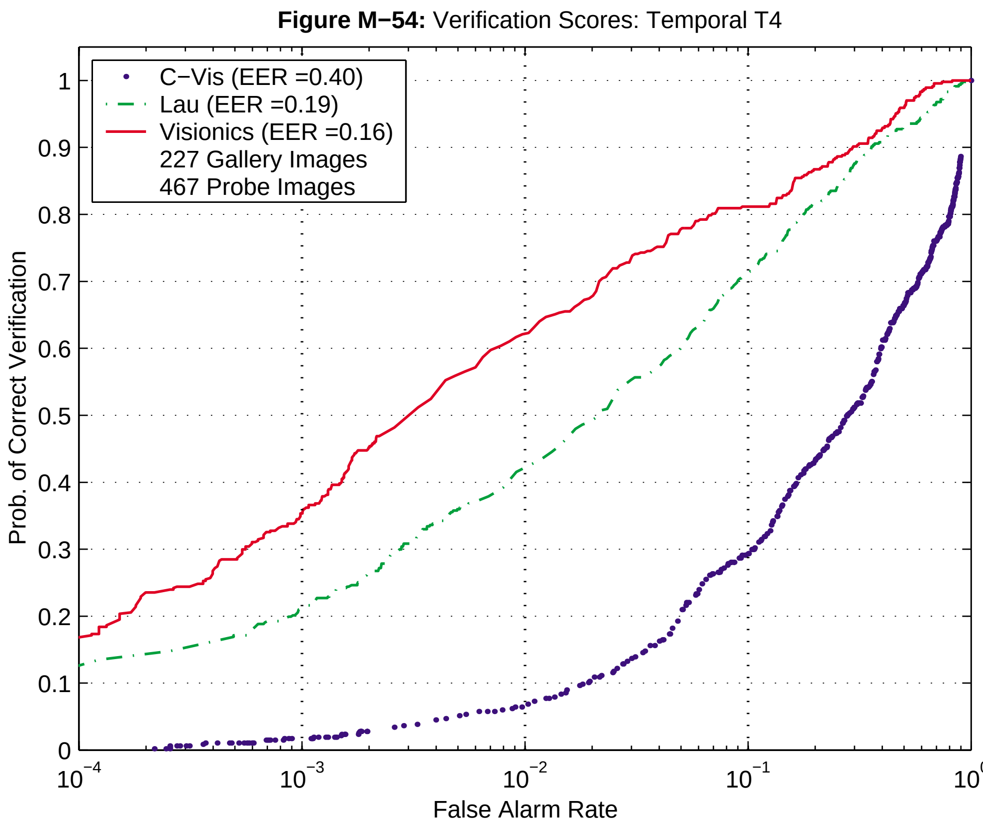


Figure M−54: Verification Scores: Temporal T4
C−Vis (EER =0.40)
Lau (EER =0.19)
Visionics (EER =0.16)
227 Gallery Images
467 Probe Images
Prob. of Correct Verification
False Alarm Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

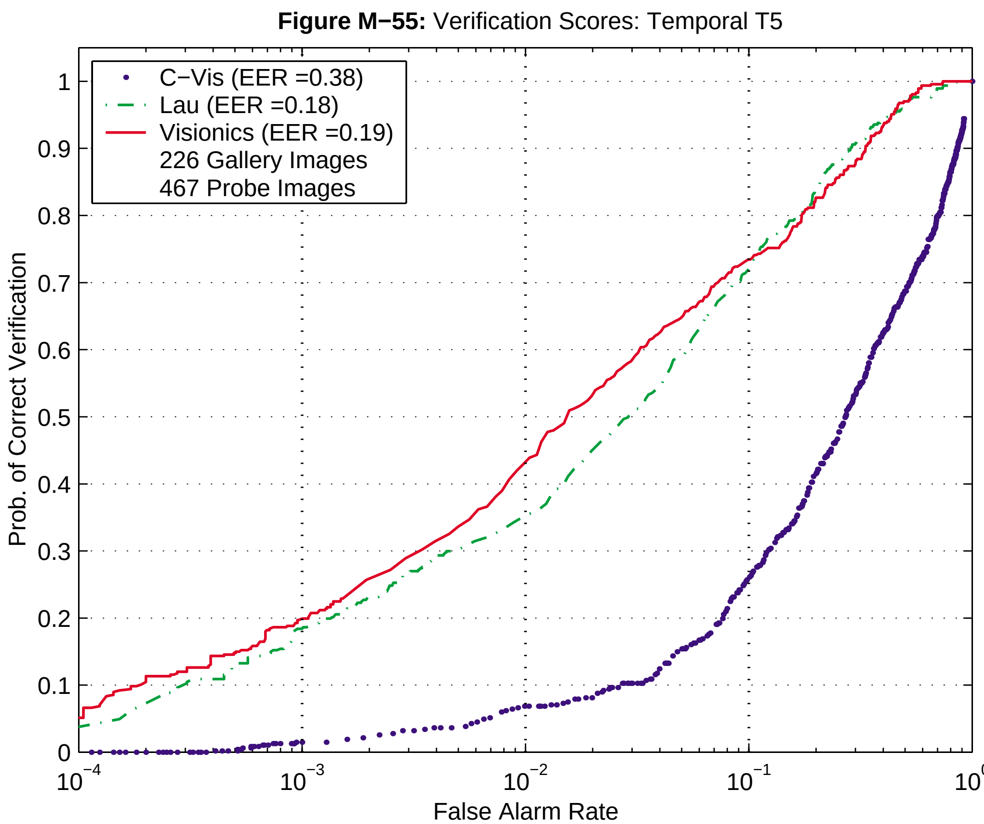


Figure M−55: Verification Scores: Temporal T5
C−Vis (EER =0.38)
Lau (EER =0.18)
Visionics (EER =0.19)
226 Gallery Images
467 Probe Images
Prob. of Correct Verification
False Alarm Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

Appendix H
FRVT 2000 Product Usability Test Results

Banque-Tec Old Image Database Timed Test Results Verification Mode Subjects Walk Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	1	7.53	No	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
2	Cooperative	12	1	6.41	No	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	9.95	No
	Indifferent	12	1	X	X	1	X	X
		8	1	5.19	No	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	3.83	No	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	4	4.87	No	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	8.37	Yes	1	9.55	Yes
		8	3	3.77	Yes	1	X	X
		4	1	X	X	1	3.35	Yes

C-Vis Old Image Database Timed Test Results Verification Mode Subjects Wal k Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	1	10.00	No	3	6.62	No
		8	1	6.63	No	1	5.90	No
		4	1	7.69	No	1	5.44	No
	Indifferent	12	1	X	X	3	7.45	No
		8	1	X	X	1	8.26	No
		4	1	6.73	No	1	6.12	No
2	Cooperative	12	1	X	X	7	4.09	No
		8	1	8.76	No	1	8.33	No
		4	1	5.09	No	1	5.18	No
	Indifferent	12	1	8.26	No	3	7.04	No
		8	1	9.46	No	1	9.13	No
		4	1	8.14	No	1	6.48	No
3	Cooperative	12	1	9.33	No	1	10.00	No
		8	1	8.13	No	1	5.44	No
		4	1	5.58	No	1	5.09	No
	Indifferent	12	1	X	X	2	8.00	No
		8	1	10.00	No	1	7.15	No
		4	1	7.37	No	1	7.58	No
1 Variability Test	Cooperative	12	3	5.66	No	4	5.51	No
		12	2	8.00	No	3	5.54	No
		12	4	5.95	No	3	5.42	No
		12	4	4.92	No	5	3.95	No
		12	3	6.56	No	1	8.28	No
		12	3	6.58	No	3	5.78	No
		12	2	6.53	No	4	5.47	No
		12	3	6.19	No	1	9.75	No
1 Photo Test	Cooperative	12	6	4.09	No	7	3.74	No
		8	2	5.50	No	4	3.80	No
		4	1	4.02	No	1	5.58	No

Lau Technologies Old Image Database Timed Test Results Verification Mode Subjects Walk Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Miros Old Image Database Timed Test Results Verification Mode Subjects Walk Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	5	3.54	Yes	7	3.07	Yes
		8	5	1.70	Yes	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	9	2.38	Yes	1	X	X
		8	1	X	X	1	X	X
		4	3	2.20	Yes	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	7	3.16	Yes	9	2.78	Yes
		12	8	2.75	Yes	6	3.71	Yes
		12	8	3.22	Yes	9	2.83	Yes
		12	7	3.80	Yes	1	X	X
		12	7	3.65	Yes	6	4.06	Yes
		12	8	2.93	Yes	6	4.94	Yes
		12	6	4.90	Yes	7	3.20	Yes
		12	5	5.85	Yes	6	5.09	Yes
1 Photo Test	Cooperative	12	5	6.03	Yes	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Visionics Corporation
Old Image Database Timed Test Results
Verification Mode
Subjects Walk Toward Camera

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	7.49	Yes	4	4.68	Yes
		8	1	8.47	Yes	1	X	X
		4	1	4.16	Yes	1	4.19	Yes
	Indifferent	12	2	6.28	Yes	5	4.42	Yes
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	2	7.40	Yes	6	4.51	Yes
		8	1	X	X	1	X	X
		4	1	6.31	Yes	1	5.04	Yes
	Indifferent	12	6	4.47	Yes	4	4.99	Yes
		8	1	X	X	1	9.64	Yes
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	8.81	Yes	1	7.23	Yes
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	4.23	Yes	1	X	X
1 Variability Test	Cooperative	12	4	4.44	Yes	4	4.88	Yes
		12	1	8.51	Yes	7	3.06	Yes
		12	2	6.47	Yes	8	3.26	Yes
		12	4	5.05	Yes	8	3.02	Yes
		12	7	3.79	Yes	8	3.46	Yes
		12	6	4.58	Yes	8	3.12	Yes
		12	1	8.99	Yes	6	4.78	Yes
		12	2	6.89	Yes	8	3.28	Yes
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	2	5.24	Yes	4	3.58	Yes
		4	1	4.46	Yes	1	5.01	Yes

Banque-Tec
Old Image Database Timed Test Results
Identification Mode
Subjects Walk Toward Camera

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	4.03	No	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	10.00	No
	Indifferent	12	1	X	X	1	X	X
		8	1	8.50	No	1	X	X
		4	1	8.87	No	1	10.00	No
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	3	5.14	No	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	8.11	No	1	X	X
		8	4	3.78	No	1	X	X
		4	4	2.74	No	1	X	X

C-Vis Old Image Database Timed Test Results Identification Mode Subjects Walk Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	3	6.31	No	2	7.91	No
		8	4	3.44	No	1	8.02	No
		4	2	3.91	No	1	7.40	No
	Indifferent	12	6	4.54	No	3	7.12	No
		8	4	3.37	No	1	X	X
		4	2	4.29	No	1	4.69	No
2	Cooperative	12	5	5.44	No	1	8.35	No
		8	1	10.00	No	1	6.66	No
		4	1	5.03	No	1	4.87	No
	Indifferent	12	6	4.85	No	2	8.38	No
		8	1	5.93	No	1	6.49	No
		4	1	6.57	No	1	7.29	No
3	Cooperative	12	1	8.33	No	7	3.89	No
		8	1	8.05	No	1	9.43	No
		4	1	5.86	No	1	8.74	No
	Indifferent	12	6	4.52	No	6	4.75	No
		8	1	9.64	No	1	9.99	No
		4	1	8.57	No	1	6.25	No
1 Variability Test	Cooperative	12	1	9.43	No	6	4.55	No
		12	1	8.56	No	6	4.50	No
		12	3	6.54	No	5	5.01	No
		12	2	7.16	No	5	5.14	No
		12	1	7.19	No	5	5.34	No
		12	3	6.31	No	3	6.67	No
		12	5	4.12	No	1	10.00	No
		12	1	10.00	No	1	8.59	No
1 Photo Test	Cooperative	12	6	5.24	No	7	4.52	No
		8	2	4.79	No	3	4.87	No
		4	1	4.26	No	1	6.92	No

Lau Technologies Old Image Database Timed Test Identification Mode Subjects Walk Toward Camera								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Miros
Old Image Database Timed Test Results
Identification Mode
Subjects Walk Toward Camera

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
1 Variability Test	Cooperative	12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
1 Photo Test	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X

Visionics Corporation
Old Image Database Timed Test
Identification Mode
Subjects Walk Toward Camera

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
	Indifferent	12	1	10.00	Yes	1	9.44	Yes
		8	1	X	X	4	4.52	Yes
		4	1	X	X	1	X	X
2	Cooperative	12	1	X	X	1	X	X
		8	1	7.19	Yes	1	10.00	Yes
		4	1	X	X	1	X	X
	Indifferent	12	1	X	X	1	X	X
		8	1	X	X	1	X	X
		4	1	X	X	1	X	X
3	Cooperative	12	1	X	X	1	X	X
		8	1	X	X	1	X	Yes
		4	1	5.97	Yes	1	X	X
	Indifferent	12	1	10.00	Yes	1	X	X
		8	1	9.51	Yes	1	X	X
		4	2	5.66	Yes	1	4.82	Yes
1 Variability Test	Cooperative	12	1	8.18	Yes	1	X	X
		12	1	X	X	3	6.49	Yes
		12	1	X	X	3	6.87	Yes
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	1	X	X
		12	1	X	X	3	7.14	Yes
		12	1	8.01	Yes	1	X	X
1 Photo Test	Cooperative	12	1	X	X	3	7.48	Yes
		8	1	X	X	1	X	X
		4	1	9.52	Yes	1	X	X

Banque-Tec
Enrollment Timed Test Results
Verification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	X	X	12	X	X
		8	8	7.95	No	8	X	X
		4	4	1.47	Yes	4	1.29	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	3.23	Yes	8	3.02	Yes
		4	4	6.96	Yes	4	1.71	Yes
2	Cooperative	12	12	X	X	12	X	X
		8	8	7.86	No	8	X	X
		4	4	1.77	Yes	4	1.39	Yes
	Indifferent	12	12	10.00	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.10	Yes	4	1.72	Yes
3	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	1.89	Yes	4	7.07	No
	Indifferent	12	12	X	X	12	X	X
		8	8	10.00	Yes	8	7.83	No
		4	4	2.64	Yes	4	X	X
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	8.12	No
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	7.47	No	4	7.92	No

C-Vis
Enrollment Timed Test Results
Verification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	10.00	Yes	12	3.94	Yes
		8	8	3.14	Yes	8	4.88	Yes
		4	4	5.92	Yes	4	8.42	No
	Indifferent	12	12	8.49	No	12	4.39	No
		8	8	3.86	Yes	8	5.73	Yes
		4	4	X	X	4	3.48	Yes
2	Cooperative	12	12	3.85	No	12	3.25	Yes
		8	8	3.06	Yes	8	6.07	Yes
		4	4	5.05	Yes	4	5.04	Yes
	Indifferent	12	12	2.85	Yes	12	3.82	Yes
		8	8	4.09	Yes	8	3.71	Yes
		4	4	5.45	Yes	4	5.69	Yes
3	Cooperative	12	12	3.81	No	12	3.25	Yes
		8	8	4.24	Yes	8	3.23	Yes
		4	4	4.01	Yes	4	4.03	Yes
	Indifferent	12	12	3.34	No	12	3.73	No
		8	8	8.19	Yes	8	8.59	Yes
		4	4	10.00	Yes	4	4.04	Yes
1 Variability Test	Cooperative	12	12	4.01	No	12	2.76	Yes
		12	12	5.00	Yes	12	5.30	Yes
		12	12	3.62	No	12	3.55	Yes
		12	12	4.04	No	12	3.50	Yes
		12	12	4.79	Yes	12	3.89	No
		12	12	2.93	Yes	12	3.86	Yes
		12	12	3.92	Yes	12	3.31	Yes
		12	12	3.48	Yes	12	2.83	Yes
1 Photo Test	Cooperative	12	12	4.88	No	12	2.85	No
		8	8	X	X	8	3.79	No
		4	4	6.63	No	4	5.69	No

Lau Technologies
Enrollment Timed Test Results
Verification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	1.78	Yes	12	1.50	Yes
		8	8	2.07	Yes	8	1.05	Yes
		4	4	1.25	Yes	4	0.90	Yes
	Indifferent	12	12	1.44	Yes	12	1.67	Yes
		8	8	1.29	Yes	8	1.37	Yes
		4	4	2.46	Yes	4	X	X
2	Cooperative	12	12	1.56	Yes	12	1.46	Yes
		8	8	1.04	Yes	8	0.93	Yes
		4	4	1.54	Yes	4	2.24	Yes
	Indifferent	12	12	1.50	Yes	12	7.68	Yes
		8	8	1.80	Yes	8	1.69	Yes
		4	4	1.85	Yes	4	1.53	Yes
3	Cooperative	12	12	3.47	Yes	12	X	X
		8	8	1.11	Yes	8	4.63	Yes
		4	4	1.18	Yes	4	1.30	Yes
	Indifferent	12	12	2.71	Yes	12	2.70	Yes
		8	8	1.19	Yes	8	1.22	Yes
		4	4	1.47	Yes	4	0.96	Yes
1 Variability Test	Cooperative	12	12	8.14	Yes	12	1.14	Yes
		12	12	X	X	12	2.76	Yes
		12	12	1.32	Yes	12	0.89	Yes
		12	12	1.19	Yes	12	1.95	Yes
		12	12	1.52	Yes	12	1.42	Yes
		12	12	2.54	Yes	12	2.08	Yes
		12	12	0.77	Yes	12	1.28	Yes
		12	12	0.96	Yes	12	1.90	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Miros
Enrollment Timed Test Results
Verification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	3.01	Yes	12	X	X
		8	8	1.57	Yes	8	1.65	Yes
		4	4	1.62	Yes	4	X	X
	Indifferent	12	12	2.10	Yes	12	3.05	Yes
		8	8	2.16	Yes	8	1.62	Yes
		4	4	3.69	Yes	4	X	X
2	Cooperative	12	12	X	X	12	10.00	Yes
		8	8	1.98	Yes	8	1.49	Yes
		4	4	8.35	Yes	4	X	X
	Indifferent	12	12	2.60	Yes	12	9.48	Yes
		8	8	2.20	Yes	8	3.22	Yes
		4	4	9.68	Yes	4	X	X
3	Cooperative	12	12	2.48	Yes	12	X	X
		8	8	1.57	Yes	8	1.33	Yes
		4	4	X	X	4	X	X
	Indifferent	12	12	3.09	Yes	12	X	X
		8	8	2.48	Yes	8	1.50	Yes
		4	4	X	X	4	X	X
1 Variability Test	Cooperative	12	12	2.37	Yes	12	10.00	Yes
		12	12	2.19	Yes	12	5.73	Yes
		12	12	1.49	Yes	12	1.72	Yes
		12	12	1.73	Yes	12	2.15	Yes
		12	12	1.82	Yes	12	2.67	Yes
		12	12	1.86	Yes	12	2.19	Yes
		12	12	1.61	Yes	12	2.21	Yes
		12	12	1.55	Yes	12	2.60	Yes
1 Photo Test	Cooperative	12	12	X	X	12	7.23	Yes
		8	8	6.45	Yes	8	X	X
		4	4	2.25	Yes	4	2.33	Yes

Visionics Corporation Enrollment Timed Test Results Verification Mode Subjects Stand In Place								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	12	6.59	Yes	12	X	X
		8	8	3.19	Yes	8	4.62	Yes
		4	4	2.91	Yes	4	3.89	Yes
	Indifferent	12	12	7.62	Yes	12	X	X
		8	8	2.54	Yes	8	4.83	Yes
		4	4	7.51	Yes	4	9.48	Yes
2	Cooperative	12	12	2.94	Yes	12	3.50	Yes
		8	8	3.02	Yes	8	2.58	Yes
		4	4	2.84	Yes	4	3.04	Yes
	Indifferent	12	12	2.87	Yes	12	3.39	Yes
		8	8	2.63	Yes	8	2.85	Yes
		4	4	2.99	Yes	4	2.78	Yes
3	Cooperative	12	12	3.27	Yes	12	3.54	Yes
		8	8	2.89	Yes	8	2.72	Yes
		4	4	3.01	Yes	4	2.90	Yes
	Indifferent	12	12	3.85	Yes	12	2.63	Yes
		8	8	2.63	Yes	8	2.76	Yes
		4	4	2.88	Yes	4	3.08	Yes
1 Variability Test	Cooperative	12	12	3.35	Yes	12	X	X
		12	12	2.48	Yes	12	X	X
		12	12	3.93	Yes	12	3.39	Yes
		12	12	3.01	Yes	12	6.67	Yes
		12	12	X	X	12	8.07	Yes
		12	12	4.24	Yes	12	X	X
		12	12	6.54	Yes	12	X	X
		12	12	2.72	Yes	12	9.37	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Banque-Tec Enrollment Timed Test Results Identification Mode Subjects Stand In Place								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	2.62	Yes	4	2.37	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	3.08	Yes	8	3.00	Yes
		4	4	1.70	Yes	4	2.53	Yes
2	Cooperative	12	12	8.63	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.09	Yes	4	1.58	Yes
	Indifferent	12	12	7.32	No	12	X	X
		8	8	X	X	8	X	X
		4	4	2.57	Yes	4	2.64	Yes
3	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	3.61	Yes	4	2.91	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	2.48	Yes	4	10.00	No
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
1 Photo Test	Cooperative	12	12	X	X	12	8.19	No
		8	8	X	X	8	7.60	No
		4	4	X	X	4	7.58	No

