

---

MITRE TECHNICAL REPORT

# Technology Assessment for the State of the Art Biometrics Excellence Roadmap

## Volume 2 (of 3)

### Face, Iris, Ear, Voice, and Handwriter Recognition

**March 2009; v1.3**

James Wayman (Editor, Face, Voice)

Nicholas Orlans (Editor, Face, Ear)

Qian Hu (Voice)

Fred Goodman (Voice)

Azar Ulrich (Handwriting)

Valorie Valencia (Iris)

**Sponsor:** Mr. Scott Swann  
**Dept. No.:** G063, G551, G064, G023

**Contract No.:** J-FBI-07-164,  
**Project No.:** 14008FC09-LA

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

This document was originally published June 2008, and reflects the state-of-the-art as of that date.

**This software (or technical data) was produced for the U. S. Government under contract J-FBI-07-164, and is subject to the Rights in Data-General Clause 52.227-14 (JUNE 1987)**

© 2008 The MITRE Corporation. All Rights Reserved

**MITRE**

This page intentionally left blank.

## Executive Summary

This report presents the technology assessment portion of the State of the Art Biometrics Excellence Roadmap (SABER) study which was conducted over a 10 month period in 2007-2008. The study included an extensive survey of biometric technologies, current products, systems, independent performance evaluations, and an overview of select research activities. The MITRE team was provided access to FBI laboratories where discussions with analysts and scientists contributed enormously to understanding the breadth of forensic biometric applications and how they are used. The MITRE team also had support from senior external consultants. The team visited representative federal, state, and local booking environments, a state detention facility, and saw large surveillance systems used for security and gaming. The site visits provided a valuable perspective on the constraints and challenges that must be considered for the FBI to fully realize the Next Generation Identification (NGI) system. The proposed roadmap recognizes FBI's leadership in fingerprint technology as a solid foundation for expansion, and seeks a pragmatic course using cost-effective supporting technologies.

### The Daubert Challenge

All commercial and government application developers seek biometric technologies that are accurate and cost effective. However, biometrics and other identification methods used by the FBI for law enforcement purposes are unique; they may be subjected to additional standards and scrutiny based on Daubert criteria. In the US Supreme Court case "*Daubert vs. Merrell Dow Pharmaceuticals* (92-102), 509 U.S. 579 (1993)," the Court suggested criteria for determining if scientific evidence was reliable and hence admissible:

1. Is the evidence based on a testable theory or technique?
2. Has the theory or technique been published and peer reviewed?
3. For a particular technique, does it have a known error rate and standards governing its operational use?
4. Is the underlying science generally accepted within a relevant community [Daubert vs. Merrell, 1993]?

These Daubert criteria apply in all U.S. federal courts and but only in some state courts. However, the FBI should strive to meet the Daubert standards for biometric evidence used in all prosecutions. For this reason, additional scientific development is needed in biometric technologies and for supporting testimony from scientific experts.

The investigative applications of biometrics are not subject to Daubert criteria; therefore, biometrics can be used in investigations, regardless of their scientific development. Between investigation and prosecution lies the area of warrants. The required scientific defensibility of technical methods is not always clear with warrant actions. It is prudent for the FBI to pursue Daubert compliance, and seek to elevate the usability of technical evidence from investigations to warrants and prosecutorial needs.

Toward partial fulfillment of MITRE support to the FBI Criminal Justice Information Systems (CJIS) Technology Evaluation Standards Test (TEST) unit, the State-of-the-Art Biometrics Excellence Roadmap (SABER) Technology Assessment document contains assessments of multiple biometric technologies. The biometric technologies are assessed in general terms and considered within the FBI's Center of Excellence. The Technology Assessment is organized in large volumes. Volume I contains fingerprint, palm, vascular recognition, and standards. Volume II (this volume) contains face, iris, ear, and handwriter recognition, and voice modalities. Volume III contains DNA. Overarching recommendations for technology development are contained in Volume I, and more modality specific recommendations for technology gaps and FBI recommendations occur within each section.

### **Trends and Issues**

Biometric technologies such as face, iris, voice, and handwriting recognition are maturing. If effectively integrated (fused), additional biometric technologies offer promise for improved performance and an expanded application base for searching and identity resolution.

### **Recommendations and Challenges**

#### **For Face Recognition:**

- There is no widely accepted common training and minimum proficiency for human surveillance operators who also perform identification. The FBI should develop and provide common training material for human face examiners to fill this need (preliminary initiative in this area is currently underway from Forensic Audio Visual Image Analysis Unit (FAVIAU)).
- Provide a more quantified understanding of facial landmarks and dermal characteristics as they appear over time and through different media.
- The inconsistent face image quality from mug shot environments performs well below the current NIST evaluation results on idealized, high quality images. Where ever possible, the FBI should require the minimal adoption of Subject Acquisition Profile 40 and encourage continued progress toward profiles 50/51 (refer to the June 2008 MITRE report Certified Product List (CPL) Way Ahead for additional details).
- Face recognition depends on successful face segmentation (or face detection), which is known to suffer performance degradation due to imaging and orientation factors. The FBI should evaluate the use of research tools for face detection against relevant media to include video sources, uncontrolled images, mug shots, and civil identification photos. As there currently is no known evaluation dataset for face detection performance, the FBI should consider developing a "challenge dataset" that represents their face detection and forensic (quality) needs in partnership with appropriate existing research programs such as the Intelligence Advanced Research Projects Activity (IARPA).

- Methods and techniques for searching database with sketches and composite images constructed by forensic artists or computer generated composite.

#### **For Iris Recognition:**

- Before future NGI integration of iris technology, the FBI should explore the use of iris recognition within smaller, controlled pilot programs. Examples of possible uses include:
  - Training programs to familiarize examination and analysis community
  - Prisoner registration and visitor identification
  - Registered sex offenders and probation cases
  - Mobile ID and counter gang policing
- In support of science and technology, the following recommendations speak to Daubert related issues:
  - Recommend that the FBI begin a multi-year, multi-spectral data collection effort on a small number of long-term (10 year) volunteers (~100) to determine the stability of the iris pattern at different wavelengths over time.
  - Recommend that the FBI invest in research on iris recognition from both low and high resolution visible-wavelength color imagery obtained through common photographic methods.
  - Recommend that the FBI begin a program for developing, documenting, and testing methodologies for human-aided recognition of irises that will lead to Daubert-admissible testimony to support the results of automated iris comparison systems and high resolution photography.
  - Recommend intra-governmental cooperation (e.g., with Department of Homeland Security Science and Technology (DHS S&T) and IARPA) in developing and testing iris recognition systems capable of operating at distances of many meters with walking data subjects. The same technology that enables robust collection also will improve acquisition time and usability for semi-cooperative subjects in controlled application environments.

#### **For Ear Recognition:**

- Recommend that the FBI start a data collection effort for a diversity of ear prints and ear images, the latter over multiple angles and illumination conditions to support research into distinctiveness and stability.

- Recommend that the FBI begin a research effort into describing and quantifying individual ear features, with supporting statistical metrics developed across a variety of ear images, toward the goal of Daubert admissibility.
- Recommend that, upon advancement of the above tasks, the FBI develop a training and testing program for forensic ear and ear print examiners as a component to augment forensic face recognition.

### **For Speaker Identification:**

- Direct funding to National Institute for Science and Technology Information Access Division (NIST IAD) for broadening Speaker Recognition Evaluation (SRE) to include test protocols of operational interest to the FBI.
- Direct funding to Linguistic Data Consortium (LDC) to establish test and development databases supporting forensic applications of speaker recognition technology.
- Fund industrial and academic groups already actively involved in the NIST SRE to continue their involvement. Such groups have been working without U.S. government funding, but cannot be expected to increase their output or performance without some level of government or commercial support.
- Create robust data collection protocols and “best practices” involving both telephone and office environment speech aimed at lowering error rates.
- Leverage the relevant international work to support scientific acceptance of forensic speaker recognition technologies, such as that by the Forensic Science Service, the University of Lausanne, and the Netherland Forensic Institute.
- Develop a plan for integrating speaker data with other modalities.
- Develop rapid hardware/software systems for “real time” processing of speech data against a large number of recognized target speakers. Commodity hardware such as GPUs or multi-core processors hold promise for making high performance processing cost effective.
- Develop additional “chain of custody” protocols and standards applicable to speech data collected by a variety of agencies, most outside the FBI.
- Develop forensically acceptable pre-processing algorithms for enhancing speech data, including robust activity detection and noise suppression.
- Work with the DoD, DHS and other agencies with a mission of combating terrorism to develop policies and procedures for implementing data collection protocols for speaker identification.
- Develop in-house capability for expert testimony at trial regarding the results of speaker recognition technologies.

- Begin a series of workshops with relevant stakeholders (DoD, DHS, DNI, NIST, NSA, LDC, and foreign allies) to outline a specific path forward and develop a timeline and a budget for this work.

## **For Handwriter Recognition:**

### **0 to 2 Years**

- Baseline handwriter recognition performance for questioned documents by conducting comparative analysis between systems; fund and leverage the experience of NIST for evaluating recognition performance of handwriter recognition as well as the underlying feature extraction processes.
- Propose standard feature representations derived from leading research and current prototype systems, and advance these through NIST.
- Collect progressively larger known test sets for training, development, and testing of existing and future systems.
- Request case feedback to better establish ground truth and performance metrics (human and automated).
- Refine support tools for human visualization, mark up, and verification of features.

### **2 to 5 Years**

- Integrate writer recognition with character, text, and language recognition.
- As non-handwritten communications become more prevalent, such as blogging, text messaging and emails, there is a growing need to identify writers not by their written script, but by analysis of the typed content. Currently, there are some studies in the area of writer's colloquial analysis that may lead to the emerging technology of writer identification in the "blogosphere." These technologies could possibly create a profile and even identify a writer's identity. Similar to colloquial speech analysis, studies have shown that bloggers and chatters use a colloquial form of writing instead of a standard form when blogging, chatting, or text messaging. Recommend investment in scientifically-based text-independent e-mail and blog writer identification and document linking.

### **5 to 10 Years**

Consider, for investigative use, integrating automated services in Next Generation IAFIS for handwriter recognition. An initial form of integration could be the cross referencing of confirmed samples (solved questioned documents) to their corresponding criminal files.

Recommendations for technology development partnerships for consideration by the FBI occur in separate documents from The MITRE Corporation. An assessment and recommendations for

renewed commitment to the Certified Product List occur in the Certified Product List Way Ahead report.



## **Acknowledgements**

This report was actively supported by many dedicated individuals and experts within the FBI and the National Institute of Standards and Technology (NIST). The authors wish to acknowledge special thanks to Tom Hopper, Dr. Hiro Nakosone, Richard Vorder Bruegge, and Dr. Nicole Spaun. We also thank Dr. John Butler of NIST for his review and comments.

# Table of Contents

<b>1</b>	<b>Face Recognition</b>	<b>1-1</b>
1.1	Introduction and Background	1-1
1.2	Face Detection	1-3
1.3	Performance and System Evaluations	1-3
1.4	Human Recognition and Automated Recognition	1-4
1.5	Face Standards	1-4
1.6	3-D Face Recognition	1-5
1.7	3-D Face Acquisition	1-6
1.8	Technology Gaps and Challenges	1-7
<b>2</b>	<b>Iris Recognition</b>	<b>2-1</b>
2.1	Background	2-1
2.2	State of the Industry	2-6
2.3	Growth and Markets	2-7
2.3.1	Brief History	2-7
2.3.2	High-Profile Implementations	2-8
2.3.2.1	UK IRIS	2-8
2.3.2.2	Schiphol Airport Privium System	2-9
2.3.2.3	Afghan Repatriation Program	2-10
2.3.2.4	UAE Iris Expellees Tracking and Border Control System	2-11
2.3.2.5	Iris Recognition in the DoD	2-12
2.4	Performance	2-13
2.4.1	IBG ITIRT	2-14
2.4.2	NIST Iris Challenge Evaluation	2-18
2.4.2.1	ICE 2005	2-18
2.4.2.2	ICE 2006	2-19
2.4.3	Authenti-Corp IRIS06	2-24
2.4.4	NIST IREX 08	2-27

2.5	Standards	2-27
2.6	Research and Development	2-28
2.6.1	Data Acquisition	2-28
2.6.2	Iris Segmentation	2-29
2.6.3	Texture Encoding	2-29
2.6.4	Iris Comparisons	2-29
2.6.5	Test Databases	2-29
2.7	Human Issues	2-30
2.7.1	Usability	2-30
2.7.2	Safety	2-31
2.7.3	Privacy	2-32
2.8	Forensic Capabilities	2-32
2.9	Vulnerabilities	2-33
2.10	Future Capabilities	2-34
2.11	Technology Gaps and Challenges	2-34
2.12	Vendors for Iris Recognition Hardware	2-36
<b>3</b>	<b>Ear Recognition</b>	<b>3-1</b>
3.1	Background	3-1
3.2	Accuracy	3-3
3.3	Standards	3-4
3.4	Spoofing and Vulnerability	3-4
3.5	Privacy	3-4
3.6	Future Capabilities	3-5
3.7	Technology Gaps and Challenges	3-5
3.8	Recommendations	3-6
<b>4</b>	<b>Speaker Identification</b>	<b>4-1</b>
4.1	Introduction	4-1
4.2	Background	4-1
4.3	Government Involvement	4-5

4.4	Evaluation Standards	4-7
4.5	Recent Scientific Advances	4-8
4.6	State of the Industry	4-9
4.7	High-Profile Implementations	4-11
4.7.1	Distinctions Between Speaker Verification and Speaker Identification	4-11
4.7.1.1	Speaker Verification Systems	4-11
4.7.2	Speaker Identification System for Forensic Applications	4-12
4.7.2.1	FASR	4-12
4.8	Standards and Interoperability	4-12
4.8.1	Forensic Capabilities	4-13
4.8.2	Vulnerabilities	4-14
4.9	Technology Objectives	4-15
4.10	Addressing Daubert Admissibility for Voice	4-15
4.11	Recommendations	4-16
<b>5</b>	<b>Handwriter Recognition</b>	<b>5-1</b>
5.1	Technology Background	5-2
5.1.1	Individuality of Handwriting	5-4
5.2	State of the Industry	5-4
5.2.1	Data Collection	5-4
5.2.2	Automated Handwriter and Handwriting Recognition (OCR) Technologies	5-4
5.3	Growth and Markets	5-5
5.3.1	Data Collection Tools	5-5
5.3.1.1	ForensicXP	5-5
5.3.1.2	Foster and Freeman	5-5
5.3.2	Handwriter Identification Tools	5-6
5.3.2.1	European FISH	5-6
5.3.2.2	U.S. Secret Service FISH	5-7
5.3.2.3	CEDAR-FOX	5-7
5.3.2.4	FLEX-Tracker and FLEX-Miner	5-9

5.3.2.5	FLASH ID	5-9
5.3.2.6	European Script	5-9
5.3.2.7	Arkansas State Crime Laboratory	5-10
5.4	Performance and Accuracy	5-10
5.5	Standardization and Interoperability	5-12
5.5.1	ANSI/NIST ITL 1-2007	5-12
5.5.2	WANDAML	5-13
5.6	Data	5-13
5.6.1	CEDAR	5-13
5.6.2	The U.S. Government Agencies	5-13
5.6.2.1	National Institute of Standards and Technology (NIST)	5-13
5.6.2.2	The United States Secret Service (USSS)	5-14
5.6.2.3	Federal Bureau of Investigation (FBI) Files	5-14
5.6.3	Requested Writings	5-14
5.7	Technology gaps and Challenges	5-16
5.8	Conclusions	5-17
5.9	Addenda	5-18
5.9.1	Applicable Organizations and Conferences	5-18
<b>Appendix A</b>	<b>Speaker References</b>	<b>A-1</b>
<b>Appendix B</b>	<b>Acronyms</b>	<b>B-1</b>
<b>Appendix C</b>	<b>Revisions</b>	<b>C-1</b>

## List of Figures

Figure 1-1. Local and Global Face Image Features	1-1
Figure 1-2. Elastic Bunch Graph Matching	1-2
Figure 1-3. Reduction in Error Rates in FRVT [Phillips, et al, 2007]	1-4
Figure 1-4. 2-D and 3-D Representations (from Chang, 2004)	1-6
Figure 2-1. Structures of the Iris	2-1
Figure 2-2 Grayscale Image and Texture of Iris	2-2
Figure 2-3. Schematic of Iris Recognition Operation	2-3
Figure 2-4. Eye Image with the Iris Area Segmented.	2-4
Figure 2-5 UK IRIS Kiosk at Manchester Airport	2-9
Figure 2-6. Primum Kiosk at Schiphol	2-9
Figure 2-7. PRIVIUM Iris Recognition	2-10
Figure 2-8. Afghan Repatriation Program	2-11
Figure 2-9. UAE Iris Expellees Tracking and Border Control System	2-12
Figure 2-10. U.S. Marine Corp Entry Control Point in Fallujah, Iraq	2-13
Figure 2-11. U.S. Soldier use Iris Recognition to Verify Identity	2-13
Figure 2-12. ITIRT Intra-Device DET Curves	2-16
Figure 2-13. ITIRT Cross-Device DET Curves	2-17
Figure 2-14. ICE 2005 ROC Curves	2-19
Figure 2-15. FRVT/ICE 2006 Boxplot Results	2-21
Figure 2-16. ICE 2006 Left and Right Eye Boxplots	2-24
Figure 2-17. IRIS06 Attempt-Level Native and Interoperability ROC Curves	2-26
Figure 2-18. IRIS06 Attempt-Level Performance with Glasses	2-27
Figure 3-1. Iannerelli's Ear Anatomy and Measurements [Burge, et al., 1998]	3-1
Figure 3-2. Acoustic Waveform Probe with Receiving Microphone [Philips Research, 2005]	3-2
Figure 3-3. Microphones for Ear Acoustics [Philips Research, 2005]	3-3
Figure 3-4. Cauliflower Ear	3-5
Figure 4-1. 3-D GMM Mixture Diagram	4-3

Figure 4-2. Illustration of Gaussian Mixture Model Training	4-4
Figure 4-3. SVM Training Process	4-9
Figure 5-1. “London Letter”	5-15
Figure 5-2. “CEDAR Letter”	5-16

## List of Tables

Table 2-1. Types of Iris Recognition Camera Products	2-6
Table 2-2. FRVT/ICE 2006 Results	2-22
Table 3-1. Reported Ear Recognition Performances [Hurley, 2007]	3-3
Table 3-2. Reported Ear Acoustic Equal Error Rates [Philips Research, 2005]	3-4
Table 4-1. Industry Vendors for Speaker Recognition Systems	4-10
Table 5-1. Writer-Independent Methods with 1 and 16 Training Samples	5-11

This page intentionally left blank.

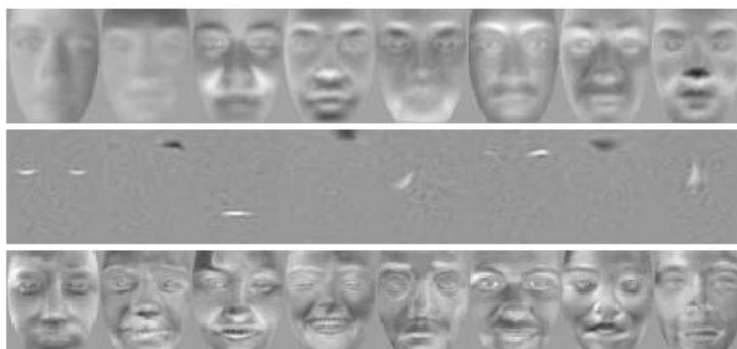


# 1 Face Recognition

## 1.1 Introduction and Background

Early work on automated and semi-automated face recognition methods during the 1960s used the geometrical relationships between facial landmarks, such as eyes, tip of nose, and corners of the mouth as metrics for recognition [Bledsoe, 1966]. Finding these landmarks in a precise and consistent way across variation in pose, illumination, and facial expression remains an intractable problem. By the late 1980s, this approach was abandoned for general image processing approaches unspecific to facial recognition. Under U.S. government funding, Sirovich and Kirby [Sirovich, 1987] developed an image decomposition method based on Principal Component Analysis (PCA), a general mathematical method known since the early 20<sup>th</sup> century. This method required only the manual positioning of the eyes. By 1993, these semi-automated methods were becoming robust enough for the Army Research Laboratory to begin a testing program.

In 1996, Penev and Atick [Penev, 1996] modified the PCA method, using “local features” as opposed to the “global features” of the earlier work. Figure 1-1 shows “local” and “global features,” with the “local features” in the middle row. The “global” features have non-zero (non-gray areas in the image) values over all of the face area. The “local” features have non-zero (non-gray) values only over localized parts of the face.



**Figure 1-1. Local and Global Face Image Features<sup>1</sup>**

The “local” and “global” methods find the combination of “features” shown above that add up to any input face. Any face can be approximated by some weighted combination of these features; the particular weightings becoming the code that represents the face with respect to this set of features. The methods of “decomposing” a face image into component “features” dominated automated facial recognition into the first decade of the 21<sup>st</sup> century. By the early 2000s, all

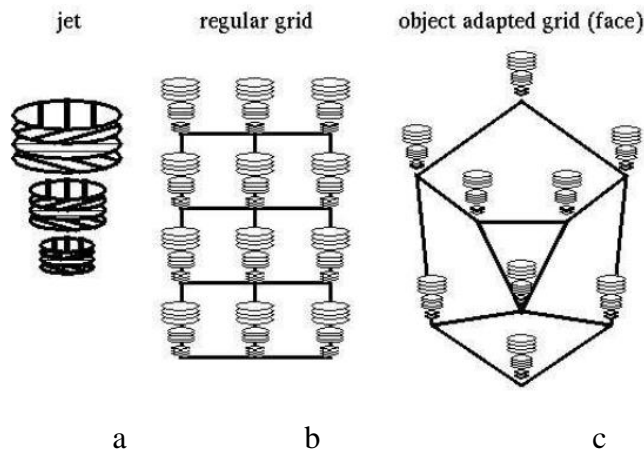
---

<sup>1</sup> Published with permission by J. Ross Beverage, Colorado State University.

aspects of the methods, even locating the eyes, were automated. No manual marking of the face images was required.

By 2005, however, a new technique was gaining popularity in commercial applications, based on performance results in U.S. government-conducted tests [Phillips, 2000, 2006]. This technique, “Elastic Bunch Graph Matching” (EBGM), was based on work funded by the Office of Naval Research (ONR) [Wiskott, et al, 1997]. Rather than attempting to decompose each face into “global” or “local” features, the EBGM method placed small blocks of numbers (“filters”) over small areas of the image, multiplying and adding the blocks with the pixel values to produce numbers (i.e., “jets”) at various locations on the image. Those locations could be adjusted to accommodate minor changes in pose angle and facial expression. Figure 1-2a and Figure 1-2b show these jets, and the jets placed on a rectangular graph or grid.

