Evaluating Biometrics for Airport Security
An Overview

October 4, 2001

Executive Summary

Because of the events of 11 September 2001, the FAA should expect several calls for immediate installation of biometric technologies in airports to improve security. Many of these calls will result in near-term in-situ demonstrations throughout the United States using a variety of biometric types and vendors. Some of these demonstrations could succeed while others will fail because of the rush to install and/or the lack of capability of the technology, vendor, or entity performing the demonstration. Whatever the result, these in-situ demonstrations will produce measurables that will help the FAA determine how they could conceivably implement a biometrics solution nationwide. It will be impossible to fully understand the measurables derived from these in-situ demonstrations without results from technology evaluations. Unfortunately, these evaluations are not available for most biometric types.

This paper presents an overview of how to evaluate biometric systems and is provided to give the FAA an understanding of the level of effort required to accurately judge biometric system capabilities. The paper begins with an overview of the biometric industry and ideals that need to be followed in an evaluation. The paper then provides an outline of how to perform a set of technology and scenario evaluations.

1. Introduction

Even before the terrorist attacks of 11 September 2001, the air industry was starting to investigate the utility of biometrics to improve airport security. For example, Chicago's O'Hare airport had installed a system using fingerprint biometrics for increasing speed and security for cargo truck drivers delivering to the airport. Also, Charlotte/Douglas International airport, in cooperation with US Airways, had a pilot program using iris recognition technology to verify employees entering secure areas. Israel's Ben Gurion airport uses hand geometry to speed people through customs, while Iceland's Keflavik airport uses face recognition for surveillance applications.

The discussions of using biometrics to increase airport security has greatly increased in the press and biometrics and aviation communities in the weeks after the attacks. Senator John Edwards sponsored a bill (S. 1429 - A bill to provide for the improvement of security at airports and seaports) on 14 September 2001 that requires airports to strengthen airport security "by using biometric or similar technologies that identify individuals based on unique personal characteristics." Rep. Jerry Weller introduced on 25 September 2001 a bill (H.R.2970 - "To amend the Internal Revenue Code of 1986 to allow businesses to expense qualified security devices") that would allow businesses to list security devices, such as biometrics, in their itemized deductions.

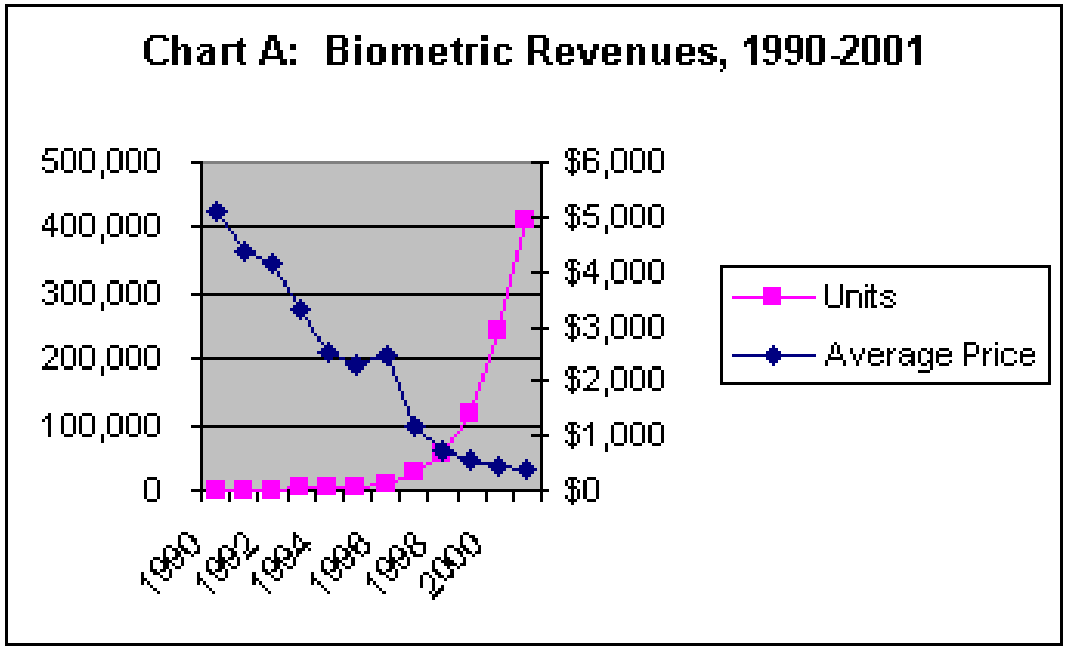
Numerous airports, airlines, vendors, system integrators, and government agencies have been contacting the FAA to offer their services for demonstrations or installations of biometric technology to increase security. The FAA itself has established a working group to study the feasibility of this idea. This working group will find that the biometrics community has not prepared itself for the position they now find themselves in. The market for biometrics, although increasing (see section 3.1), was small compared to the "gold mine" they now perceive as possible through the air industry. The vendors were fighting each other for every dollar, which meant marketing took a secondary role to industry-building. The consequence is that there have been very few attempts at accurately gauging the overall state of the art for the technology, and very few airline industry - specific evaluations performed by non-vendors.