C-Vis Enrollment Timed Test Results Identification Mode Subjects Stand In Place								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	12	3.80	Yes	12	3.69	Yes
		8	8	4.42	Yes	8	4.79	Yes
		4	4	6.51	Yes	4	6.28	No
	Indifferent	12	12	7.81	Yes	12	3.71	Yes
		8	8	5.50	Yes	8	4.11	Yes
		4	4	10.00	Yes	4	4.94	Yes
2	Cooperative	12	12	3.67	Yes	12	3.78	Yes
		8	8	3.93	Yes	8	5.43	Yes
		4	4	3.72	Yes	4	6.80	Yes
	Indifferent	12	12	3.98	Yes	12	4.85	Yes
		8	8	4.02	Yes	8	4.69	Yes
		4	4	5.06	Yes	4	6.20	Yes
3	Cooperative	12	12	4.80	Yes	12	4.88	Yes
		8	8	6.33	Yes	8	4.38	Yes
		4	4	X	X	4	10.00	Yes
	Indifferent	12	12	6.49	Yes	12	5.04	Yes
		8	8	9.03	Yes	8	5.43	No
		4	4	10.00	Yes	4	8.72	No
1 Variability Test	Cooperative	12	12	4.41	Yes	12	3.76	Yes
		12	12	4.45	Yes	12	5.36	Yes
		12	12	5.06	Yes	12	3.71	Yes
		12	12	3.78	Yes	12	4.32	Yes
		12	12	4.33	Yes	12	3.78	Yes
		12	12	6.56	Yes	12	4.76	Yes
		12	12	10.00	Yes	12	3.77	Yes
		12	12	4.20	Yes	12	4.00	Yes
1 Photo Test	Cooperative	12	12	5.56	No	12	5.07	No
		8	8	6.24	No	8	6.61	No
		4	4	8.50	No	4	7.22	No

Lau Technologies Enrollment Timed Test Results Identification Mode Subjects Stand In Place								
Subject ID	**Behavior Mode**	**Start Distance**	**Backlighting Off**			**Backlighting On**		
			Final Distance	**Acquire Time**	**Correct Match?**	**Final Distance**	**Acquire Time**	**Correct Match?**
1	Cooperative	12	12	3.15	Yes	12	2.90	Yes
		8	8	2.10	Yes	8	1.67	Yes
		4	4	2.43	Yes	4	1.32	Yes
	Indifferent	12	12	2.17	Yes	12	2.21	Yes
		8	8	1.96	Yes	8	5.44	Yes
		4	4	6.47	Yes	4	X	X
2	Cooperative	12	12	2.27	Yes	12	1.47	Yes
		8	8	1.60	Yes	8	1.52	Yes
		4	4	1.81	Yes	4	1.23	Yes
	Indifferent	12	12	2.33	Yes	12	X	X
		8	8	2.23	Yes	8	1.38	Yes
		4	4	2.59	Yes	4	1.29	Yes
3	Cooperative	12	12	2.09	Yes	12	2.30	Yes
		8	8	1.57	Yes	8	3.42	Yes
		4	4	2.29	Yes	4	2.25	Yes
	Indifferent	12	12	1.78	Yes	12	4.54	Yes
		8	8	2.02	Yes	8	X	X
		4	4	2.40	Yes	4	2.39	Yes
1 Variability Test	Cooperative	12	12	1.97	Yes	12	3.04	Yes
		12	12	2.24	Yes	12	4.50	Yes
		12	12	2.25	Yes	12	2.68	Yes
		12	12	3.95	Yes	12	4.88	Yes
		12	12	2.57	Yes	12	3.32	Yes
		12	12	2.83	Yes	12	3.06	Yes
		12	12	2.73	Yes	12	3.15	Yes
		12	12	1.87	Yes	12	3.17	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Miros
Enrollment Timed Test Results
Identification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	5.06	Yes	12	6.12	Yes
		8	8	6.50	Yes	8	2.97	Yes
		4	4	3.45	Yes	4	5.21	Yes
	Indifferent	12	12	4.67	Yes	12	X	X
		8	8	9.96	Yes	8	4.70	Yes
		4	4	6.72	Yes	4	4.73	Yes
2	Cooperative	12	12	X	X	12	4.74	Yes
		8	8	3.41	Yes	8	2.63	Yes
		4	4	5.43	Yes	4	8.89	Yes
	Indifferent	12	12	4.59	Yes	12	X	X
		8	8	7.01	Yes	8	5.58	Yes
		4	4	4.68	Yes	4	X	X
3	Cooperative	12	12	5.66	Yes	12	X	X
		8	8	3.65	Yes	8	6.69	Yes
		4	4	6.43	Yes	4	X	X
	Indifferent	12	12	4.48	Yes	12	X	X
		8	8	3.49	Yes	8	3.39	Yes
		4	4	X	X	4	X	X
1 Variability Test	Cooperative	12	12	5.29	Yes	12	3.62	Yes
		12	12	6.67	Yes	12	3.14	Yes
		12	12	3.75	Yes	12	7.50	Yes
		12	12	4.63	Yes	12	X	X
		12	12	4.76	Yes	12	X	X
		12	12	7.30	Yes	12	4.13	Yes
		12	12	3.89	Yes	12	5.74	Yes
		12	12	6.39	Yes	12	7.96	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Visionics Corporation
Enrollment Timed Test Results
Identification Mode
Subjects Stand In Place

Subject ID	Behavior Mode	Start Distance	Backlighting Off			Backlighting On		
			Final Distance	Acquire Time	Correct Match?	Final Distance	Acquire Time	Correct Match?
1	Cooperative	12	12	X	X	12	X	X
		8	8	8.09	Yes	8	8.74	Yes
		4	4	X	X	4	8.28	Yes
	Indifferent	12	12	X	X	12	X	X
		8	8	6.59	Yes	8	5.66	Yes
		4	4	8.79	Yes	4	X	X
2	Cooperative	12	12	X	X	12	9.04	Yes
		8	8	8.88	Yes	8	9.23	Yes
		4	4	10.00	Yes	4	9.66	Yes
	Indifferent	12	12	8.64	Yes	12	8.52	Yes
		8	8	9.32	Yes	8	7.67	Yes
		4	4	8.20	Yes	4	X	X
3	Cooperative	12	12	X	X	12	X	X
		8	8	8.38	Yes	8	8.25	Yes
		4	4	8.12	Yes	4	8.87	Yes
	Indifferent	12	12	8.36	Yes	12	8.72	Yes
		8	8	9.19	Yes	8	7.54	Yes
		4	4	9.77	Yes	4	9.80	Yes
1 Variability Test	Cooperative	12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	8.60	Yes	12	X	X
		12	12	9.57	Yes	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	X	X
		12	12	X	X	12	8.70	Yes
		12	12	X	X	12	9.81	Yes
1 Photo Test	Cooperative	12	12	X	X	12	X	X
		8	8	X	X	8	X	X
		4	4	X	X	4	X	X

Appendix I
Case Study - A Participant Withdrawals

What follows is an excerpt of communications between one of the vendors and the FRVT 2000 sponsors regarding the methodology chosen for the test. Our intent in providing this information is not to judge this particular individual or his company's views. Rather, this is an interesting case study for anyone that wishes to perform any future evaluations because it provided some idea of the issues they should expect to encounter. We believe this shows that differing views do exist in the biometrics community and, by including this alternate view, we will spark further discussions about evaluation methodologies that will improve all future biometric technology evaluations.

The vendor was the first to sign up to participate in FRVT 2000 only two (weekend) days after it was announced to the public. The day after signing up, the vendor wrote in a message distributed via the Biometric Consortium's listserv that they were "provisionally entering the FRVT 2000 facial recognition vendor test, subject to our acceptance that the test is hard enough." In the same forum they said they "regard the previous FERET test with enormous skepticism. The problem is that the test protocol was *too easy*."

A few weeks passed before the vendor expressed concern that some of the vendors who had participated in previous FERET tests would have an unfair advantage since they have seen some of the images. The sponsors did not feel this was an issue because the FERET program did not involve any vendors in the FERET evaluations, the FERET images used for FRVT 2000 had not been made available to anyone, and a representative of this vendor had previously been given the FERET development database, which was available to all the other vendors. (This is question 7 in the restricted area FAQ.) The sponsors did not anticipate this complaint from this representative because previously he had stated that the FERET evaluations were too easy. He also claimed that the live tests would be unfair because the live images of the subjects would not be exactly the same for each vendor and proposed that we use prerecorded video clips instead. (This is question 10 in the restricted area FAQ.)

Approximately one week later, the vendor submitted a signed request to participate in the evaluation and was given their ID and password to access the restricted area of the FRVT 2000 web site, and hence, the Image Development Set and API documentation. The next day the vendor asked "what proportion of the recognition test set are of very small faces, such as 'i00011'? Our system will not return a similarity score for such images. It is a perfectly reasonable response for a system to 'abstain' when it does not consider the input data to be reliable enough to give an accurate similarity score." (This is question 20 in the restricted area FAQ.) He also asked, "Why separate performance and usability tests? How can the results be combined and assessed? If I were choosing a system to buy, I'd want to know the recognition performance on real-world images - that is, those from the usability test. Performance and usability are not separable so it seems scientifically dangerous to test them separately." (This is question 19 in the restricted area FAQ.) The answers to these inquiries were provided to all vendors the next day.

The following week, which was one week before the detailed Test Plan was released to all participants, the vendor sent a letter withdrawing from the FRVT 2000 evaluations. The primary cause cited for requesting the withdrawal was a disagreement with the evaluation methodology used for the FRVT 2000, which is explained in Section 3 of the Executive Overview of this report. The vendor requested a more narrow evaluation that concisely ranked the vendor systems. The sponsors rejected this approach as it would significantly limit the usefulness of the evaluation to only those with a planned usage that exactly matched the narrow evaluation methodology.

Appendix J
Data Collection

NIST HumanID Database

Collection and Processing

P. Jonathon Phillips

Patrick Grother

Nicole Snyder

William W. Klein

Karen Marshall

National Institute of Standards and Technology (NIST)
Information Access and User Interfaces Division
Visual Image Processing Group
100 Bureau Drive
Gaithersburg MD 20899-8940

1.0 Introduction

This document describes the design, acquisition, and preparation of the NIST HumanID Database as it stands in March 2000. This database has been prepared to facilitate robust evaluation of human recognition systems. It contains heterogenous data; both still and video imagery of several thousand individuals obtained using different sensors under different lighting conditions in different locations over a period of years. In its current form the database contains only face imagery; it is designed to include arbitrary biometric profiles. All images and video sequences are accompanied by ground truth files that contain various attributes that together define the scope of the database, and allow subsets of the database to be appropriately selected for use in different recognition scenarios.

1.1 History

The original Face Recognition Technology (FERET) program conducted multiple tasks over a three-year period starting September 1993. That program was sponsored by the Defense Advanced Research Projects Agency (DARPA). The HumanID database is sponsored by DARPA and the Department of Justice (DOJ). The original program (from 1993-1996) was responsible for the initial database. This present database is a continuous of the work done in that three-year period. Although that work is not specified here, much of the groundwork seen in this database was done during that time period. In addition, it is important to note that the original FERET images are included in HumanID database.

1.2 Purpose

The object of the FERET project was to develop face recognition systems that can assist intelligence, security, and law enforcement personnel[2]. A possible scenario in surveillance applications is face recognition and identification from a video sequence. This database is being used as the source of a distribution by NIST to the face recognition community to allow for government-monitored testing and evaluation of face recognition algorithms using standardized test procedures. The HumanID database consists of three components: a developmental subset, a sequestered subset, and an evaluation subset. The developmental subset is sent to the researchers for algorithm development and training. The developmental subset is a subset of the evaluation subset and consists of a handful of images. The evaluation subset will be used to make an assessment of the current state-of-the-art in face recognition and identification algorithm development through the usage of standardized tests and test procedures. The sequestered subset is the images that are not apart of the evaluation subset.

1.3 Overview of the Project

The imagery was collected at several locations within the USA from 1993 through to 2000. In all instances human subjects were required to sign a consent form allowing their images to be captured and distributed. The imagery is anonymous; the identities of the subjects are keyed solely as integer ids unlinked to the subjects' names. The database contains a large volume of still and video images. It contains images, captured at different stations, of people whose appearance may or may not have changed over time. A few of the individuals from one site were recorded at another site. Since lighting conditions may effect the darkness or reflectivity of certain types of eyeglasses, when possible, people wearing eyeglasses were captured both with and without them. The database contains variations in illumination, scale, pose, and background composition for each subject.

The digital video data consists of files containing 10-15 second sequences recorded in MPEG-2 (Moving Picture Experts Group). There are from one to three digital video files for each person contained in the database. Each digital video file contains information that is a representation of National Television System Committee (NTSC) frames, 720 X 480 pixels in height and in width. A user of this database will need an MPEG-2 video player/decoder to access atomic frames contained within the video data files and a still image viewer. For each person contained in this database, analog, digital, video frames, and still-from-video badging system still image(s) are provided using the Joint Photographic Experts Group [1](JPEG) format.

2.0 Image Capture

This section of the document explains the recruitment and data acquisition procedures. for collection all needed materials were gathered. The room for inside collection was rearranged so that a subject first was recorded, then in the case at NIST went to the badging station, then the still image station, then the inside video station, then went outside.

2.1 Equipment Used

The following is a table of the equipment used to gather the data and at which site the equipment was used.

2.1.1 Cameras

	Cannon[i] Video Camera (ours)	**Cannon[i] Video Camera (theirs)**	**Sony[i] Digital Model 91**	**Sony[I] Digital Mavica**	**Minolta[i] X-700 35mm**	**still-from-video badging**	**Olympia[i] 35mm**
station used	outside video sequence	inside video sequence	inside digital still	outside digital still	inside analog still	inside still	inside analog still
site(s) used	Dahlgren, VA (99)	Dahlgren, VA (99)	Dahlgren, VA (99)	Dahlgren, VA (99)	Dahlgren, VA (99)		
	NIST, MD (00)	NIST, MD (00)	NIST, MD (00)	NIST, MD (00)		NIST, MD (00)	NIST, MD (00)

2.1.2 Other Equipment

	For: Cannon[i] Video Camera (ours)	*For: Cannon[i] Video Camera (theirs)*	*For: Sony[i] Digital Model 91*	*For: Sony[i] Digital Mavica*	*For: Minolta[i] X-700 35mm*	*For: Olympia[i] 35mm*
charger	**1**	**1**	**1**	**1**		
tripod	**1**	**1**	**1**	**1**	**1**	**1**
batteries	**2**	**1**	**1**	**1**		

ac adapter	**1**	**1**

For the still-to-video badging system: computer, monitor, keyboard, mouse, EBACS[i] MK3 MOD 4 badging software, Integral Technologies FlashPoint[i] 3075 PCI video frame grabber card. Imaging Technology Corporation[i] CCD 100 video camera, Lowel iLIGHT[i] portrait lighting system, and cabling. (EBACS[i] software was developed by Crane. The capture board, camera, light, cabling, and software drivers to control the camera were all purchased as a package from Imaging Technology Corporation[i].)

Other: 3 light stands, 3 extra light bulbs, 3 photofloods, and 3 power strips.

2.2 Subject Recruitment and Enrollment

Subjects appearing in the database appeared voluntarily for the data collection in response to local advertising of the session. Each subject was informed of the objective of the project, that being to capture still and video images to be included in the NIST HumanID database. It was pointed out to the prospective participant that the images could possibly be used in example illustrations contained within a research paper and/or published within a research journal

If at any point during the recruitment process the subject expressed a desire not to participate in the data collection project, the subject could terminate participation in this project without any repercussions.

Once the subject agreed to participate in data collection, the Consent Form was presented to the subject and they were told to carefully read the contents of the form. Any questions with regard to the project and the Consent Form were answered.

After the subject read and signed the Consent Form, the recruiter signed as official witness and one of the staff that was authorized as a investigator signed as so. On the order of 5% of participants declined to continue at this stage. For all other their Consent Form was filed by date. A tracking session id was given to the subject to allow the enrollment of images from all capture stations to be included in the database with a unique and correct identity. It was also explained to each subject the session ID was not associated in any way with the subject's signed Consent Form and the only usage of this number would be to allow for the images captured at each collection station to be grouped together after processing. Additionally, the subject was informed that the session ID would not be associated with the final number used to index the database.

The recruiter then recorded on the log beside that session ID number the date and gender of the subject. If data had been taken on previous occasions at the site the subject was asked if they had their picture taken previously. If so they were instructed to identify themselves in a folder of the previously gathered images presented as thumbnails. If they did find themselves their existing integer identifier number alongside the picture was also recorded with the current session ID to allow correct integration into the HumanID database. Also if it was the second or later day of the collection the recruiter asked if the subject had their picture taken on a previous day. If so, the date they said had their previous picture taken was also recorded next to the session ID.

2.3 Image Capture

All image capture stations built paper logs of session Ids, time, date and sequence number to allow ground truth construction once the images were recovered from the various media. In some instances the sensors automatically recorded date and time. This was used to provide consistency checking during the database preparation. The preprinted log sheets containing the date and the station (specifying if the station was inside or outside). On the log sheet the station operator recorded the session ID and time for every instance an image was taken. At no time would any media or log sheet have more than one date collection nor would the collection of one station of a person be on more than one of the same media. A request was made to subjects wearing glasses that the glasses should be removed, although this was not a requirement.

2.3.1 Badging Station

The setup used was identical to the system used to create badges at Crane, IN and other customer sites.

At this station a subject stood in front of a wall and gave the station operator their session ID. The station operator then input that session ID into the image capture program. At which point the light above the video would turn on and the station operator asked the subject to look at the camera. Next the station operator would manually move the camera and zoom in or out until the person's head filled up 50% of the screen. Thereupon the station operator captured one frame of the video and saved it to disk. The subject was then instructed to go to the next station.

2.3.2 Inside Still Image Capture

Both analog and digital mugshots conformed to the guidelines made at the Mugshot and Facial Image Workshop held at NIST on October 1995. Those guidelines mandated that a full-face or frontal pose mugshot image shall be captured using a gray background and uniform lighting[3]. This setup can be observed in Figure 1.

Specifically, each still image was captured using an 18% gray background and three studio lights[4]. A background of approximately 1.21 meters (4 feet) in height and 1.21 meters in width was hung on the wall. The background is high quality paper that is wide toned, seamless, and is colored thunder gray. The subject was seated approximately .91 meters (3 feet) on center in front of the gray background and the camera was placed approximately 2.73 meters (9 feet) on center from the background.

Three floodlights provide uniform studio lighting. Each floodlight was mounted on a light stand 1.82 meters above the floor. Two floodlights were positioned approximately 1.21 meters (2 feet) from the center of the background; the distance from each floodlight to the background was be approximately 2.42 meters (8 feet). Another floodlight was positioned directly in back of the camera at a distance of approximately 3.33 meters (11 feet) from the gray background.

The studio lighting was arranged so that the subject's shadow was not visible in the background of the captured image and facial/eyeglass reflection of the lighting was minimized or not visible.

For this reason a 18% gray background (27-1253) paper, manufactured by Savage Universal Corporation[i], was purchased at Penn Camera[i]

Six different images of one subject were taken. In all of these pictures the subject was facing the camera (frontal pose). They are as follows:

- 2 images using the above described mugshot lighting
  - 1 with a FA (regular) expression
  - 1 with a FB (some other i.e. a smile) expression
- 2 images using FERET lighting (the above lighting minus the floodlight behind the camera)
  - 1 with a FA (regular) expression
  - 1 with a FB (some other i.e. a smile) expression
- 2 images using no floodlights only the lighting of the room
  - 1 with a FA (regular) expression

- 1 with a FB (some other i.e. a smile) expression

In order to capture all six images with the least amount of man power 3 power strips were used. The first power strip which was plugged into the wall had the digital camera and the second power strip plugged into it. The second power strip (plugged into the first power strip) had the two floodlights next to the camera and the third power strip plugged into it. The third power strip (plugged into the second power strip) had the one floodlight behind the camera plugged into it. Therefore, the operator of this station would:

1) **Turn on** the power to the second and third power strip

   (The first power strip **remains on for the entire day**)

2) With all *three floodlights* on, ask the subject for the FA expression then the FB expression
3) **Turn off** the power to the third power strip
4) With the *two side floodlights* on, ask the subject for the FA expression then the FB expression
5) **Turn off** the power to the second power strip
6) With *no floodlights* no, ask the subject for the FA expression then the FB expression

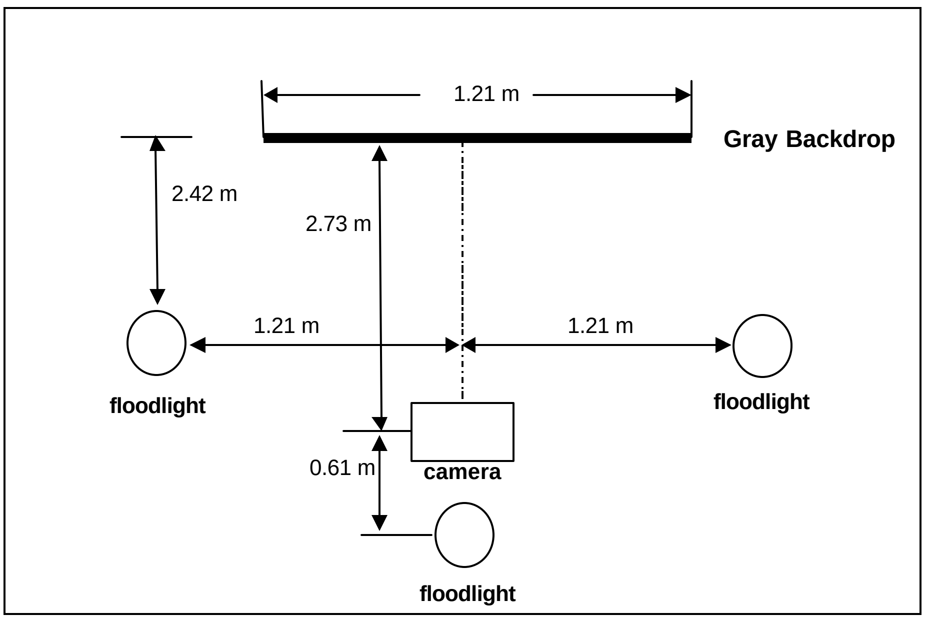


Figure 1 - Still Image Capture Set

2.3.2.1 Inside Digital Image Capture

Digital still images were captured on a 3.5inch diskette using a digital camera. Each image was captured in landscape mode of operation and will be 1024 by 768 pixels in width and in height.