Recently, the trend has been to place those filters around identifiable landmarks on the face, such as eyes, nose, and mouth corners. Although it is not possible to precisely find those landmarks, they can be found close enough so that the filters can be placed near them. Figure 1-2c shows how the grid holding the filters can be placed near landmarks on the face. This new technique has greatly enhanced facial recognition performance under variations of pose, angle, and expression. New techniques for illumination normalization are also being used on the images prior to application of the filters.



**Figure 1-2. Elastic Bunch Graph Matching<sup>2</sup>**

Recognizing faces in general uncontrolled images, including complicated scenes, requires “segmentation”—detecting the presence and location of faces and separating them from the background or non face regions. Most approaches for segmenting still images use “color and

---

<sup>2</sup> Published with permission of L. Wiskott.

shape” methods. Here, color imagery is required. Human skin seems varied in color, but when considered in the context of all possible occurring colors in a scene, naturally-toned faces are limited in their coloration. Consider a color picture of a human face, where each pixel is some combination of red, green, and blue. If we graph each pixel in a 3-D coordinate system with red on the “x axis,” green on the “y-axis,” and blue on the “z-axis,” the color of the pixel places it somewhere in this graph. Pixels of natural human skin colors fall into a limited region within the total color space if imaged under natural illumination conditions. For example, no facial pixel can be purely on the red, green, or blue axes. If many pixels neighboring each other in an image all have a round shape and fall within the permissible area of the “skincolor space,” and near the center of those pixels we find two horizontally related areas resembling eyes, we can conclude that this group of pixels may represent a face. Those pixels can be segmented away from the rest of the image to separate the face from the background.

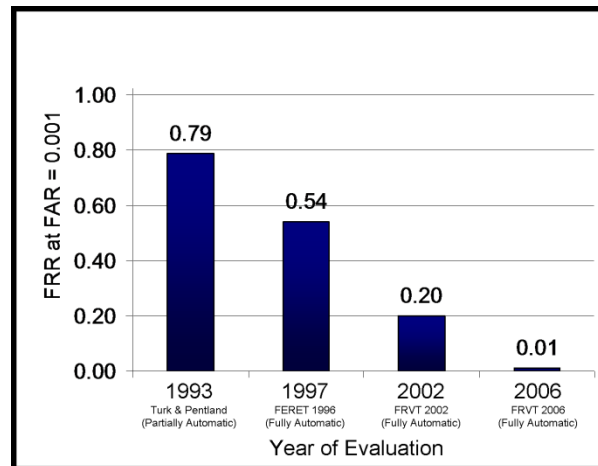
Under funding from the U.S. government (ONR, Technical Support Working Group [TSWG], and the Intelligence Community [IC]), face segmentation algorithms have improved significantly over the last few years, but errors can still occur. If background colors are too similar to skin tones, or eyes cannot be found, or the lighting is stark or of an unnatural color, the system can fail to find faces in the image. Face-like objects also can be incorrectly segmented to be faces. With video sources, motion information also can be considered from frame-to-frame to improve face detection performance.

## **1.2 Face Detection**

Although face recognition is pre-conditional on the performance of face detection, there have been only limited organized, government research projects and evaluations that isolate the face detection problem. Within the past few years, face detection capabilities have been introduced in commercial cameras from several leading manufactures. *MITRE recommends that the FBI consider organizing a “face detection challenge,” which would include a carefully organized data collection effort.* The desired outcome would be to better understand the performance and limitations of face detection across various media from still and video sources; robust face detection systems are those that perform well and are invariant to illumination, orientation, and camera distance. A second potential outcome would be to encourage providers to make face detection available as a utility that is independent to face recognition, such as face quality assessment for collection scenarios or unattended capture requirements.

## **1.3 Performance and System Evaluations**

Face recognition systems have advanced considerably over the past decade. There are several products from leading vendors that have reached a level of maturity suitable for use on portrait type images commonly used for identification card programs and travel documents. As reported in the Face Recognition Vendor Tests (FRVT) and subsequent National Institute of Standards and Technology (NIST) Face Recognition Grand Challenge (FRGC), systems are now capable of achieving fully automated recognition rates in the high 90th percentile at a false accept rate of .001 on high resolution, high quality data and within certain constrained environments.



**Figure 1-3. Reduction in Error Rates in FRVT [Phillips, et al, 2007]**

In real-world applications scenarios, face recognition still encounters difficulties performing consistently in the presence of limited image resolution or large variations in pose, illumination, facial expression, time delay (aging), and possible occlusions. In other studies where similar face recognition technology has been evaluated against images of lower resolution or quality, automated facial recognition performance has been less operationally acceptable.

## **1.4 Human Recognition and Automated Recognition**

Automated techniques currently do not utilize the methods of human forensic examiners. For example, moles, scars, and blemishes that humans may use as a basis for exclusions are largely ignored by automated techniques. The forensic face recognition community requires more consistent training material and research to quantify how distinctive and permanent these features really are. Preliminary forensic face recognition course material is being developed by the FBI at the Forensic Audio, Video and Image Analysis Unit (FAVIAU) which is within the Operational Technology Division (OTD) Digital Evidence Section (DES). This course material represents a positive step toward articulating how to advance the practice and supporting science. Details of this work have been published by Dr. Nicole Spaun in the proceedings of the IEEE Biometric Theory, Applications, and Systems 2007, *Forensic Biometrics from Images and video at the Federal Bureau of Investigation*.

## **1.5 Face Standards**

Unlike other image-based biometric modalities that have specialized sensors, face recognition relies on “all purpose” commercial cameras. While the resolution and imaging capabilities of commercial cameras are sufficient for many face recognition applications, the variance in quality is due not only to the camera, but also the acquisition environment. The relevant international standard is ISO/IEC 19794-5: 2005, “Information technology–Biometric data interchange formats–Part 5: Face image data.” This standard was developed to support requirements of the

International Civil Aviation Organization (ICAO) for images to be placed on electronic passports with limited storage capacity. It should not be seen as replacing the mug shot and facial image standards in ANSI/NIST ITL 1-2000/2006 “Best Practice Recommendation for the Capture of Mugshots,” Version 2.0, September 23, 1997.

Additional discussion on the ISO/IEC 19794-5 standard and how it pertains to law enforcement applications is contained in the June 2008 MITRE report, Certified Product List Way Ahead, also part of SABER.

## **1.6 3-D Face Recognition**

Since 2002, with help from government investments, 3-D face and hybrid 2-D to 3-D techniques have demonstrated modest promise toward improving the robustness and utility of face recognition in less constrained environments. The general approach stems from 3-D computer graphics techniques pioneered in the 90s to render (or reconstruct) images based on changing the orientation and illumination conditions. Blanz and Vetter applied these techniques to the face recognition problem to compensate for some of the pose and illumination effects [Blanz and Vetter, 2003]. A sufficiently detailed 3-D face model can be used to compensate for pose, illumination and expression – factors that degrade classic 2-D methods. 3-D models can also be used to generate large sets of reference images for each subject that include anticipated variances; thus boosting recognition ‘performance.’ Recent NIST studies have shown that 3-D methods can augment, but not replace, 2-D facial recognition, as 3-D facial recognition requires the presence of standard 2-D “texture” images to work effectively. Consequently, the addition of 3-D processing methods and the use of 3-D sensors adds cost and complexity to standard 2-D facial recognition techniques.

Face data can be stored in 2-D, 3-D, or a combination of both formats. Pure 3-D data is shape only and can be represented with points, meshes, or range images. Normally, both shape and texture is concurrently imaged and the 2-D image is applied or “textured” onto the 3-D surface by virtue of sensor registration. With alignment of common landmarks, the 2-D and 3-D components can be manipulated in a coherent fashion.

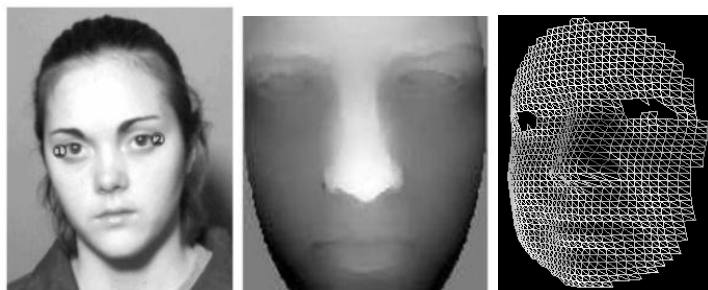


Figure 1-4. 2-D and 3-D Representations [from Chang, 2004]

## 1.7 3-D Face Acquisition

Currently, there are four technologies available for acquiring 3-D face data:

- Stereo imaging
- Structured light sensor
- Laser sensors (e.g., Lidar)
- Hybrid techniques.

**Stereo imaging** is the process in which two cameras are mounted with known fixed distance between them (parallax). The distance is used with machine vision techniques to triangulate data on the image planes and estimate range information for each pixel. Stereo graphics have been used in other disciplines and are a reasonable means for approximating depth information. These techniques are hard to sustain at high frame rates and accuracy can fluctuate depending on the distance of the subject to the camera and other environmental factors. An example of a commercial product using this technique is Geometrix's ALIVE's FaceVision 200 series system.<sup>3</sup> Note – There is potential to use multiple uncontrolled still (or video) images to generate 3-D models, while not “stereo” per se, it involves the same process, using a profile and front view to create a 3-D view.

**Structured light** techniques use a camera and light projector that projects a known “structured” pattern onto the face target. The resulting distortion of the light pattern is used to compute depth information while it concurrently images 2-D texture information. An example of a commercial product using this technique is the Konica Minolta Vivid system. This was the system used by University of Notre Dame for the 3-D data collections for NIST's FRGC.

---

<sup>3</sup> Geometrix, an ALIVE Tech Corporation, [www.geometrix.com](http://www.geometrix.com)

**Laser scanners** are theoretically the most accurate, but also the most expensive modality with potentially the slowest scan rates. Depending on the product and its intended purpose, the scanning process may take up to 30 seconds to produce 3-D point data. While this is appropriate for site surveys and scanning inanimate objects, it is not practical for scanning live, moving faces. Some 3-D Lidar sensors have demonstrated improvements in this area, but remain mostly as research prototypes and not commercial products.

**Hybrid techniques** combine one or more of the above methods. For example, the 3-DMD “3-DMDface” system<sup>4</sup> is a commercial product that combines stereo graphics and structured lighting. According to the vendor, this sensor captures a full face image from ear to ear and under the chin in 1.5 milliseconds with geometric distortion less than 0.15 mm.

## 1.8 Technology Gaps and Challenges

FBI has many applications for face recognition involving a variety of video and still image sources, such as closed circuit television surveillance, broadcast video, mug shots, identification cards, badges, yearbooks, and personal “snapshots.” Some of the major challenges and related recommendations are presented below.

- There is no widely accepted common training and minimum proficiency for human surveillance operators who also perform identification. FBI should develop and provide common training material for human face examiners to fill this need (preliminary initiative in this area is currently underway from FAVIAU).
- Provide a more quantified understanding of facial landmarks and dermal characteristics as they appear over time and through different media.
- The inconsistent face image quality from mug shot environments performs well below the current NIST evaluation results on idealized, high quality images. Where ever possible, FBI should require the minimal adoption of Subject Acquisition Profile 40 and encourage continued progress toward profiles 50/51 (refer to the June 2008 MITRE report Certified Product List [CPL] Way Ahead for additional details).
- Face recognition depends on successful face segmentation (or face detection), which is known to suffer performance degradation due to imaging and orientation factors. FBI should evaluate the use of research tools for face detection against relevant media to include video sources, uncontrolled images, mug shots, and civil identification photos. As there currently is no known evaluation dataset for face detection performance, the FBI should consider developing a “challenge

---

<sup>4</sup> 3-DMD, a 3Q company, 3-DMDface™ System, <http://www.3-Dmd.com/Products/3-DSystems.asp>

dataset” that represents their face detection and forensic (quality) needs in partnership with appropriate existing research programs such as IARPA.

- Develop methods and techniques for searching a database with sketches and composite images constructed by forensic artists or computer generated composite.



## 2 Iris Recognition

### 2.1 Background

Alfonse Bertillion understood in the 1880s that individuals could be recognized based on their iris images, specifically the distinctive patterns and textures in the irises created by various structures, known in the modern ophthalmologic literature as crypts, furrows, frills, ridges, ligaments, freckles, coronas, and collarettes (Figure 2-1).<sup>5</sup> The iris is the only internal organ that can be seen outside the body. Human irises carry a stable information density of more than 3.2 measurable bits per mm<sup>2</sup>.<sup>6</sup> It is widely held, although not scientifically established, that after youth, the iris' patterns do not change over an individual's lifetime. Further, it is commonly believed that iris texture patterns are different for each person and for each eye of the same person.<sup>7</sup> If these assumptions are true, irises would be ideal for biometric identification.

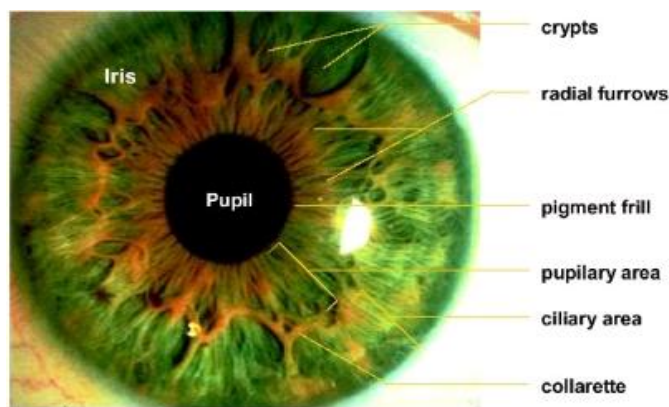


Figure 2-1. Structures of the Iris

Iris recognition realizes some of the advantages of fingerprint and facial recognition, while simultaneously minimizing some of their disadvantages. For example, fingerprint recognition is known for its low error rates; however an individual is typically required to place his or her finger on a sensor to be recognized. Physical contact is required. The error rates for current facial recognition technologies are typically higher than for fingerprint technologies. However, facial recognition is often preferred because its operation is non-contact. Iris recognition combines the low error rates of fingerprinting and the non-contact operation of facial recognition and, as such, may prove valuable for many criminal justice applications.

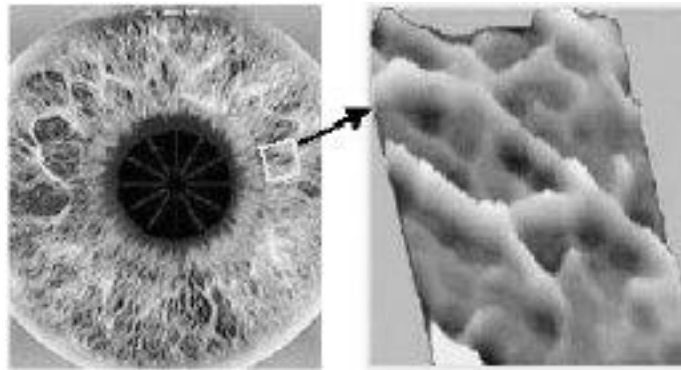
<sup>5</sup> Image from <http://www.rdecom.army.mil/rdemagazine/200305/index.htm>. (accessed September 1, 2007).

<sup>6</sup> Bakk. Medien-Inf. Tilo Burghardt, "Inside Iris Recognition," Master Course in Global Computing and Multimedia, University of Bristol, November 2002, p. 1,

<http://www.cs.bris.ac.uk/home/burghard/SecurityReport.pdf> (accessed May 25, 2008).

<sup>7</sup> <http://www.cl.cam.ac.uk/~jgd1000/genetics.html> (accessed May 26, 2008).

One constraint of iris recognition is that a modicum of cooperation is required on the part of the individual user: the user's eyes must be open and facing in the direction of the camera. Facial images can be collected covertly from uncooperative users and matched even when the eyes are closed (though the user still must be facing in the direction of the camera). An uncooperative user's finger can be placed on a fingerprint sensor via brute force. However, it is difficult to force an uncooperative user to open their eyes. As such, iris recognition may be most applicable to user groups that are either cooperative or unaware that the technology is being employed.



**Figure 2-2 Grayscale Image and Texture of Iris**

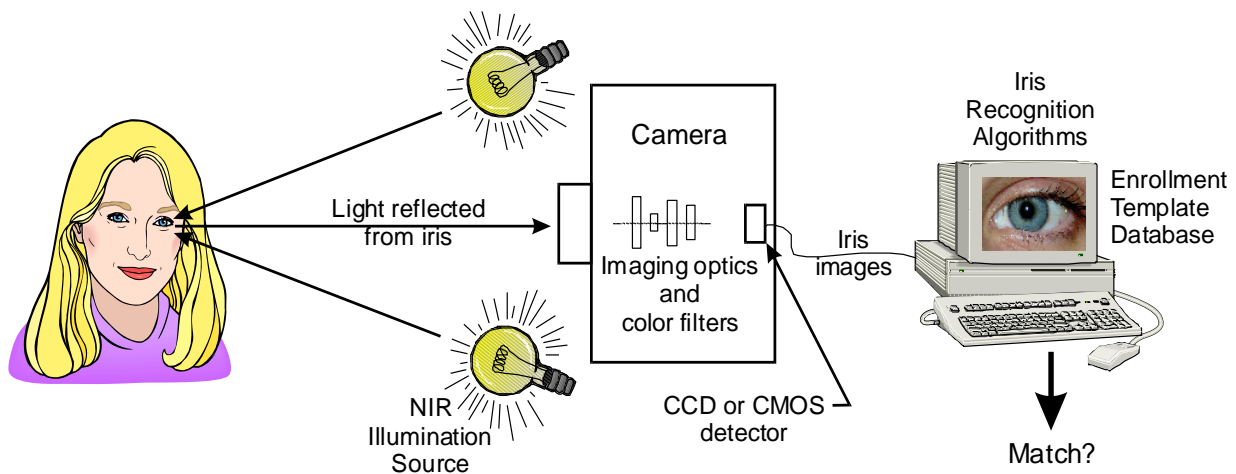
Most currently deployed automated methods for iris recognition systems are based on algorithms and methods developed by Professor John Daugman at the University of Cambridge. Daugman's method, originally patented by Iridian in the early 1990s, is the basis for many practical and effective applications. Iridian was purchased and merged into L-1 Identity Solutions in 2006, the same year the original patent expired. However, the strength of that patent had hampered multiple-vendor competition.

Other groups have conducted significant development over the last few years and second wave of iris capture devices and recognition software is emerging (see tables 2-3 and 2-4). Recent advances for acquiring iris images has extended use of iris from being predominantly access control to a variety of other applications, and several significant technology evaluations have been conducted within the past five years.

Most commercial iris recognition products identify individuals using images of iris patterns captured in the near-infrared (NIR) portion of the optical spectrum. NIR illumination is used for several reasons. First, ambient visible light can be filtered from the images to prevent undesirable environmental reflections in the iris image. The orientation of NIR light sources in iris recognition cameras is well-controlled to inhibit undesirable NIR reflections in the iris images. In addition, NIR light is not intrusive to the human eye. Most people cannot see NIR light. Finally, the iris texture of darker irises is more fully revealed with NIR illumination. Melanin pigment, which is present in larger amounts in dark eyes, absorbs much of the light in the visible spectrum, but

reflects much of the light in the NIR portion of the spectrum. As such, NIR iris images are almost always used for iris recognition purposes.

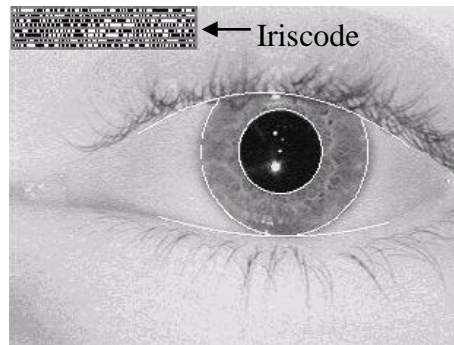
In the NIR, the visible “color” of the iris is not observed, and a monochromatic grayscale representation of the iris is used (Figure 2-2). Current iris recognition cameras photograph irises at object distances from several inches to several meters in front of the eye. Typical iris recognition cameras for access control yield a 640x480 image of the eye, where the diameter of the iris is approximately 180 to 200 pixels.



**Figure 2-3. Schematic of Iris Recognition Operation**

The basic operation of an iris recognition system follows four general steps: data acquisition, iris segmentation, texture encoding, and matching. The data acquisition process is illustrated schematically in Figure 2-3. Light from an NIR illumination source, such as NIR light emitting diodes (LEDs) or a flash lamp, is reflected off the individual’s iris, and an NIR image of the iris is collected with a camera. The goal of data acquisition is to obtain a high-quality image that can readily be used for iris recognition purposes.

The iris image is then transferred to a computer where a segmentation process locates the pupil and iris centers, radii, and boundary regions interior to the white of the eye (sclera) and the eyelids. The goal of iris segmentation is to isolate the iris region in the image. Segmentation locates the pupillary boundary between the pupil and iris, and the limbic boundary between the iris and the sclera; these boundaries are not concentric. Early systems modeled these boundaries as circles; current state-of-the-art systems are modeling the boundaries as ellipses. The segmentation process is illustrated by the white outlines around the iris in Figure 2-4.



**Figure 2-4. Eye Image with the Iris Area Segmented.<sup>8</sup>**

The segmentation step is one of the more challenging and crucial steps of the iris recognition process. If an iris is segmented incorrectly, it is difficult, if not impossible, to correctly match that iris in future recognition attempts. Segmentation is difficult because the influence of various features in the image that occlude the iris (e.g., eyelids, eyelashes, specular reflections, and shadows) vary widely, and sometimes drastically, over the diverse human population, over different data acquisition environments, and between different camera systems. Segmentation algorithms must be sophisticated enough to take these differences into account and accurately locate the iris regardless. In addition, different camera systems utilize different illumination and detection spectral (wavelength) bandwidths, which can influence contrast between the pupil, iris, sclera, eyelashes, eyelids, and reflections, making iris segmentation more challenging. A significant amount of intellectual property is invested in iris recognition segmentation algorithms.

After segmentation, the texture encoding step converts the iris pattern into a bit vector code, typically by applying a filter. A variety of filters have been used to encode iris texture information.<sup>9</sup> Professor Daugman uses a Gabor (log) filter. In Professor Daugman's approach, the segmented iris region is normalized and converted from Cartesian coordinates into polar coordinates. Multiple 2-D Gabor filters are applied to the image to encode it into a 256-byte binary code known as an "iriscodes." Specifically, the normalized image is divided into a grid of regions. The dot product is computed between complex Gabor filter and each region over which the filter is placed, and the phase angle of the resulting complex dot product is then quantized to two bits. The resulting two bits are assigned to that region. The results for each region are then assembled to create the 2048-bit iriscodes. In this fashion, using multiple Gabor filters of different sizes, the texture of the iris is encoded in terms of wavelet phase information spanning several

---

<sup>8</sup> The iris code of 0s and 1s is depicted in the upper left corner with 1 indicated by white and 0 by black.