The FAA should expect several immediate in-situ demonstrations from several different sources, many of which will produce valuable "lessons learned". These lessons learned will not be totally sufficient for determining national implementations without simultaneous technology evaluations. This paper presents an overview of how to evaluate biometric systems and is provided to give the FAA an understanding of the level of effort required to accurately judge biometric system capabilities.

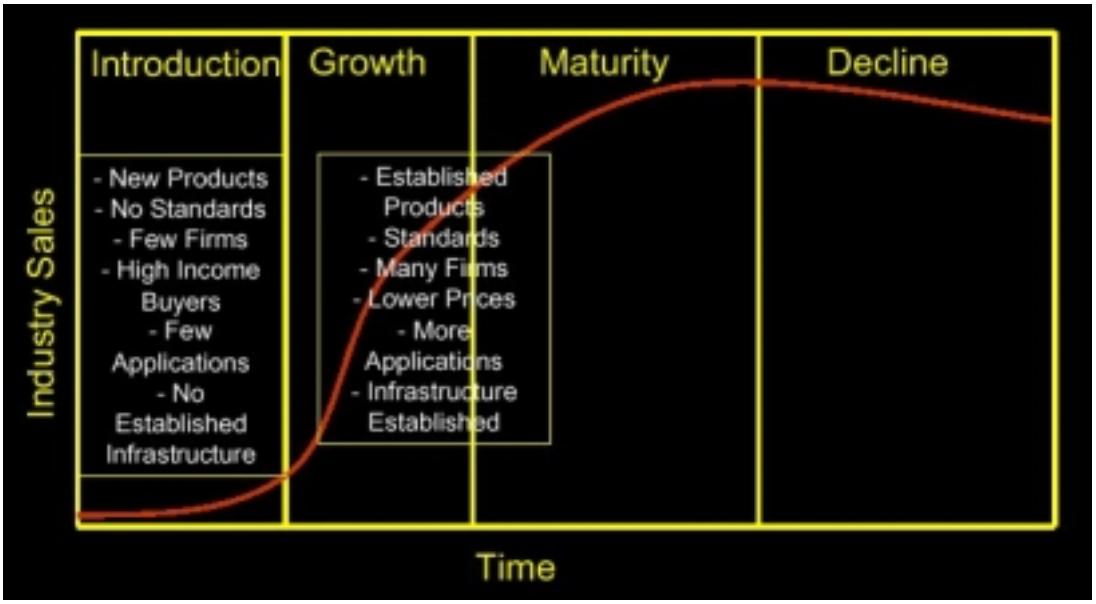
1.1 Status of the Biometrics Industry

For years, people have been anticipating the arrival of biometric technologies. There are numerous applications for the technology, if everyone understood where they work and where they don't work. Even without a full understanding of the technology, law enforcement is beginning to embrace the technology as an aid in identifying or verifying individuals.