The subject's head width must be 50 % of the final cropped image width. Therefore, the subject was centered in the LCD viewfinder manually.

As the digital images are captured on the diskette, the operator recorded on the digital still image captured log sheet the subject's session ID number, the time and date taken, and the diskette number. The same information was recorded on the diskette label once all images for that subject were taken. . When all six poses were captured the operator ejected and replaced the diskette.

2.3.2.2 Inside Analog Image Capture

Analog still images will be captured by a 35-mm camera with a 49-mm lens using Kodak Royal Gold[i] 400 ASA colored film. Normally, rolls of film that contained enough space to capture 36 exposures were used. The camera, in automatic exposure mode of operation, captured a landscape exposure. The operator took for each subject the six poses. The operator recorded next to the frame number on the log sheet, the subject's session ID number, and the date and time. Then placed the roll number on the top of the log and the roll's canister. Never was one subject taken on two different rolls. Therefore, not all of the 36 exposures were used.

2.3.3 Video Movie Images

In all video sequences, only one person appears in the video. The subject was instructed to walk to the camera operator in a straight line, focusing his/her eyes at the camera's zoom lens and then to peel off at less than one meter from the camera. As the subject walked the route to the camera, facial features (eyes, mouth, and nose) only become apparent/distinguishable whenever the subject is in the last third of the course.

2.3.3.1 Inside Video Images

One indoor sequence for each subject was taken. This sequence used nothing more than the lighting provided by the overhead lights. From the operator's view the subject started one meter to the right of the walking route facing left. The subject then walked one meter, turned to face the camera, and then traversed the rest of the route. The distance traversed was approximately 8 meters depending on the size of the room. Over last third course, the camera's zoom lens was set to allow for capture of the entire facial region. Therefore, the camera's focus and zoom were set manually. The aperture was set at 2.8 with the ND option off. The stab option off and set for manual focus.

2.3.3.2 Outside Video Images

To allow for different scenarios, two different outdoor video sequences were attempted. The first sequence is as the one described above where the distance traversed was approximately 15 meters. This distance was greatly influenced by the size of the site, location of the site, location of the camera, and length of the power extension cord. The other sequence was the subject walking perpendicular to the camera. The subject was instructed to start outside of the view of the camera and then walk across the view of the camera. This was approximately 15 meters. These video sequences were shot using varied backgrounds. When available, an outdoor background containing trees was given preference. For these sequences the camera was set at fully automatic with the exception of the stab option being off.

2.3.4 Outside Digital Still Images

For the outdoor still images two poses per subject were taken. The first pose was a full frontal with the person's head filling 50% of the viewfinder. The second pose was the subject looking approximately 45 degrees away from the camera.

2.3.5 Colorcards

Colorcards were used throughout the collection process at each of the sites in order to 1) maintain the integrity of the database 2) to have a ruler in which to measure the results. The Colorcards used are a ColorChecker made by Gretag Macbeth[i]. This particular Colorcards is a checkerboard array of 24 colored squares (4 squares down and 6 squares across). The colors on the card are designed to help evaluate the many variables associated with color reproduction and range from black to white. The Colorcards were taped to an erasable board on which the date and time of capture. In addition, the Colorcards for inside also had the roll number used at the time.

2.4 Layout of the Sites

The following are diagrams of the layout of each site.

2.4.1 Dahlgren, VA (year: 1999)

2.4.2 NIST, MD (year: 2000)

2.5 Miscellaneous

2.5.1 Number of Operators

The minimum number of operators needed to be able to collect the data is three. In this instance one operator would be outside and two would be inside. However the ideal is one for each station. At Dahlgren in 1999 there was five operators (three inside and two outside). At NIST in 2000 there was also five operators (four inside and one outside).

2.5.2 Errors

- On the third day of being at Dahlgren in 1999 the analog camera jammed. As a result four subjects have no analog images.
- At Dahlgren in 1999 and then at NIST in 2000 it was fairly cold and some people chose not to participate in the images taken outside.

3.0 Image Processing

Four different types of processing took place. Processing digital still images, analog still images, video clips, and frames from video sequences.

3.1 Processing of Digital Images

The images received on floppies were from the digital cameras, both inside and outside. The images captured from the digital cameras were included in the database under the control of an interractive perl script developed at NIST. The scripts prompted the user for session IDs and a floppy disk media. The script for the inside floppies assumed that there were six images for each session ID entered. While the script for the outside floppies assumes that there were two images for each session ID entered. If this assumption was not met the script copied the entire floppy to disk and told the user to examine those images and determine why that was not met and then press enter to resume processing. In the case that there was less than expected the script assumed that the one missing was an overhead image.

3.2 Processing of Film Images

The film was sent to a commercial laboratory for processing. After the 35mm film was developed, the laboratory converted each developed negative to a digital Kodak[i] PhotoCD format file, **pcd**. Each **pcd** file contains 5 multiresolution images. NIST used the highest resolution entry (2048 by 3072) pixel resolution image to be the source image for this CD-ROM distribution.

Once the CD was received at NIST, enrollment proceeded using a perl script. This script opened each image on the CD. It then asked the user to enter the session ID for this person. It then opened the digital image of this person and asked if it was visually the same person. If it was the user entered whether the lighting was (m)ugshot, (f)eret, or (o)verhead. It then copied the file into the appropriate directory converting from kodak's pcd format to standard JPEG/JFIF using the ImageMagick convert program.

3.3 Processing Video

There are from one to three digital video files for each person. Each video file contains approximately 20 to 30 seconds of video recorded in MPEG-2 format. The digital video was captured using a Canon[i] XL1 digital video camera. This camera recorded an interlaced National Television System Committee (NTSC) video stream on a digital videocassette tape. Using the videocassette recorder contained in the camera and Pinnacle[i] Systems personal computer hardware and software, the recorded digital video was selected and captured from the cassette and transferred to the personal computer, de-interlaced, and stored with a file recorded within an Audio Visual Information (AVI) format. Next, the AVI file was converted from digital files (containing luminance (Y), chroma red (Cr), and chroma blue (Cb) values for each pixel) to frame level files (containing red (R), green (G), and blue (B) values for each pixel) that were recorded in portable pixel map (PPM) format. Then, the sequence of frame files was input to the MPEG-2 encoder to generate a movie.

The MPEG-2 compression format was chosen for this distribution because it was designed to take advantage of correlation between neighboring frames and compensate for the motion when compressing the video data[5]. On average, this results in an approximately 6.07 reduction from the original video stream size (AVI file) and thereby allows us to store more video data.

3.3.1 Software Used

Figure 5 - Digital Video (MPEG-2) Creation Process Flow illustrates the hardware and software packages that were used to create the Human ID database. These packages were resident on an IBM-compatible personal computer that used Windows/NT[ii] 4.0 as its operating system. NIST authored PERL scripts and C programs to provide a batch processing facility used to create the database.

3.3.1.1 Digital Video Acquisition at the Personal Computer

An operator used a Canon[i] XL1 camera to record digital video sequences using its internal videocassette recorder (VCR) on a 4-mm digital videocassette (DVC). The video sequences contained on a DVC were transferred to the personal computer using the Pinnacle Systems[i] DV300 Version 1.5 product. This product contains a graphical user interface (GUI) that allows a user to select and to transfer a video sequence(s) to the personal computer for storage in an AVI file(s).

This transfer uses hardware, a personal computer FireWire™[1] board that conforms to the IEEE-1394 interface specification, and allows personal computer software to issue Twain interface commands to the digital camera's VCR component to enable playback of the DVC to the personal computer. During the transfer from the camera to the personal computer, the two interlaced NTSC frame fields are combined to produce a composite frame sequence. After transfer and conversion, the digital video data stream is stored in an AVI type of file.

3.3.1.2 Video File Conversion to Frame Level Files

Conversion from an AVI file to frame level PPM files, suitable for input to the MPEG-2 encoder, was accomplished by using a NIST created conversion program, **nistavitoppm.** This program converts the digital video stream to frame level RGB file(s) that are in PPM format. It requires a Sony[i] compressor/decompressor (CODEC), contained on the IEEE-1394 interface board, to be installed on the personal computer that decodes the digital video data stream contained within the AVI file.

3.3.1.3 MPEG-2 Video

MPEG-2 software was obtained from the MPEG Software Simulation Group (MSSG). A user may download software from the MSSG site using the www.mpeg.org universal resource locator (URL) address. From this site, the following software may be obtained:

3.3.1.4 Encoding

The **mpeg2encode** [6] program is the MSSG implementation of an ISO/IEC DIS 13818-2[7] encoder. It converts an ordered set of uncompressed input picture frames into a compressed bit stream compliant with the standard. It contains 54 different parameters that control the encoding of the set of frames into a compressed MPEG-2 file. Most of these parameters were the MSSG's suggested defaults for compressing images that originated from NTSC video signals. The MSSG's default values for the group of pictures (GOP) parameter is a value of fifteen frames and the I/P frame distance (IPFD) parameter is a value of three frames. By changing the frame values to thirty frames for the GOP parameter and to fifteen frames for the IPFD parameter, NIST observed an approximate twenty two per cent decrease in the time required to encode a MPEG-2 video file with no detectable loss in quality of the video. In the process of transforming an AVI file to a MPEG-2 file, a 6 to 1 reduction in video file size was observed.

---

[1] FireWire™ is a registered trademark of the Apple Computer Inc.

3.3.1.5 Decoding

The **mpeg2decode**[8] program is the MSSG implementation of an ISO/IEC DIS 13818-2[7] decoder. It allows for conversion of MPEG-1 and MPEG-2 video bit streams into uncompressed video frame sequences. Using the decoder, a video file can be decoded into frame files formatted in either separate Y (luminance), U (chrominance blue), and V (chrominance red) files or a single PPM file that contains R, G, and B values for each pixel. Figure 4 provides an illustration of the output of computed frame luminance differences between the first and thirtieth frames that were extracted from a video; this allows motion to be illustrated.

3.3.1.6 Playback

The **mpeg2ply** program is the MSSG implementation of a GUI that displays an MPEG-2 video file that is decoded by the mpeg2decode program.

3.4 Video Frames

During the process of converting AVI format to MPEG-2 four video frames were copied and converted from PPM format to JPG format. Therefore, one AVI file was converted into PPMs as described in "Video File Conversion to Frame Level Files" above. After all of the frames were in PPM format the user, using an image viewing program, manually went through the frames and choose four. The first frame chosen was the frame in which the subject was the farthest away while facing the camera. The second frame chosen was the frame in which the subject had walked approximately 3 meters. The third frame chosen was the frame in which the subject was halfway between the second frame and the last frame chosen. The last frame chosen was the last frame in which the user could clearly see both of the subject's eyes. Then the frames chosen were copied into JPG format.

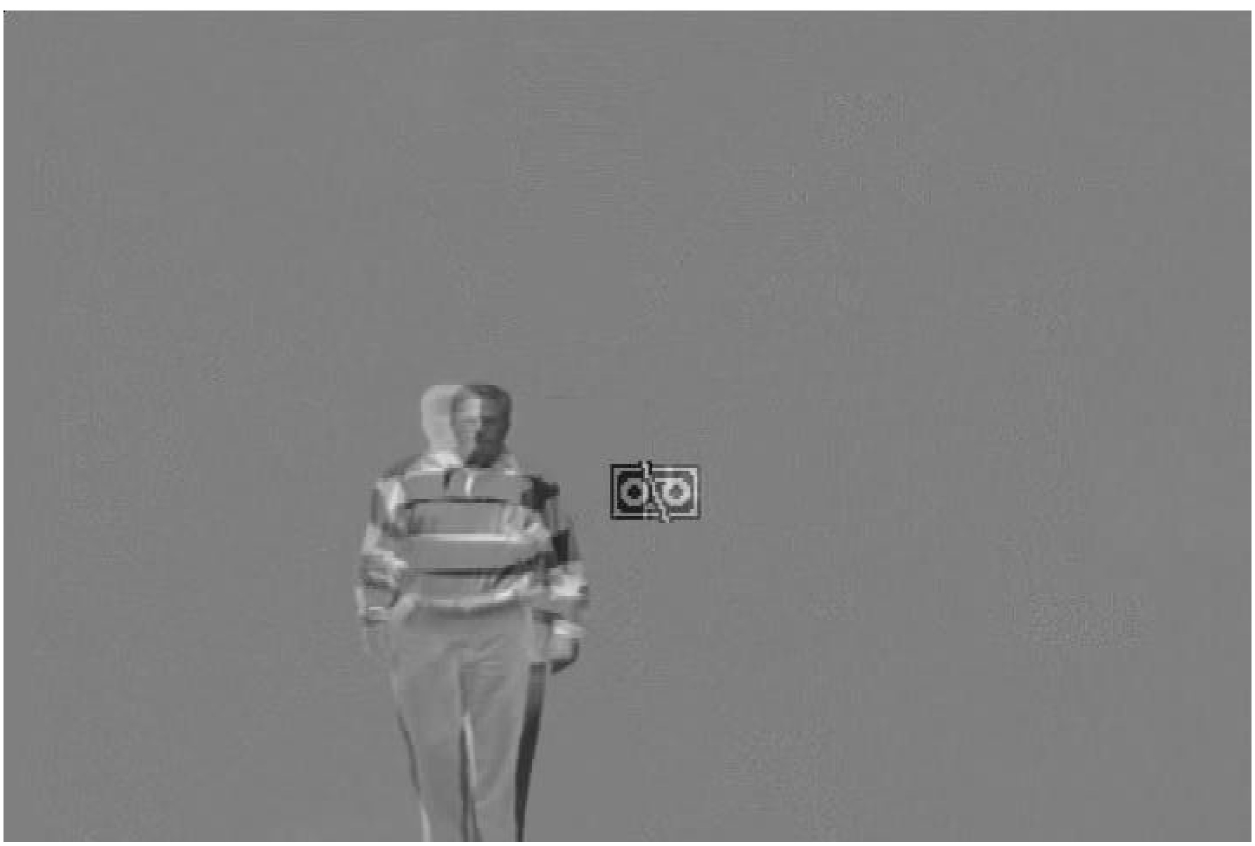


Figure 4 - Frame Luminance Differences

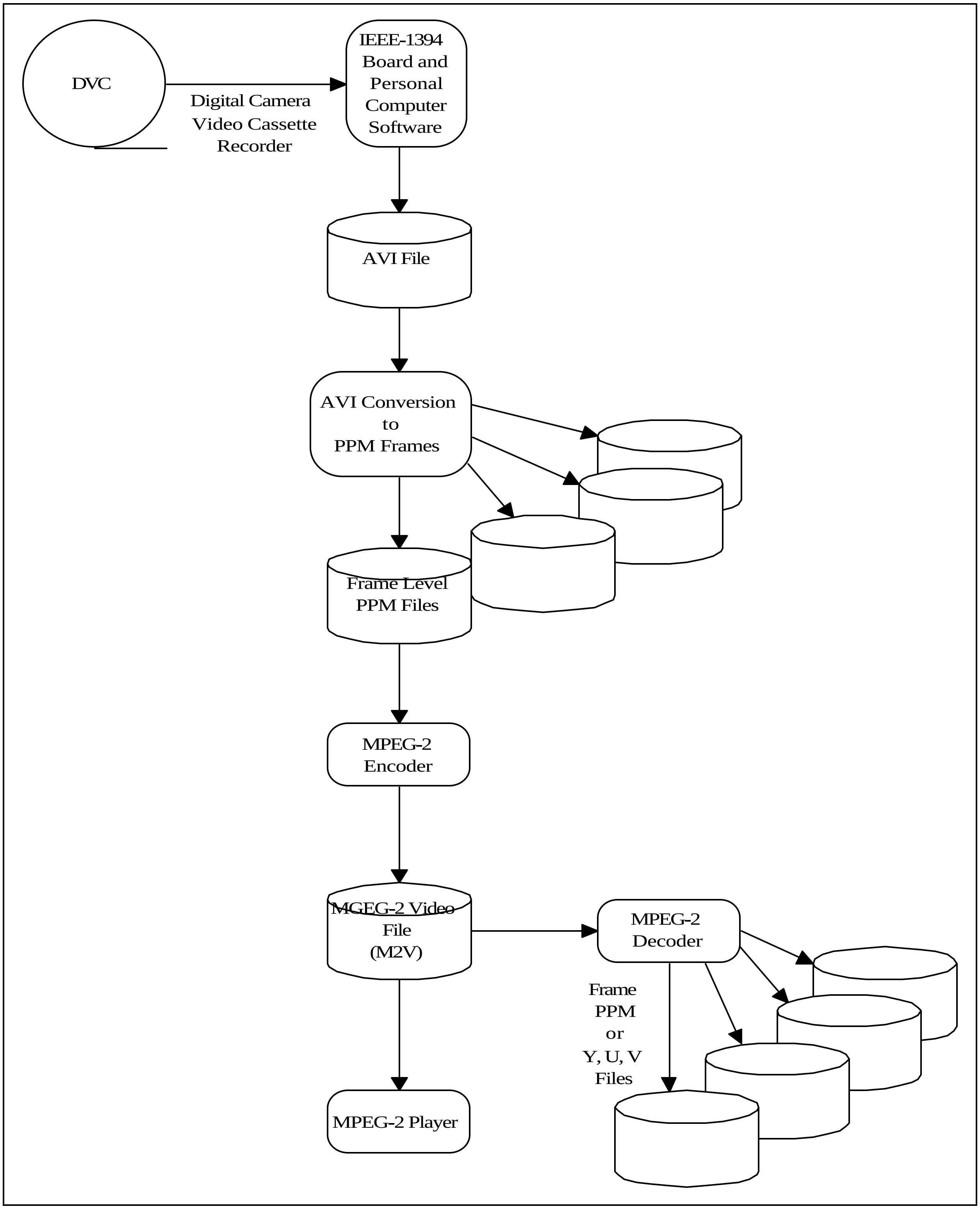


Figure 5 - Digital Video (MPEG-2) Creation Process Flow

4.0 Database Organization and Content

All directory and file names in the NIST Human ID database conform to the ISO-9660 standard naming conventions. All file names and file extension names have "8.3" format names of the form **NNNNNTSS.EXT** where:

NNNNN is a uniquely assigned subject number that ranges from 00001 to 99999 that identifies each indivual.

T denotes the type of file and may contain the following values:

a denotes an image file captured using an analog still camera

b denoted an image captured using the at the still-from-video badging system

d denotes an image file captured using a digital still camera

f denotes one of the original FERET images captured using an analog still camera

v denotes an image file captured using a digital video camera

z denotes still images that have been processed in some way described in the ground truth f ile.

SS denotes a uniquely assigned sequence number that ranges from 00 to 99.

EXT denotes the type of file name extension and may contain the following values:

gnd denotes a ground truth file that contains ASCII data which describes the contents of a like named image file

jpg denotes an image file that contains a still image recorded in JPEG/JFIF format

m2v denotes an image file that contains a video recorded in MPEG-2 format

The above naming convention allows each file name and extension to be unique within and outside of the NIST Human ID Database NN directory structure.

4.1 Ground Truth Files

For each image file there is a corresponding ground truth file that provides a summary of the image characteristics. Each ground truth field is encoded with an ASCII value(s) that is not a space value and is delimited by a new line character. The fields can appear in any order, although the history field, date_taken, and gender_at_site generally appear on the first couple of lines.