<sup>9</sup> For a list of filters that have been applied to iris recognition and their associated references in the open literature, refer to Table 2.1 in Xiaomei Liu's PhD dissertation, "Optimizations in Iris Recognition," Notre Dame, 2006 (accessed May 26, 2008).

scales of analysis.<sup>10</sup> The resulting code cannot be directly interpreted in terms of the iris structures, but rather a mathematical abstraction of the iris image. Interestingly, the amplitude of the resulting complex dot product (the “real” part of the complex result) is not used. As such, Daugman’s approach is fairly tolerant to amplitude variations in the image (brightness and contrast).

Regardless of which filter and encoding approach is used, the goal of the encoding step is to obtain a representation of the iris texture pattern that can be subsequently used to compare irises. For example, the iriscodes illustrated in Figure 2-4 is essentially a non-invertible digital representation of the iris texture in Figure 2-2 impacted by the iris features illustrated in Figure 2-1.

Two irises can be compared by finding the distance, or more commonly the Hamming Distance (HD), between the iriscodes. A Hamming Distance is the count of the number of bits that are different between two vectors of 0s and 1s that are of equal length. A “normalized” Hamming Distance divides this count by the total number of bits compared, and therefore will be a measure between 0 and 1. Unusable bits, such as those that are masked in the segmentation process due to obscuration (e.g. by reflections, eyelashes),, are ignored when computing the distance between two iriscodes resulting in an adjusted HD based on the fraction of valid bits compared. The adjusted HD (referred to as HD in the discussion below) is then compared to an application-dependent threshold (typically 0.32, but usually adjusted in operational systems to meet the competing error rate requirements of the system management) to determine if the irises match. An HD of 0.32 means that when two iriscodes are compared, 32 percent of the bits disagree. As such, a lower HD indicates that the templates are more similar, and a higher HD indicates that the templates are less similar. In other words, lower HD scores indicate better matching performance for Daugman-based algorithms.<sup>10</sup>

Daugman’s approach to iris recognition possesses several key advantages. First, computing bit differences between two iriscodes is very fast. Second, it is easy to handle rotation of the iris; one iriscodes is simply shifted relative to the other, and the two codes compared again. Finally, matching results can be interpreted using a statistical test of independence.<sup>11</sup> Many other approaches have been proposed, however the Daugman approach remains the most popular. Most commercial iris recognition systems utilize Professor Daugman’s algorithms.

---

<sup>10</sup> For a detailed technical description of Professor Daugman’s segmentation, Iriscodes generation, and matching algorithms, refer to “How Iris Recognition Works,” John Daugman, PhD, OBE, <http://www.CL.cam.ac.uk/users/jgd1000/csvt.pdf> (accessed September 1, 2007).

<sup>11</sup> J. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” IEEE Transactions on Pattern Analysis and Machine Intelligence **15**(11), pp.1148–1161 (1993).

## 2.2 State of the Industry

The current generation of commercial iris recognition products is designed for a variety of operational scenarios. For example, some iris recognition systems use single-eye cameras and other systems use dual-eye cameras. For single-eye cameras, the left and right eyes are presented to the camera separately (the camera optics collect only one iris image per user presentation). For dual-eye systems, the left and right eyes are presented to the camera simultaneously (the camera optics collect both left and right iris images during one user presentation).

In addition, different levels of user participation (predominantly active or predominantly passive) are required to interact with different systems. Most iris recognition systems require the user to look directly into the center of the camera (on-axis presentation) from within a zone known as the collection volume. The size and location of this collection volume depends upon the design specifications for the camera. In many cases, the user is responsible for placing themselves within this collection volume, which may be small or large. Often the camera will provide visual or auditory cues to help the user find the appropriate location. In some cases, the user must purposefully align their eyes in the camera; mirrors are often employed to help the user provide an on-axis presentation (active user effort). In other cases, the user need only look in a specified direction once they are located within the collection volume (nominal user effort). For some systems, the user looks straight ahead and a trained operator aligns the camera (minimal user effort) to obtain an on-axis presentation.

**Table 2-1. Types of Iris Recognition Camera Products**

Type	Collection Volume	User Effort	Eye Configuration
Access Control	Moderate	Active User places eyes in appropriate location	Typically Dual-Eye, some models Single-Eye
Single-Eye Handheld	Small	Minimal Operator aligns camera with user's eye	Single-Eye
Dual-Eye Visor	Small	Active User aligns visor with eyes	Dual-Eye
Stand Off	Large	Nominal User looks toward camera	Dual-Eye, some with combined face

## 2.3 Growth and Markets

### 2.3.1 Brief History

The use of the iris to identify people was first proposed by French criminologist Alphonse Bertillon in 1885.<sup>12</sup> He developed a system to classify the pigment and arrangement of the aureole around the pupil and of the periphery of the iris.<sup>13</sup> The field of iris recognition was furthered in the 1930s when ophthalmologists observed that each iris had a detailed and unique structure that remained unchanged over decades. Ophthalmologist Frank Burch proposed using iris patterns for human identification in 1936.<sup>14</sup> The idea was also conveyed in the 1983 James Bond movie “Never Say Never Again.” In 1986, ophthalmologists Aran Safir and Leonard Flom patented the concept of iris recognition.<sup>15</sup> In 1989, they asked Professor John Daugman, who was at Harvard University at the time, to develop recognition algorithms based on the iris. Professor Daugman’s original work, which combines computer vision, wavelet theory, and statistical pattern recognition, was published in 1993.<sup>16</sup> Professor Daugman further patented his work in U.S., European, and international patents in 1994 and 1995. A competing approach was patented by Sarnoff Laboratory in 1996.

A company called IriScan was the Assignee for the Flom and Safir concept patent, and both the Daugman and Sarnoff implementation patents. IriScan successfully commercialized Professor Daugman’s algorithms through partnerships with several device integrators, with the first commercial products becoming available in 1995. With protection afforded by the concept patent and interoperability assured by the Daugman patent, iris recognition systems were broadly deployed without the competition and testing that was common for facial and fingerprint biometric modalities. Even today, most existing commercial iris recognition technology is based on Daugman’s work.

Patent issues have historically characterized the iris recognition sector.<sup>17</sup> IriScan merged with Sensor in 2000, and changed its name to Iridian Technologies. Iridian owned and was very

---

<sup>12</sup> A. Bertillon, “La couleur de l’iris,” *Revue scientifique*, **36**, p.65 (1885). [http://www.authentic-corp.com/resources/Bertillon\\_Couleur\\_de\\_L'Iris.pdf](http://www.authentic-corp.com/resources/Bertillon_Couleur_de_L'Iris.pdf) (in French) (accessed March 5, 2008).

<sup>13</sup> <http://chnm.gmu.edu/courses/magic/plot/bertillon.html> (accessed March 5, 2008).

<sup>14</sup> National Center for State Courts of the United States of America, The Court Technology Laboratory <http://ctl.ncsc.dni.us/biomet%20web/BMIris.htmlhttp://www.cs.bris.ac.uk/home/burghard/SecurityReport.pdf> (accessed March 3, 2008).

<sup>15</sup> L. Flom and A. Safir, U.S. Patent No. 4641349 (1986), International patent WO8605018A1 (1986).

<sup>16</sup> John G. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15** (1993) <http://www.cl.cam.ac.uk/~jgd1000/PAMI93.pdf> (accessed March 3, 2008).

<sup>17</sup> See <http://www.iris-recognition.org/> for a complete list of iris recognition software and hardware patents through March 2002.

protective of the Flom and Safir concept patent, which made it difficult for alternative solution providers to enter the iris recognition market. For example, Iridian filed legal suits against a small South Korean company, IriTech, over patent disputes in 2002. The cases were resolved in 2004. Iridian dropped LG Electronics from its list of licensees in 2004, due to licensing agreement disputes. Iridian officials assailed a small UK company, Smart Sensors, when Smart Sensors announced their alternative iris recognition algorithms in 2004. These and other similar legal disputes allowed Iridian to be the governing iris recognition technology provider, and all dominant providers of the technology relied on the patented, intellectual property from a single vendor. As a result, the iris recognition competitive market was stifled for many years.

The key Flom and Safir concept patent expired in the U.S. in February 2005 and opened the doors to other implementations. L1 Identity Solutions acquired Iridian Technologies in 2006 and assumed patent rights to the Daugman algorithms. LG Electronics and Iridian (L-1) resolved their licensing and intellectual property dispute in May 2008. The market is starting to show signs of healthy competition, which is no longer limited to integrators who license technology from Iridian (L-1). As a result, a variety of alternative iris recognition algorithms and a wide variety of camera systems are now available. The patent on Daugman's specific implementation of iris recognition expires in 2011.

### **2.3.2 High-Profile Implementations**

A variety of successful, high-visibility implementations of iris recognition have demonstrated the effectiveness of the technology. Several are described below.

#### **2.3.2.1 UK IRIS**

The UK Iris Recognition Immigration System (IRIS) went live at London Heathrow Airport in 2005 and has since become operational at Manchester, Birmingham, and Gatwick airports. Specialized IRIS enrollment offices are in place at each of those airports. Those wishing to use the system must register in advance by presenting their passport and having images of their irises taken. Incoming passengers make no claim of identity as they enter the IRIS kiosk (Figure 2-5)<sup>18</sup>. Passengers look into a camera, which images both of their irises. If a search of those irises against the iris database of all enrolled travelers reveals very similar irises, the passenger is assumed to be that enrollee. An immigration receipt is printed with the name of the "identified" traveler as listed in the database and the border crossing is recorded. As of this writing, there are over 100,000 people enrolled in UK IRIS and about 12,000 transactions per week. Each transaction requires that the submitted iris images be searched against those in the entire database, so there are on the order of one billion comparisons per week. Most of those comparisons are to enrolled irises not matching the data subject, so there are approximately one billion opportunities per week for a

---

<sup>18</sup> [http://newsimg.bbc.co.uk/media/images/41556000/jpg/\\_41556080\\_iris203.jpg](http://newsimg.bbc.co.uk/media/images/41556000/jpg/_41556080_iris203.jpg)



false match. The false match error rates must be extremely low to support this application. Any user not recognized by IRIS is referred to the primary immigration queue for processing.



**Figure 2-5 UK IRIS Kiosk at Manchester Airport**



**Figure 2-6. Primum Kiosk at Schiphol**

### **2.3.2.2 Schiphol Airport Privium System**

The Privium system, a fast-track border passage program, was launched at Schiphol airport in Amsterdam, Netherlands in October 2001. Privium is a service for frequent travelers and allows subscribers to clear immigration using iris recognition and a smart card as proof of identity. Upon enrollment, subscribers are issued a smart card that contains a digital representation of their iris. In

the airport, subscribers insert their smart card into a reader at the kiosk (Figure 2-6)<sup>19</sup>, and then proceed to present their iris to an LG IrisAccess 2200 reader (Figure 2-7)<sup>20</sup>. If the presented iris matches the data on the card, the turnstile opens and the subscriber can proceed. There is no central database for the Privium system, iris data is stored only on the smart card.



**Figure 2-7. PRIVIUM Iris Recognition**

Privium is open to European Economic Area passport holders and basic subscription costs €99. The program, which has over 30,000 members, offers additional benefits to its members, such as priority parking and business class check-in.

### **2.3.2.3 Afghan Repatriation Program**

The United Nations High Commissioner for Refugees (UNHCR) started to use iris recognition to help stem fraud during repatriation of Afghan refugees in 2002. Registered refugees receive an assistance package upon arrival in Afghanistan that can include a monetary grant, currently USD\$100, food, and some non-food items like shelter materials and agricultural kits. To prevent refugees from doubling back across the Pakistan border to claim repatriation packages multiple times, UNHCR implemented an iris recognition system. The iris recognition systems, which use LG IrisAccess 2200 equipment, are set up at several fixed screening locations in Pakistan, and mobile units are available for use in remote areas. When returnees are screened, their irises are enrolled anonymously in the database and checked against all other irises previously enrolled in the database. If there is no match, the refugee is registered and given clearance to receive the assistance package upon arrival in Afghanistan. If the comparison reveals that the returnee's irises are already in the database, the person is refused a second assistance package.

---

<sup>19</sup> <http://www.flyvln.com/emc.asp?pageId=547>

<sup>20</sup> <http://www.dartagnan-biometrics.com/?id=24>



**Figure 2-8. Afghan Repatriation Program**

To meet the cultural needs of the refugees, tests on women and children are done by female refugee agency workers. In addition, only the eye is seen onscreen so traditional objections to photographing a women's face is not an issue<sup>21</sup>. To protect privacy, no information that can identify the refugee (e.g., name, age, or destination) is recorded in the iris database.

UNHCR has operated a voluntary repatriation drive each year since 2002. As of February 2008, over three million Afghans have returned home during that time, while two million registered (and an unknown number of unregistered) Afghans remain in Pakistan.

#### **2.3.2.4 UAE Iris Expellees Tracking and Border Control System**

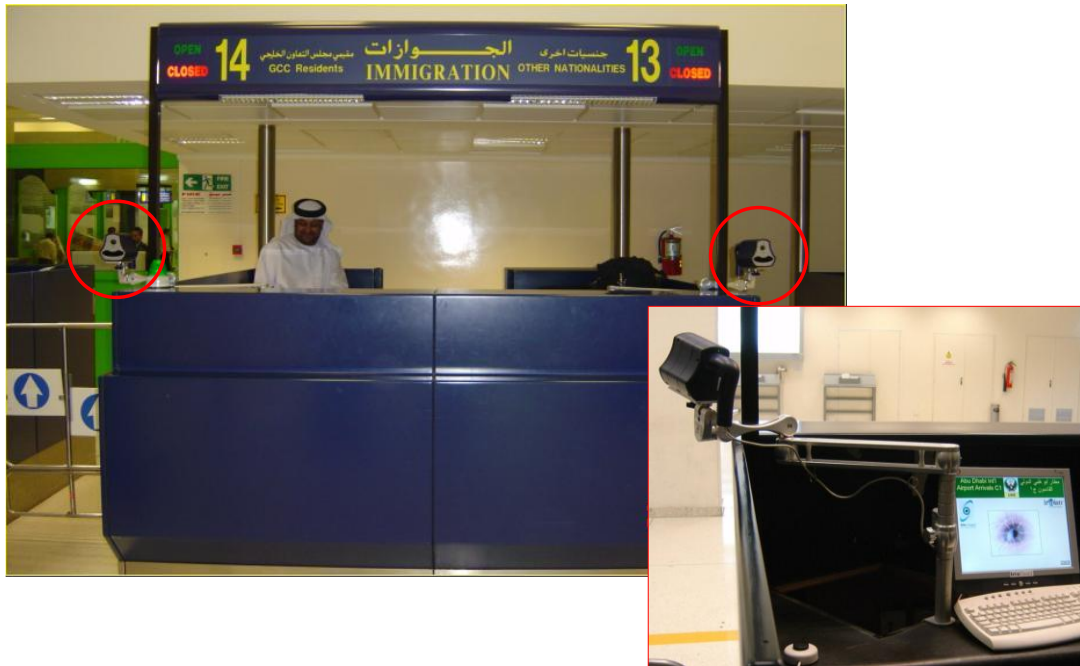
Abu Dhabi Police in the United Arab Emirates (UAE) piloted an iris-based expellees tracking and border control system in 2001 and rolled the system out nationally in 2003. Irises from prison inmates and all foreigners expelled from the UAE are enrolled and merged into one central database. When a passenger arrives at any UAE air, land, or sea border point, their irises are compared in real time via internet links to the irises in the database to reveal any person who was previously expelled from the country or who spent time in a UAE prison. If a match is not found, the passenger is cleared to enter the UAE. If a match is found, entry into the UAE can be denied.

Multiple detached deportation centers and border point centers are geographically distributed throughout the UAE. The border point systems are integrated within passport control, as

---

<sup>21</sup> <http://www.unhcr.org/cgi-bin/texis/vtx/home/opendoc.htm?tbl=NEWS&id=3f86b4784&page=news>

illustrated in Figure 2-9. The systems architect, application provider, and integrator for the system is IrisGuard, and IrisGuard's IG-H100 cameras are used.



**Figure 2-9. UAE Iris Expellees Tracking and Border Control System**

As of March 16, 2008, 1,504,432 irises representing 160 nationalities are enrolled in the database—the largest and most searched iris database in the world. Over 15,528,600 searches and 10 trillion cross comparisons have been carried out, with 216,047 past expellees revealed. According to the Ministry of Interior, all matches have been confirmed by other records.<sup>22</sup> The turnaround time for an exhaustive search through the database is less than 2.0 seconds.

### **2.3.2.5 Iris Recognition in the DoD**

The Department of Defense (DoD) has used iris recognition for detainee population management, personnel screening for access to bases and facilities (Figure 2-10)<sup>23</sup>, mobile identification (Figure 2-11)<sup>24</sup>, and intelligence analysis. The first application of iris recognition in the DoD used Securimetrics Pier devices for systems in fixed locations. The Pier devices were initially fielded as

<sup>22</sup> <http://www.cl.cam.ac.uk/~jgd1000/deployments.html>

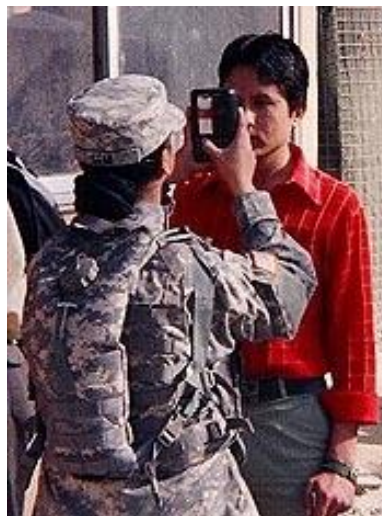
<sup>23</sup> <http://www.securimetrics.com/solutions/gfx/iraq.jpg>

<sup>24</sup> [http://www.biometrics.gov/images/DoDBTF/DoD\\_1.jpg](http://www.biometrics.gov/images/DoDBTF/DoD_1.jpg)

part of the Biometrics Automated Toolkit (BAT) in 2003 and were used for detainee management. The Biometrics Identification for Secure Access (BISA) systems uses similar Securimetrics devices, and is used for enrolling applicants seeking access to U.S. bases. The Securimetrics Handheld Interagency Identification Device (HIIDE) incorporates an iris camera, a single fingerprint sensor, and a regular camera to collect facial images. HIIDE is used for mobile identification and collection applications. Although there are areas for improvement and optimization, the DoD has made successful use of iris recognition technology in a variety of environments.



**Figure 2-10. U.S. Marine Corp Entry Control Point in Fallujah, Iraq**



**Figure 2-11. U.S. Soldier use Iris Recognition to Verify Identity**

## **2.4 Performance**

Early performance studies of iris recognition technology were performed in 1996 by the U.S. Department of Energy Sandia National Laboratories, in 1997 by British Telecom, in 2000/2001

by the UK National Physical Laboratory, and in 2001/2002 by the U.S. DoD Army Research Laboratory. These studies were performed on early-to-market iris recognition products and often with pre-standardized test protocols. As such, results of these tests are difficult to interpret and not indicative of the performance of today's mature commercially-available iris recognition products.

Three publicly-accessible evaluations have been performed in the last few years. The International Biometrics Group (IBG) performed the "Independent Testing of Iris Recognition Technology (ITIRT)"<sup>25</sup> in the 2004 timeframe; the National Institute of Standards and Technology (NIST) conducted the two-phase "Iris Challenge Evaluation" in 2005 and 2006 (ICE 2005 and ICE 2006);<sup>26</sup> and Authenti-Corp performed the "Standards-Based Performance and User Cooperation Studies Of Commercial Iris Recognition Products" study in the 2006 timeframe, also known as the Iris Recognition Study 2006 or IRIS06.<sup>27</sup> The salient results of these iris recognition performance studies are outlined below. A meta-analysis of these three studies was performed by NIST.<sup>28</sup>

#### **2.4.1 IBG ITIRT**

ITIRT was sponsored by the U.S. Department of Homeland Security (DHS). Data was collected from live human test subjects between October and December 2004 in New York City using iris recognition cameras from LG, OKI, and Panasonic. Iris images were collected from 1,224 test subjects using each of the cameras. Additional iris images were collected from 458 of these test subjects when they returned for a second session three to five weeks later. The final report was released in May 2005.

IBG used PrivateID development software toolkits, provided by Iridian, to build custom acquisition applications for the OKI and Panasonic devices. These devices were operated through Iridian's PrivateID Application Programming Interface (API). IBG used the LG IrisAccess 3000 SDK v3.00 to build acquisition applications for the LG device.

Iris samples from all cameras were processed offline (subsequent to data collection) using a single, shared implementation of Iridian's matching software. IBG used Iridian's KnoWho Original Equipment Manufacturer (OEM) SDK v3.0 to build the custom matching application. Iridian provided IBG with a custom utility that converted LG samples into PrivateID format so that LG samples could be processed within Iridian's proprietary KnoWho OEM SDK environment.

---

<sup>25</sup> <http://www.biometricscatalog.org/itirt/ITIRT-FinalReport.pdf>

<sup>26</sup> [http://iris.nist.gov/ice/ICE\\_Home.htm](http://iris.nist.gov/ice/ICE_Home.htm)

<sup>27</sup> <http://www.authenti-corp.com/iris06/report/>

<sup>28</sup> Elaine M. Newton and P. Jonathon Phillips, "Meta-Analysis of Third-Party Evaluations of Iris Recognition," NISTIR 7440 [http://iris.nist.gov/ice/IrisComparisonY070820\\_NISTIR.pdf](http://iris.nist.gov/ice/IrisComparisonY070820_NISTIR.pdf) (accessed 30 May 2008).

IBG measured match rates, enrollment and image acquisition rates, and levels of effort, such as the duration of the transactions. Failure to enroll (FTE) rates ranged from 1.6 percent to 7.1 percent for the three devices tested. An FTE was declared if neither the left or right eye could be enrolled. IBG did not specify the number of attempts allowed before an FTE was declared; however, mean enrollment transaction durations, including those that failed to enroll, ranged from 56 to 74 seconds. Failure to acquire (FTA) rates<sup>29</sup> ranged from 0.3 percent to 0.7 percent. An FTA was declared if neither the left or right eye could be acquired given three attempts with each eye. The mean acquisition transaction duration, including FTAs, ranged from 7.1 to 18.7 seconds. The false non-match rates (FNMRs), which were measured offline using the Iridian algorithm at HD=0.33, ranged from 0.583 percent to 1.57 percent. A false non-match was declared if neither the left nor the right eye matched after three attempts with each eye.