Biometric industry watchers agree that biometrics is about to burst into the public consciousness and will begin to be used in many applications. There are numerous indicators that show this. The first is a government and industry led effort to establish a uniform Application Programmers Interface (API) so that different biometric systems can communicate with one another and a mixture of biometric types can easily be used together. Secondly, a number of mergers and acquisitions have taken place recently, signaling that the small start-up companies of a few years ago are succeeding and are joining forces to increase the market. Finally, and most importantly, the purchase of biometric devices has increased significantly over the past couple of years, as shown below. [1]



Over the last several years, biometrics has worked its way out of the introduction stage of the industry life cycle and into the growth stage [2]. Within the next five to ten years, industry watchers anticipate the technology will move into its maturity stage where revenues are stable and industry practices are set. If the FAA is to help direct this technology for the maximum benefit of its end-users, it must do so now.



1.2 Advantages of Evaluations

Evaluations of advancing technology provides two types of benefits - short-term and long-term. The short-term benefits are obvious. First, end-users have the ability to determine which technology will best assist their efforts now. Purchasing and implementation decisions are based on factual evaluation results rather than having non-technical field agents sort through vendor claims to determine which system is best. Second, federal Program Managers obtain an assessment of the capabilities of individual systems as well as an assessment of the current state-of-the-art. This enables them to make informed decisions for future development efforts.

The long-term benefits of continuous evaluations of advancing technology may not be obvious. Continuous evaluation of technology fosters the advancement of that technology, even when development funding has not been provided. An excellent example is NIST's speech recognition evaluations [3]. Comparison of score results from 1996 tests with those from 1999 tests show a considerable improvement, even though NIST did not fund any of the participants to improve their algorithms.

2. Evaluation Methodology

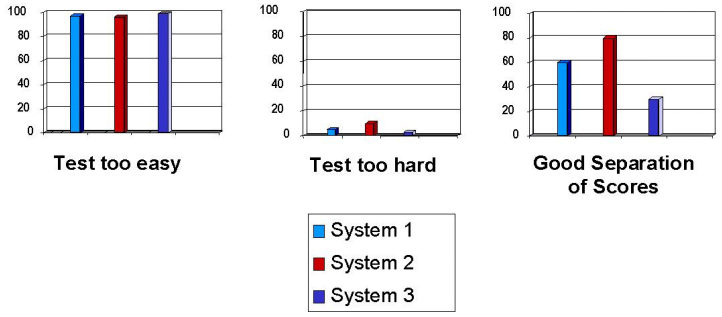
Besides funding, the biggest obstacle in evaluating technologies with numerous applications was the evaluation methodology. Until recently, no set methodology has existed that would provide results that would be useful for different applications of the same technology. A new methodology to perform these evaluations has been developed [4,5] and successfully demonstrated [6]. This methodology contains several ideals and a three-step evaluation plan discussed below.

The first ideal 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 “independent evaluations” may actually be funded by vendors, integrators or resellers.

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.



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 is that the evaluator needs to understand the true requirements for the 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. Someone may have a desired goal of 90% on some measurable. If they 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 their true requirement. If their evaluations show the technology can hit at a 70% rate and they are measuring against their desired requirement, it would appear as if the technology would fail miserably. However, if you use their true requirement, the evaluations would show that the addition of technology would be a resounding success!

In addition to the ideals presented above, references 4 and 5 provide 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, the evaluator will not truly understand why unless they have the technical information from the Technology Evaluations to analyze with the Scenario Evaluation results.

3. Evaluation Plan

3.1 Overview

The best way to structure these evaluations would be to have one individual (Evaluation Program Manager) with previous experience in conducting biometric evaluations lead the overall effort.

This individual would oversee the work of several Test Agents that perform the actual evaluations. Preferably, these Test Agents would also have experience evaluating biometric systems but the lack of qualified government personnel will probably cause this to not be possible. Therefore, Test Agents with somewhat related experience and demonstrated ability could be chosen (from one or several different government organizations). The Test Agents will be responsible for writing a detailed test plan (based on the evaluation methodology and goals established by the Program Manager), proctoring the tests, scoring the results, and writing the evaluation reports. The Program Manager would give direction and approve recommendations from each Test Agent at each critical step.