Each ground file contains minimally the following fields:
history= [image name];[another image name]

time_taken= [hours]:[minutes]:[seconds] GMT # or in the instance of the video frames the frame number:
video_frame_range= (closest, closer, farthest, or farther)

date_taken= [day] [month] [year]
gender_at_site= (M, F or N)
lighting= (indoor_mugshot, indoor_feret, indoor_overhead, outdoor_daylight or badging_system)
site= (Dahlgren, VA ; NIST, MD ; or Crane, IN)
media= (film_still, digital_still, digital_video, video_with_frame_grabber, or video_still)
id= [database ID]

The following fields were captured by an internal NIST program called 'NewTruthWriter'. (Description of this program is below.) Each image should have these fields; however, the videos and other such media will not. In those instances, as many fields as possible should be recorded.

pose= (frontal, or nominal45)
eye_glasses= (yes or no)
left_eye_coords= [x coords] [y coords]
right_eye_coords= [x coords] [y coords]
nose_coords= [x coords] [y coords]
mouth_coords= [x coords] [y coords]

The following field is only included when applicable:

expression= (FA or FB)

The following fields are not yet included in the ground truths:

color_card= [image name]
gender_retrospective = (M, F, N, or U)

4.1.1

4.1.2 In addition, for certain images the ground truths contain special fields that describe certain image processing treatments applied to the image. In these cases the image was derived from another "parent" image already present in the database. The new ground truth file contains the name of the parent and the processing details.

4.1.3 4.1.1 Capturing Coordinates

The program used in order to capture coordinates is an internal NIST program named 'NewTruthWriter'. This program first loads an image. Then a user manually clicks on the subject's left eye, then the subjects right eye, then the tip of the subject's nose, and lastly the center of the subject's mouth. Also, the user chooses the subject's pose and whether the subject is wearing eyeglasses or not. That information is put into a file which is the image name with the ending of '.cor' rather than '.jpg'. This file is of ASCI characters and is one line in length with 0 for no glasses or 1 for glasses, then the pose, the x and y coordinates starting with the person's right eye and ending with the person's mouth**.**

5.0 Looking Towards the Future

More data collection is planned for the near future and will continue to take place both at the sites already used and at new sites.

6.0 References

[1] J.L. Mitchell, W. B. Pennebaker, "JPEG Still Image Compression Standard", Van Nostrand Reinhold, New York, NY, 1993.

[2] P. J. Phillips, P. J. Rauss, S. Z. Der, " FERET (FACE Recognition Technology) Recognition Algorithm Development and Test Results", ARL-TR-995, Army Research Laboratory, Adelphi, MD, October 1996.

[3] "Best Practice Recommendation for the Capture of Mugshots", Version 2.0, www.nist.gov/itl/div894.03/face/bpr_mug3.html, Mugshot and Facial Image Workshop, Gaithersburg, MD, September 1997.

[4] M. Rubinfeld, C. L. Wilson, "Gray Calibration of Digital Cameras to Meet NIST Mugshot Best Practice, NISTIR 6322, March 1999.

[5] J.L. Mitchell, W. B. Pennebaker, "MPEG Video Compression Standard", Chapman & Hall, New York, NY, 1996.

[6] "mpeg2encode: MPEG-2 Video Encoder", Version 1.1, MPEG Software Simulation Group, June 1994.

[7] "Generic Coding of Moving Pictures and Associated Audio", Part 2: Video, ISO/IEC 13818-2 International Standard, 1995.

[8] "mpeg2decode: MPEG-2 Video Decoder", Version 1.1, MPEG Software Simulation Group, June 1994.

---

[i] Certain equipment/software may be identified in order to adequately specify or describe the subject matter of this work. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the equipment identified is necessarily the best available for the purpose.

Appendix K
Evaluation Preparations

K.1 Similarity File Check

The sponsors of FRVT 2000 wanted to make sure that the output produced by vendor software during the Recognition Performance Test could be read successfully and processed by the sponsor-developed scoring software. The goal was to resolve any potential problems before testing began. Participating vendors were required to compare each of the 18 images in the Image Development set with each of the other images in the set and create similarity files according to the format described in the API document. These similarity files were e-mailed to the sponsors for compliance verification. The software tried to read each of the ASCII files containing similarity scores and returned error messages if any compliance problems were found. A few vendors had errors in their similarity files and were asked to resubmit modified similarity files. All participating vendors eventually submitted correct similarity files and were notified of this.

K.2 Room Preparation

Several weeks before the tests began, the testing room was prepared. The arrangement of the different test stations is described in Appendix C. Figures K1 and K2 show a detailed layout of the room and the locations of the overhead fluorescent lights.

K.3 Backlighting

Backlighting was used for some trials in the timed tests. This was to simulate the presence of an outside window behind the subject in a controlled and repeatable manner. To accomplish this, a custom lighting device was built. It consists of a track lighting system with fixtures arranged in a 4 x 4 grid. The lights used for this device were manufactured by Solux and chosen because they have a spectral power distribution that closely mimics that of daylight. The particular model used for this application has a beam spread of 36 degrees and a correlated color temperature of 4,700 degrees Kelvin. Power requirements for each bulb are 50 watts at 12 volts. The 4 x 4 light grid was mounted inside a box facing toward the camera. The inside of the box was covered with flat white paint. The front side of the box, which faced the camera, was 4 ft. x 4 ft. The material used on the front side is a Bogen Lightform P42 translucent diffuser panel. The lights were arranged so the beams overlapped on the surface of the front panel for even illumination.

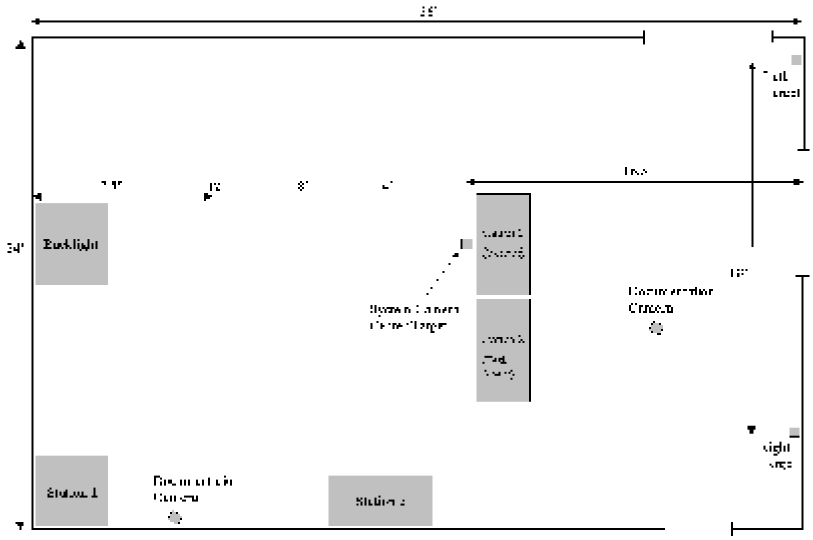


Figure K1 - Testing Room Layout

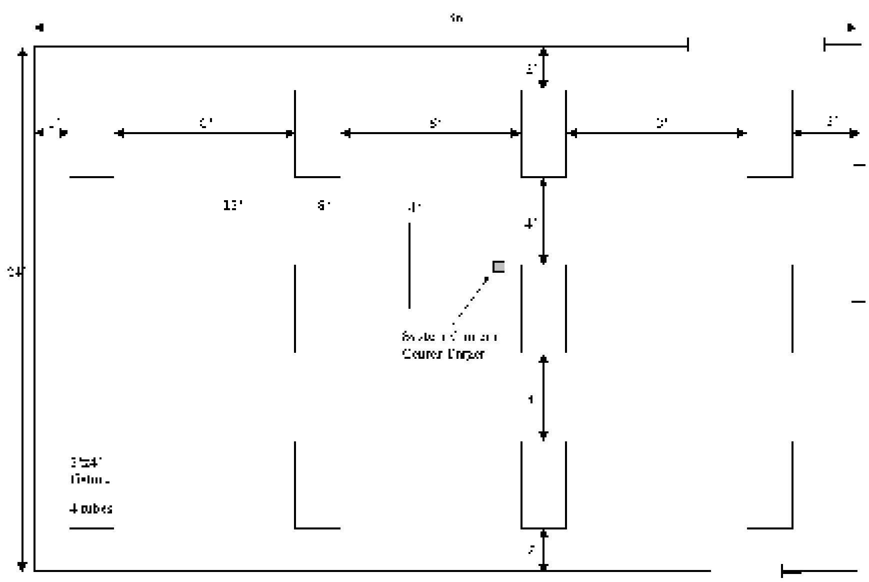


Figure K2 - Fluorescent light layout for testing room

K.4 Subject Training

In the weeks leading to the first test date, the test agent met several times with the three test subjects in the room where the testing would take place. The purpose of these meetings was to explain the Product Usability Test procedures described in the test plan, let the subjects practice their roles to achieve consistent behavior before the tests began and uncover any problems with the test plan procedures. The subjects practiced walking in front of a camera about 15 times each at the first meeting. During this session, a few procedural improvements were suggested and implemented by the subjects.

- Use a metronome set to 60 beats per minute to synchronize walking cadence and head movement, giving more consistent results with each trial.
- Draw more attention to the stop marker placed one foot in front of the camera so the subjects could more easily detect this location while walking and turning their heads during the indifferent trials.
- Begin identification trials with bodies one-quarter turned from the camera path to help ease the awkwardness of the 180-degree turn specified in the original test plan.

To accomplish these improvements, a metronome was purchased. Two tripods were placed at the stop marker with yellow caution tape stretched between them at a height of 3 feet for added visibility using peripheral vision. The test plan was updated to specify facing 90 degrees from the camera path at the beginning of identification trials.

After the improvements were made and the test procedures were updated, two more practice sessions were held. Each session lasted approximately one hour, and each subject participated in about 20 to 25 trials. Both sessions were held the week before the first vendor test to keep the procedures fresh in the subjects’ minds.

Appendix L
Modifications to the Recognition Performance Test Scoring Algorithm

Scoring Algorithm Modification

The similarity file scoring algorithm, used for the Recognition Performance portion of the FRVT 2000 evaluations, was originally developed for the FERET program. After the FERET program concluded, NIJ and DARPA co-funded an update to the algorithm so it can use the C/C++ programming language and a revised ground-truth format. The scoring algorithm was updated again for the FRVT 2000 evaluations so it could function with a less than complete set of similarity files. The new scoring algorithm was validated using three different methods.

The first validation method used the baseline PCA algorithm developed for the FERET program to develop similarity files using the same set of images used in the September 1996 FERET evaluations. The images were then scored using the new scoring algorithm and the resulting CMC curves were compared to the original results.

The second validation method the sponsors used was to write an algorithm that synthesizes a set of similarity files from a given CMC curve. The new scoring algorithm then scored the similarity files and the results were compared to the original curve for validation.

The third validation method was to provide the participating vendors with a set of similarity files derived from a baseline algorithm using FERET images, the scoring software and the results from the scoring software. Participating vendors were then asked to study the validity of the scoring code and provide feedback to the evaluation sponsors if they found any software implementation errors. The vendors did not report any errors.

Appendix M
Minor Reporting Practice Changes

M.1 Access Control System Interface Test

Only one vendor opted to take the access control system interface test, which was part of the Product Usability Test. During the test, it was noted that there was not enough information available about the access control system to make a proper signal connection with the vendor system. Some proprietary details were needed that could not be obtained within the time allowed for the test. To connect the systems, the facial recognition vendor needed to obtain details on the WIEGAND interface from the access control vendor. Since the WIEGAND protocol has many parameters that vary between systems, the facial recognition system could not be connected to the access control system without custom configuration. As a result, the Access Control System Interface Test was abandoned and no further results will be published in this report. Our conclusion is that anyone who wants to connect a facial recognition system to an access control system at this time should expect the process to include some custom development work.

M.2 FERET Images

Three of the major technical objectives of the Facial Recognition Vendor Test 2000 were to provide a comparison of commercially available systems, provide an overall assessment of the state of the art in facial recognition technology and measure progress made since the conclusion of the FERET program.

The comparison of commercially available systems needed to be designed and administered so that all vendors were on a level playing field and inadvertent advantages were not given to any participants. One of the methods used to ensure this in FRVT 2000 was to administer the test using sequestered images from the FERET program that had not been included in any previous evaluations. Any image set that was established for testing, however, has a certain life cycle associated with it. Once it has been used extensively and results using the data set have been published, developers start to learn the properties of the database and can begin to game or tune their algorithms for the test. This is certainly true of the FERET database; portions of it have been used in evaluations since August 1994. The FERET database has also been used in numerous other studies. To ensure a fair and just evaluation of the commercial systems in FRVT 2000, individual results for each vendor will be given using only those images that had been collected since the last FERET evaluations.

Another objective of the FRVT 2000 was to provide the community a way to assess the progress made in facial recognition since the FERET program concluded. There are two ways to measure progress. The best is to have the algorithms used in previous evaluations subjected to the new evaluation. Unfortunately, this was not an option for the FRVT 2000. The next best solution is to have the previous evaluation included in the current evaluation. This appears to be at odds with the goal of having an unbiased evaluation because those who participated in previous evaluations would have an advantage over those who did not. Because the goal is to measure progress and not necessarily individual system results, we can work around the potential conflict by reporting the top aggregate score from the experiments that used the FERET database.

The third goal - an overall assessment of the state of the art in facial recognition technology - can be inferred by looking at the combined results from the commercial system evaluation and the results using the FERET data.

M.3 Reporting the Results

For the Recognition Performance portion of this evaluation, the vendors were asked to compare 13,872 images to one another, which amounts to more than 192 million comparisons. The vendors were given 72 continuous hours to make these comparisons and then told to stop making their comparisons. C-VIS, Lau Technologies and Visionics Corp. successfully completed the comparison task. Banque-Tec completed approximately 9,000 images, and Miros Inc. (eTrue) completed approximately 4,000 images in the time allowed.

The complete set of 13,872 images and the corresponding matrix of 13,872 x 13,872 similarity scores can be divided into several subsets that can be used as probe and gallery images for various experiments. Probe images are presented to a facial recognition system for comparison with previously enrolled images. The gallery is the set of known images enrolled in the system.

Banque-Tec and Miros Inc. (eTrue) completed only a small number of the FRVT 2000 experiments and submitted only partial responses to several more. This forced the evaluation sponsors to decide how to accurately provide results from the FRVT 2000 experiments. The following options were considered.

Option 1 was to only release the results from the experiment that all five vendors completed (M2). This was rejected because this one experiment does not adequately describe the current capabilities of the commercial systems.

Option 2 was to release results from all of the FRVT 2000 experiments and only show the results from the vendors that completed each experiment. This would show the results for C-VIS, Lau Technologies and Visionics Corp. for all experiments and add the results from Banque-Tec and Miros Inc. (eTrue) for the M2 experiment. The sponsors chose not to do this because of the possibility that these two vendors may have received an added advantage in this category because they took more time to make the comparisons. Although the data collected does not support this hypothesis, the sponsors felt it would be better to not allow this argument to enter the community's discussion of the FRVT 2000 evaluations.

Option 3 was to change the protocol of the experiments so, for example, the D3 category only used the probes that all five vendors completed rather than the entire set. This option was rejected for the same reasons stated in Option 2.

Option 4 was to show the results from C-VIS, Lau Technologies and Visionics Corp. based on the full probe sets for each experiment and the results from Banque-Tec and Miros Inc. (eTrue) based on the subset that they completed. This option was rejected for the same reason stated in Option 2.

Option 5 was to fill the holes in the similarity matrices of Banque-Tec and Miros Inc. (eTrue) with a random similarity score or the worst similarity score that they had provided to that point. This option was rejected because the results generated would be horrendous and significantly skew the results that had been provided.

Option 6 was to show the results from C-VIS, Lau Technologies and Visionics Corp. and ignore the results from Banque-Tec and Miros Inc. (eTrue) for the FRVT 2000 experiments. This option was selected because it was the only one that was fair and just to those that had finished the required number of images and those that had not.

Appendix N
FRVT 2000 Evaluation Announcements

Sent: Friday, February 11, 2000 3:34 PM
Subject: Facial Recognition Vendor Test 2000

Numerous advances have taken place in the field of facial recognition since the last FERET test was performed in March of 1997. One of the most important of these advancements has been the introduction of facial recognition systems into the commercial marketplace. The competitiveness of the open market has brought forth numerous technological modifications to the algorithms that were available for the FERET program, and has also lowered the cost of the systems significantly. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications? This is one of the questions potential users most frequently ask the sponsors and the developers of the FERET program.

Although literature research has uncovered several mentions of recent system tests, none has been both open to the public and of a large enough scale to be completely trusted. This revelation, combined with inquiries from numerous government agencies on the current state of facial recognition technology, has prompted us to establish a new set of evaluations that will be performed in 2Q 2000.

The purpose of these evaluations is to accurately gauge the capabilities of facial recognition biometric systems that are currently available for purchase. The sponsoring agencies, as well as other government agencies, will use this information as a major factor when determining future procurement or development efforts. Participation in these evaluations is open to all facial recognition systems on the US commercial market.

Major sponsors of these evaluations include the DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), the Defense Advanced Research Projects Agency (DARPA), and Naval Sea Systems Command (NAVSEA). More information about the evaluations, as well as application forms to participate, can be found on the Facial Recognition Vendor Test 2000 web site at http://www.dodcounterdrug.com/facialrecognition.

Results List | Modify Posting

Modified: 2/11/00 12:30:00 PM

Category: Government Tests and Deployments

Sub-category: Facial Recognition

Vendor: Vendor

Title: Facial Recognition Vendor Tests 2000

Description: The DoD Counterdrug Technology Development Program Office, the National Institute of Justice (NIJ), the Defense Advanced Research Projects Agency (DARPA), NAVSEA Crane Division, and NAVSEA Dahlgren Division are sponsoring a test of commercial off the shelf (COTS) facial recognition products. The purpose of these tests is to accurately gauge the capabilities of facial recognition biometric systems that are currently available for purchase. The sponsoring agencies, as well as other government agencies, plan on using the results of these tests for both near-time acquisitions and future development efforts.

Participating vendors will run their own tests under direct supervision from the Government. Vendors will be allowed to make any adjustments they would like (while following test guidelines) on their systems prior to (each sub-) test initiation. The test will take place in the spring of 2000 at a location in the continental United States (the final location has yet to be decided).

Two categories of tests will be conducted: Recognition Performance Tests and Product Usability Tests. For each category, multiple tests will be performed to measure system performance in verification mode and in identification mode. The Recognition Performance Tests will use the FERET test methodology with a new database of images. The Product Usability Tests will evaluate performance in both low and medium security access control operational scenarios.

Final reports from these tests will be made available here in mid-2000.

Additional References

Related sites: Facial Recognition Vendor Tests

I would like more information on this technology

<< | < | > | >>

2 of 2

The DoD Counterdrug Technology Development Program Office (CDTDPO) began the Face Recognition (FERET) program in 1993. Dr. P. Jonathon Phillips was the assigned Technical Agent. At the time, very few individuals believed that facial recognition could become a viable operations support technology. Dr. Phillips and the CDTDPO Program Executive, could foresee the potential benefits of facial recognition technology and decided that those benefits far outweighed the development risk.

The FERET program consisted of three important parts. First was sponsoring research that advanced facial recognition from theory to working laboratory algorithms. Many of the algorithms sponsored by FERET form the foundation of today's commercial systems (FERET Transition). Second was the collection and distribution of the FERET database, which contains approximately 14000 facial images of 1200 individuals. The DoD Counterdrug Technology Development Program Office still receives requests for access to the FERET database, which is currently maintained at the National Institute of Standards and Technology. Portions of the FERET database have been distributed to over 100 groups outside the original FERET program. The final, and most recognized, part of the FERET program was the FERET evaluations that compared the abilities of facial recognition algorithms using the FERET database. The most recent reports from the FERET program are available for download on the FERET listing in the Counterdrug Technology Information Network.

The test methods used in the FERET evaluations form the foundation of an overall biometric evaluation methodology that was authored by Dr. Phillips, et. al., and published in the February 2000 edition of IEEE Computer. This evaluation methodology has been incorporated into the UK Biometrics Working Group in their "Best Practices in Testing Performance of Biometrics Devices". As clearly shown, the FERET program continues to have a profound effect on the facial recognition community today.

The biggest change in the facial recognition community since the last FERET evaluation in 1997 has been the introduction of facial recognition products to the commercial market. The competitiveness of the open market has brought forth numerous technological modifications to the algorithms that were available for the FERET program, and has also lowered the cost of the systems significantly. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications?

Repeated inquiries from numerous government agencies on the current state of facial recognition technology have prompted the DoD Counterdrug Technology Development Program Office to establish a new set of evaluations. The Facial Recognition Vendor Test 2000 (FRVT 2000), is co-sponsored by the DoD Counterdrug Technology Development Program Office, the National Institute of Justice, and the Defense Advanced Research Projects Agency (DARPA). The FRVT 2000 will be administered in April-May 2000 and will assess the capabilities of facial recognition systems that are currently available for purchase on the U.S. market. Results from the FRVT 2000 will be made available to the public. This evaluation will provide the counterdrug community and Government agencies with information that will assist their efforts of determining where facial recognition technology could best be used in the field. The results will also provide a blueprint of needed development efforts for the government and the vendor community.