An interesting IBG finding was that “false match rates decrease much more rapidly than false non-match rates; an order of magnitude increase in FNMR frequently corresponds to a five order of magnitude decrease in FMR,” as illustrated in the Detection Error Tradeoff (DET) curves shown in Figure 2-12. In other words, the operating point for the KnoWho OEM SDK algorithm (based on Professor Daugman’s algorithm) can be shifted to achieve very low FMRs with only a slight increase in the FNMR rates. As such, iris recognition, via Daugman-based algorithms, is ideally suited for identification applications where low FMRs are of paramount importance.

Figure 2-12 displays the match results when the enrollment and recognition images were collected from the same camera, which IBG terms the “intra-device” error rates. Figure 2-13 displays the “cross-device” error rates, where enrollment and recognition images were collected from different cameras. Comparing Figures 2-12 and 2-13, we note that cross-device error rates are higher than intra-device error rates (note the difference in the vertical scale between the two figures). In other words, the interoperability performance is not as good as the native performance for these products.

---

<sup>29</sup> Here, FTA rates correspond to the percentage of transactions that failed to collect a satisfactory iris image during a recognition (verification or identification) transaction, and the percentage of transaction that failed to achieve a successful enrollment.

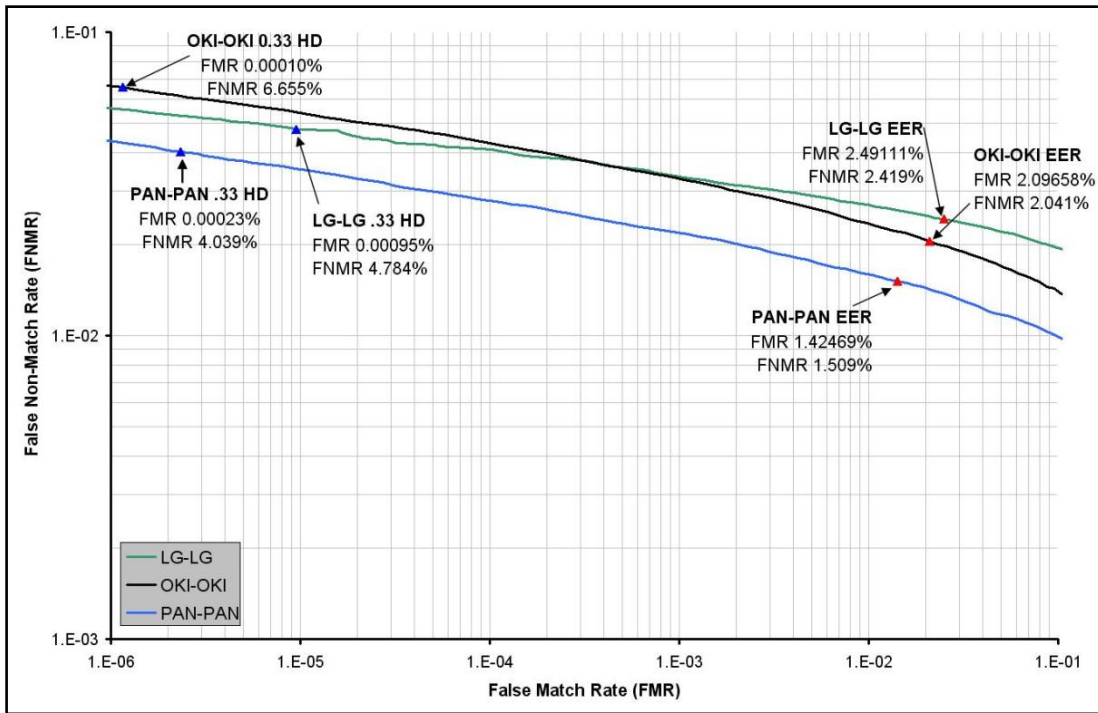
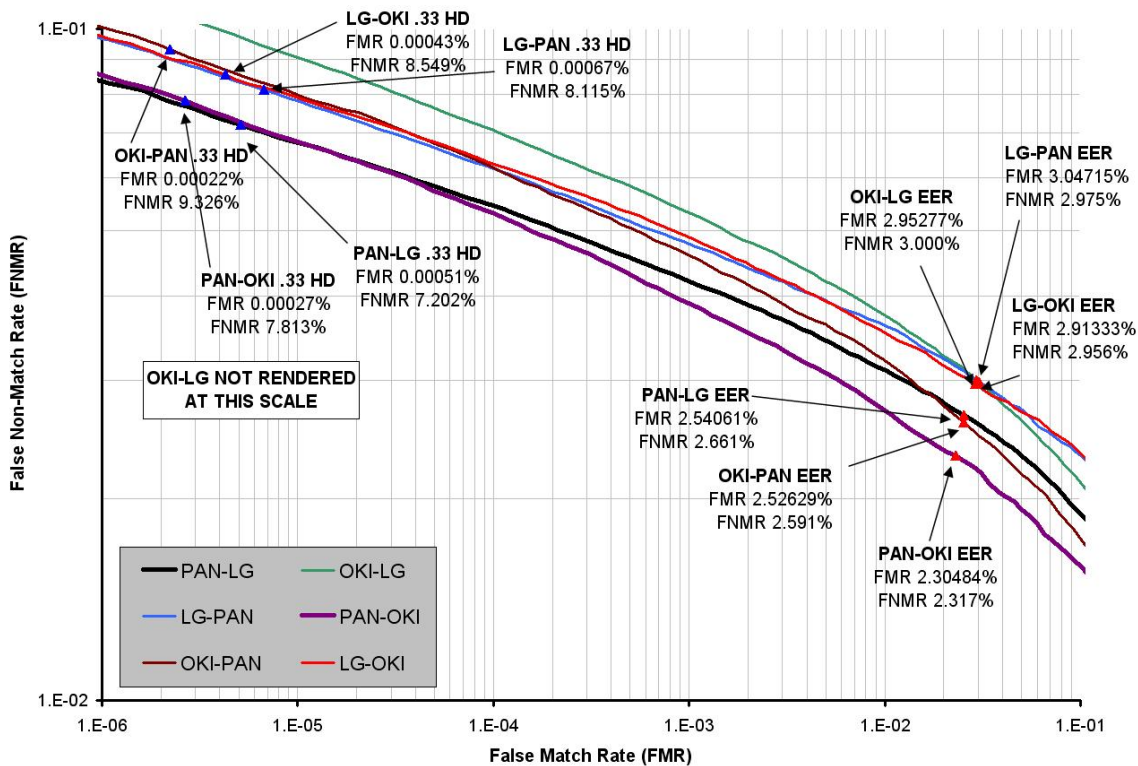


Figure 2-12. ITIRT Intra-Device DET Curves





**Figure 2-13. ITIRT Cross-Device DET Curves**

Another interesting result from the ITIRT test suggests that left and right irises from the same individual are more likely to match than irises from different individuals. This indicates that irises from the same person, much like fingerprints from the same person, are correlated to some degree. This is in contrast with prior analyses presented by Professor Daugman. Further studies are warranted to determine the degree of correlation between an individual's left and right eyes. Much work has been performed in this area for fingerprints.

IBG also performed an experiment where they compressed and decompressed the images (within the bounds detailed in ANSI INCITS 379-2004 Iris Image Interchange Format standard) and then repeated the matching experiments. They found that in nearly 50 percent of the ~4,600 comparisons they executed, the "HDs for processed (compressed-decompressed) sample comparisons were lower (or indicative of a stronger match) than non-processed (no compression-decompression operation) sample comparisons." Further study is required to determine if this is a random effect or if it is a systematic effect due to factors yet to be determined, such as loss of high frequency information in the compression-decompression process. In this vein, the NIST Iris Recognition Exchange 2008 (IREX 08) evaluation is studying how compression influences iris matching performance.

## **2.4.2 NIST Iris Challenge Evaluation**

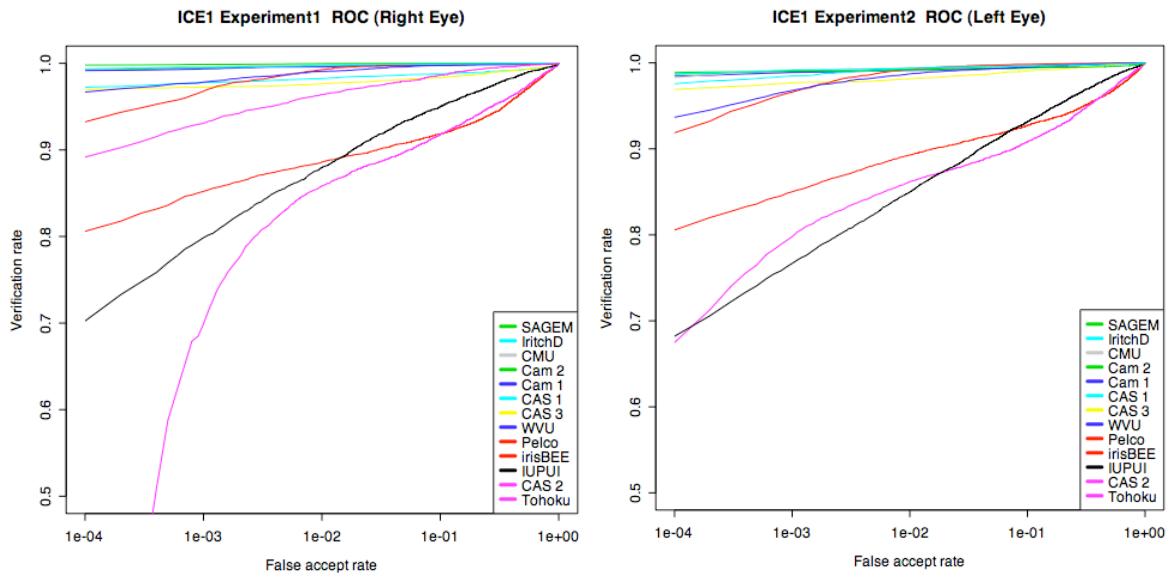
The two-phase ICE program was conducted by NIST and its support contractors, including the Schafer Corporation, the University of Notre Dame, Colorado State University, University of Texas at Dallas, and SAIC. The program was sponsored by multiple U.S. government agencies including DHS' Science and Technology Department and Transportation Security Administration, the Director of National Intelligence Information Technology Innovation Center (now IARPA), the Federal Bureau of Investigation (FBI), the National Institute of Justice, and the Technical Support Working Group (TSWG). The broad goals of the program were to facilitate the development of iris recognition technology and to assess the matching performance of iris recognition technology.

### **2.4.2.1 ICE 2005**

In ICE 2005 (the first phase of the program) iris images from 132 test subjects collected with an LG 2200 camera at the University of Notre Dame were distributed to test participants. Nine groups participated in ICE 2005 from domestic and international universities, research institutes, and commercial companies representing six different countries. The iris image data was distributed in September 2005, and the participants submitted their self-generated similarity matrices to NIST in March 2006. NIST generated Receiver Operating Characteristic (ROC) curves, as shown in Figure 2-14.<sup>30</sup> Figure 2-14 illustrates the relative flatness of the ROC curves over a wide range of FAR values for the best performing algorithms, as was also observed in the ITIRT study. The results also indicated slightly better matching performance for right eyes over left eyes. However, this might be explained by the fact the test protocol instructed all test subjects to present their left eye to the camera first. After some practice with the left eye, test subjects might have had better success interfacing with the camera when it was time to collect images with the right eye. Thus, right eye image quality and right eye matching performance might have improved due to test subject habituation (training).

---

<sup>30</sup> [http://iris.nist.gov/ICE/ICE\\_2005\\_Results\\_30March2006.pdf](http://iris.nist.gov/ICE/ICE_2005_Results_30March2006.pdf)



**Figure 2-14. ICE 2005 ROC Curves**

Another interesting result of the ICE 2005 effort was presented by the University of Bath/Smart Sensors team.<sup>31</sup> They noted that because blue and brown iris structures differ, their researchers could tune the Smart Sensors algorithm to perform well with certain ethnic groups, but not across all ethnic groups at the same time.

#### 2.4.2.2 ICE 2006

For the second phase of the program, ICE 2006, participants were required to submit their algorithms to NIST for an independent evaluation performed with sequestered iris images. Iris images for the evaluation were collected from 240 test subjects by the University of Notre Dame using the LG 2200 camera during the Spring 2004, Fall 2004, and Spring 2005 semesters. Notre Dame used modified acquisition software provided by Iridian to collect three iris images during each iris presentation when at least one of the three images passed LG 2200's built-in quality checks. Since the LG camera's internal iris image quality control algorithm was overridden, about one-third of the iris images used for the ICE 2006 data analysis met acceptable image quality standards, and two-thirds were below acceptable image quality standards.

Eight groups from domestic and international universities, research institutes, and commercial companies representing six different countries participated in the ICE 2006 test. Each group delivered executables to NIST in June 2006; the final report was released in March 2007, in

<sup>31</sup> [http://iris.nist.gov/ICE/UnivBath\\_ICE\\_Brief.pdf](http://iris.nist.gov/ICE/UnivBath_ICE_Brief.pdf)

combination with results from FRVT 2006. Results were presented in the final report for three of the iris recognition participants: the Sagem-Iridian team, the University of Cambridge, and Iritech. “For the ICE 2006 submissions, analysis was restricted to algorithms that could complete the large-scale iris experiments in three weeks of processing time on a single Intel Pentium 4 3.6GHz 660 processor.”<sup>32</sup> Unfortunately for the participants, this timing requirement was not announced *a priori*. The report noted that the Cambridge algorithm took six hours, and the Sagem-Iridian and Iritech algorithms took approximately 300 hours to complete the ICE 2006 large scale experiments. Participants were not told that execution time would be measured or reported. Had participants been notified that timing would be measured and that minimum execution times were required, they may have been able to optimize their algorithms for speed, potentially allowing performance results for more participants to be included in the final report.

A salient result of the FRVT/ICE 2006 report was that “on the FRVT 2006 and the ICE 2006 datasets, recognition performance was comparable for all three biometrics.”<sup>33</sup> The boxplots shown in Figure 2-15 (Figure 9 in the FRVT/ICE 2006 report)<sup>34</sup> illustrate that at false acceptance rate (FAR)=0.001, false rejection rate (FRR) results for iris, very-high-resolution face, and 3-D face are comparable. The FAR=0.001 operating point does not take into account a key strength of Daugman-based iris recognition technology—relatively flat ROC curves over a wide range of FAR values. Recall the IBG ITIRT finding discussed above where the operating point for Daugman-based algorithms can be shifted to very low false match rates (FMRs) with only a slight increase in the FNMR rates.<sup>35</sup> The ICE 2005 effort reported similar relatively flat ROC curves for the best performing algorithms. The FRVT/ICE 2006 report did not provide ROC curves, however results at FAR=0.0001 were provided in separate boxplots in the appendices. Table 2-2 below presents the median FRR results for the leading iris, very-high resolution face, and 3-D face algorithms shown in Figure 2-15 at FAR=0.001 and at FAR=0.0001. FRR values for three leading high-resolution face algorithms are also presented. The FRR values for iris were taken from Tables IV and V in NIST’s iris recognition meta-analysis report.<sup>28</sup> FRR values for the other algorithms were estimated from the boxplots presented in the FRVT/ICE 2006 report.

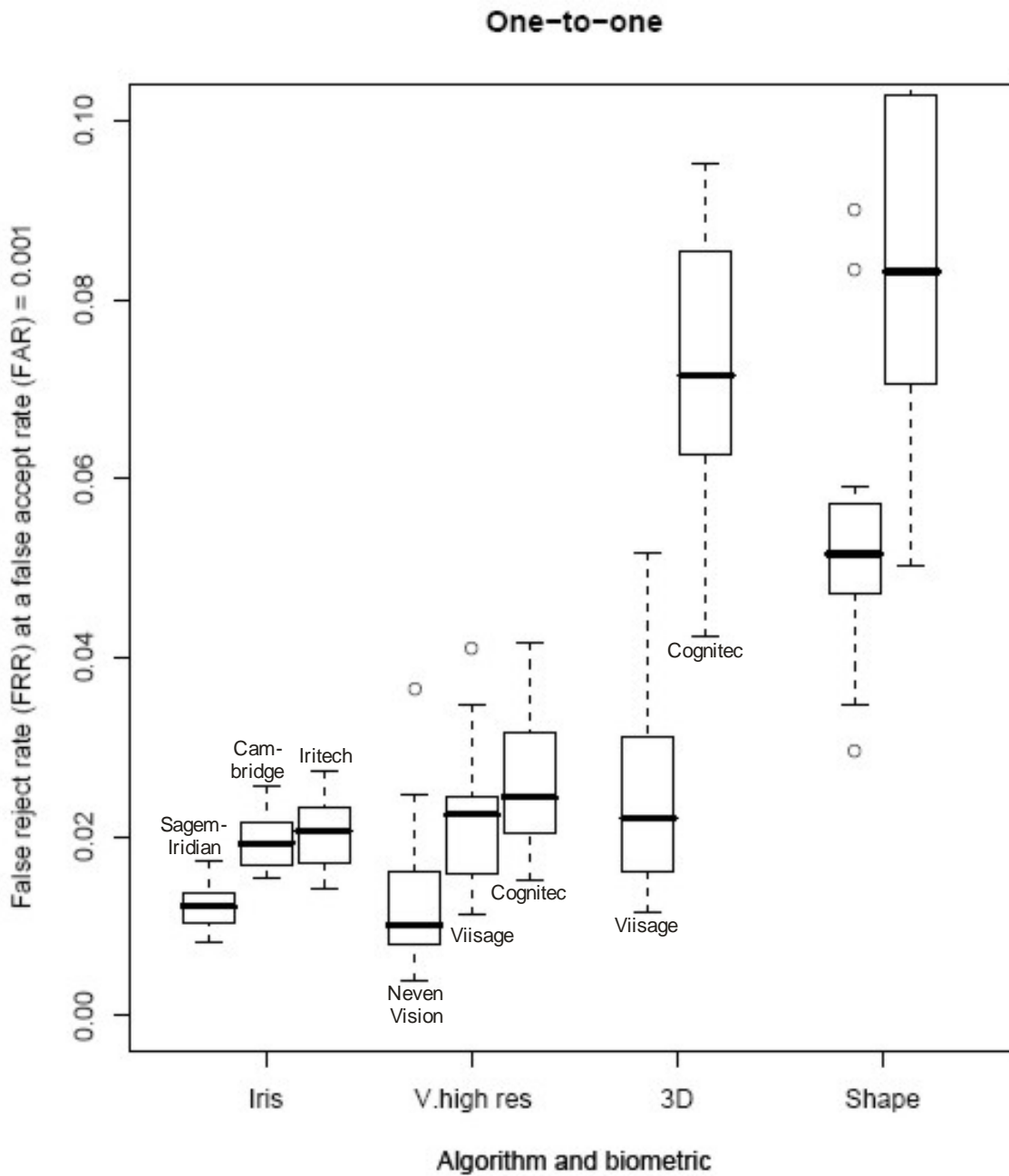
---

<sup>32</sup> P. Jonathon Phillips, *et. al.* “FRVT 2006 and ICE 2006 Large-Scale Results,” NISTIR 7408 (2007) p. 7 <http://iris.nist.gov/ice/FRVT2006andICE2006LargeScaleReport.pdf> (accessed May 30, 2008).

<sup>33</sup> *Ibid.*, p. 1.

<sup>34</sup> *Ibid.*, p. 22 (We added the algorithm names to this plot for the convenience of the reader.)

<sup>35</sup> Note that the FRVT/ICE 2006 study reports false accept rate (FAR) and false reject rate (FRR) though the terms false match rate (FMR) and false non-match rate (FNMR) are more apropos given that FTE and FTA are not taken into account.



**Figure 2-15. FRVT/ICE 2006 Boxplot Results**

**Table 2-2. FRVT/ICE 2006 Results**

Modality/Algorithm	Median FRR (%) @ FAR=0.001	Median FRR (%) @ FAR=0.0001
<b>Iris</b> (image quality poorer than typical operational images)		
Sagem-Iridian	1.22	1.41
Cambridge	1.93	2.46
Iritech	2.06	2.67
<b>Very-high-resolution face</b> (400 pixels between eye centers, image quality substantially better than typical operational images)		
Neven Vision (no longer commercially available)	1.0	1.2
Viisage	2.2	4.5
Cognitec	2.4	5.4
<b>3-D face</b>		
Viisage	2	7
Cognitec	7	16
<b>High-resolution face</b> (350 pixels between eye centers, image quality better than typical operational images)		
Neven Vision (no longer commercially available)	2.4	2.6
Viisage	3.2	4.7
Cognitec	3.5	5.1

We observe in Table 2-2 that at FAR=0.001, the FRR values for iris algorithms and for the top three very-high-resolution face algorithms are comparable. The Viisage 3-D face and the Neven Vision high-resolution face algorithms are also comparable with iris at this FAR value. We note that the very-high-resolution images and the high-resolution images had 400 and 350 pixels between eye centers, respectively. The ISO/IEC 19794-5 data interchange format for face image data in passport applications requires a minimum of 90 pixels between eyes and recommends 120 pixels between eyes. As such, the very-high-resolution and high-resolution images used for the NIST analysis had substantially higher resolution than images that would be used in operational

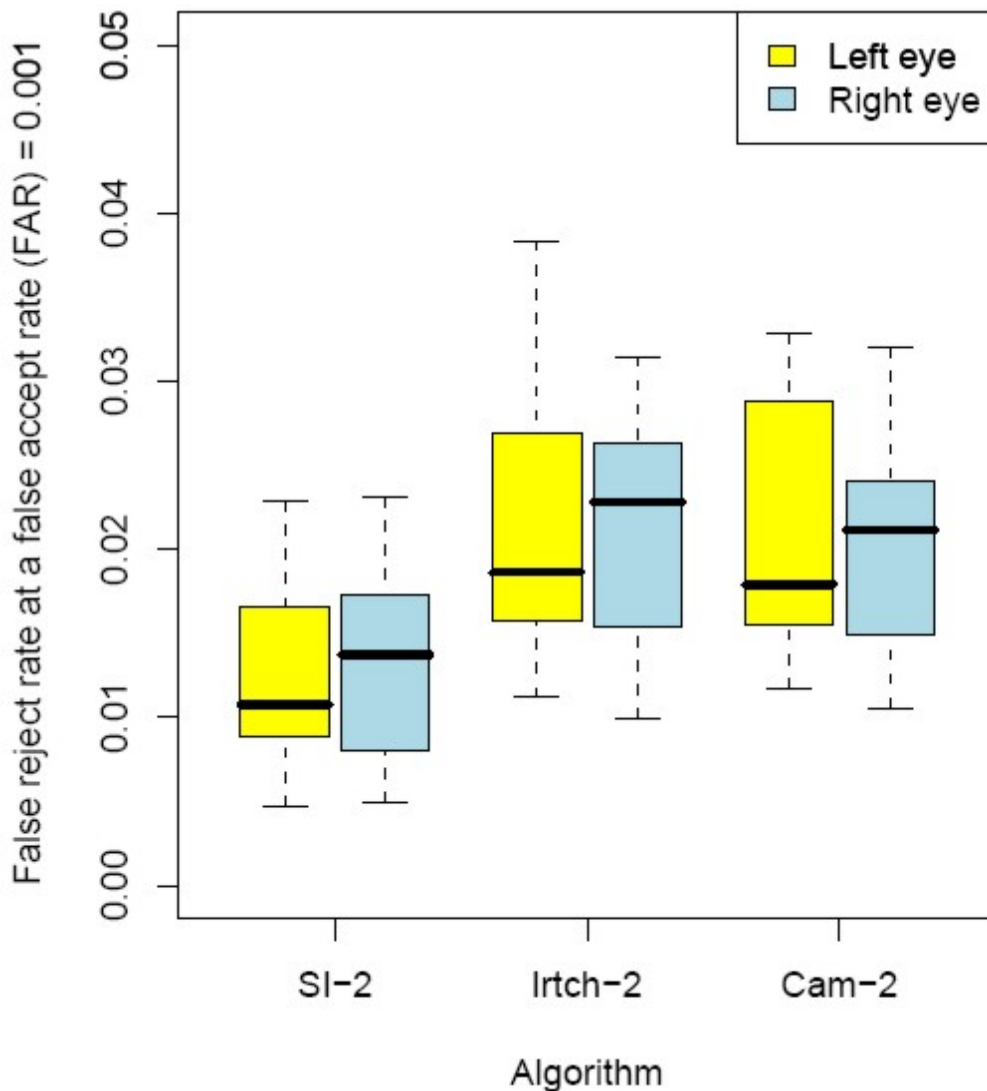
systems, such as passport systems. In contrast, two-thirds of the iris images had lower image quality than images in operational systems. As such, the result that “recognition performance was comparable for all three biometrics” will probably not be observed in operational systems at FAR=0.001.<sup>36</sup> The performance of iris recognition systems using operational iris images will likely outperform (have lower match error rates) facial recognition systems using operational face images. Further studies are required to confirm this hypothesis.