Technologies to be evaluated in this program include facial recognition, speaker recognition, fingerprint, hand geometry and iris. This evaluation program will include technology evaluations followed by several scenario evaluations at FAA end-user sites. Some of the immediate demonstrations being currently proposed to the FAA could be modified for scenario evaluations without much difficulty.

Technology evaluations will be determined for each technology according to their perceived strengths and weaknesses in much the same manner as the FRVT 2000 evaluations. Newer types of data would also be introduced. For example, facial recognition would need to include both still images as well as video images.

3.2 Evaluation Organization

The evaluation would need to be broken into three phases. Preliminary work (Phase I) could begin immediately, and would be followed with a Technology Evaluation (Phase II) and several Scenario Evaluations (Phase III) in the future

3.2.1 Phase I Tasks

3.2.1.1 Data Collection

Collecting data for the technology evaluations, as well as the scenario evaluations, will require the usage of human subjects. As such, all human subjects must be volunteers who have signed "Informed Consent" forms.

All data that is collected will be ground-truthed according to the XML standards being developed and used in DARPA's Human Identification at a Distance Program. Using this format allows the FAA to either use the scoring software originally developed for the FERET program and modified for the FRVT 2000 evaluations, or to develop a new scoring software that uses a uniform interface. Using this format also enables the FAA to place their evaluation data (post-evaluation) into the HumanID database for controlled distribution to system developers, so that the data has lasting benefit beyond that year's evaluation.

3.2.1.2 Scoring Code Review

Also required in Phase I is a statistical review of the scoring code used in the FRVT 2000 evaluations. We must determine if this scoring code is accurate for biometric types other than facial recognition. If it is not, we will need to develop new scoring code.

The second task that needs to be performed is the development of the overall evaluation methodology (i.e. what do we need to test for each biometric type). This requires an investigation into the technology from both a technical and practitioner viewpoints, and the development of some basic applications.

The third task is to begin to collect new data for the evaluations. The final task is to develop a web site that will function as the public gateway to the project.

3.2.2 Phase II Tasks

Continue data collection. Data collection will be performed by the same people that collected data in Phase I, as well as individual test agents.

Phase I will mark the beginning of this program en masse via technology evaluations. Each technology will have a test agent that is responsible for performing Phase II's technology evaluation. If necessary or beneficial, a test agent may be responsible for multiple technologies. Each test agent will work with the Program Manager to ensure the evaluations are performed properly and the evaluation report for each technology is complete.

Each test agent will also work with the Program Manager and the integration test agent to prepare an additional evaluation report that analyzes the biometric types and their evaluation results against each other.

The Program Manager will need to work on determining applications for Phase III scenario evaluations.

3.2.3 Phase III Tasks

Continue data collection for future technology evaluations.

Phase III would be application-specific scenario evaluations. The selection of the technologies and applications to evaluate will be determined via a comparison of NIJ user needs and technical capabilities as determined in Phase II.

Investigate new data formats and scoring algorithm modifications for future Technology Evaluations.

4. Summary

This document provided an overview of the state of the biometrics industry and accepted technical evaluations. Although the FAA will need to start performing several immediate in-situ demonstrations for political reasons, as well as for valuable “lessons learned”, the demonstrations by themselves will not be totally sufficient for determining national implementations.

A thorough technical analysis followed by scenario evaluations, which could use the systems from the in-situ demonstrations, will be required for a full understanding of system and architecture capabilities. The cost and schedule of the work proposed in this document is dependent on the breadth of the evaluations deemed necessary and the timeframe required to perform them. A cost/schedule estimate could be easily derived once the overall parameters for the evaluations has been determined.

References.

[1] IBIA Biometrics Advocacy Report. Volume III, Number 7, 20 April 2001. Available on the internet at: http://ibia.vwh.net/newslett04201.htm

[2] E. Bowman. Overview of the Biometric Identification Technology Industry. Presentation at Defending Cyberspace 1999.

[3] http://www.nist.gov/speech/index.htm

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

[5] D.M. Blackburn. "Evaluating Technology - Three Easy Steps to Success". *Corrections Today*, July 2001.

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