Dr. Phillips also continues to be very active in the facial recognition and biometric community, as he has been named the Program Manager for DARPA's new HumanID program. Dr. Phillips was also an advisor in the development of the Facial Recognition Vendor Test 2000, and views this evaluation as one of the major transitions from FERET to HumanID. The HumanID program is a four year $50 million effort that aims to significantly improve the recognition capabilities of numerous types of biometric systems. By funding high-risk high-reward development efforts, HumanID will move biometric technology to its next logical step - the recognition of non-cooperative subjects with high accuracy. The DoD Counterdrug Technology Development Program Office is serving as a strategic partner for the HumanID program.

The FERET program was a highly successful effort that provided direction and credibility to the facial recognition community. We are just now beginning to uncover how important the program was during the infancy of facial recognition technology. As FERET nears the end of its transition from active program to a historical program, the DoD Counterdrug Technology Development Program takes great pride on the imprint it has left on the biometrics community, and even greater pride that the FERET ideals and evaluation methods are being used by current programs both inside the Program Office and by other Government agencies.

Appendix O
Sample Images

The following images were taken from the database used for FRVT 2000. They are shown here as a representative sample of the different methods used to capture images for the test. There were three basic variables in the image collection process: expression, lighting and media.

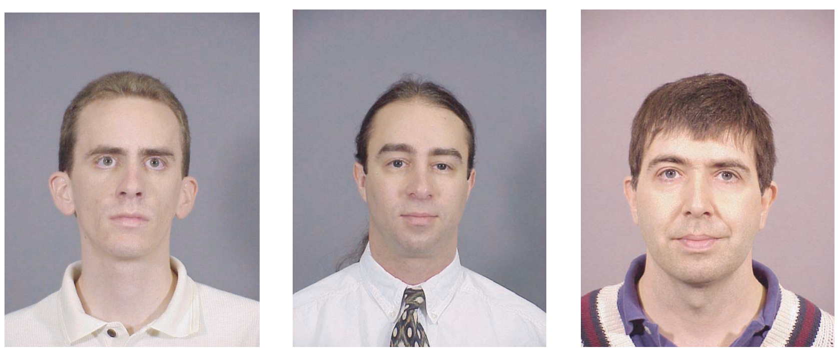
Subjects posing for the images were asked to use two different facial expressions. A normal expression, called *fa*, was used for some images. Another expression of the subject's choosing, called *fb*, was used for other images. The *fb* expression could be a smile, frown, grimace or other expression.

There were five types of lighting used for image collection. In the mugshot style, three flood lamps were used to illuminate the subject. In the FERET style, two flood lamps were used. A single flood lamp was used for the badge system lighting. Overhead fluorescent lighting was used for some images, while other images were taken outdoors using available daylight.

Images were collected using several types of media, including a digital still camera, a 35mm film camera, and a DV video camera. The 35mm film was later scanned to obtain digital images. Footage from the DV video camera was transferred digitally to a computer and still frames were selected for the database. For the badge system, the camera's analog video signal was captured using a computer with an installed frame-grabber card.

The techniques used to collect images are discussed in more detail in Appendix G.

FA Expression, FERET Lighting, Digital Still Media



FA Expression, FERET Lighting, Film Still Media

FA Expression, Mugshot Lighting, Digital Still Media

FA Expression, Mugshot Lighting, Film Still Media

FA Expression, Overhead Lighting, Digital Still Media

FA Expression, Overhead Lighting, Film Still Media

FA Expression, Badge Lighting, Framegrabber Media

FB Expression, FERET Lighting, Digital Still Media

FB Expression, FERET Lighting, Film Still Media

FB Expression, Mugshot Lighting, Digital Still Media

FB Expression, Mugshot Lighting, Film Still Media

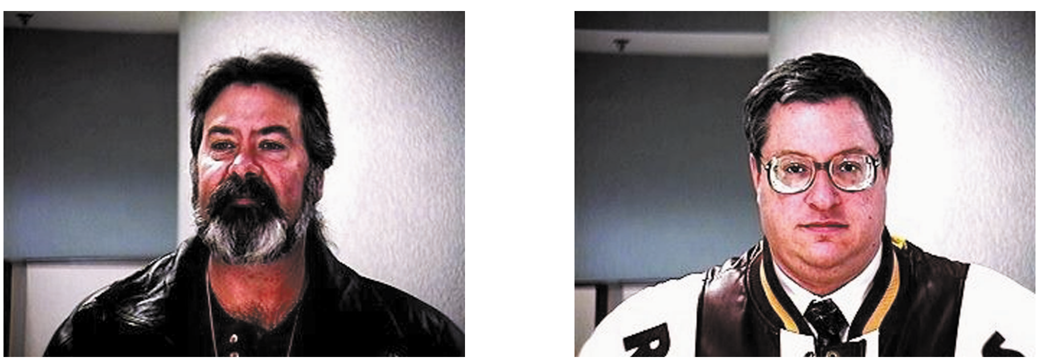
FB Expression, Overhead Lighting, Digital Still Media

FB Expression, Overhead Lighting, Film Still Media

Outside Daylight Lighting, Digital Still Media

Outside Daylight Lighting, Video Still Media

Overhead Lighting, Digital Still Media, December 1998



Overhead Lighting, Digital Still Media, November 1999

Appendix P
Vendor Product Descriptions

BANQUE-TEC

International Pty Limited

ACN 072 066 641 ABN 85 072 066 641

Summary Description of FRVT System
Banque-Tec International

Overview of System

The system constructed for FRVT consists of both hardware and software components. The hardware is essentially a PC equipped with a frame grabber and a video camera. Two software programs were provided to cover the FRVT 2000 requirements *RPTprog* for the *Recognition Performance Test* and *Eidolon* for the *Product Useability Test.* Both make extensive use of CSIRO's SQIS API. Each component is discussed in more detail below.

Hardware

The hardware used in the FRVT 2000 is listed in Table 1. Considerable computer power was required due to the nature of the tests and the more modest requirements of our commercial system are shown for comparison.

	FRVT	**Commercial**
Computer	Dell Precision 420 dual 600 MHz Pentium III	Minimum 266 MHz dual Pentium II
Memory (Ram)	512 Mb	Minimum 128 Mb
Frame Grabber	Matrox Meteor	Matrox Meteor II
Camera	Sony SSC-DC50AP	Any PAL CCD camera
Lens	Avenir SLO8551	Customer choice

Table 1: Hardware configuration

It is readily apparent that the requirements of the commercial system are somewhat less critical than those of the test system. The primary reason for this is that a real door access system is not required to capture faces over the range 4í to 12í so the additional computer power required for face location at longer ranges is not required.

A full costing of the system used in FRVT is as follows:

1. Dell Precision 420 Workstation with 512 Mb Ram: AUS$8816.00
2. Matrox Meteor II frame grabber: AUS$1420 (Meteor: AUS$1086)
3. Sony SSC-DC50AP : AUS$1629
4. Avenir SLO8551 Lens: AUS$480

Unit 5, 12-18 Victoria Street East, Lidcombe, NSW 2141 Australia PO Box 347, Lidcombe, NSW 1825, Australia

Telephone: (02) 9749 4999 Facsimile: (02) 9749 5100 Email: iallen@banquetec.com http://www.banquetec.com

World Business Centre, Newall Road, London Heathrow Airport, Middlesex TW6 2RJ, United Kingdom

Telephone: +44 (0) 181 263 2770 Facsimile: +44 (0) 181 263 2701

Software

SQIS API

The SQIS API has face detection; identification and verification capabilities designed in an object-oriented fashion and implemented using industry standard C++. Although the SQIS API is currently only available for the PC platform it has been designed to be readily ported to most other hardware. Figure (1) illustrates the basic functionality of the API. The Face Locator module (FLM) provides functions to locate a potential human face in a video stream or a still frame and to then locate the eyes of any found faces. The Face Verification module (FVM) provides the core functions for comparing one face against another and the functions required for enrolling Operators into the system. The Database module (DM) provides a convenient mechanism for dealing with face databases and comparisons against multiple faces. Routines are provided to convert a facial image into a representation suitable for comparison against other faces. This process is known as *encoding.*

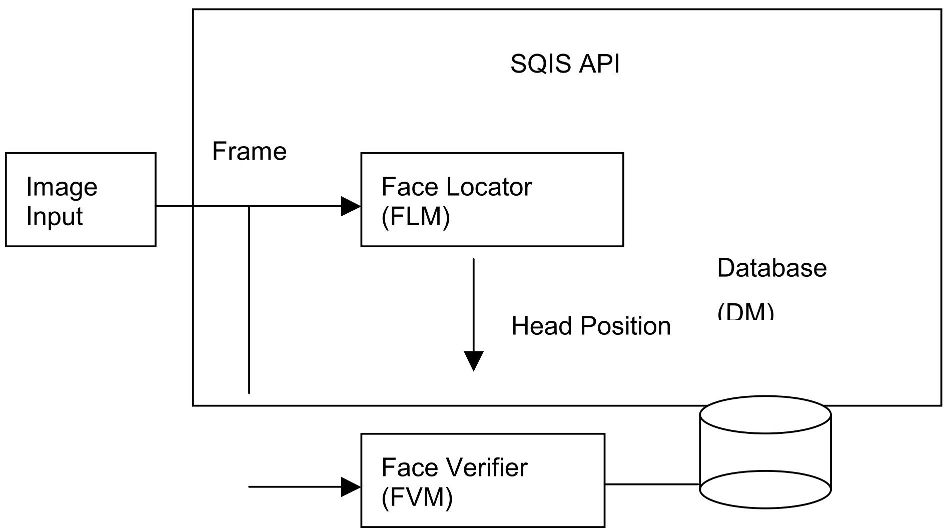


Figure 1: Block diagram for SQIS API

RPTprog

The *RPTprog* program is designed to read all JPEG images in a designated directory and perform one-to-one comparison between each image. The output is in the form of SIM files as specified in the FRVT 2000 documentation. It is hardened against system crashes and can be restarted at any point. The basic operation is as follows. All images in the designated directory are read, eye-located and encoded. The encoded images are then stored in the database. Test images are then read, eye-located, encoded and compared against the database. Finally, the output SIM file is written.

BANQUE-TEC International

BANQUE-TEC

International Pty Limited

ACN 072 066 641 ABN 85 072 066 641

Eidolon

The *Eidolon* program is a GUI based face recognition application designed to process incoming video streams at near frame rate. Such performance is possible through the use of multi-threaded programming techniques that make optimal use of the two system processors. Control of the program is through interaction with various dialogue boxes that allow certain system parameters to be set and modified on the fly. Success or failure of the system to complete a match is indicated visually. The operation of *Eidolon* may be broken up into three parts:

1. Enrolment; which may be further divided into two subcategories

    (a) Enrolment from a set of still images in which images are read in, eye-located, encoded and stored in a database, and

    (b) Enrolment from a live video stream in which the system operator selects asks the subject to stand in front of the camera and the captured video frames are eye-located, encoded and stored in a database. The enrolled face images are then presented to the operator who can carry out a manual quality check, deleting those images that are deemed to be below standard.

2. Verification (one-on-one comparison) is the mode of operation normally used in an access control system. A person desiring entry presents an identifying credential to the system which then checks the personís identity based on data derived from that credential. In the case of *Eidolon* the derived data is the enrolled Operators encoded face. The verification process proceeds as follows:

    Operator modifies system parameters from defaults if required

    Operator selects subject to be verified from database list

    Operator starts verification process

    Verification stops when a match with the subject is obtained, a preset number of frames have been tried or an operator set time limit is reached. In either of the latter cases a non-match is recorded.

3. Identification (one-on-many comparison, ie, database search) is similar to verification except that the only credentials presented to the system are captured images of the subjectís face. The *Eidolon* system then checks whether that Operator has previously been enrolled in the system by performing a search over all enrolled Operators. The basic operation is as follows:

    Operator modifies system parameters from defaults if required

    Operator starts identification process

    Identification stops when a match with the subject is obtained, a preset number of frames have been tried or an operator set time limit is reached. In either of the latter cases a non-match is recorded

Mr. Chris Burke

Director (Banque-Tec International Pty Limited)

Unit 5, 12-18 Victoria Street East, Lidcombe, NSW 2141 Australia PO Box 347, Lidcombe, NSW 1825, Australia

Telephone: (02) 9749 4999 Facsimile: (02) 9749 5100 Email: iallen@banquetec.com http://www.banquetec.com

World Business Centre, Newall Road, London Heathrow Airport, Middlesex TW6 2RJ, United Kingdom

Telephone: +44 (0) 181 263 2770 Facsimile: +44 (0) 181 263 2701

FaceSnap Recorder – General Overview

The FaceSnap recorder digitizes standard video sources and performs real-time pattern recognition on the captured images. The FaceSnap recorder detects human faces in the video stream and stores a user-adjustable image region around each face. A powerful image database facilitates easy navigation and search on large numbers of stored images. For analyzing the face image database, interactive and automatic search functions are provided. As an option, the FaceSnap recorder can perform real-time face recognition (requires two product units). The FaceSnap recorder is a revolutionary solution for person-oriented video surveillance, access monitoring and other applications requiring face image recording and face recognition.
There are several modes of operation:

1. Standard recording mode
   The FaceSnap recorder screen shows the live camera image and a number of most recently recorded persons. The user can switch between a number of different video inputs. The captured face images are displayed in a normalized format and can be stored in real-time on the local disk.

2. Replay mode
   In replay mode, the user can easily navigate through a large amount of image data. Images are time stamped and can be directly accessed through their time keys or the graphical event histogram that provides an overview of the temporal distribution of recorded images.

3. Search mode
   In search mode, the user can define image groups as a general purpose tool for data management. In addition, image groups are used to define training sets for face identification. Automatic functions for image pattern recognition are provided to search the face image database and support content-based data management.

4. Training mode
   The FaceSnap recorder can be trained to recognize faces based sample images collected by the user. In training mode, an image group can be imported to establish or extend a training set for a particular person. There are two hyper classes of persons: a) the class of all persons known to the FaceSnap recorder and b) the class of all persons the FaceSnap recorder is supposed to look for.

5. Selective recording mode
   Face images captured by a FaceSnap recorder operating in recording mode can be sent to a second FaceSnap recorder over a high-speed 100BaseTX network connection in real-time. The second system, which has to be put in the selective recording mode, receives the stream of face images and tries to identify all persons belonging to the hyper class of "active/interesting" people. Only those images for which the similarity value exceeds a user-defined threshold are displayed and optionally stored in the database.

Technology

The most important technology built into the FaceSnap recorder is the automatic selection of human faces in an image. The underlying technical problem is to detect and localize face-like patterns in a digitized image of an arbitrary scene. Moreover, the computer has to complete this search within fractions of a second since the person may not look straight toward the camera much longer than that. An additional requirement FaceSnap was designed to meet is the independence of object cues like motion or color. FaceSnap looks for face patterns in an image, not for regions showing a certain color range or being in motion relative to the image background. FaceSnap solves these problems by using a fuzzy abstract face model in combination with a neural net classifier. The outcome of the face finding stage are face image windows which are ranked according to a faceness value indicating how much the selected image portion resembles a human face
The face detection (face spotting) of the FaceSnap recorder works uses the FaceSnap technology as described above. The model-based pattern recognition has a processing rate of up to 25 images per second, depending on CPU type and size of the source image. Frontal views of human faces are preferred with a rotational tolerance of ca. +/- 15 degrees. Tolerance to deviations of the frontal face pose can be adjusted.

The face identification of the FaceSnap recorder uses the FaceCheck technology. The pattern recognition for face identification is based on user-provided training sets. The facial feature extraction is designed for frontal face views. Compensation of image rotation works up to +/- 15 degrees. Based on the location of prominent facial features, the face images are first geometrically normalized and then normalized with respect to the gray level distribution. From the normalized images, feature vectors are produced and then used for classification.

Screenshots of standard recording mode (left) and replay mode (right).

Screenshots of search mode (left) and training mode (right).

The submitted system consists of the following components:

Pos.	Quantity	Component	Price in Euro (total)
1.	2	FaceSnap recorder standard unit	10.123,60
2.	2	CD-Writer, incl. SCSI adapter	562,40
3.	1	High resolution camera (XC-8500), option package	1.508,30
4.	1	FaceCheck real-time option	1.789,50

The individual components are described in the attached price list.

Price list FACESNAP RECORDER

⇨ FACESNAP-Recorder

(Package according to specifications* without monitor)

€ 5.061,80
(DM 9.900,00)

Order options:

- CD-Writer (Teac CD-R58) as backup medium, incl. SCSI adapter. — € 281,20 (DM 550,00)
- MS WindowsNT4.0 (Workstation) Boot option, incl. additional harddrive for the NT-System (4.0GB). — € 383,50 (DM 750,00)
- Monitor, TFT 15", Type Goldstar 500LC. — € 1.073,70 (DM 2.100,00)
- Colour camera with zoom and autofocus, Type CCD-400E, Output: Composite video + S-video, remote control via RS-232. — € 741,40 (DM 1.450,00)
- Colour camera with zoom, autofocus, and pan/tilt function, Type Sony EVI-D31, Composite video + S-video, remote control via RS-232. — € 1.099,30 (DM 2.150,00)
- High resolution progressive scan black/white CCD-camera, Type Sony XC-8500CD, lens, software extension, additional framegrabber card — € 1.508,30 (DM 2.950,00)
- FACETRACK option, special twin camera with pan/tilt function and additional camera, software extension, additional framegrabber card — € 2.454,20 (DM 4.800,00)
- FACECHECK real-time option, software extension for online face recognition (two FaceSnap recorder units required for operation) — € 1.789,50 (DM 3.500,00)
- Additional video inputs: 2x Cinch/RCA for PAL/NTSC/SECAM (composite), 1x Mini-DIN for S-video. — € 127,80 (DM 250,00)

Price list as of 01.04.2000, C-VIS reserves the right to change prices without notice. F.O.B. shipping point. Buyer is responsible for all tariffs, customs, taxes and transport & shipping costs. Delivery follows after payment by either credit card (Eurocard/ Mastercard), cheque or cash. Delivery time: 2 weeks after receipt of written order.

*** Technical specifications** (minimal delivery features, C-VIS reserves the right to change product specifications without notice):

Hardware platform: ATX Desktop PC, Dual Pentium III System (2 x 600 MHz), 64MB SDRAM, 20GB EIDE HDD (18GB useable for image data), LS-120 FDD, SVGA Graphics Card, PCI 10/100-Mbps Ethernet Card, PCI Video Capture Card, 3-Button computer mouse.
Video inputs: 2 x Cinch/RCA-Connector for PAL/NTSC/SECAM (composite), 1 x Mini-DIN for S-video.
Video output: VGA-Connector, Output resolution 800 x 600, 64K colour @ 72 Hz,
Network connection: RJ45-Connector for LAN-Integration (Ethernet 10/100Mbps).

Face spotting: FACESNAP® Technology, Model-based pattern recognition, Processing rate up to 25 images per second, Tuned to frontal views of human faces, Rotational tolerance ca. ± 15°, Adjustable tolerance to deviations from frontal face pose.

Face identification: FACECHECK® Technology, Pattern recognition based on user-provided training sets, Facial feature extraction designed for frontal face views, Compensation of image rotation up to ca. ± 15°.

User Manual: English or German, Online help

Face Recognition Vendor Test 2000
eTrue TrueFace Overview

eTrue, Inc.
144 Turnpike Road, Suite 100
Southboro, MA 01772-2121 USA
Phone: 508-303-9901
Fax: 508-303-9902
Email: info@etrue.com
Web: http://www.etrue.com

Submission: TrueFace API 4.0 SDK
Test Date: May 15-19, 2000

TrueFace Overview

The eTrue (formerly Miros) TrueFace API version 4.0 Software Development Kit (SDK) represents the latest face recognition technology from eTrue. It offers image acquisition and manipulation from Video-for-Windows (VFW) compatible video sources as well as reading standard image file formats. TrueFace is the only face recognition product certified by the International Computer Security Association (ICSA).

The TrueFace SDK gives applications the ability to find faces in images and to perform facial verification and identification. The software runs on Windows 9x, NT, and 2000. To speed development, the software includes several working sample applications written in Microsoft Visual C++ and Visual Basic along with source code.

The TrueFace SDK can be purchased with one of three different licenses:

1. **Locate** – For applications that only want to find faces in images.
2. **Verify** – For applications that want to verify a user's face against a claimed identity . Includes the Locate functionality.
3. **Identify** – For applications that want to identify a user's face w ithin a database of users. Includes both the Locate and Verify functionalities. The Identify licensing is priced based on the desired number of users in the database.

The hardware requirements for the TrueFace SDK are as follows:

1. A PC running Windows 95, 98, NT or 2000.
2. The preferred processor is an Intel Pentium III or higher. The Intel Pentium II and the Celeron family also will work as well as all comparable AMD processors.
3. At least 32 MB of RAM.
4. A Video-For-Windows (VFW) compatible camera and driver. This includes all USB, parallel-port, and frame-grabber acquisition sources that have VFW drivers.

An overview of the face finding and matching capabilities of the TrueFace SDK is given in the following sections.