At FAR=0.0001, only the Neven Vision algorithm using very-high-resolution and high-resolution images is comparable in performance to the iris recognition algorithms. All other very-high-resolution and high-resolution face results, and all of the 3-D face results, do not perform comparably with iris at the lower false accept rates where iris recognition algorithms shine. We note that, as above, the facial images being used by the Neven Vision algorithms in the NIST analysis have much higher resolution than face images used in typical operational systems, and that the iris algorithms performed well in spite of the abnormally low image quality of two-thirds of the iris images. We further note that, after being purchased by Google Inc. in August 2006, the Neven Vision algorithms are no longer commercially available for face recognition. We conclude that iris technologies will provide better recognition performance than facial technologies for applications requiring low false accept rates, such as when biometric data is used to search existing databases for a prior criminal history or for alternate identities.

An interesting result of ICE 2006 is illustrated in Figure 2-16, which shows that the right eye performs slightly better than the left eye; the opposite was observed in the ICE 2005 test. Presumably, the data collection protocol for ICE 2006 was modified from that used for ICE 2005 to change the presentation order of left and right eyes; however, the details of data collection protocol have not been publicly released.

---

<sup>36</sup> The NIST ICE test was a “technology” test; results of technology tests do not necessarily reflect the actual performance that will be observed in real operational systems.



**Figure 2-16. ICE 2006 Left and Right Eye Boxplots**

### 2.4.3 Authenti-Corp IRIS06

The standards-based IRIS06 effort was sponsored by the U.S. National Institute of Justice and the U.S. Department of Homeland Security. Data was collected from live human test subjects May through December 2006 in Phoenix, Arizona, using three commercially available iris recognition cameras. Names of the cameras were not released. Multiple sets of ISO/IEC 19794-6 compliant iris images were collected from 295 test subjects using each of the cameras during the first visit, and 264 of the same test subjects when they returned for a second visit two to eight weeks later. The draft final report was released in May 2007 for public review; the final report is dated September 2007.



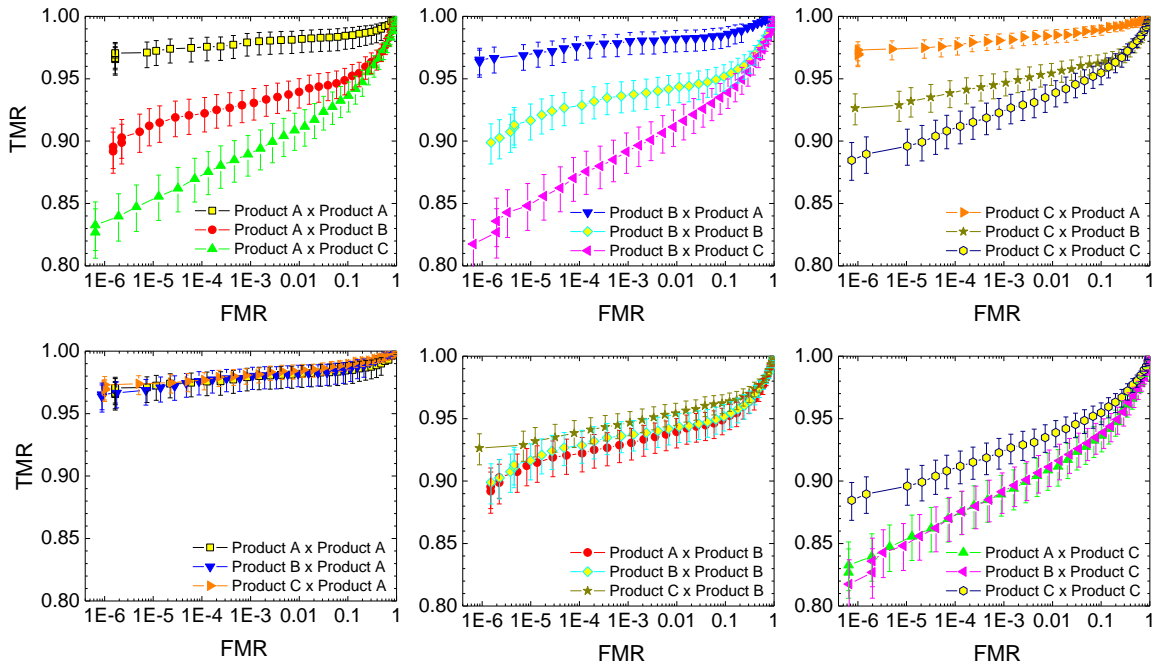
IRIS06 was conducted using the ANSI INCITS 358-2002, BioAPI 1.1 biometric application programming interface and was in accordance with the ISO/IEC 19795 standard for biometric performance testing. Online and offline performance metrics, such as true and false match rates, generalized true and false accept rates, and enrollment and recognition transaction times, along with the associated confidence intervals, were reported for the three products evaluated. Online metrics were obtained using each product's commercial software. Offline metrics were computed using template generation and matching algorithms provided by Professor Daugman. Off-axis gaze and pose experiments were also performed to explore user-cooperation factors.

FTE rates ranged from 0.35 percent to 3.39 percent for the three cameras tested. An FTE was declared if neither the left nor right eye could be enrolled after three attempts with each eye. Mean enrollment transaction times, including FTEs, ranged from 32.2 to 70.1 seconds. FTA rates ranged from 1.5 percent to 6.9 percent, where an FTA was declared if neither left nor right eye could be acquired after three attempts with each eye. The mean recognition transaction times, including FTAs, ranged from 7.9 to 21.4 seconds. These metrics are comparable to those obtained in the ITIRT test with the exception of the FTA rates. The ITIRT FTA rates are substantially lower than those measured during the IRIS06 effort.

The online FNMRs ranged from 0.0 percent to 1.8 percent, and the offline FNMRs (at HD=0.32) ranged from 0.3 percent to 2.7 percent. A false non-match was declared if neither the left nor right eye matched after three attempts with each eye. Recall that the online comparisons were performed with each camera's native commercial algorithm, while the offline comparisons were performed with the algorithm provided by Professor Daugman. While not identical, the online and offline FNMRs are comparable with each other and with the ITIRT offline FNMRs obtained under similar conditions.

We note that NIST performed an all-to-all comparison of the ICE iris images to generate the genuine and impostor score distributions. This approach emulates an attempt-level analysis. The results presented here for ITIRT and IRIS06 apply transactional intelligence to the analyses by allowing multiple (three) attempts and using the lowest resulting score to generate the genuine and impostor distributions. Transaction-level error rates reflect real-world performance when multiple match attempts are permitted, which is typical in operational systems. Transaction-level FNMRs are typically lower than attempt-level FNMRs.

The IRIS06 attempt-level native and interoperability ROC curves are presented in Figure 2-17. In the figure, "Product M x Product N" indicates that Product M enrollment samples and Product N recognition samples are compared. Native curves are indicated by yellow-filled symbols. The first row of curves in the figure shows performance when enrollment samples from one product are compared to recognition samples from all three products. The second row of curves shows performance when recognition samples from one product are compared to enrollment samples from all three products. (This is the same data presented in the first row but organized differently.)



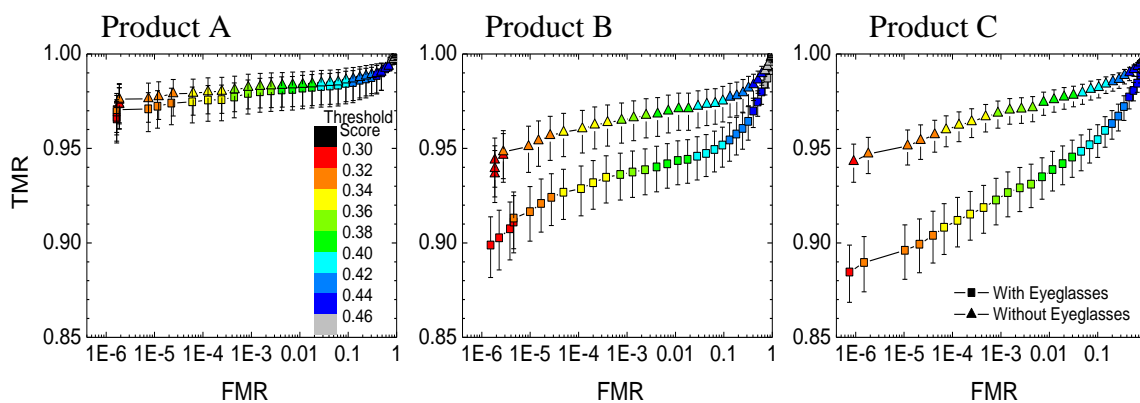
**Figure 2-17. IRIS06 Attempt-Level Native and Interoperability ROC Curves**

As with ITIRT and ICE 2005, we observe that FNMR, specifically the true match rate ( $TMR=1-FNMR$ ), is stable over a wide range of FMR values. We also observe that interoperability matching performance is better than native matching performance in some cases. The best matching performance is obtained using enrollment images from Product C and recognition images from Product A. Recall that in the ITIRT study, all interoperability performance was poorer than native performance.

Additional IRIS06 findings of interest include:

- Eyeglasses degraded matching performance for two of the products tested but not for the third, see Figure 2-18.
- Right and left eyes exhibited statistically similar matching performance.
- Time separation between enrollment and recognition attempts (from 15 minutes to almost eight weeks) did not have a measurable influence on performance, indicating that iris recognition technology is suitable for non-habituated (non-trained) users. (Temporal effects over longer periods are not yet empirically determined.)
- The products tested demonstrated tradeoffs between speed, collection volume, image quality, and match rates. Higher match rates required longer transaction times; faster transaction times resulted in lower match rates. High quality images required longer transaction times; shorter transaction times were obtained with cameras that had larger collection volumes.

- The evaluated products generally performed well with yaw and roll angles of  $\pm 20^\circ$  or more when the test subjects were located at manufacturer-designated distances from the camera. The products also performed better when test subjects gazed upward (with neutral pose) or faced upward (with neutral gaze) relative to the camera rather than downward.



**Figure 2-18. IRIS06 Attempt-Level Performance with Glasses**

#### 2.4.4 NIST IREX 08

The NIST Iris Recognition Exchange 2008 (IREX 08) evaluation is just getting underway. IREX will study two issues: 1) the influence of compression on matching performance; and 2) the interoperability performance of compact iris data formats, such as polar and region-of-interest compression formats. IREX is designed to support identity management applications where compact size and interoperability are important. Additional information about IREX and the current status of the program can be found at <http://iris.nist.gov/irex/>.

#### 2.5 Standards

There are two primary standards that address collection, storage, and exchange of iris data: ISO/IEC 19794-6:2005 Biometric Data Interchange Format—Part 6: Iris image data and ANSI/NIST-ITL 1-2007 (Type 17). ISO/IEC 19794-6:2005 defines a rectilinear format for iris images that can be raw or compressed and two variations of a polar image format (“iris only” and “no [outer] boundary”). The polar image format requires specific preprocessing and segmentation steps, which can be raw or compressed. The iris only polar image format is more compact than the rectilinear format, however, the narrowly defined region of interest excludes information about the eye region that might be of interest to forensic applications. The standard also defines data structures and headers to facilitate interoperability among vendors. The domestic version of this standard, ANSI/INCITS 379-2004 Iris Interchange Format, was withdrawn by ANSI in 2008 in favor of the international version (ISO/IEC 19794-6:2005). The ANSI/NIST-ITL 1-2007, Type 17 format is a strict derivative of ISO/IEC 19794-6:2005. ANSI/NIST-ITL 1-2007 is also specified in

FBI EBTS Version 8.0, which has been expanded over previous versions to include iris and other biometric modalities (e.g., palmprint and face) in recognition of the rapidly developing biometric identification industry.

All iris interchange formats are based on images rather than templates, as iris templates are specific to the algorithms that generate them in the current state of technology. Although the standardized polar formats are more compact than the rectilinear format, minimizing data storage requirements and the bandwidth needed for data transmission, the international biometrics standards body voted to remove all polar formats from the imminent revision of ISO/IEC 19794-6 in January 2008. The current formats were found to be critically sensitive to the consistency of segmentation and subject to sampling problems.<sup>37</sup> Two alternate compact forms are currently being investigated. A method advanced by Professor Daugman involves cropping and compressing portions of a rectilinear image based on regions of interest (ROI). This method is called the ROI-masked version. Another method proposed by Dr. Daehoon Kim of Iritech, termed the unsegmented polar version, involved defining concentric inner and outer circles, neither of which is necessarily centered on the pupil or iris. These alternate compact formats are being explored as part of NIST's IREX effort.

In February 2008, the NSTC Subcommittee on Biometrics & Identity Management published a draft "Registry of USG38 Recommended Biometric Standards" for public comment.<sup>39</sup> This document requires the use of the rectilinear image format of ISO/IEC 19794-6:2005 or ANSI/NIST-ITL 1-2007, Type 17, and limits lossy compression to a 6:1 ratio. It states that irises stored in any of the polar image formats of ISO/IEC 19794-6:2005 may be retained only if their rectilinear image parents are also retained.

## **2.6 Research and Development**

### **2.6.1 Data Acquisition**

Most available commercial iris biometrics systems require some level of cooperation from the user, as image acquisition conditions are constrained. In response to this constraint, most of the research being performed in the area of data acquisition involves making the collection of iris images less intrusive to the user. A complete discussion of approaches to engineer less intrusive image acquisition is beyond the scope of this report, however a thorough review of this work is

---

<sup>37</sup> Patrick Grother, "Iris Exchange (IREX) Evaluation 2008," 2008, p. 8,, [http://iris.nist.gov/irex/IREX08\\_conops\\_API\\_v2.pdf](http://iris.nist.gov/irex/IREX08_conops_API_v2.pdf) (accessed May 30, 2008).

<sup>38</sup> USG-United States Government.

<sup>39</sup> [http://ts.nist.gov/Standards/Biometrics/upload/Biometric\\_Standards\\_Registry\\_02042008.pdf](http://ts.nist.gov/Standards/Biometrics/upload/Biometric_Standards_Registry_02042008.pdf) (accessed May 30, 2008).

provided in Section 4.1 (p. 14-17) of a recent survey of iris biometrics performed by the University of Notre Dame.<sup>40</sup>

### **2.6.2 Iris Segmentation**

Segmentation, which locates the pupillary boundary (between the pupil and iris), and the limbic boundary (between the iris and the sclera) is a challenging and crucial step in the iris recognition process. Accurate segmentation is a prerequisite for good recognition performance. Segmentation is influenced by many factors, including features of the eye, features of the image, the acquisition environment, and the iris camera design. For example, segmentation algorithms can confuse the rims of the eyeglasses with the boundaries of the iris.

Research in this area focuses on approaches to find the pupillary and limbic boundaries and on approaches to locate occlusions, such as eyelids, eyelashes, and strong reflections. A thorough review of segmentation research efforts is provided in Section 5 (p. 21-29) of the University of Notre Dame iris survey.

### **2.6.3 Texture Encoding**

A substantial amount of research has been performed in the area of texture encoding – the process of converting iris images into numbers for comparison. Techniques include alternate approaches to Professor Daugman’s Gabor filter method for producing binary iris code, exploring various filters to represent iris texture with real-valued feature vectors, and combinations of these two approaches. A review of literature in this area is provided in Section 6 (p. 29-41) of the University of Notre Dame iris survey.

### **2.6.4 Iris Comparisons**

A variety of approaches to improve matching performance have been studied. These approaches are presented in Section 7 (p. 41-48) of the University of Notre Dame iris survey.

### **2.6.5 Test Databases**

A variety of iris test databases are available to researchers. The first widely used database was the CASIA V1.0 dataset, which contains 756 NIR iris images from 108 Chinese test subjects. In this dataset, the pupil area of each eye was “replaced with a circular region of constant intensity to mask out specular reflections from the NIR illuminators.”<sup>41</sup> This method, however, makes iris segmentation artificially simple and should not be used for segmentation experiments. The CASIA-IrisV3 dataset is now available<sup>41</sup>, which includes three subsets: CASIA-IrisV3-Interval,

---

<sup>40</sup> Kevin W. Bowyer, Karen Hollingsworth, and Patrick J. Flynn, “Image understanding for iris biometrics: A survey,” *Computer Vision and Image Understanding* **110** (May 2008), p. 281, [http://www.nd.edu/~kwb/BowyerHollingsworthFlynnCVIU\\_2007.pdf](http://www.nd.edu/~kwb/BowyerHollingsworthFlynnCVIU_2007.pdf) (accessed May 30, 2008).

<sup>41</sup> CASIA datasets <http://www.cbsr.ia.ac.cn/english/Iris%20Databases.asp> (accessed 30 May 2008).

CASIA-IrisV3-Lamp, and CASIA-IrisV3-Twins. CASIA-IrisV3 contains a total of 22,051 iris images from more than 700 subjects. All iris images are 8-bit gray-level JPEG files collected under near infrared illumination. Almost all subjects are Chinese.

Additional iris databases are available from Michal Dobes and Libor Machal (Czech Republic),<sup>42</sup> West Virginia University (USA),<sup>43</sup> NIST (collected by University of Notre Dame, USA),<sup>44</sup> Multimedia University (Malaysia),<sup>45</sup> University of Beira Interior (Portugal),<sup>46</sup> and University of Bath (United Kingdom).<sup>47</sup> Iris datasets from the Carnegie Mellon CyLab,<sup>48</sup> and the DoD may also be available.

Several researchers have investigated creating and using synthetic iris image datasets. However, with the availability of large datasets of real human irises, the need for synthetic iris images is diminished. Synthetic iris image research is summarized in Section 8.1.1 (p. 56-57) of the University of Notre Dame iris survey.

Recent studies have sought to characterize iris structures in NIR, and correlate these characteristics to their visible light manifestation. Fundamental research needs to be done here, and collection of iris data for research should include instrumented illumination from a range of IR and visible wavelengths in order to assess similarities and differences.

## **2.7 Human Issues**

All biometric modalities are enhanced by paying attention to human factors of collection and the behavior of the subject relative to the sensor. With iris recognition, these issues are particularly acute as our visual system responds to the environment and exhibits both autonomic and behavioral responses according to our attention, cooperation and feedback, and overall cognition. The next few sections address some of the human issues that are particular to iris recognition.

### **2.7.1 Usability**

As with any biometric system, some users may have difficulty or may not be able to use iris recognition technology. For example, systems mounted at normal eye height may be impossible to use by individuals in wheelchairs. Users might be required to remove hard contact lenses and glasses for enrollment, which can make it difficult to align their eyes with the camera. Eyeglasses

---

<sup>42</sup> UPOL dataset <http://phoenix.inf.upol.cz/iris/>

<sup>43</sup> Email [arun.ross@mail.wvu.edu](mailto:arun.ross@mail.wvu.edu)

<sup>44</sup> ICE 2005 and ICE 2006 datasets, email [ice@nist.gov](mailto:ice@nist.gov)

<sup>45</sup> MMU1 and MMU2 datasets <http://pesona.mmu.edu.my/~ccteo/>

<sup>46</sup> UBIRIS.v1 and UBIRIS.v2 noisy iris image databases <http://iris.di.ubi.pt/>

<sup>47</sup> <http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/database.htm>

<sup>48</sup> <http://www.cylab.cmu.edu/default.aspx?id=302>

and hard contact lenses may need to be removed during recognition as well. This adds a level of inconvenience to subjects who use vision aids.

Other eye conditions, such as blindness, nystagmus (tremor of the eyes), and strabismus (cross-eyed or wall-eyed), also can make it difficult to align the eye with the camera. Cataracts (clouding of the lens) and glaucoma (often associated with an increase of pressure inside the eye that can create spots on the iris) cause the iris pattern to be unstable. An enrolled person may not be identified at a later date because the cataract or glaucoma sufficiently changed their iris pattern. Individuals with aniridia (underdevelopment of the iris) most likely cannot be enrolled in iris recognition systems.

### **2.7.2 Safety**

The use of NIR light, which is invisible to most humans, can help individuals feel more comfortable with iris recognition products. However to ensure eye safety, iris recognition products must adhere to the illumination safety standards ANSI/IESNA RP-27.1-96 and IEC 60825-1 Amend.2, Class 1 LED, the latest standards in the NIR illumination safety.

As research initiatives have sought to increase capture distance and relax constraints on iris acquisition systems, there is a need to modify (and a desire to increase) the NIR illumination levels. Prior safety studies report that the cornea and lens should not be exposed to irradiance of more than  $10 \text{ mW/cm}^2$  for light in the 770 nm to 3000 nm range [Matey, 2006].

Iris recognition is sometimes confused with the biometric modality of retina scanning, where an NIR (non-laser light) beam is rotated over the retinal pattern on the back of the eye. Figure 2-19 shows a typical pattern that results from the reflected signal returned from the retina as the beam rotates over a circle. This somewhat invasive approach can create misconceptions for some iris recognition system users. Current iris recognition systems use NIR LEDs or a filtered flashlamp to illuminate the iris so that an external photograph of the eye can be taken, and like retinal scans, do not shine a laser into a person's eye. "Analysis of illumination levels for the Iridian-based systems shows that, even under worst-case assumptions, these are still significantly lower than the maximum permitted levels of the relevant standards"<sup>49</sup>

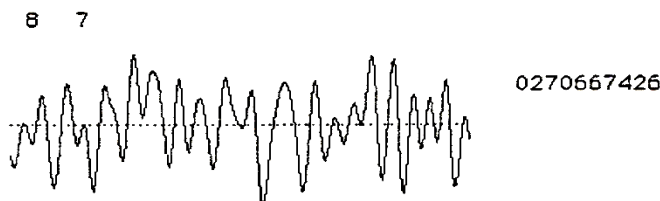
---

<sup>49</sup> BioVision: Roadmap for Biometrics In Europe to 2010, p. 109, <ftp://ftp.cwi.nl/pub/CWIreports/PNA/PNA-E0303.pdf> (accessed May 30 2008).

## Examples of Retinal Scan Waveforms

## Acceptable Quality Images

## Example 1



**Figure 2-19. Typical Retinal Scan (used courtesy of the Harris Corporation)**

### 2.7.3 Privacy

In addition to the “Big Brother” types of privacy concerns that many individuals have with biometrics in general, iris recognition poses additional medical privacy issues for individuals who believe in iridology. “Iridology is an alternative medicine technique whose proponents believe that patterns, colors, and other characteristics of the iris can be examined to determine information about a patient’s systemic health. . . . Iridologists see the eyes as ‘windows’ into the body’s state of health. Iridologists use . . . charts to highlight certain systems and organs in the body as healthy and others as overactive, inflamed, or distressed. Iridologists believe this information may be used to demonstrate a patient’s susceptibility toward certain illnesses, to reflect past medical problems, or to predict health problems which may be developing.”<sup>50</sup>

“Scientific research into iridology has shown mostly, but not entirely, negative results. However, all double blinded, rigorous tests of iridology have failed to find any statistical significance to iridology.”<sup>51</sup> Furthermore, iridologists examine the color information from an iris to determine health information; iris recognition uses NIR images of the iris, which do not reveal color features.

## 2.8 Forensic Capabilities

The ocular region contains a great deal of forensic information for individualizing subjects. In addition to the iris, the shape of the eye, and the length and “style” of eyelashes and eyebrows provide potentially discriminating information.

<sup>50</sup> <http://en.wikipedia.org/wiki/Iridology> (accessed May 30, 2008).

<sup>51</sup> <http://www.healthonhealth.com/Health-Reference/I/Iridology> (accessed May 30, 2008).



Although it is widely believed that the iris is stable over time, there are perplexing examples that contradict this assumption. One example is the story reported in the April 2002 National Geographic magazine concerning Afghan woman Sharbat Gula made famous by National Geographic photographer Steve McCurry (<http://ngm.nationalgeographic.com/ngm/0204/feature0/index.html>). Sharbat's "haunting green eyes" exhibited changes over 17 years. There is no information if there were any changes in the NIR range, although visual analysis shows considerable differences in the red-band of the color image.

Traditionally, eye color is recorded during bookings; however, the forensic community needs a richer vocabulary to reference the appearance of the iris and the ocular region of the face. As with forensic face examination, this area also needs tools and training toward a common vocabulary for human recognition as well as supporting quantified methods of comparison.

## **2.9 Vulnerabilities**

Iris recognition systems are vulnerable to impersonation and concealment. There have been a number of impersonation studies published on the internet using iris patterns printed on paper capable of spoofing systems. It has been shown helpful to remove the pupil area of the printed iris, and then place the printing in front of a real eye, or to cover the pupil area with a clear contact lens.

Some of the best internally documented work in iris impersonation has been by the Computer Electronic Security Group (CESG), a branch of the UK Government Communications Headquarters (GCHQ). Contact lenses embedded with printed patterns are commonly available for cosmetic applications. These lenses are intended to change the pattern or color of the iris. These lenses are well documented to conceal the true texture of the iris, leading to false non-matches. To determine if such lenses could be used for impersonation, CESG enrolled the right eye of a scientist covered with such a lens. To impersonate the scientist, another person would have to use the lens with the exact same radial orientation. Because of the health and safety issues involved, CESG performed a proxy experiment, using the same scientist's left eye as a proxy for another individual. After correct radial orientation of the lens, the system recognized the left iris with the lens as the right iris. Because radially-stable contact lenses are available for people with astigmatism, this experiment appears to show that iris recognition impersonation is possible using printed lenses.

CESG also has conducted a limited amount of work on concealment through pupil dilation by commonly used ophthalmology drugs Tropicamide and Phenylephrine. One of the two systems studied could recognize the dilated iris as the enrolled normal iris if the eyelid was held wide open. The second system could not recognize the dilated iris under any circumstances. Professor Daugman has claimed that algorithms have been developed to determine the extent of dilation of the iris, and to provide a warning if dilation is severe.

## 2.10 Future Capabilities

Current research efforts for iris recognition are being directed in a number of areas:

The planned NIST “Multi-Biometric Grand Challenge” is proceeding with still and video imagery of faces and irises. The iris images will be low and high resolution, and will contain “off angle” images. Algorithm developers will tune their algorithms to deal with such imagery, and combine face and iris images for better recognition accuracy. This effort is being funded, in part, by the FBI.

Increasing the distance at which the iris can be imaged has been an important issue since before the DARPA “Human ID at a Distance” (HID) program from 2000-2003. Imaging the iris at a distance involves finding a face, locating the eyes, and then focusing an iris capture system on the eyes. In the late 1990s, the Sensor Corporation (which was owned by Sarnoff Labs) built multi-camera units with capability of imaging the iris at a distance of about a meter. More recent “Iris on the Move” units built by Sarnoff, based on technologies developed under the DARPA HID program, can extend that distance to about three meters. Other companies working in this direction include AOptix and Honeywell.

Just as high resolution, color images taken under visible light conditions were available and important in the National Geographic study of Sharbat Gula, the “Afghan girl,” such imagery could have important forensic applications if automated comparison techniques were well understood. At least one group has been looking at automated recognition of color iris images collected under visible lighting conditions.

The optimal NIR illumination wavelengths and bandwidths for iris recognition are not well understood. It appears that the best wavelength may be dependent on the eye color being imaged. In 2004, NIST discovered these issues in attempting to develop their own iris collection system for the MBARK (Multimodal Biometric Acquisition Research Kiosk, since renamed Multimodal Biometric Application Research Kit). Several groups are proposing work in this area.

The stability of the iris over time is not well understood. As discussed in this article, there is anecdotal evidence that significant changes can occur over time, but there are no vertical databases upon which to test this. Research groups have proposed such efforts.

## 2.11 Technology Gaps and Challenges

FBI is currently not collecting iris data. While iris recognition technology is maturing rapidly and there are use cases that make sense for FBI and law enforcement community to consider, there is currently neither a strong interest nor a clear adoption path within major AFIS collection platforms. Some of the major challenges and related recommendations are presented below.

Before future NGI integration of iris technology, the FBI should explore the use of iris recognition within smaller, controlled pilot programs. Examples of possible uses include:

- Training programs to familiarize examination and analysis community

- Prisoner registration and visitor identification
- Registered sex offenders and probation cases
- Mobile ID and counter gang policing.

In support of science and technology, the following recommendations speak to Daubert related issues:

- Recommend that the FBI begin a multi-year, multi-spectral data collection effort on a small number of long-term (10 year) volunteers (~100) to determine the stability of the iris pattern at different wavelengths over time.
- Recommend that the FBI invest in research on iris recognition from both low and high resolution visible-wavelength color imagery obtained through common photographic methods.
- Recommend that the FBI begin a program for developing, documenting, and testing methodologies for human-aided recognition of irises that will lead to Daubert-admissible testimony to support the results of automated iris comparison systems and high resolution photography.
- Recommend intra-governmental cooperation (e.g., with DHS S&T and IARPA) in developing and testing iris recognition systems capable of operating at distances of many meters with walking data subjects. The same technology that enables robust collection also will improve acquisition time and usability for semi-cooperative subjects in controlled application environments.

## 2.12 Vendors for Iris Recognition Hardware

Table 2-3. Industry Vendors for Iris Recognition Hardware

Company	Product/Model	Attributes	Comments and Websites (viewed March, 2009)
<b>Access Control Devices</b>			
LG Iris	IrisAccess 4000 (iCAM4000/4010, iCAM4100/4110)	Dual-eye, visual user interface with audio prompts, motorized height adjustment, integrated face camera with built-in illuminators	Third generation product, captures images of both eyes (nearly) simultaneously, optional embedded SmartCard reader, automatically adjusts camera height when smartcard presented, iCAM4100 incorporates 16-element keypad, flush or recess mounted <a href="http://www.lgiris.com/ps/products/irisaccess4000.htm">http://www.lgiris.com/ps/products/irisaccess4000.htm</a>
OKI	IRISSPASS-M	Dual-eye, visual user interface with audio prompts, large collection volume	Compliant with BioAPI (ANSI INCITS 358-2002) and ISO/IEC 19794-6, integrated software (Iridian PrivateID® v2.3) <a href="http://www.oki.com/jp/FSC/iris/en/">http://www.oki.com/jp/FSC/iris/en/</a>
Panasonic	BM-ET330	Dual eye, manual tilt camera head, can operate in identification or verification mode, voice prompts and visual cues to align users eyes in a mirror, integrated video camera can be used to capture facial images of users	Administration software (BM-ES330) uses Iridian Private ID and KnoWho technology, purchase user licenses separately, can be integrated with smart card reader <a href="http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?displayTab=O&amp;storeId=11201&amp;catalogId=13051&amp;itemId=88595&amp;catGroupId=21552&amp;surfModel=BM-ET330">http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?displayTab=O&amp;storeId=11201&amp;catalogId=13051&amp;itemId=88595&amp;catGroupId=21552&amp;surfModel=BM-ET330</a>
Panasonic	BM-ET200	Dual eye, manual tilt camera head, voice prompts and visual cues to align users eyes in dual mirror, one mirror for each eye	Administration software (BM-ES200) uses Iridian Private ID and KnoWho technology, purchase user licenses separately, can be integrated with smart card reader <a href="http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?storeId=11201&amp;catalogId=13051&amp;itemId=111038&amp;catGroupId=14468&amp;surfModel=BM-ET200&amp;displayTab=O">http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?storeId=11201&amp;catalogId=13051&amp;itemId=111038&amp;catGroupId=14468&amp;surfModel=BM-ET200&amp;displayTab=O</a>
Iritech	Neoris 2000	Captures both irises and face at the same time	Uses Iritech-patented iris recognition algorithms, BioAPI compliant, long focal length <a href="http://iritech.com/product_1-4.htm">http://iritech.com/product_1-4.htm</a>

Iritech	IrisCAMM	Auto focus	Uses Iritech-patented iris recognition algorithms, designed for mobile iris recognition devices, optional USB 2 interface, available June 2008 <a href="http://iritech.com/product_1-4.htm">http://iritech.com/product_1-4.htm</a>
Jiris	JNC-1000	Single eye, designed for identification applications	In development <a href="http://www.jiristech.com/english/products/iris.php">http://www.jiristech.com/english/products/iris.php</a>
Rehoboth Tech	Irikon Access Control	Automatically detects iris position as user approaches and moves up and down to correctly acquire iris image, auto focus	1:N authentication, 60 cm iris capture range Patent-pending optical technology and recognition algorithms, full system embedded in one chip, no need to connect to external computing device <a href="http://www.rehobothtech.com/sub2_5.asp">http://www.rehobothtech.com/sub2_5.asp</a>
Beijing eyesight Information Technology	Iris Recognition Access Control System	Iris registration and recognition	Camera has USB interface, separate processor unit with RJ45 interface, automatic acquisition Company-patented core iris recognition technology, products certified by Chinese Ministry of Public Security <a href="http://translate.google.com/translate?hl=en&amp;sl=zh-CN&amp;u=http://www.hongmoshibie.cn/&amp;sa=X&amp;oi=translate&amp;resnum=1&amp;ct=result&amp;prev=/search%3Fq%3Dhongmoshibie%26hl%3Den%26sa%3DG">http://translate.google.com/translate?hl=en&amp;sl=zh-CN&amp;u=http://www.hongmoshibie.cn/&amp;sa=X&amp;oi=translate&amp;resnum=1&amp;ct=result&amp;prev=/search%3Fq%3Dhongmoshibie%26hl%3Den%26sa%3DG</a>
Alpha Engineering	Veri-Iris	Iris enrollment and identification	Uses algorithm developed by Dr. Noh (Yonsei University, Seoul), also used for time and attendance, enrollment station and USB remote optical unit, video-based images captured at 19-21 inch distance, provides a network video camera for video paging and conferencing <a href="http://www.ntq.com.vn/Biometric/veri_iris.htm">http://www.ntq.com.vn/Biometric/veri_iris.htm</a>
<b>Single-Eye Handheld Devices</b>			
L-1 Identity Solutions (Securimetrics)	PIER 2.3,	Portable, provides enrollment and identification in the field, operator aligned	Uses Daugman $2\pi$ algorithm, incorporated into the U.S. Army's Biometric Application Toolset (BAT) <a href="http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=38&amp;Itemid=174">http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=38&amp;Itemid=174</a>
L-1 Identity Solutions (Securimetrics)	PIER-T	Tethered, connects to a host PC or laptop, similar to the PIER 2.3	Similar to the PIER 2.3, lower-cost <a href="http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=33&amp;Itemid=335">http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=33&amp;Itemid=335</a>

L-1 Identity Solutions (Securimetrics)	HIIDE Series 4	Portable, multi-modal, provides enrollment and identification using iris, finger, and face when connected to a host PC or when operating in the field un-tethered, operator aligned	Uses Daugman $2\pi$ algorithm, unit also features a single fingerprint sensor and a camera for collecting facial images <a href="http://www.11id.com/index.php?option=com_content&amp;task=view&amp;id=34&amp;Itemid=170">http://www.11id.com/index.php?option=com_content&amp;task=view&amp;id=34&amp;Itemid=170</a>
IrisGuard	IG-H100	Hand held but can also be tripod, wall, or desk mounted, visual and auditory alignment cues, operator aligned	Used to prevent United Arab Emirates expellees from re-entering country, uses Iridian Private ID software, USB 2 interface to PC <a href="http://www.irisguard.com/pages.php?menu_id=29&amp;local_type=0">http://www.irisguard.com/pages.php?menu_id=29&amp;local_type=0</a>
Jiris	JHC-1000	Hand held unit for PC and terminal verification applications	In development, small webcam style with hand-grip placement <a href="http://www.jiristech.com/english/products/iris.php">http://www.jiristech.com/english/products/iris.php</a>
Jiris	JMC-1000	Hand held unit developed specifically for cellular phone and other portable mobile devices	In development <a href="http://www.jiristech.com/english/products/iris.php">http://www.jiristech.com/english/products/iris.php</a>
xVista		Hand held iris scanning and identification system designed for portable computing devices	Developed through a £1.8 million, six-year partnership with the University of Sussex, iris data is captured and processed with a unique algorithm, designed specifically to operate on low-power computing devices such as a camera-equipped mobile phones, scanned irises registered in central database, 256 Mb mobile phone memory card can hold over 250,000 separate iris templates <a href="http://www.xvista.co.uk/html/our-solution/">http://www.xvista.co.uk/html/our-solution/</a>

Beijing eyesight Information Technology	Handheld iris collector	Compact USB device	Handheld unit uses USB 2.0 interface, separate controller uses USB 2.0 or RJ45 interface, auto/manual acquisition Company-patented core iris recognition technology, products certified by Chinese Ministry of Public Security <a href="http://translate.google.com/translate?hl=en&amp;sl=zh-CN&amp;u=http://www.hongmoshibie.cn/&amp;sa=X&amp;oi=translate&amp;resnum=1&amp;ct=result&amp;prev=/search%3Fq%3Dhongmoshibie%26hl%3Den%26sa%3DG">http://translate.google.com/translate?hl=en&amp;sl=zh-CN&amp;u=http://www.hongmoshibie.cn/&amp;sa=X&amp;oi=translate&amp;resnum=1&amp;ct=result&amp;prev=/search%3Fq%3Dhongmoshibie%26hl%3Den%26sa%3DG</a>
<b>Dual-Eye Visor</b>			
Retica	Mobile-Eyes	See-through dual-iris handheld collection device for identity verification	Mobile device, looks similar to binoculars <a href="http://www.retica.com/site/technology/index.html">http://www.retica.com/site/technology/index.html</a>
Crossmatch	I SCAN 2	Handheld, compact dual iris capture scanner	USB-powered, ANSI INCITS 379-2004 and ISO/IEC 19794-6 compliant, compatible with known iris matching algorithms, included Iris Snap SDK and driver software enable image finding and stabilization, pupil segmentation, image quality assessment and auto capture functionality, integrates into Cross Match Jump Kits, collects images but does not perform matching <a href="http://www.crossmatch.com/I_SCAN_2.html">http://www.crossmatch.com/I_SCAN_2.html</a>
<b>Stand Off</b>			
Aoptix	Iris At A Distance	Uses real-time imaging corrections via adaptive optics technology to collect high quality iris images	Curvature adaptive optics corrects for subject motion thus minimizing motion blur and providing in-focus images, real-time tracking and steering enables use of narrow-field, high-magnification objective lenses resulting in high spatial resolution images, active tracking finds subjects in field of view, multiple images of the subject's eye are captured in one second or less, can send "best" image to matching algorithm, claims to be matching algorithm agnostic, capture volume approximately 1 cubic meter at 2 meter distance. Subject must gaze at target for about 1 second but no other participation required, estimated throughput is 20 subjects per minute, integrates facial imaging, production to start mid-2008, <a href="http://www.aoptix.com/pdf/BreakthroughBio.pdf">http://www.aoptix.com/pdf/BreakthroughBio.pdf</a>

Sarnoff	Iris on the Move (IOM)	Identifies users as they walk through a portal at a normal walking pace	User does not need to stop and look at camera, can identify up to 20 subjects per minute, system typically comprises four high quality cameras and infra red lighting, images are captured at a rate of 15 frames/second, proprietary algorithms detect the iris and perform a match on each image, system operates optimally at distances up to three meters, enrollment images typically acquired with a high-image-quality camera system, IOM also available in over-the-door and drive-through configurations <a href="http://www.sarnoff.com/products/iris-on-the-move">http://www.sarnoff.com/products/iris-on-the-move</a>
Retica	Eagle-Eyes	Designed to track and dual-iris identify multiple moving subjects several meters away	Under development, designed to scan a crowd and store iris data for many people at once, designed to use high-quality video cameras and software to capture and check iris images from people as they move from distances of up to 20 meters away <a href="http://www.retica.com/site/technology/index.html">http://www.retica.com/site/technology/index.html</a> <a href="http://masshightech.bizjournals.com/masshightech/stories/2007/08/06/story3.html">http://masshightech.bizjournals.com/masshightech/stories/2007/08/06/story3.html</a>
Hoyos Group, Global Rainmaker	Hbox	Acquires iris and face in real-time as a person moves towards a door at a distance of about 5 feet for identification or verification	Proprietary software suite for the acquisition and matching at a distance and in motion of multimodal biometrics (SAMBI) includes face acquisition algorithm, facial matching algorithm, iris acquisition algorithm, iris matching algorithm; installs over door or free standing; estimated throughput up to 30 people per minute; LCD flat panel monitor on unit displays customizable information; demo unit installed at Unisys in Reston, VA, <a href="http://www.hoyosgroup.com/">http://www.hoyosgroup.com/</a> <a href="http://www.hoyosgroup.com/loadfolder/hbox/hbox_presentation.pdf">http://www.hoyosgroup.com/loadfolder/hbox/hbox_presentation.pdf</a>
<b>Miscellaneous</b>			
Panasonic	BM-ET100US Authenticam	For PC access applications	Discontinued but still available from distributors, bundled with Iridian Private ID software for use with stand-alone PC <a href="http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?displayTab=O&amp;storeId=11201&amp;catalogId=13051&amp;itemId=63725&amp;catGroupId=16817&amp;surfModel=BM-ET100US">http://catalog2.panasonic.com/webapp/wcs/stores/servlet/ModelDetail?displayTab=O&amp;storeId=11201&amp;catalogId=13051&amp;itemId=63725&amp;catGroupId=16817&amp;surfModel=BM-ET100US</a>




JIRIS	JPC Series	For PC and terminal verification applications	Small webcam style models with different acquisition distances (8 cm for JPC-1000, 30 cm for JPC-1500), USB interface <a href="http://www.jiristech.com/english/products/iris.php">http://www.jiristech.com/english/products/iris.php</a>
Rehoboth Tech	Irikon system security module	For PC access applications	Appears to be under development, 6 cm iris capture range, can register up to 200 irises Patent-pending optical technology and recognition algorithms, full system embedded in one chip, no need to connect to external computing device <a href="http://www.rehobothtech.com/sub2_1.asp">http://www.rehobothtech.com/sub2_1.asp</a>

Two additional companies, Senex Technology (<http://www.epsys.no/senex/senex.htm>) and Evermedia (<http://www.stsevermedia.com>), have provided commercial iris recognition products in the past but do not appear to be commercially active at this time (no response to inquiries regarding their products). These products employ an alternative algorithm developed by Dr. Shinyoung Lim, which is described in Table 2-4 below.