Face Finding

When processing an image, TrueFace first finds the face using a combination of very efficient neural networks. The image can be either color or black-and-white in any standard image format and of any size. When processing a video stream, TrueFace tracks the movement of the faces in the field of view, thereby increasing system throughput. After finding the face and locating the eyes, TrueFace generates a binary face template that can range from 500 to 2000 bytes in size, depending upon the accuracy desired.

Face Matching

The matching algorithm in the 4.0 version of the TrueFace API is based on an adaptive matching technique using neural networks. Matching times between two face templates range from under 0.1 ms to 4 ms on a Pentium III 800 MHz PC, depending on the template size. All of the face templates for a single person can be aggregated together into a person template. A person template allows the common properties of the individual face templates to be used to speed up the matching process further. In addition, person templates can reduce the time needed to update the enrollment database when upgrading to a new TrueFace version.

Face Recognition Vendor Test 2000

For the FRVT2000, the core face recognition engine in the TrueFace SDK was used for both the Recognition Performance Test and the Product Usability Test. However, the test application software used to access the TrueFace SDK was different between the two tests because of the requirements for the two tests. For both tests, the test application software is described in more detail in their respective sections below. For both tests, full-size face templates were used for the highest accuracy.

The PC platforms used in both tests were not identical but were very comparable. Both PCs used the same processor (Pentium III at 800 MHz), had a 133 MHz front-side bus (FSB), and had 256 MB of RAM. However, the type of RAM differed between the two machines (SDRAM vs. RDRAM). They both used a 20 GB Ultra ATA-66 (7200 RPM) hard drive and ran Windows NT 4.0 Workstation with Service Pack 5. On sample runs for the Recognition Performance Test, both PCs performed nearly the same.

Recognition Performance Test

The test application for this test was a custom console-based (command line) Win32 application written in C++. It first generated face templates for all the images in the test set. Then, it performed the matching between the face templates and collated the match output scores according the FRVT2000 specification for this test. Finally, it generated the required output similarity files.

The component list for the system submitted for this test is given in the following table:

COMPONENT	DESCRIPTION	COST
Software	eTrue TrueFace API v. 4.0 Software Development Kit with an Identify engine with a database limit of 100,000 users. Driven by a custom command-line program.	See pricing note below
Computer	Dell Dimension XPS B800R, 800 MHz Pentium III, 133 MHz FSB, 256 MB 266 MHz RDRAM, 20 GB ATA-66 7200 RPM hard disk, 17 inch monitor, WinNT 4.0 SP5.	$2750

Product Usability Test

The test application for this test was TrueFace ID version 2.5.

The test scenario is an access control application. TrueFace ID is intended for use as a tool to identify people under suspicion, usually in public settings, and to notify security personnel if a possible match has been found. Because fraud by photograph is typically not a problem in public settings, TrueFace ID has no photograph detection capability. A "Verify" mode allows for verification in addition to identification.

Enrolling images into TrueFace ID is easy. It can accept images in files (any standard format) or previously captured video images. It also can be used to enroll cooperative subjects in a controlled setting and uncooperative subjects under surveillance. For a cooperative subject, the person simply needs to look at the camera for a few seconds while the software captures 8 images of the person. For an uncooperative subject under surveillance, TrueFace ID stores the last 20 faces found, any of which can be enrolled into a person's database record.

When identifying faces in a stream of video images, TrueFace ID scans the input images and attempts to continuously identify the best faces found. Visually, the software displays the whole image, the cropped-out facial image, and the list of possible matches from the database, sorted in order of decreasing match score. All matches above a low threshold are displayed. Any matches above a high threshold can generate an audible alert for notifying security personnel. Furthermore, all matches above either the high or low threshold can be saved to a database event log for later review.

The hardware requirements for running TrueFace ID are the same as for the TrueFace SDK. However, the minimum amount of memory for TrueFace ID is 128 MB of RAM. For a database size larger than 1,000 people, additional RAM is recommended.

The component list for the system submitted for this test is given in the following table:

COMPONENT	DESCRIPTION	COST
Software	eTrue TrueFace ID v. 2.5 (uses TrueFace API v. 4.0 with an Identify engine) with a database limit of 100,000 users.	See pricing note below
Computer	Micron Millennia MAX GS133, 800 MHz Pentium III, 133 MHz FSB, 256 MB 133 MHz SDRAM, 20 GB ATA-66 7200 RPM hard disk, 17 inch monitor, WinNT 4.0 SP5.	$2300
Video Capture	Coreco Bandit frame grabber and display card with camera cable.	$870
Camera	Hitachi VCC-151 color camera.	$625
Lens	Computar T6Z5710-CS 5.7-34.2 mm 1:1.0 zoom lens or Fujinon-TV CF12.5A 12.5 mm 1:1.4 fixed focal length lens (CS-mount).	$205

Pricing Note

eTrue was asked to supply the price of its TrueFace software for this report. Unfortunately, we were unable to comply with this request for the following reasons: The price of the TrueFace software varies considerably depending on which application it will be used for, which country we sell it to, what support we provide with the product, how many units are ordered, and other considerations we negotiate with our customers. The price can fluctuate dramatically over a short time frame in this competitive market place, especially during a period of quickly changing market demand that we are in today. To be fair with all our corporate customers, we provide our price on an individual customer basis. However, our aim is to be the least expensive provider of face recognition, with our range of performance, in the entire industry.

Overview of HUNTER™

A Facial Recognition Surveillance System

LAU Technologies
30 Porter Road
Littleton, MA 01460

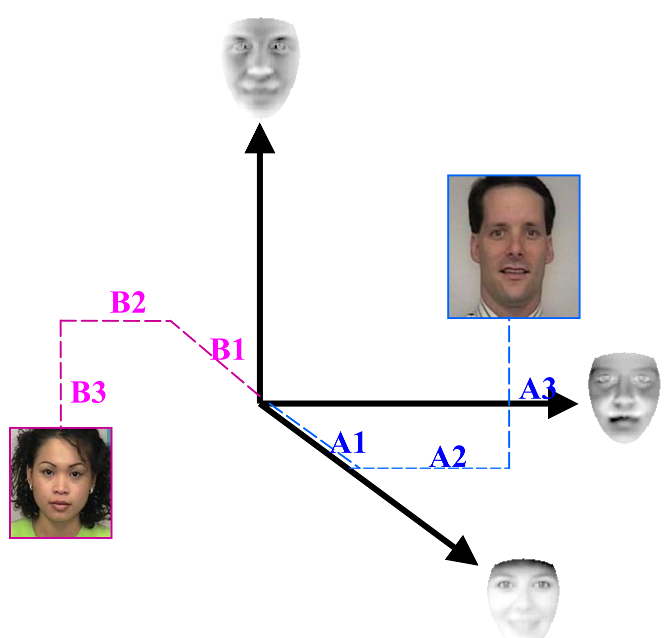
1.0 Lau Technologies Core Facial Recognition Technology

The facial recognition technology developed at the Massachusetts Institute of Technology (MIT) and exclusively used by LAU Technologies employs "Eigen faces", which are characteristics of a person's face, and maps the facial image into a multi-dimensional face space. Using special techniques developed by LAU the Eigen faces are used to provide high speed facial matching to one or many candidate faces in a database.

Figure 1.1: Eigen Faces

The multidimensional Eigen space itself is determined through a separate process at Lau Technologies that is typically done only once. This process begins with a large diverse population of thousands of facial images. For each of these images, the head and eyes are located; the image is standardized, and then masked. The resulting ensemble of localized, standardized and masked images are then processed with a mathematical technique called Principal Components Analysis. The result of Principal Components Analysis a set of face-like images called Eigen faces. Each Eigen face is mathematically independent/orthogonal to all others, and is an independent degree of freedom for describing faces. In other words, these face-like images are the most efficient set of building blocks needed to build any face. Figure 1.1 shows the first five Eigen faces that result when Principal Components Analysis is performed on a large sample of face images.

Figure 1.2: Face Space

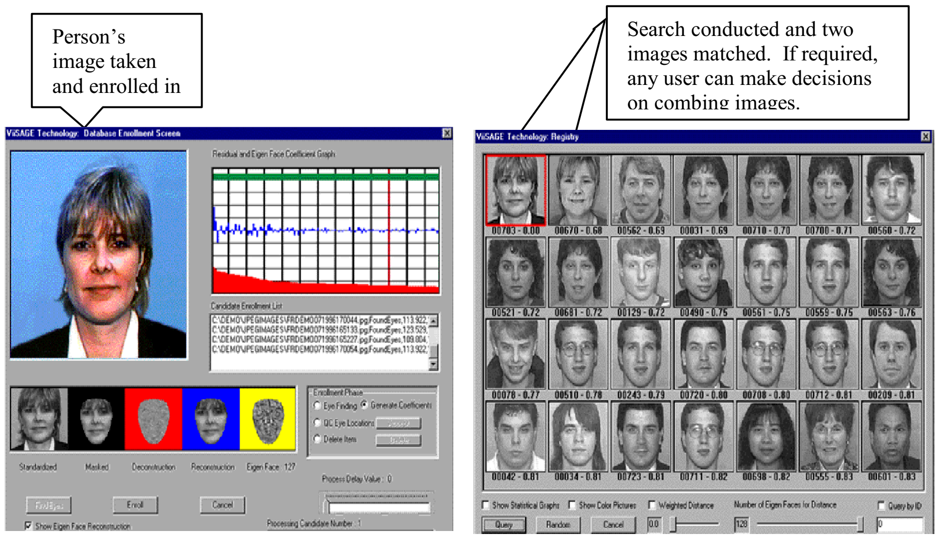


After standardization and masking, the image is projected into the multidimensional Eigen space of facial recognition as shown in Figure 1.2. The result of this projection of the facial image onto the Eigen face templates is a set of Eigen coefficients, which together form an Eigen vector. The multidimensional Eigen space is constructed of 128 mathematically orthogonal coordinates and each coordinate is representative of a single characteristic Eigen face.

The first coefficient of the image being enrolled is calculated through the projection of that image onto the primary Eigen space coordinate, which is also referred to as the average Eigen face. Once determined, the first coordinate projection is subtracted form the original image in order to produce a residual image. This residual image is then projected onto the second designated Eigen space coordinate and thus the second coefficient is obtained. Then the second projection is subtracted from the previous residual image in order to produce a new residual image. Successively, each new residual image is projected onto the next coordinate. Each projection subtracted produces a further deconstructed residual image.

This process of projecting the resulting residual image onto each coordinate produces a total of 128 characteristic Eigen coefficients. Together this set of characteristic coefficients represents a complete vector projection in the facial recognition Eigen space. The combination of these 128 coefficients with respect to their corresponding Eigen face images produces a reconstructed masked image that can be viewed and verified as being visually very similar to the original masked image.

The captured or scanned digital image is converted to Eigen coefficients (simply described as facial features) and submitted to the Facial Recognition Search Engine, which returns results in real-time to the requesting client. These results comprise the closest matches found for the individual. They are ranked in order with the closest match shown first.



2.0 Product Usability Test

Figure 2.1 shows a block diagram of a Facial Recognition Access Control system, operating in verification (1 to 1 matching) mode. The user is prompted for his or her ID number in the Access Control Entry Screen. Although this example shows a screen where an ID number is typed on a keypad, alternative means of asserting identity include magnetic strip card, or RF proximity badge. Once the ID is asserted, the system continuously searches for heads and eyes for a fixed time period. Each time that a face is found, it is standardized and converted to a set of Eigen face coefficients, and compared to the database of one or more reference images corresponding to that ID number. If the captured face matches one or more reference images with a degree of confidence, which exceeds a customer-defined threshold, then a GO decision is made, and access is granted. Otherwise, face images will continue to be captured until the timeout period. If no match is found before the timeout period, then a NO GO decision is made, and access is denied. In either case, an entry is stored in a log file for immediate or subsequent review by a system administrator. For each transaction, the log file stores a pair of images showing both the enrollment image used for the match, and the live capture image. This information can serve as a valuable audit trail to understand events where unauthorized access was attempted.

Facial Recognition Access Control System

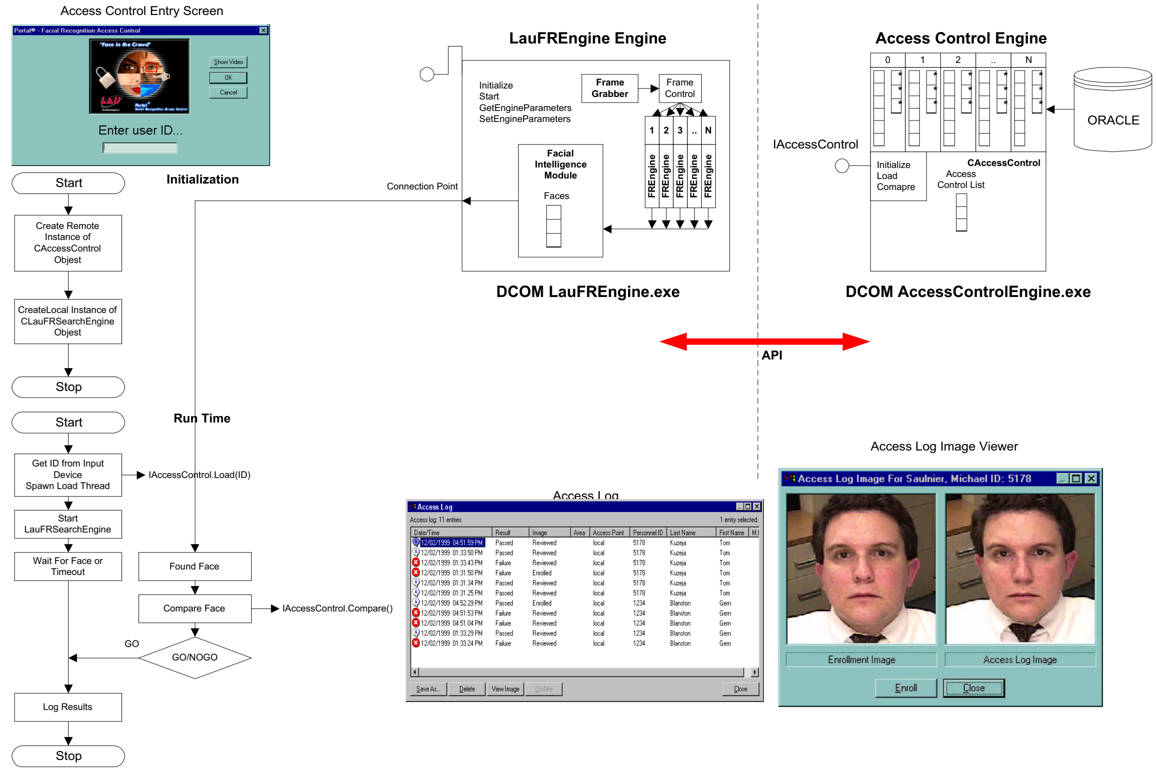


Figure 2.1 Access Control Flow Diagram

3.0 Hunter Surveillance Product

The Hunter system automatically acquires (finds the face) and recognizes subjects in real time. The surveillance application continuously searches the camera's field of view for heads. When a head is found the eye locations are calculated, and the face is converted to an Eigen vector. For each successful acquisition the subject's face is displayed in the top half of the GUI. If a face is captured that is sufficiently close to a face on the watch list (pre-enrolled subjects), then the potential match is display for consideration by the operator.

The facial biometric data stored will include Eigen coefficients, compressed standardized images and eye locations of the original image. Since each coefficient is 2 Bytes in size, a set of Eigen coefficients for a single face is a total of 256 Bytes. These characteristic coefficients are used solely for the purpose of facial recognition searches and verifications. The eye locations and standardized images are used for review and future enhancement purposes, but they are no longer required for the performance a facial biometric match. This biometric data can be stored with relational links to corresponding demographic data. This demographic data can be utilized to enhance or modify facial search results

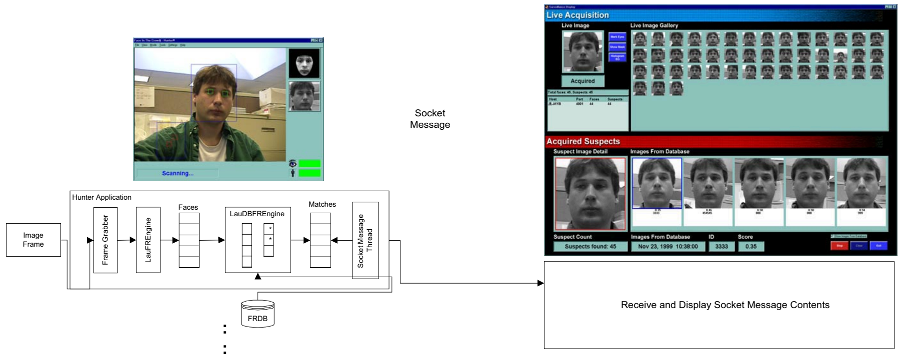


Figure 3.1 Surveillance Block Diagram

4.0 Products and Pricing

Hunter Surveillance System

The entry level Hunter software product is designed to operate as a stand alone surveillance system. The product is shipped on a CD with a decoding software dongle. The price of Hunter in this configuration is $1000 U.S.

Hunter runs on a PC running Windows NT platform, and uses COTS video hardware. Lau will be pleased to make hardware recommendations for frame grabbers, cameras and computers.

Cost of system components used in FRVT2000.

Recognition Test

Dell XPSB1000r	$3908
Lau Technologies Hunter™ Surveillance System	$1000
	$4908

Usability Test

Dell XPS B866	$2117
Lau Technologies Hunter™ Surveillance System	$1000
Hitachi VK-C77 video camera	$1000
Matrox Meteor II Frame Grabber	$ 500
	$4617

Minimal Configuration

Dell L66RN	$1158
Lau Technologies Hunter™ Surveillance System	$1000
Sony EV-400	$ 400
Matrox Meteor II Frame Grabber	$ 500
	$2958

VISIONICS CORPORATION
FaceIt® FACE RECOGNITION TECHNOLOGY

One Exchange Place, Suite 800
Jersey City, New Jersey 07302

FaceIt® Technology used in Facial Recognition Vendor Test 2000

All Visionics Corporation's products are derived from the algorithms in the FaceIt®Identification Software Developers Kit (SDK).

Recognition Performance Test

The system used for the Recognition Performance Test was comprised of an application written with the FaceIt® Identification SDK and a computer running the Windows NT operating system. The FaceIt® application performs one-to-many and many-to-many facial matches on stored images. It is a high-speed, high-accuracy engine designed for checking the integrity of databases and for prevention of identity fraud.

The application used for the Recognition Performance Test includes algorithms for automatic face segmentation from an image (face finding and alignment), facial template creation, and template-to-template matching. This functionality is provided to developers using the SDK in two COM objects: FaceItLocate and FaceItRecognize. The console application built with these objects consists of four simple modules that thinly wrap the SDK objects. The first module does not actually rely on the SDK, but is used to read images from a directory and create a text database with links to each image location, as required for the Facial Recognition Vendor Test 2000. The next module uses the FaceItLocate object to find faces in the images and pinpoint the location of the eyes. The remaining modules use the FaceItRecognize object. The third module creates facial templates, the biometric FacePrints. The fourth and final module performs matching operations and generates the similarity files. The code for this application is available with purchase of the Identification SDK.

Face Finding and Template Creation
On a single 500MHz Pentium III computer, up to 60 images can be pre-processed (face finding, alignment and template creation) per minute; very complex images may require more pre-processing time. Visionics face finding technology is able to find heads at a very wide range of sizes and in complex, real-world scenes.

FaceIt® technology employs **Local Feature Analysis** (LFA) to represent facial images in terms of local, statistically derived building blocks. Identity is determined not only by which elements are characteristic of a particular face, but also by the manner in which they are geometrically combined (i.e. their relative positions). LFA is a mathematical technique that enables high accuracy facial matching and is robust with respect to variations in lighting, facial expression, hairstyle and pose.

Facial Matching Speed
LFA operates in two modalities: Vector, which uses a very compact representation of the face, and Intensive, which uses a more rich representation. Search speed using the Vector modality is 47,000,000 matches per minutes on a single 500 MHz Pentium III computer. Search speed using the Intensive modality is 10,000 matches per minutes on the same CPU. The accuracy of the two modes is equally high, except when image quality is poor. In this case, the Intensive mode of LFA may provide superior performance.

For large database searching applications where some images are poor quality, we recommend use of a two-pass search strategy in order to optimize both speed and accuracy. First, a rapid pass is performed over all records using the Vector mode. Results are sorted in order of the confidence that the comparison was a match. Then a second pass is performed using the Intensive mode to search some fraction of images that yielded the highest confidence of a match in the first pass. The fraction used in the second pass is a tuning parameter that enables one to trade-off between speed and accuracy when image quality is sub-optimal. We utilized this technique in the Recognition Performance Test, specifying that the top 15% of the images from the first pass be searched again in the second pass.

Component List
- Application built from FaceIt® Identification Software Developers Kit
- Dell PowerEdge 6300 computer (only a 400 MHz single processor required)

Cost Breakdown
The total cost of the system used for the Recognition Performance Test was $26,660***** in April, 2000. This figure includes the cost of the SDK used to create the facial recognition application.