**Table 2-4. Industry Vendors for Iris Recognition Software**

Company	Product	Attributes	Comments and Websites
L-1 Identity Solutions (formerly Iridian, Iriscan, Sensor)	SIRIS SDK	Enrollment and matching platform	Utilizes the latest Daugman $2\pi$ 2007 algorithm, developed by Professor John Daugman (University of Cambridge, UK). The $2\pi$ algorithms use a Gabor Transform approach to extract features. <a href="http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=71&amp;Itemid=188">http://www.l1id.com/index.php?option=com_content&amp;task=view&amp;id=71&amp;Itemid=188</a>
Iritech	IrisSDK	Development kit designed for large-scale system integration	Uses Iritech's patented iris recognition algorithms, variable multi-sector analytic method that selectively utilizes only the good portions of the captured image, developed by Dr. Daniel Kim (Korea), SDK normally supplied to IriTech's strategic partners, operates in verification and identification modes, includes iris image quality assessment <a href="http://iritech.com/product_1-3.htm">http://iritech.com/product_1-3.htm</a>
Smart Sensors	MIRLIN software library	Feature extraction, diagnostics and matching functions	Uses discrete cosine transform (DCT) approach to extract features, developed by Professor Don Monro at University of Bath (UK), joint venture with University of Bath <a href="http://www.smartsensors.co.uk/solutions.htm">http://www.smartsensors.co.uk/solutions.htm</a>

Company	Product	Attributes	Comments and Websites
JIRIS	JIRIS-SDK V2.0	Application programming interface to support JIRIS cameras, C++ libraries for image capture, template extraction and matching functions	Uses patented shape-based pattern and curvature scale-space filtering approaches to extract features, developed by Woong-Tuk Yoo (Korea), extracts features only from three tracks (concentric circles) of the iris closest to the pupil (rather than the 8 tracks between the pupil and sclera used by the Daugman $2\pi$ algorithm) and thus only requires a declaration of the boundary between the pupil and iris, declaration of the iris-sclera boundary is not needed, International Patent Publication Number WO/2005/024708 <a href="http://www.jiritech.com/english/products/sdk.php">http://www.jiritech.com/english/products/sdk.php</a>
Qriteck	SDK for IRIBio Mouse	SDK and API documentation and sample source code in VC++ and VB	Patented algorithm designed to support 1:1 and 1:N authentication with less than 40% of whole iris image (to take epicanthal fold and downward pointing eyelashes that cover much of upper half of iris, common in Asians, into account), must execute NDA before ordering <a href="http://www.qritekna.com/biometric/developer.html">http://www.qritekna.com/biometric/developer.html</a>
Retica	Iris recognition software	Claims to be capable of analyzing images captured using all commercial iris cameras	Patented algorithms for single-eye and Dual-eye verification and identification, algorithms can be licensed for incorporation into existing biometric systems, also have patented retina matching and fusion algorithm <a href="http://www.retica.com/site/technology/index.html">http://www.retica.com/site/technology/index.html</a>
Carnegie Mellon University (CMU)	Research, beta product		Matching algorithms employ correlation filters and perform matching in Fourier space.
Dr. Shinyoung Lim, Korea			Uses the Daubechies Wavelet Transform to extract features, developed by Dr. Shinyoung Lim, requires less storage (81 Bytes) than Duagman's $2\pi$ algorithm (256 Bytes), does not appear to be commercially available at this time. <a href="http://www.stsevermedia.com/id8.html">http://www.stsevermedia.com/id8.html</a> <a href="http://www.cise.ufl.edu/~lim/cv/CV_Shinyoung_Lim_Full.pdf">http://www.cise.ufl.edu/~lim/cv/CV_Shinyoung_Lim_Full.pdf</a> <a href="http://www.springerlink.com/content/kyb1wuctdtxqmwbfl/">http://www.springerlink.com/content/kyb1wuctdtxqmwbfl/</a>
Dr. Seung-			Uses multi-resolution independent component analysis

Company	Product	Attributes	Comments and Websites
In Noh, Yonsei University, Seoul			(ICA) approach to extract features, results in smaller iris code size compared to Gabor wavelet (Daugman $2\pi$ ) approach <a href="http://cherup.yonsei.ac.kr/">http://cherup.yonsei.ac.kr/</a> <a href="http://64.233.179.104/scholar?hl=en&amp;lr=&amp;q=cache:hREXegP_2-QJ:www.kmutt.ac.th/itc2002/CD/pdf/19_07_45/FA1_OC/4.pdf+2002+International+Technical+Conference+on+Circuits/Systems+noh">http://64.233.179.104/scholar?hl=en&amp;lr=&amp;q=cache:hREXegP_2-QJ:www.kmutt.ac.th/itc2002/CD/pdf/19_07_45/FA1_OC/4.pdf+2002+International+Technical+Conference+on+Circuits/Systems+noh</a> <a href="http://ietisy.oxfordjournals.org/cgi/content/abstract/E88-D/11/2573">http://ietisy.oxfordjournals.org/cgi/content/abstract/E88-D/11/2573</a>
LG Iris	iData SDK	Basic development tool for LG IrisAccess 4000 series product	Presumed to use Iridian version of Daugman $2\pi$ algorithm <a href="http://www.lgiris.com/ps/products/idata/idatasdk.htm">http://www.lgiris.com/ps/products/idata/idatasdk.htm</a>
LG Iris	iData Eclipse	Development tool for applications requiring non-segmented polar images	Presumed to use Iridian version of Daugman $2\pi$ algorithm, claims ISO-compliant cross platform interoperability <a href="http://www.lgiris.com/ps/products/idata/idaaeclipse.htm">http://www.lgiris.com/ps/products/idata/idaaeclipse.htm</a>
Oki	Prototype	Iris recognition middleware for mobile phones and mobile communication devices with built-in cameras	Operates with a cell phone's existing camera in the visible spectral region, requires about 200KB of phone's memory and about another 200KB when in operation, requires camera ability of at least 1-megapixel, not clear what type of algorithm is used, apparently based on an original iris recognition algorithm developed by Oki  <a href="http://www.oki.com/en/m_iris/">http://www.oki.com/en/m_iris/</a> <a href="http://www.oki.com/en/m_iris/product/">http://www.oki.com/en/m_iris/product/</a> <a href="http://www.reghardware.co.uk/2007/07/25/oki_iris_scan/">http://www.reghardware.co.uk/2007/07/25/oki_iris_scan/</a> <a href="http://techon.nikkeibp.co.jp/english/NEWS_EN/20061128/124572/">http://techon.nikkeibp.co.jp/english/NEWS_EN/20061128/124572/</a>

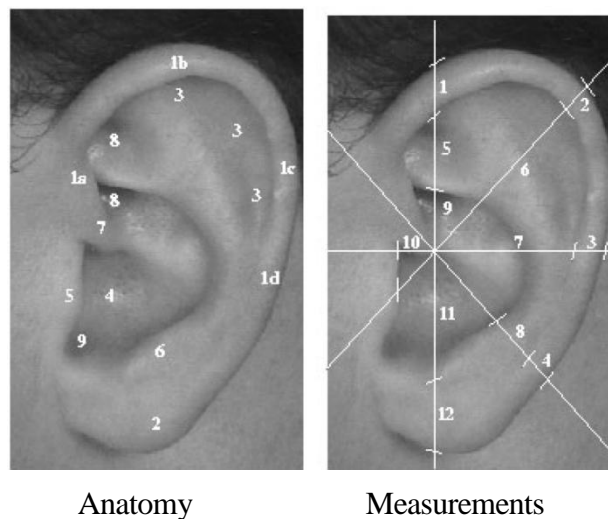
This page intentionally left blank.

# 3 Ear Recognition

## 3.1 Background

The basis for ear recognition is the assumption that the shape and details of the ear are distinguishing and stable. Ears can be obscured from view by hair, hats, and off-angle viewing. As with other visible biometrics, image detail varies according to resolution, viewing angle, and illumination conditions. Ears provide a potential benefit in that they are a mostly rigid structure and not susceptible to elastic deformations (i.e., facial expression). The intuitive understanding is that the ear does enlarge slowly with age, but the overall shape and proportions remain stable over time.

A manual classification system was developed by Alfred Iannerelli in the late 1950s and early 1960s. Iannerelli published his system in 1964 and a revised edition in 1989 [Iannerelli, 1989]. His system consists of dividing a photograph of the ear into 45 degree segments and establishing geometric ‘earmarks.’ The anthropometric measurements are represented in Figure 3-1. Anatomy in the left frame illustrates: 1a-1d) the helix rim; 2) lobule; 3) antihelix; 4) concha; 5) tragus; 6) antitragus; 7) crus of the helix; 8) triangular fossa; and 9) incisures intertragica (or intertragic notch). Measurements are shown in the right frame.



**Figure 3-1. Iannerelli’s Ear Anatomy and Measurements [Burge, et al., 1998]**

The European Commission established a research program entitled “Forensic Ear ID (FEARID)” that ran from February 2002 to May 2005. The FEARID program involved nine academic and police partner organizations that researched and reported on the scientific and technical basis for ear recognition from ear prints, latent impressions of ears left after contact [The European Commission, 2005].

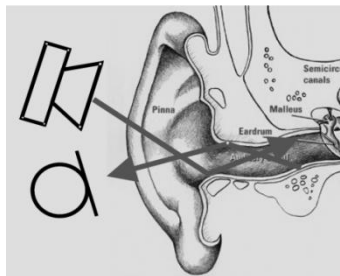
Based on assumed individuality of ears, the use of latent ear prints for evidence has been considered by forensic research [Meijerman, 2006]. Advocates assert that latent ear prints are conceptually similar to latent fingerprints. They provide residual prints that can be attributed to individual burglars or eavesdroppers who pressed their ear against the surface of a door or window.

Ear prints as forensic evidence have been used in convictions; their use also has been debated, challenged, and, in several cases, overturned. In “Another Ear Print Conviction Reversed,” Law professor Andre Moenssens notes several cases where ear prints were involved in convictions, and later questioned and reversed [Moenssens]. Moenssens’ summary of the expert testimony in the 1998 Dallagher case is quoted as follows:

*According to the appeals court decision, “Dr. Champod’s conclusions seems to be that at the present time ear print comparison can help to narrow the field, and may eliminate, but cannot alone be regarded as a safe basis on which to identify a particular individual as being the person who left one or more prints at the scene of a crime. He points out that neither the Forensic Science Service in the United Kingdom nor the Federal Bureau of Investigation in the United States carry out ear print comparisons.”*

Professor Van Koppen’s testimony ran along similar lines. His report concluded, *“The validity of ear identification is unknown. The research that is necessary to say anything on the validity of ear identification has not been conducted. On top of that, the method used by Van der Lugt and Vanezis is subjective to an extent that they are unable to explain how they came to their judgment that there is a match between the ear mark found at the crime scene and the ear print from the suspect.”*

The acoustic properties and features of the inner ear have been examined for containing potentially distinguishing characteristics. Philips Research has published on the topic [Akkermans, 2005], and Sandia Corporation was awarded a patent in July 1998 [Bouchard, 1998]. The publication from Philips Research investigates how the acoustic properties of the pinna (outer ear flap) and ear’s auditory canal can be measured with low cost microphones embedded in applications that use headphones, cell phones, and ear pieces.



**Figure 3-2. Acoustic Waveform Probe with Receiving Microphone [Philips Research, 2005]**

The Philips researchers assert the restriction to low cost speakers and microphones can easily generate and measure sounds from 100Hz up to 15KHz. If they can resolve features on the order of 1/10 of a wavelength, this equates to a spatial resolution on the order of 2mm.



**Figure 3-3. Microphones for Ear Acoustics [Philips Research, 2005]**

### 3.2 Accuracy

Several research databases of modest size exist and a variety of automated recognition techniques have been proposed and researched. An update to Hurley’s summary of reported ear biometric performance [Hurley, 2007] is presented in Table 3-1. The summary indicates the largest ear image dataset is quite small, slightly over 300 subjects, and that automated performance has been reported on these datasets with accuracies in the range of 85 percent to 97.8 percent.

**Table 3-1. Reported Ear Recognition Performances [Hurley, 2007]**

Researcher	2-D/3-D	Technique	Performance	Dataset
Moreno	2-D	Neural Net	93%	168
Hurley	2-D	Force Field	92.2%	252
Mu	2-D	Geometric	85%	308
Yan	3-D	ICP	97.7%	302
Chen	3-D	ICP	90.4%	104

The accuracy of Philips Research developed ear acoustic matching varies according to the frequency of the probe waves and the receiver devices used. The placement and orientation of the receiver microphone in the ear affects accuracy. As reported by Philips, the ‘worst case’ equal error rate results (on 17 subjects) with and without Fisher Linear Discriminate Analysis (LDA) are shown in Table 3-2.

**Table 3-2. Reported Ear Acoustic Equal Error Rates [Philips Research, 2005]**

	Headphone	Earphone	Mobile Phone
No LDA	8.0%	8.4%	15%
LDA	1.4%	1.9%	7.2%

### **3.3 Standards**

At this time, there are no activities or proposed activities for ear images, prints, or internal acoustic features within current national or international standards bodies. Facial profile images may contain visible ear information, particularly the higher resolution level 40 and level 50/51 Subject Acquisition Profiles that have been recently updated for facial records (type 10 records).

There was no evidence of consensus in the research on what an appropriate acquisition resolution should be for the ear as a biometric. Some research databases cropped images to 400 by 500 [Chang, 2003] or approximately 200 pixels per inch (ppi), while others appear to be 50 ppi or less.

The lack of standards is consistent with the lack of products and commercial interest. The general belief is there may not be a sufficient commercial market for the technology, even if more mature products existed.

### **3.4 Spoofing and Vulnerability**

As ears are not widely used for identification, there consequently have not been published accounts of people obscuring or altering their ears for the purpose of avoiding identification. However, a person could easily wear hair over their ears or have them altered, if they were motivated to do so.

### **3.5 Privacy**

A considerable assortment of chromosomal disorders and medical conditions are associated with ear shape.<sup>52</sup> Down Syndrome and cleft palate can cause abnormal shaped ears. A possible symptom of DiGeorge Syndrome is low set ears with a notched ear fold. Macrotia (large ears) may result from chromosomal disorders as well as from certain forms of mental retardation. Whether large ears are due to growth disorders or natural variation, cosmetic ear reduction and alteration procedures are offered by some plastic surgeons.

Cauliflower ear, or boxer's ear, is an acquired deformity resulting from blunt force trauma to the cartilage structure. With prolonged damage over time, blood clotting between the ear cartilage and

---

<sup>52</sup> Wrongdiagnosis.com. [http://www.wrongdiagnosis.com/symptoms/ear\\_shapes/causes.htm](http://www.wrongdiagnosis.com/symptoms/ear_shapes/causes.htm).



skin prevents normal blood circulation. The ear shrivels and deforms into the classic cauliflower-like appearance. The condition is observed in impact sports such as wrestling, boxing, rugby, judo, and mixed martial arts. One study [Kordi, 2007] reports over 75 percent of Iranian wrestlers afflicted with cauliflower ears refuse treatment. Kordi presumes this is due to the wrestler's pride and belief that the condition is a badge of honor.



**Figure 3-4. Cauliflower Ear<sup>53</sup>**

### **3.6 Future Capabilities**

The use of ear shape and its positioning on the head is of interest for supporting human examination and forensic identification. Academic studies with small, well-constrained datasets have demonstrated some automated techniques also may be worth pursuing. Future studies may assist in maturing ear recognition as an identification science and as a practical technology. While not exhaustive, example topics may include:

- Combined acquisition and processing of 2-D and 3-D full face, profile, and ear
- Higher resolution ear collections as will be available under SAP 50/51 for face imaging.

### **3.7 Technology Gaps and Challenges**

The following technology gaps stem from the CJIS need for forensic quality data in light of Daubert criteria.

---

<sup>53</sup> Image Source: University of Wisconsin, Athletic Injury Digital Image Library (AIDIL), <http://www.uwec.edu/kin/majors/AT/aidil/images/cauliflowerear.JPG>.

- There is a need for order-of-magnitude larger datasets than current academic collections; this data will help establish accuracy rates and target subject acquisition profiles with known resolution and automated segmentation tools.
- Human examiners require common training material and a common taxonomy with quantitative methods to describe and compare ears; that is, a supporting science for describing individuality of ears (and faces) [Spaun, 2007].
- There is a need for common training material for human examiners on how to objectively compare faces, ears, and other facial landmarks as part of facial examination process.

### **3.8 Recommendations**

The understanding and use of ears for recognition is considered a reasonable means for augmenting face recognition (in situations where the ear is visible – no additional collection requirements). Some challenges and possible activities to address them are presented below.

- Recommend that the FBI start a data collection effort for a diversity of ear prints and ear images, the latter over multiple angles and illumination conditions to support research into distinctiveness and stability.
- Recommend that the FBI begin a research effort into describing and quantifying individual ear features, with supporting statistical metrics developed across a variety of ear images, toward the goal of Daubert admissibility.
- Recommend that, upon advancement of the above tasks, the FBI develop a training and testing program for forensic ear and ear print examiners as a component to augment forensic face recognition.

## 4 Speaker Identification

### 4.1 Introduction

Speaker recognition has been dominated by two applications for about 40 years: the commercial problem of cooperative verification for access control, and the government-oriented problem of identifying uncooperative (or unaware) speakers. The former is “easy” in that the speaker provides a unique pass-phrase, but it is difficult in that the cost of a false-accept could be enormous. The latter problem of identifying text-independent speakers over unknown channels of telephone bandwidth speech (300–3700Hz) is made even more difficult by the sparseness of speech samples in some applications. The result has been an uncomfortable tradeoff between false alarms and missed detections. The commercial players have faded with time, as many applications have proven unprofitable. In 1996, NIST began a worldwide competition in the text-independent telephone application that has proven successful, both in number of competitors (more than 40 this year) and advancements in technology. Error rates below five percent are now routinely achieved in NIST tests. Cell phones and landlines are mixed, and multilingual testing has shown no particular difficulties with any language. Even cross-language testing (train in one language and test in a second) has not shown a large increase in errors.

There are many methods for doing text-independent speaker recognition, but one has predominated over the last 15 years—Gaussian Mixture Models (GMM) using acoustic features based on the short-time (~20 msec) cepstrum of speech. Most competitors include a GMM subsystem in their overall recognition system. However, when additional training data is available (i.e., more than a single phone call), results improve by using additional higher-level features including phonetic sequences, prosodics (the rhythmic patterns of speech), and word selection (ideolectics). When eight training calls and matched channels are used, error rates below two percent have been achieved.

Most recently, NIST and other government sponsors have begun to investigate new applications, where far-field microphones capture speech in an interview room situation. Long neglected speech technology problems are now beginning to be re-examined (e.g. room acoustics). Other current research areas include temporal effects (age of the speaker and time lapse between speech samples), and the effects of vocal effort on recognition performance. While progress has been made on the research side, there still remain numerous challenges for speaker ID to be used for operational forensic applications. Some major obstacles include cross-channel effects, mixed and unknown microphones, insufficient speech and poor audio quality data, and the complexity of obtaining and establishing ground truth.

### 4.2 Background

Speaker recognition technology dates back to work in “visible speech” by Bell Laboratories during WWII. Visible speech used spectrograms created by pens on paper, directed by analog filter banks, to show intensity variations of sound frequencies over time. Humans were trained to recognize words and speakers from these spectrograms, which became known as “voiceprints” in

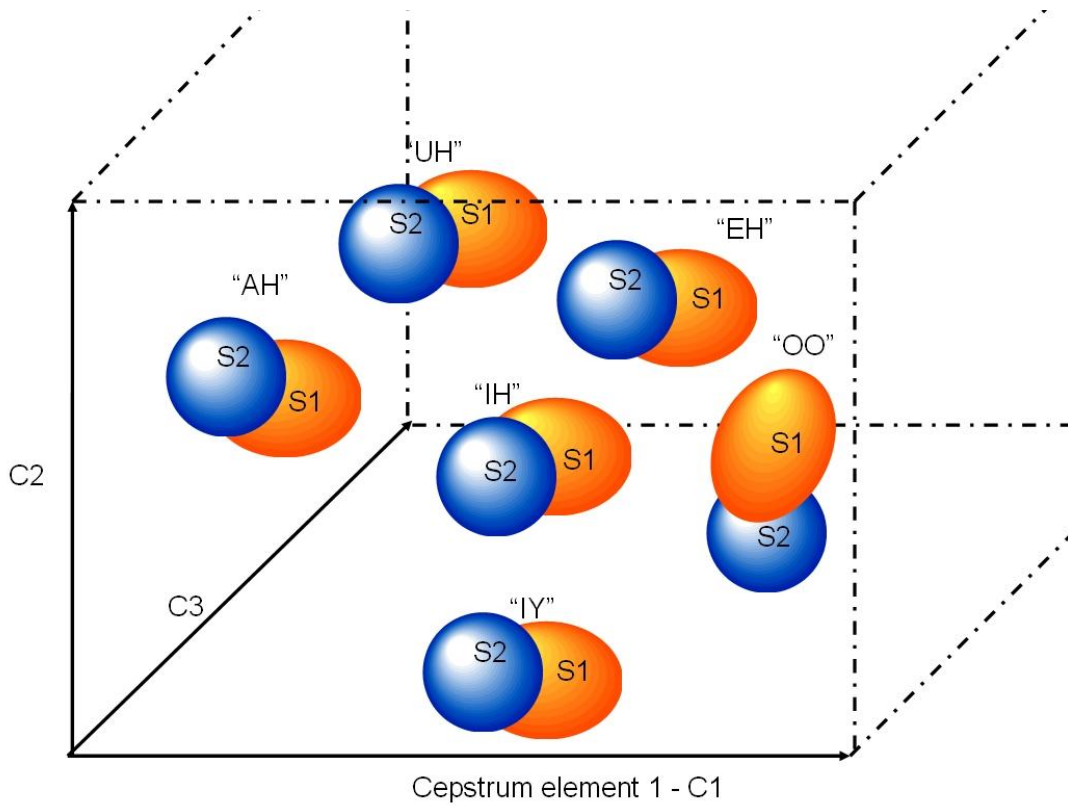
1962 with the publication of a famous paper by Lawrence Kersta in the journal *Nature*. Forensic applications of voiceprints were immediately controversial and were the subject of numerous critical studies.

By mid-1960, the scientific community had moved away from the use of spectrograms for speaker recognition and on to more statistically-based techniques, in an effort to build completely automated systems for speaker recognition. In 1969, “cepstral coefficients” were proposed as the appropriate features for speaker recognition. To this day, they remain the dominant mathematical features used for speaker recognition. The term “cepstrum” is derived by reversing part of the word “spectrum”—a play on words illustrating in a crude way the technical meaning. The cepstrum is derived by finding the spatial frequencies in a graph of log-energy frequency components of the speech. The cepstrum shows harmonic relationships in the sound energy.

In 1992, the GMM technique using cepstral coefficients as features was introduced; it quickly became the dominant classification approach in speech and speaker recognition. The GMM approach assumes that each person has a limited number of vocal tract “states,” which varies widely depending on the application (generally between 128 and 2,048). Each sound or “phone” emitted by a person is assumed to come from one of those states. Each sound results in some pattern of cepstral coefficients, so the cepstral coefficients characterize that state.