The cost of the FaceIt® Identification SDK from Visionics is $9,995.

The computer used was a Dell PowerEdge 6300 with four Pentium Xeon 550 MHz CPUs, 512 Cache, 1GB RAM, an 18GB SCSI Hard Drive, CD-ROM drive, 15 inch monitor and a 2GB Iomega Jaz Drive. This system was priced at $16,665 from Dell's website. *****Note that while we chose to use a relatively high-end computer for the purpose of performing nearly 200,000,000 matches in 28 hours, the Recognition Performance Test can be run on any Pentium class computer. Typically we recommend a single Pentium III 400 MHz processor as the best compromise between speed and price. However, FaceIt® technology is fully scalable as shown by our ability to run our many-to-many engine on a quad processor.

Product Usability Test

The system used for the Product Usability Test was comprised of an application called **FaceIt® Surveillance**, a computer running the Windows NT operating system and a Video for Windows compatible video capture system. FaceIt® Surveillance utilizes the same core algorithms contained in the Identification SDK, but is a product designed for real-time face finding in video and automatic searching for facial matches in a watch list. This application is available for purchase from Visionics Corporation.

FaceIt® Surveillance includes algorithms for automatic face segmentation from a video image (face finding and alignment), facial template creation, and template-to-template matching. FaceIt® Surveillance is an intelligent software solution that is designed to compliment and enhance existing CCTV systems by automating and improving the routine and arduous surveillance tasks performed by a human operator. FaceIt® Surveillance accepts as input either live or recorded (archived) video and performs one-to-many searches for the purpose of identification and to alert operators as to potential matches with members of a watch list.

Face Finding and Template Creation and Facial Matching Speed
Please refer to information above, in description for system used in Recognition Performance test.

In order to perform the Verification portion of the Product Usability test, we created a facial database containing one or more images of a single person. However, FaceIt® Surveillance was not designed for the purpose of one-to-one matching. Also, because this application is designed for automated surveillance, rather than access control, it does not include the **Liveness Testing** mechanisms that are employed by Visionics' technology for computer information security. These proprietary mechanisms enable FaceIt® to distinguish real faces from photographs of faces in access control verification scenarios.

Component List

- FaceIt® Surveillance
- Dell Precision 210 Workstation
- Canon VC-C3 camera
- Winnov Videum capture card

Cost Breakdown
The total cost of the system used for the Product Usability Test is $14,675.

The cost of FaceIt® Surveillance from Visionics is $9,995 - $24,995, depending on database size. A similar product, FaceIt® Sentinel, allows search "on-demand" (click on face of interest to initiate search for a match) and is available for $4,995 - $9,500, depending on database size. The most economical versions of Surveillance and Sentinel handle up to 1000 records each.

The computer used was a Dell Precision 210 with two Pentium III 600 MHz CPUs, 512 Cache, 384 MB RAM, an 18GB IDE Hard Drive, a CD-ROM drive and a monitor. A comparable machine, the Precision Workstation 220, is currently priced at $3131 from Dell's website. FaceIt® Surveillance was designed to run on a dual processor.

The camera used was Canon VC-C3 Communication Camera, which costs $1400.

The Winnov Videum VO PCI capture card costs $149.

For further information, please contact:
Kirsten Rudolph Nobel, Ph.D.
Government Liaison
kirsten@visionics.com
201-332-9213, ext. 207
www.visionics.com

Appendix Q
Vendor Position Papers on
FRVT 2000 Evaluation Report

The sponsors of the FRVT 2000 sent a copy of the Evaluation Report to the participating vendors on February 8, 2001. The sponsors were given the option to prepare and submit a position paper to be included in this appendix. The deadline for including their position papers was 9 A.M. EST on February 16, 2001.

The submitted position papers are included in this appendix without modification. Furthermore, the sponsors have decided not to comment on these papers except to correct one misconception. In the lighting section on page one of the Lau Technologies position paper, Mr. Cusak states, "An option was made available to participants to make use of auxiliary lighting equipment in the manner they saw fit. Most participants enthusiastically embraced this offer." Surprisingly, this was not the case becase none of the participating vendors used auxiliary lighting equipment.

Comments from C-VIS on the Facial Recognition Vendor Test 2000

The FRVT 2000 was the first independent, truly open, and published evaluation of commercial facial recognition systems. However, the preconditions for participation excluded research projects and implicitly required a number of advanced technical capabilities that are not found in all commercial systems on the market. We believe that the FRVT 2000 is an important milestone in the evolution of facial recognition technology and for the development of markets for its applications. As the sponsors correctly point out, the FRVT 2000 report cannot be used as a "buyer's guide for facial recognition". It does, however, give valuable insights on both the relative strengths and weaknesses of the tested systems and on the state of the art in general.

Our comments on the Facial Recognition Vendor Test 2000 focus on three questions:
1. Which results of the test are most significant for practical applications?
2. What do the results tell us about our product FaceSnap RECORDER?
3. What should be additional issues in future facial recognition tests?

Feasibility of Practical Applications

There are numerous existing and potential applications of facial recognition. We do not attempt to rank them according to some criteria of importance. However, there are two fundamentally different situations for which customers may want to have a facial recognition system: In the one category (A), both the reference image(s) (gallery set) and the probe set are taken from the same image source. This is typical for biometric computer login programs, for example. In the other category (B), a gallery set consists of images taken with different cameras and at different locations than the images of the probe set. In the FRVT 2000 report, one example for this situation is cited as a "mugshot vs. subsequent video surveillance scenario".

There is another fundamental distinction across all applications of facial recognition: Do we need a human supervisor or not? That is, does the customer need an interactive facial recognition system for assistance to a human operator, or is there a compelling need for a fully automatic system?

The FRVT 2000 results of the "Enrollment Time Test" and a number of experiments in the "Recognition Performance Test" clearly show that with current technology both interactive and automatic systems are feasible for viable commercial solutions to be operated in a category A situation.
If we look at automatic systems to be operated in a category B situation, the FRVT 2000 results clearly suggest that viable commercial solutions are not feasible today. This must be concluded from the results of the "Old Image Database Timed Test" and the "Distance Experiments" (D1-D3). As an example, a false alarm rate of 10% with only a ca. 50% chance of correct verification (average from the best system / Figure 22) is certainly not acceptable in most professional applications (just imagine an industrial inspection system with those error rates!).
If we look at interactive systems, the identification performance and the false alarm rate are much less critical in practice. In fact, a human operator searching a database does not gain significant convenience or time saving when presented with only 20 identification hits instead of 50, for example. That is, for category B situations other features and capabilities of a facial recognition product may be more important in practical applications than what has been tested in the Recognition Performance Test.

FaceSnap RECORDER

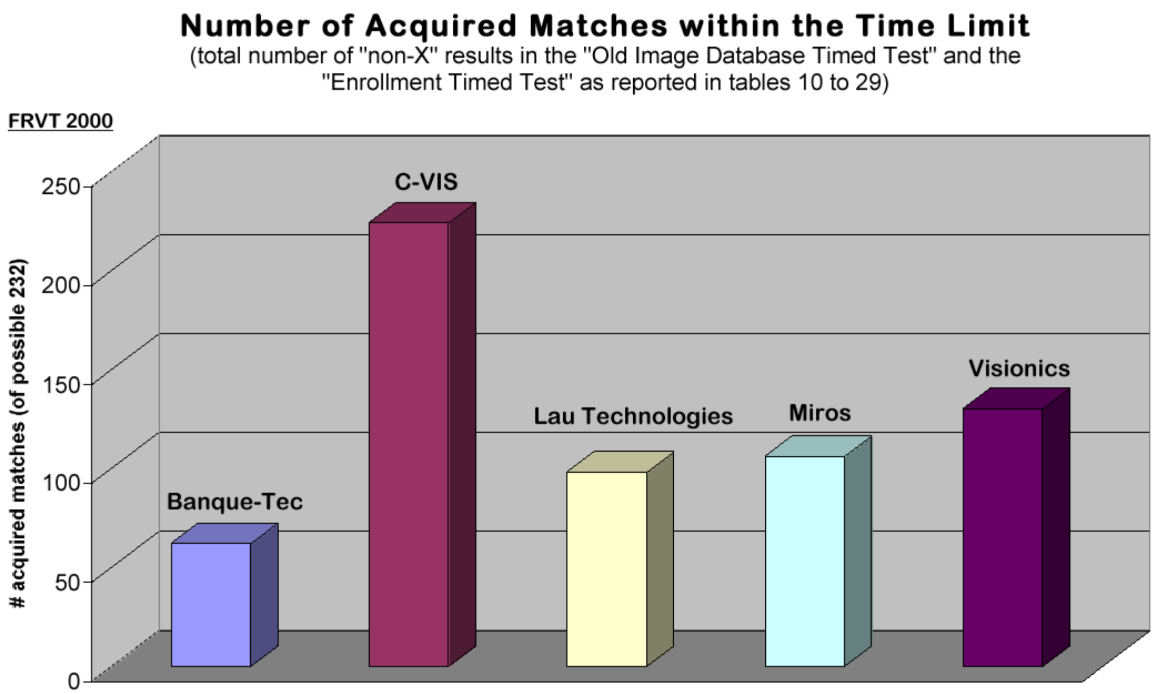
C-VIS submitted its product FaceSnap RECORDER to the Facial Recognition Vendor Test 2000. We felt that the addition of the Product Usability Test to the FERET style evaluations targeted products like the FaceSnap RECORDER. However, the FaceSnap RECORDER was not designed for access control and it is neither specialized on automatic database searches. The FaceSnap RECORDER is a digital video recorder with built-in real-time face capturing. The FaceCheck subsystem for face image search usually

Contact: Prof. Dr. Thomas Zielke • zielke@c-vis.com • Information: www.c-vis.com

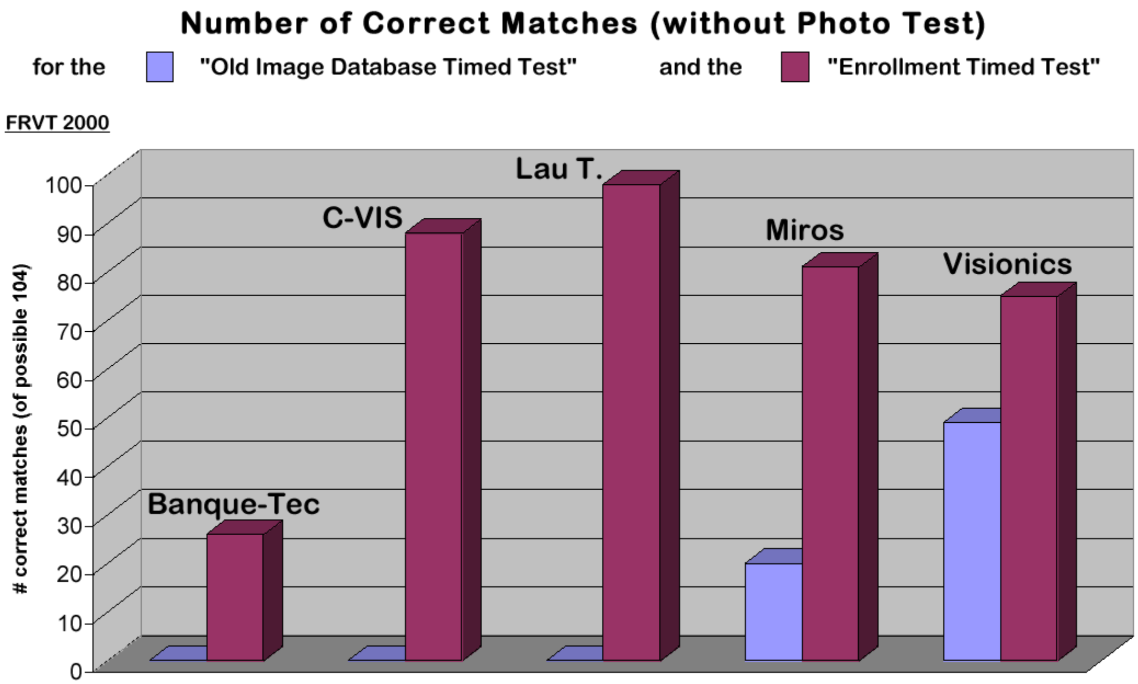
requires several sample images of each reference face. For the FRVT Recognition Performance Test it had to run in an atypical operational mode in which a direct comparison of just two face images is possible. Actually, this requirement of the FRVT 2000 initiated the development of a new software generation that has eventually become operational only in the beginning of 2001.

The FaceSnap RECORDER is being used for video surveillance, access monitoring, post-event analysis and a number of applications in law enforcement. The product philosophy is based on the realistic assessment that in most surveillance applications reliable automatic facial identification is not feasible given the current state of the art in facial recognition. Instead, the FaceSnap RECORDER delivers reliable real-time performance on the detection and recording of face images. Using this capability, police officers can find persons in time-lapse video recordings within minutes, as opposed to hours it takes them without a FaceSnap RECORDER.

In the FRVT 2000, the superior real-time performance of the FaceSnap RECORDER has become visible only in the results of the Product Usability Test. The report contains many comprehensive charts for the results of the Recognition Performance Test. Unfortunately, the reader of the report is left with the raw result data (tables 10 to 29) when it comes to the Product Usability Test. For this reason, we produced two charts that visualize the overall performance of the participating vendors in the Product Usability Test. The data were taken from the tables number 10 to number 29 on pages 46 to 55.
The first chart given below shows the total number of acquired matches in all experiments of the Product Usability Test. This number is an indicator of the reactiveness of the systems in the face capturing phase. In the FRVT 2000 experiments, the FaceSnap RECORDER was set up to always capture at least three face images before a recognition decision was made. We believe that the FaceSnap RECORDER is currently the best system for face image recording. Although the FRVT 2000 lacks any separate test for face finding, the results of the Product Usability Test suggest a superior performance of the FaceSnap RECORDER.



The Old Image Database Timed Test and the Enrollment Timed Test gave significant different results for all vendors. Again, the report contains the raw data from the experiments only, no charts. Below we show a chart that summarizes the total numbers of correct matches in the Old Image Database Timed Test and the Enrollment Timed Test. For this chart, the trials in the Photo Test were not included because some vendors treated faces on photos as normal recognition objects and others tried to check for liveness.



In a typical application of the identification mode of the FaceSnap RECORDER many face images of visitors are routinely collected by a stationary surveillance camera at some point in a building. From these images a selection of 3 to 20 images of the same person is used to train the system. After that, the user can choose to record only images that have similarities with "known" persons or he can choose to record all images that do not look alike any of the "known" persons. In any case, the identification mode of the FaceSnap RECORDER is currently restricted to operate in category A situations only (see above).

Suggestions for future facial recognition tests

The FRVT 2000 appears to focus primarily on the "Recognition Performance Test" which is an attempt to evaluate a technology by a set of "designed test data". In our opinion, this approach has to be accompanied with an evaluation of an extensive set of application scenarios. In a real application scenario, the success or failure of a particular system may not depend primarily on the performance of a single "core technology". Instead, the quality of interaction of all system components and the suitability of the system design for a particular purpose becomes crucial. For an access control scenario, like the one in the Product Usability Test, a number of additional test issues should be added, e.g.:

- Rejection of Imposters (Is there an active decision of the system on imposters, or does the system just "don't recognize"? What is the FRR in what reaction time?
- Does the system recognize the case that more than one person is present and imposters may walk in together with an authorized person?
- Does the system recognize other attempts to fool it? (e.g. by a photo, does it recognize the fact that there is photo or just don't recognize the face?)

There are a number of important real world applications for which customers seek a complete and viable solution. A commercial system should be tested against its claims with respect to particular applications. For future tests, it would be desirable to exactly identify the most important applications of facial recognition and to tailor test procedures to them.

Contact: Prof. Dr. Thomas Zielke • zielke@c-vis.com • Information: www.c-vis.com

eTrue
(Formerly Miros, Inc.)
144 Turnpike Rd.
Southborough, MA 01772

Response to the FRVT 2000 Report
February 14, 2001

In reading the Face Recognition Vendor Test 2000 report, it is clear that a lot of effort and resources were applied to the preparation, collection, testing, analysis, evaluation and reporting of commercial face recognition technology. Overall, we appreciate the thoroughness and rigor of the approach (given the assumptions that were made) and reporting of the results. Unfortunately, the results that are reported do not show the complete state of the accuracy of face recognition technology today.

The test team made certain assumptions in preparing the test which were intuitive prior to the test, but which led to some unexpected surprises during the test. Using a homegrown face recognition algorithm as a guide, the team estimated the number of images that a commercial face recognition solution should be expected to process in a few days of testing. Providing some buffer beyond what their home grown solution performed, the test team decided on a base number of a little less than 200,000,000 face pair comparisons, with several hundred to thousand subset test pairs for the analysis.

By the definition of this type of test, it eliminated commercial face algorithms that could not perform 200,000,000 face pair comparisons in a few days to a week, but which may have been more accurate. Our company has no idea how accurately our face recognition algorithm performed on this test, because we have not been told and because the results of our participation in the major accuracy portion of the test are not reported.

When we asked the test team about reporting partial results (eTrue completed about a third of the test), they responded that the number of sample test points were too small (tens instead of hundreds) for a comparison between full results and partial results to be meaningful. It is unfortunate that based on the decision to use a homegrown algorithm as a guide and the subsequent choice of 200,000,000 face pair comparisons, this test mostly shows the accuracy of fast algorithms. Furthermore, this level of speed is simply irrelevant in one-to-one face verifications which represent a majority of real world applications. Many of these applications just need to complete one face pair comparison in under 1 second.

If there was an algorithm available (not necessarily ours) that could only do 1,000,000 face image pair comparisons in a few days but had 10X lower false accept errors and 10X lower false reject errors than the results report, this test would never show it. The government test team could have taken guidance from other third party organizations that have done tests in which both the accuracy and speed of face recognition technology were quickly and cheaply determined. It's too bad that the public or private sector will not have the opportunity to decide between faster versus potentially more accurate face recognition in a government sponsored test with the FRVT 2000 report, because potentially more accurate face recognition results were simply ignored.

Sincerely,

Michael Kuperstein, Ph.D.
Chairman, eTrue, Inc.

30 Porter Road
Littleton, MA 01460

February 15, 2001

BlackburnDM@nswc.navy.mil
jphillips@darpa.mil
Bone_Mike@crane.navy.mil

Dear FRVT2000 Sponsors:

Thank you for providing Lau Technologies with the opportunity to participate in this important test program. Our team on site found the administrators to be very helpful, cooperative and sincerely committed to conducting the test by the most proper and objective means possible.

Please find attached, in PDF format, our response to the FRVT2000 test program. In our response, we included plots that were created from tabular data within your report. We would like to suggest, outside of the scope of the formal response, that perhaps you also see the benefit of presenting this data in plot format and consider appending your final release.

On behalf of Lau Technologies, I thank you for your professionalism. I have been involved first hand in a number of facial recognition tests over the last few years, ranging from Heathrow airport to the deserts of the occupied West Bank. From my experience, the testers and the tested always learn important lessons from intense efforts like this one. We look forward to working together with you to collectively move this emerging field forward, and bringing facial recognition a step closer to meeting the challenge of the most demanding real world applications.

Congratulations on a test program very well done.

Sincerely,

Francis J. Cusack Jr.

Francis J. Cusack Jr.
Director, Biometric Business Development
Lau Technologies

Lau Technologies' Position Paper

Introduction

Lau Technologies was pleased to participate in this important test effort. We share in the hope of the test creators and administrators that objective and profession third party test data will be invaluable to potential consumers in this emerging field. We point out here some of our insights into the test results, and suggestions for future testing presented very much in the spirit of furthering the development of facial recognition products that will be of practical value to industry and government.

Lighting

An option was made available to participants to make use of auxiliary lighting equipment in the manner they saw fit. Most participants enthusiastically embraced this offer. While lighting has historically been a fundamental limitation to robust recognition performance, Lau made a conscious decision to not conduct any of our tests with any supplemental lights of any kind. This decision reflects our commitment to build, and submit for test, products that are practical and robust. We have put a high priority on developing and fielding algorithms that actively mitigate the effects of real world lighting variations, and while there is still much room for improvement, have realized what we believe to be a unique and highly effective solution.

Surveillance

Many of the most compelling real world facial recognition applications are those that require automatic subject acquisition and recognition. Biometric surveillance systems that can acquire, identify and track subjects autonomously are now well within the realm of what can be successfully deployed, as we demonstrated at Super Bowl Thirty Five in Tampa this year. For many of the applications we discuss with partners and customers, this automatic real time video recognition is central. Furthermore, as this capability is developed and refined, new applications will present themselves. In light of these trends, and the intrinsic advantage facial recognition has in this area over other biometrics, we would like to see more third party testing with a strong emphasis on recorded digital video sequences. This will allow repeatable, reliable experiments testing recognition for a variety of conditions and scenarios typical of real world requirements for biometric surveillance systems.

February 15, 2000

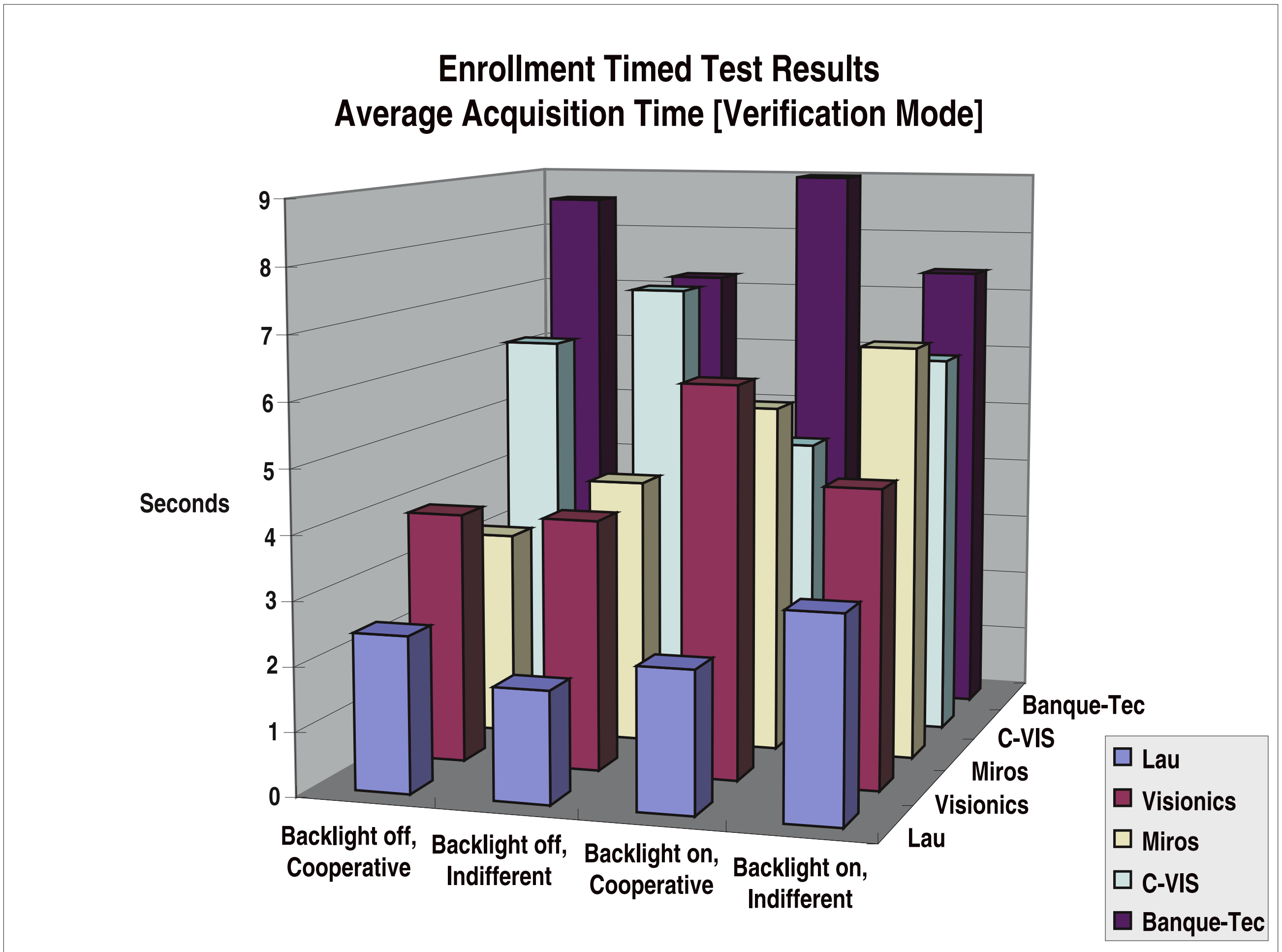
Lau Technologies

Speed

Lau has placed a high priority on developing technologies and systems to specifically meet challenging surveillance applications. This is evidenced by our results in the distance experiment featuring color images at a variety of distances with overhead and outdoor lighting, and the enrolled usability test. The recognition response time is also of particular importance in surveillance, as the probability of capturing suitable images from uncooperative subjects is proportional to the rate at which the video can be processed. Again, Lau's understanding of these dynamics, rooted in real world products and experiences, is evidenced by the relevant data in this report.

A comparative chart was constructed for all participants in the Enrollment Timed Test. Since the maximum allowed was 10 seconds, failures to acquire and acquisitions with wrong identities were counted as taking 10 seconds for the purpose of computing these averages. It may be of interest to readers to note that the Lau Technologies software ran on a single processor Pentium III at 866 MHz, while other participants chose to run quad Xeon processor platforms for some of the testing.

Plotted below is the Verification data from Tables 20 to 24 on page 51 to 53 of the report.

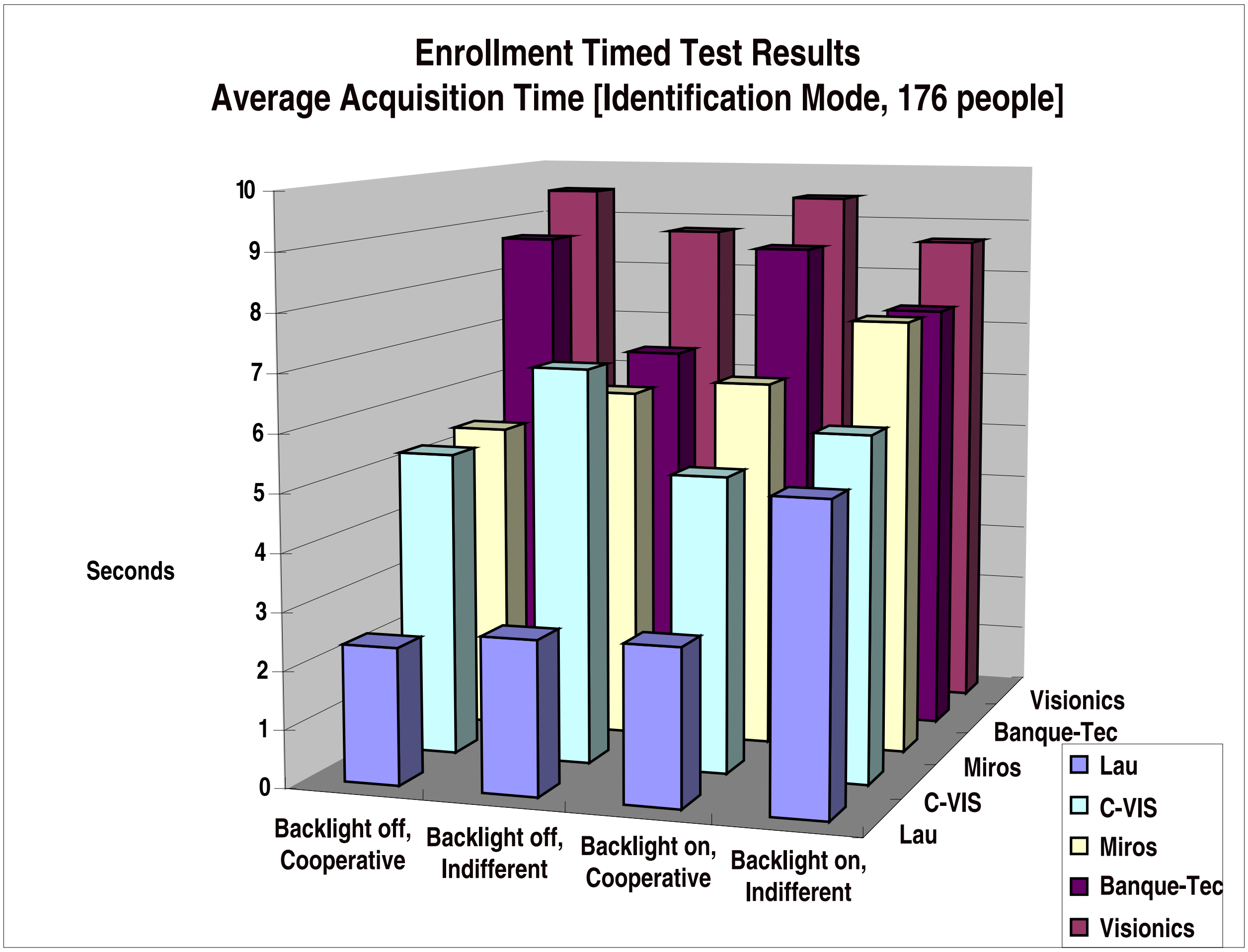


Verification Mode Average Acquisition Time

February 15, 2000

Lau Technologies

Plotted below is the Identification data from Tables 25 to 29 on page 53 to 55 of the report.



Identification Mode Average Acquisition Time

VISIONICS CORPORATION
FaceIt® FACE RECOGNITION TECHNOLOGY

One Exchange Place, Suite 800
Jersey City, New Jersey
07302

Comments on FRVT2000 Test Results

Visionics Corporation FRVT2000 Team

Executive Summary

The Facial Recognition Vendor Test 2000 (FRVT2000) was a state-of-the-art evaluation of facial recognition solution providers for the year 2000. It was divided into two categories: the Recognition Performance Tests and the Access Control Product Usability Tests. Overall both categories paint an extremely strong technology picture for Visionics, which ranked number one in 6 out of 7 of the Recognition Performance tests categories, as well as in most of the access control Product Usability tests. In most cases, Visionics significantly outperformed the nearest competitor.

The single Recognition Performance study in which Visionics did not rank first as well as some of the non-optimal timing results in the Product Usability test, can be explained by a primary constraint imposed by the test protocol. Namely, each of the tests had to be performed using a *single off-the-shelf implementation of our technology.* Because of this constraint, we chose products that had the closest compatibility with the most test evaluation criteria, but the choice could not be perfect. In the real world, a product implementation that is optimized for a specific task at hand would be supplied, unlike in these tests, where a single compromise had to be made to cover all tasks. Therefore, as we explain in this response document, these results should be viewed as anomalies and do not indicate a failure of our technology.

It is important to note that there were a number of vendors that participated in the FRVT2000 but whose technology failed to complete the tests on time and that there were others that chose not to take the test at all or pulled out after criticizing it. Our position from the beginning has been to promote honest communication and assessment of our technology, so we have embraced this test and its objectives. We have also encouraged many partners to test and validate our technology in their specific applications and environments.

The list of real world applications in which FaceItÆ has already been adopted is a testament to the fact that it is a serious commercial technology that continues to be the state of the art. Our in-house research team is tasked with the continued advancement of the technology and its maintenance ahead of the competition. Our record shows that we have released significant enhancements of the technology at the rate of once every one or two quarters.

It is worth noting that among all the vendors that participated in the FRVT2000, Visionics is the only vendor that was selected to participate in DARPAís Human ID at a Distance program. Our goal in this program is to develop complete systems for identification up to 200 feet. This has already resulted in major advancements in our capabilities.

Recognition Performance Test Paradigm Comments

The FRVT2000 recognition test as a whole does not correspond to a test of face recognition as it would be used in a particular market implementation, such as duplicate searching or surveillance. Instead, FRVT2000 is an attempt to test the boundaries of the technology and meaningfully differentiate between vendors—which is very valuable to the industry as a whole and to the adopters of technology. It sets apart what is genuine from what is marketing hype. As such, equal error rates quoted for each test are meaningful between vendors but do not indicate the typical performance of a particular product.

Both the Recognition Performance and the Product Usability tests measured a combination of two different technologies: face finding and face recognition. This is the correct approach because for almost all uses of face recognition, automatic face finding is required as a necessity (large scale database searching) or as something extremely useful (access control). In many cases, failure to acquire can be the limiting factor in the performance of the product. This point was not made explicitly in the FRVT2000 evaluation report.

Interpreting Results - Recognition Performance Tests

Introduction

Visionics technology performed extremely well in the Recognition Performance test. A brief summary of results is shown in the table below.

Test	Visionics	Lau	C-VIS	Miros	Banque-Tec
Expression	1	2	3	X	X
Illumination	1	2	3	X	X
Pose	1	2	3	X	X
Media	1	2	3	X	X
Distance	2	1	3	X	X
Temporal	1	2	3	X	X
Resolution	1	2	3	X	X

Table 1: Recognition Tests results shown as rank order where 1 is best. (X) indicates unable to complete entire test.

Table 1 shows the order of results for each category of test. 1 indicates the best performance and 3 indicates the worst. To be concrete, the rank order can be determined by comparing the top match results in the identification experiments for each category, and averaging over different experiments. However, the rank order is basically independent of all good measures used to determine it. For example, ordering the equal error rates for each experiment (with the lowest EER indicating the best performance) will yield the same results. The "X" indicates that the vendor was not able to complete the entire test. Overall, one can see from this table that Visionics performed exceptionally well in recognition performance.

A quick way to analytically gauge performance is to average the ROC curves for all trials together for each test and present the results:

At FAR = 10^{-3} (1 in 1,000) the probability of verification (1-FRR) is...					
Test	Visionics	Lau	C-Vis	%Gain over Lau	%Gain over C-Vis
Expression	.925	.72	.21	**+28%**	**+340%**
Illumination	.71	.55	.11	**+29%**	**+545%**
Pose	***.12***	.02	.02	**+500%**	**+500%**
Media	.965	.88	.335	+10%	**+188%**
Distance	***.12***	.12	.01	0%	**+1100%**
Temporal	***.32***	.22	.02	**+45%**	**+1500%**
Resolution	.905	.67	.17	**+35%**	**+432%**

Table 2: Averages of the probability of verification at FAR 1 in 1,000

The **bold** values are where we clearly outperformed the competition. The ***bold, underlined, italic*** values indicate areas in which we did as well (distance) or, in other cases (pose and temporal), better than the competition, but feel that we could have performed better. These latter areas are discussed in the sections below.

Pose Experiments

It is important to note that in this test the two images to be compared were frontal and at pose 40 degrees. While we far outperformed the competition, the technology provided for this particular test was not designed to perform well matching at poses differing by this degree. It is important to note, however, that we do provide better performance when matching fixed pose against the same fixed pose. Also the technology that was tested did not provide for pose compensation through morphing which is now a technological element that we offer.

Distance Experiments

The distance experiment results are different than all of the other experiments in that Visionics did not outperform all other vendors. We believe that the fundamental reason is that the test software used for this experiment was a poor fit to the task at hand. In fact, we believe that the experiment itself does not fit into a product category.

The software used by Visionics for the recognition performance testing was based on our Identification SDK. This SDK is designed for database searching applications. This was the appropriate choice for most of the recognition tests, as one is comparing many images against many images (a many-to-many identification search).

For the distance experiment, however, this software was not the right choice. According to the sponsors themselves, the Distance study, "...may be thought of as mimicking a low-end video surveillance scenario...". However, the design of the Distance study was not the most appropriate one for testing technology geared toward the task of video surveillance.

As discussed above, the Recognition Performance test can be thought of as a combination of two different tests: a face finding test and a face recognition test. The critical point is this: for surveillance, we perform face finding quite differently than we do for ID systems. For surveillance, we find multiple faces in the video frame (multi-face), and allow for the possibility that some faces found, particularly small faces, may be artifacts of the poor image quality. (These faces are never recognized, of course.) In ID systems, we only attempt to recognize one face, assuming that the one real face in the image has a much stronger signal than any artifacts.

Our analysis of the distance results indicates that the face artifacts generated by the poor image quality were in many cases those being matched against the gallery database. This explains the roughly linear ROC curves in experiments D2-D7, as this linearity is a signal that many of the scores are random in nature. It also explains why Visionics performed so well on the resolution experiments R1-R4; there were not artifacts in these images. If the multi-face technology that is used today in our surveillance products was used in tests D1-D7, we are confident we would have found the real faces and hence the results would have been significantly better.

There is an additional reason that we believe this study was not a real test of surveillance. In surveillance, image processing time is a crucial factor, as video is being fed into the system. In many cases it is better to be able to process multiple frames quickly rather than process a single frame thoroughly as one gets many chances to identify a different image of the person this way. One can also combine information between frames to improve the face finding and face recognition accuracy.

Motion detection is a very good pre-filter that can straightforwardly reduce the face artifacts input into the search system. The ability to perform motion detection requires video input, which was not provided in this test.

In the future, we would recommend that static images not be used to test surveillance. Instead, sequences of pre-recorded video would offer a more realistic testing scenario and allow the most accurate assessment of technology that has been explicitly designed to take advantage of video input.

Temporal Experiments

In a real world application such as access control, we recommend enrolling multiple (5) images/templates of the same person and design our systems to work in that manner. This dramatically decreases the FRR. Obviously, the Recognition Performance test did not allow us to take advantage of this real world systems engineering technique.

In addition, we recommend that the application perform a dynamic update of the templates. Dynamic update enables the facial recognition system to ìle arnî the changes in a face that occur over time.

Nonetheless, we managed to outperform the competition in this study.

Additional Caveats

We would like to stress that the test did not really measure performance in real time. Vendors were not penalized for taking longer (unless they could not finish within three days). For the record we finished the test in two days; leaving more than 24 hours unused. The software was set on a fast setting which uses the least computing time. Had we set the software to "intensive" mode, we could have done even better than we did. However, had we done so, we would have used up the additional 24 hours which was allowed by the test. We erred on the side of finishing early since we wanted an extra day in case of unforeseen problems occurred or in case the test consisted of more images than expected. (The number of images was not set to a fixed number prior to the test.)

Since the time of the test, new and improved face finding and face recognition algorithms have been developed. At that time, the new code was not yet mature.
In addition, performance in the pose tests would have benefited significantly from our latest technology that finds faces at arbitrary pose.

Interpreting Results- Product Usability Tests

The second half of the FRVT2000 testing was a simulation of a hands-off access control system.

The system used by Visionics however, was not intended for access control applications. This caused two problems to occur in the testing. Firstly, in the simulation of the access control scenario where a single image is loaded into the surveillance system, there is a few-second startup time that comes from the surveillance software design. This time lag appears as a constant added to each acquire time because we had to re-start the surveillance application for each trial. This led to artificially large face acquire times in tables 14, 19, 24, and 29. This startup time lag would be negligible in a properly designed access control system using Visionics technology.

Secondly, the surveillance product used for the test has no "liveness" (photo test) technology. The surveillance product was not designed to include a mechanism for distinguishing between a live subject and a photograph, because in real world surveillance scenarios, there is no expectation that a person would attempt to be identified using a photograph. (Note that Visionics has advanced "liveness" technology that has been used in IT security products for a number of years.)

There were several problems with both the test design and with our performance. For instance, the tests are not repeatable nor do they accurately reflect the results. In most tests the subject walked towards the camera and the distance from the camera was used as a measure of how the recognition works. Due to the use of our surveillance product, the subject was typically at the close limit (1 foot) by the time the first image (12 foot) was recognized. So we are on record as recognizing the person at 1 foot even though the recognition was from a picture of the subject acquired 12 feet away. Because of the height difference in the subjects (5'5" - 6'2") we could not even see their faces at 1 foot away from the camera. Therefore, the results are misleading.

Nevertheless, one can roughly summarize the test results by counting the number of correct match results obtained in the Old Image Database Timed Tests and the Enrollment Timed Tests. We have excluded the photo test for the reason given above. This analysis is shown in Table 3.

Correct Matches for the Timed Tests					
	Visionics	Lau	C-Vis	Miros	Banque-Tec
Old Image Database Timed Test					
Verification	**33**	0	0	20	0
Identification	**16**	0	0	0	0
Enrollment Timed Test					
Verification	***<u>45</u>***	49	41	40	13
Identification	***<u>30</u>***	49	48	41	13
Total	124	98	89	81	26

Table 3: Timed Test Correct matches

The **bold** values are where we clearly outperformed the competition. The ***bold, underlined, italic*** values indicate unexpected performance. Both are discussed in the sections below.

Old Image Database Timed Test

The old image timed test was the more difficult of the two timed tests, because the enrollment was not performed in-situ. While the enrollment procedure was far from optimal, it is clear from Table 3 that Visionics far outperformed all other vendors. In a real-world access control application, we believe that this performance difference would increase because some, if not most, of the non-matches were due to a time lag attributable to the design of the surveillance software.

Enrollment Timed Test

Overall, most groups performed well on this test because of its relative ease -- the recognition of users occurred directly after the enrollment procedure. Failure of Visionics to score well on the Enrollment/Identification test was due *solely* to the time-lag in the surveillance software used and does not reflect the true performance of Visionics technology in the context of access control.

Vita

Mr. Duane M. Blackburn serves as a Deputy Program Manager for the Department of Defense (DoD) Counterdrug Technology Development Program Office. The goal of the program is to develop technology and prototype systems to enhance the counterdrug capability of the Department of Defense and civilian law enforcement agencies consistent with the goals of the National Drug Control Strategy and the DoD mission.

In addition to his responsibilities within the Program Office, Mr. Blackburn is on a part-time detail at the National Institute of Justice (NIJ). NIJ is the R&D arm of the U.S. Department of Justice. The goal of the detail is to maximize cooperation between the two agencies by finding projects with similar projects and/or technologies and working together to leverage the development efforts.

Mr. Blackburn holds a Bachelor of Science degree in Electrical Engineering (1996) and a Master of Science Degree in Electrical Engineering (2001) from Virginia Polytechnic Institute and State University.