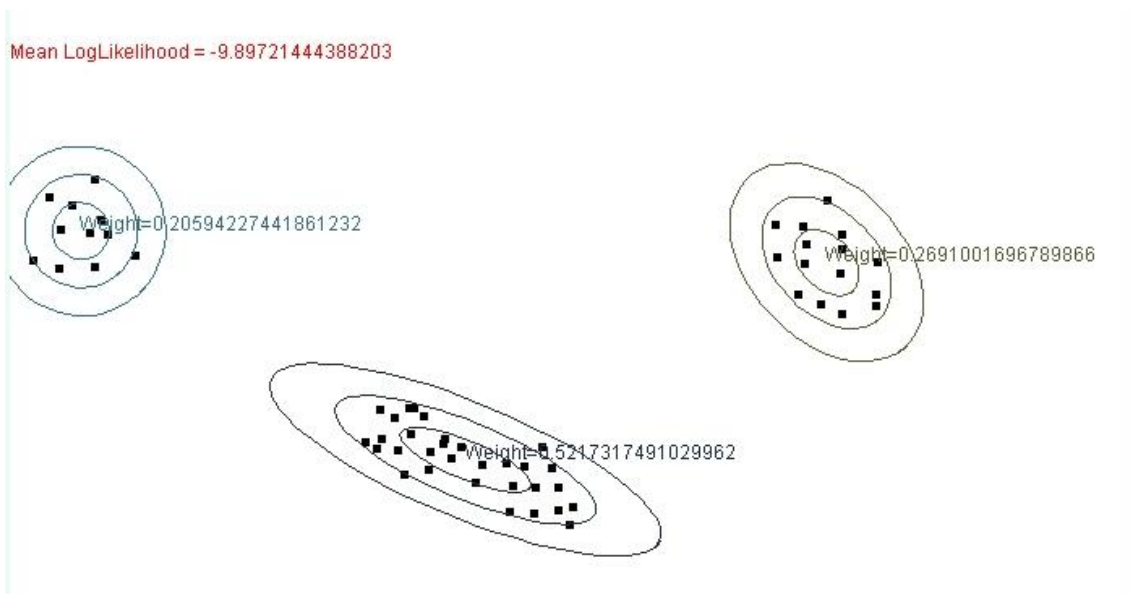
To help visualize GMM, assume, for example, that there are only three cepstral coefficients instead of the usual 12 to 18. (In practice time difference terms are usually appended to the cepstral coefficients giving a sort of velocity measure, or first derivative, for the cepstrum). Thinking of a 3-D coordinate system (x,y,z), the x-axis will be the value of the first coefficient, the y-axis the second, and the z-axis the third. The cepstral coefficients of each vocal tract state can be represented as a point in this 3-D space. If there are 64 vocal tract states, there will be 64 points in our space. But, each time a sound is heard from a particular state, the cepstral coefficients will be slightly different; instead of having 64 points in the space, there will be 64 clouds of points (each cloud is called a mixture). When someone is enrolled in the speaker recognition system, cepstral coefficients are computed at about 50 times per second to create 64 mixtures. The location, shape, and percentage of total points in each of these 64 mixtures will become the enrollment “model” for that speaker. When the voice of an unknown speaker is obtained, the resulting clouds of points can be compared to those of any known speaker

Figure 4-1 illustrates the situation. The 3-D cepstral space shows the mixtures associated with two different speakers (S1 and S2), for six different vowel sounds. If the incoming cepstral vectors are nearest to the S1 “clouds,” the system declares S1. If nearest to S2, S2 is declared. Or the incoming vectors may be far from both, in which case no positive classification can be made.



**Figure 4-1. 3-D GMM Mixture Diagram**

Figure 4-2 illustrates the training process for a GMM, where instead of 64 mixtures, three are shown (with only two Cepstral coefficients to make the display understandable). Each mixture is characterized by its mean, its variance in each dimension, and weighting data.



**Figure 4-2. Illustration of Gaussian Mixture Model Training**

One problem with GMMs is that they can generate poor initial values due to different background effects. As researchers became aware of this, a solution emerged. Rather than run the GMM process on every speaker from scratch, speaker recognition systems have employed “Universal Background Models” (UBMs) in various ways. A UBM is a model created using a variety of speakers of the same type and under the same conditions as expected by the operational system. For example, if the system will be used primarily on English-speakers with an American dialect sampled over telephones, the UBM will contain a large number of such speakers under these conditions. To avoid the starting point problem in modern Speaker ID systems, the UBM is trained first. From this process, 64 common mixtures are obtained, which become the starting points for finding all individual speakers. Once the UBM is trained, the models for the enrolled speakers can be computed. By starting from common points, the way in which each speaker’s mixtures move away from the UBM, and away from each other, can be seen. The UBM also is used at scoring time. The score of the individual enrolled speaker is compared to the score of the UBM. Unless the score is significantly closer to the individual (than it is to the UBM), it is declared to be a no-match.

The above GMM method is used for “text independent” speaker recognition—the case in which each speaker is talking freely. For “text dependent” speaker recognition, each speaker recites a password or “pass-phrase.” The procedure is the same, but now there is an expected ordering to the transition from vocal tract state to vocal tract state. For a speaker to be recognized, not only must the states be the same as the enrollment model, but the transitions through the states must be the same. The approach designed to handle this case is called a “Hidden Markov Model” (HMM) and is similar to a GMM, except for the addition of defined states and their sequences (along with state transition probabilities). “Text independent” speaker recognition systems are nearly

always based on GMM alone and “text dependent” systems on GMM in tandem with HMM. Both methods are “acoustic,” meaning that they rely only upon the sound frequency (spectral) characteristics, not the content of the speech, and both are based on “short term” measures (20-30 msec of data producing each set of cepstral coefficients). Both methods are language-independent – both GMM and HMM could be trained on babble, for instance. A standing research question asks, “Can a GMM trained on a bilingual speaker in one language, be used to recognize the same speaker using another language?” “There are many research studies reporting high performance, but effort continues to determine exactly how good it is. A related problem is, “Can GMM or HMM be trained to recognize the language being spoken?” There has been work in this area. Although GMMs and HMMs are not predominant in this area, they have been quite useful. Automatic language identification is an active area of research with current recognition rates better than 95 percent for approximately 12 languages.

Some additional factors complicate the above simplified description. Every microphone has its own “transfer function” – specific frequency characteristics of the microphone. These characteristics change the locations in space of the cepstral coefficients representing the vocal tract states. If the enrollment model was created with a different microphone than the sample speech, it will be hard to correctly compare the sample to the model. Various techniques are available for “channel normalization,” which attempt to estimate the transfer function of the microphone in use and to correct for it accordingly. Moreover, cell phones have distortions associated with the speech coding (e.g., GSM for European phones) that are different than conventional landlines, which use  $\mu$ (mu)-law coders. These differences are also handled under channel normalization.

With the incremental improvement by different algorithms and combination of algorithms and biometric modalities, Automatic Speaker Identification technology has been making its way to forensic applications, although there are still many technical and user-interface challenges affecting its performance in real applications.

### **4.3 Government Involvement**

The U.S. government has taken an active role in supporting speaker recognition research since the “visible speech” activities of WWII. Since 1996, NIST has been coordinating annual text-independent speaker recognition evaluations (NIST SRE). Since 2003, NIST has publicly acknowledged that funding for this effort comes from the National Security Agency. According to the NIST website ([http://www.nist.gov/speech/tests/sre/2008/sre08\\_evalplan\\_release3.pdf](http://www.nist.gov/speech/tests/sre/2008/sre08_evalplan_release3.pdf)), “These evaluations are an important contribution to the direction of research efforts and the calibration of technical capabilities. They are intended to be of interest to all researchers working on the problem of text independent speaker recognition. To this end, the evaluation is designed to be simple, focus on core technology issues, fully supported, and accessible to those wishing to participate.”

Participation in the evaluation is open internationally to all, but only participants can attend the evaluation workshops and results are not openly published. Since 1996, over 40 research sites,

including some in Europe and Israel, have participated in NIST SRE. In 2008, there were more than 40 submissions.

The Linguistic Data Consortium (LDC) at University of Pennsylvania was established by DARPA in 1992 to develop speech datasets for research applications. The NIST SRE has used the text-independent, telephone-bandwidth data sets developed by LDC from volunteers, who are largely college students. For the first time, the 2008 test will include “microphone channel” speech collected by the LDC in interview scenarios, in addition to the telephone datasets.

Telephone data for the NIST SRE tests are labeled by gender of the speaker, type of transmission channel (cellular, cordless, or land-line), and type of instrument (hand-held, speaker phone, head set, ear bud). Speech segments for enrollment or testing vary in length from 10 seconds to over eight minutes. All of these factors are known to impact the error rates for telephone speaker recognition. The NIST SRE can lead to estimates of performance of the various competing algorithms over all combinations of gender, data length, and type.

The NIST SRE does not sponsor tests of text-dependent speaker recognition systems (using passwords or pass-phrases), which is the approach of commercial systems for access control. When publicly questioned on the need for text-dependent testing within NIST SRE, NIST administrators invariably cite “sponsor requirements” in determining test protocols. The implication is that government funding sources are only concerned with text-independent applications and not applications for access control.

However, the FBI initiated a Forensic Automatic Speaker Recognition (FASR) evaluation (Nakasone 2001) which did include text-dependent evaluation and speaker modes such as spontaneous speech and read speech using FBI Forensic Voice Dataset collected during 1985-1989.

The UK National Physical Laboratory conducted a comparative test of eight biometric technologies in an access control application. One of the best performing technologies was a text-dependent speaker recognition system used in a quiet environment with the same land-line phone for enrollment and verification.

NIST held tests similar to SRE for automated Language Identification (LID) in 2003, 2005, and 2007. The 2007 test included 26 languages and dialects. In another area, the NIST “Rich Transcription” tests measure algorithm performance in converting speech to text.

In 1992, DARPA, NSF, and the DoD established the Center for Language and Speech Processing (CLSP) at the Johns Hopkins University (JHU). In 2002, the DoD also established the Human Language Technology Center of Excellence at Johns Hopkins. JHU reports that a minimum of \$48.4M in funding is involved through 2015.

Two other significant U.S. government organizations in automatic speaker identification algorithm development and evaluation and applications are the U.S. Air Force Research Lab in Rome, NY and the FBI in Quantico, VA. The latter has been an active player in shaping the Automatic Speech Recognition (ASR) technology to FASR (Nakasone, 2001, 2003, 2004)



IARPA is sponsoring the collection of new SID corpora as well as new research in dealing with the issues of live microphone data.

#### **4.4 Evaluation Standards**

In NIST-run evaluations of speaker identification systems, the two basic evaluation metrics are miss rate and false alarm rate. The interpretation of a system's output can be modified to result in lower false alarm rates but higher miss rates or vice versa. Because of this flexibility, NIST combines false alarm rate and miss rate, weighted by the "costs" of these errors, into a "Detection Cost Function," which serves as the performance measure for all systems in the test. This performance measure has been controversial because it requires competitors to set the threshold making the trade-off between false alarm and miss rates prior to seeing the full data set. NIST justifies this requirement by saying, "the task of determining appropriate decision thresholds is a necessary part of any speaker detection system and is a challenging research problem in and of itself." In common discussion, the term "equal error rate," although not of operational significance, is used as a single metric of performance—the EER is the point at which the miss rate and false positive rate are equal.

For example, say that there are 10,500 audio clips in a test corpus and, of those, 500 clips spoken by the target. A speaker ID system with a 5 percent equal error rate would correctly identify 475 clips as spoken by the target, correctly identify 9,500 clips as not spoken by the target, miss 25 clips, and incorrectly claim that 500 other clips are spoken by the target.

Because of an agreement between NIST and NSA regarding test protocols, results of the SRE test are not publicly reported. Individual participants are free to report their own results. Consequently, there is understanding, particularly from the MIT Lincoln Laboratory, as to how well systems are performing in these tests.

In the 1998 NIST evaluation, an EER of 3 percent would have been very good performance. In the 2002 JHU workshop, the baseline system had an EER of 3.3 percent with ~3 minutes of training and an EER of 0.7 percent with ~24 minutes of training. By combining a number of higher-level features, the researchers at the JHU workshop brought the EER to 0.2 percent with ~24 minutes of training. To return to the example above, this system would correctly identify 499 clips spoken by the target, correctly identify 9980 clips not spoken by the target, miss 1 clip, and incorrectly claim that 20 clips are spoken by the target.

The error rates reported for these systems *should be viewed as a lower-bound* of what one might expect in an operational environment. Generally, one can expect ~2 percent EER if the audio quality is good and if the training data closely matches the testing data. Note that the operator of a speaker ID system can operate at any point on the false-alarm/missed detection continuum. Where the operator chooses to operate depends on the relative cost of these two types of errors. For example, if missed detections are viewed as disastrous, detection rates can be increased; however, the operator will have to pay with more false alarms.

## 4.5 Recent Scientific Advances

The most striking advance in speaker recognition in this decade has been “ideolectics” – the introduction of linguistic elements into the process of recognizing speakers from long segments of text-independent speech. This approach was the focus of a summer-long, government-sponsored workshop at the Johns Hopkins University (JHU) Center for Language and Speech Processing (CLSP) in 2002. “Ideolectics” looks beyond short-term acoustic measures and attempt to recognize distinguishing content in each person’s speech – the frequent use of phrases such as “uh-huh,” “I mean,” and “well, uh.” Subsequent SRE evaluations increased the length of the enrollment speech segments so that these linguistic elements could be recognized for each speaker. Substantial improvement in recognition performance was noted for those systems exploiting ideolectics.

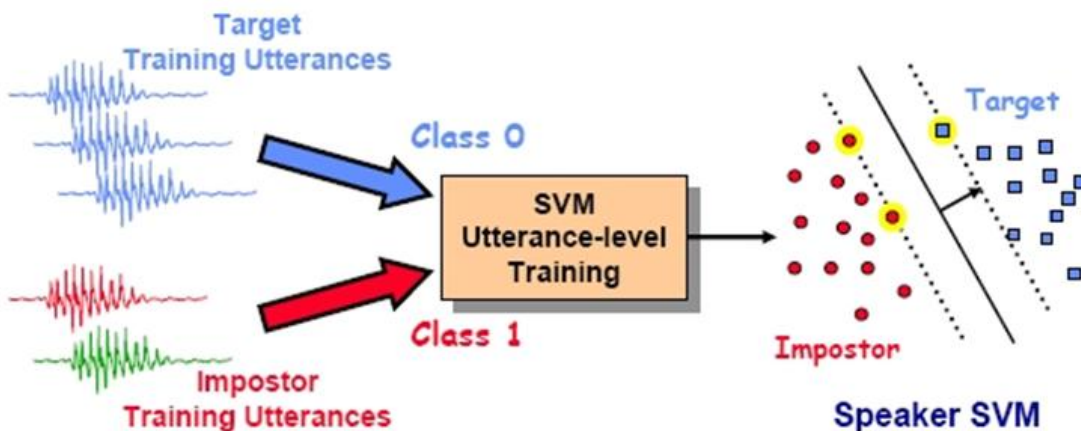
Within the last few years, research systems have started to incorporate other higher-level and slightly less frequent features. In the JHU 2002 summer workshop, researchers experimented with prosodic features (pitch and energy dynamics), phoneme features (using universal phoneme recognizers), lexical features (i.e., word choice as provided in various levels of fidelity from human transcription to noisy ASR), and conversational features (turn-based features such as phones per second, phones per word, various energy contours) under the SuperSID project at the 2002 JHU CLSP Summer Workshop (WS2002):

<http://www.clsp.jhu.edu/ws2002/groups/supersid/>. The paper in Campbell, J. P. et al. (2003) show how these novel features and classifiers provide complementary information and were fused to drive down the equal error rate on the 2001 NIST Extended Data Task to 0.2 percent—a 71 percent relative reduction in error over the previous state of the art. Of course, the improved accuracy exploring high level information comes with a computation cost; 24 minutes of training data will seldom be available in most applications.

In the last couple of years, there has been movement toward application of “Support Vector Machines” (SVM) to speaker recognition problems. The fundamental principle of SVMs is to increase the dimensionality of the feature space with the hope of more broadly separating the speech of different speakers. For example, if we have values for two cepstral coefficients  $(x,y)$ , we can artificially increase them to three variables with the transformation  $(x,y,xy)$ , whereby the product of  $x$  and  $y$  become the value on the  $z$ -axis. Perhaps with a large quantity of data, the best transforms for more broadly separating speakers may be learned.

Recent work is beginning to look at combining SVM approaches with GMM. The most popular usage is the so-called Gaussian Supervector technique. In this method, each of the Gaussian mixtures (64 were used in the earlier example) are concatenated into a single Supervector. Since each Cepstral vector is  $\sim 40$  dimensional, the resulting Supervector is of length  $64 \times 40 = 2560$ . Supervectors as large as 80,000 have been used in recent papers. After the model is created, an SVM training process is done, comparing the person’s Supervector to impostor Supervectors, and defining a decision boundary hyperplane. Figure 4-3 illustrates the SVM training process.

When unknown speech comes in, a corresponding unknown Supervector is created. This Supervector is compared to the decision hyperplane and a decision is made.



**Figure 4-3. SVM Training Process**

In their 2008 paper, “Speech Recognition as Feature Extraction...,” Stolke et al. report that; “...a great deal of progress and innovation in speaker recognition has been brought about by the use of support vector machines (SVMs) as speaker models. Through the ingenious design of features and kernels, SVMs have been applied to speaker modeling for a wide range of phenomena, from low-level cepstral observations to high level prosodic and lexical patterns.” In the Campbell, W. M., et al. (2004) paper, they show that the SVM and GMM are complementary technologies. Recent evaluations by NIST (telephone data) and Netherlands Forensic Institute (NFI) /TNO (forensic data) give a unique opportunity to test the robustness and viability of fusing GMM and SVM methods. They show that fusion produces a system which can have relative error rates 23 percent lower than individual systems alone.

#### **4.6 State of the Industry**

There appear to be fewer companies selling speaker recognition systems today than 10 years ago. Companies, such as Intellitrak, Lernout & Houspie, ITT, Keyware, Veritel, and T-Netix have either merged with other companies or gone out of business, as both the government and consumer markets for speaker recognition failed to develop as widely predicted. Some applications survived. Apple has included speaker recognition in Mac computers since OS 9. Most of the surviving vendors provide voice authentication or verification systems for access control applications. Table 4-1 lists some of the surviving commercial products and research institutes.

**Table 4-1. Industry Vendors for Speaker Recognition Systems**

<b>Company</b>	<b>Current Models</b>	<b>Attributes</b>	<b>Comments and Websites</b>
Nuance	“Verifier”	Text-dependent	<a href="http://www.nuance.com/verifier/">www.nuance.com/verifier/</a>
VoiceVault	“Caller Authentication”	Text-dependent	<a href="http://www.voicevault.com">www.voicevault.com</a>
Persay	“VocalPassword”	Text-dependent	<a href="http://www.persay.com">www.persay.com</a>
Persay	“Free Speech”	Text-independent	
Persay	“SPID”	Text-independent for intell applications	
Securus		“inmate calling” jail call monitoring services	<a href="http://www.t-netix.com">www.t-netix.com</a>
RSA		Call center	<a href="http://ww.rsa.com">ww.rsa.com</a>
Speech Sentinal	Securivox		<a href="http://www.speechsentinel.co.uk">www.speechsentinel.co.uk</a>
Anovea		Text-dependent	
IBM	Research Model	Text dependent & independent models	<a href="http://www-306.ibm.com/software/voice/">http://www-306.ibm.com/software/voice/</a>
Agnitio	Kivox	Text dependent & independent models	<a href="http://www.agnitio.es/">http://www.agnitio.es/</a>
Authenticate	Internet and Call	Text Independent	<a href="mailto:info@authenticate.com">info@authenticate.com</a>
Cellmax Systems	VIOMetrics	Text Dependent	<a href="http://www.cellmax-systems.com/">http://www.cellmax-systems.com/</a>
Diaphonics	multi-factor authentication combining voice biometrics with knowledge verification	Text dependent & independent models	<a href="http://www.diaphonics.com/">http://www.diaphonics.com/</a>

Company	Current Models	Attributes	Comments and Websites
Porticus Technology	For Financial Wire Transfer Application	Text Dependent	<a href="http://www.porticusinc.com/">http://www.porticusinc.com/</a>
Voice Verified	identity authentication	Text Dependent	<a href="http://www.voiceverified.com/">http://www.voiceverified.com/</a>
ALIZE	Open Source Toolkit to build speaker ID models	Text dependent & independent models	<a href="http://mistral.univ-avignon.fr/en/">http://mistral.univ-avignon.fr/en/</a>
SRI	Research System	Text dependent & independent models	<a href="http://www.sri.com/esd/pep/secuprds.html">http://www.sri.com/esd/pep/secuprds.html</a>
MIT LL	Research System	Text dependent & independent models	<a href="http://www.ll.mit.edu/index.html">http://www.ll.mit.edu/index.html</a>
AFRL	Speaker ID Models and Speech Enhancement tool kit	Text dependent & independent models	

## 4.7 High-Profile Implementations

### 4.7.1 Distinctions Between Speaker Verification and Speaker Identification

In speaker verification, a speaker enrolls in an authentication system by speaking a 5 word or less utterance. The next time a speaker wants access to a system, she/he speaks that same utterance. In speaker identification, analysts have a recording of an identified or labeled speaker and want the system to identify other audio clips that may contain the same speaker. Speaker identification applications may involve both text-independent and text-dependent speech samples.

#### 4.7.1.1 Speaker Verification Systems

There have been at least two high-profile attempts at commercial implementation of speaker recognition systems over the last decade. Both were immediately withdrawn. The first was a 1995 attempt by Sprint to use speaker recognition with their FONCARD service. High profile television advertisements were run for several months, in which a famous actor said that Sprint would be increasing security of the FONCARD by recognizing people through their voices. The system never went public.

In 1999, the Home Shopping Network gave several high profile talks, including one at the U.S. government's Biometric Consortium meeting, to introduce the inclusion of speaker recognition technology in its call center. The system was either never activated or immediately discontinued.

## **4.7.2 Speaker Identification System for Forensic Applications**

### **4.7.2.1 FASR**

FASR Prototype System was implemented at the FAVIAU (Nakasone 2001, 2002, 2003, 2004). The prototype was developed jointly by U.S. Air Force Research Laboratories, Rome, NY, and FAVIAU with technical inputs from MIT Lincoln Lab and BAE Systems. FASR uses robust speaker recognition algorithms including Mel cepstral coefficients, cepstral mean subtraction, or RASTA filtering and Gaussian Mixture Models with Universal Background Models. FASR is a PC-based workstation on a LAN with an efficient graphic user interface supporting: (1) data acquisition and playback; (2) signal and spectrographic display; (3) speech enhancement; (4) speech segmentation and labeling; (5) tone detection and removal; (6) speech quality measures (SNR, duration, bandwidth); (7) speaker identification and verification; (8) UBM Generation; and (9) automated computation of confidence measurements for each UBM.

Speaker recognition algorithms used in FASR have been tested on NIST single speaker and FBI Forensic Voice Database [Nakasone 2002]. The confidence measures are computed from large speaker populations with ground truth used to generate a UBM. The language types are primarily English, but include a variety of foreign languages such as Arabic, Spanish, and Chinese.

Since FASR prototype delivery in 2000, the FBI and its partners have continued experiments and research to refine performance, and established standards for data assessment and confidence measurements.

The FBI is using a PC-based FASR system (Nakasone, 2002). It improves the turnaround time over the traditional spectrographic method, but was not operational in a real time mode at the time of the paper (2002). It supports post-processing for forensic and intelligence purposes.

## **4.8 Standards and Interoperability**

There are currently no standards for voice collection that deal with equipment, speech style and content, and environment such that accuracy and interoperability can be assured. Given the broad range of models used (GMM, HMM, SVM) and the variation in the "background" data used in their generation, standardization at the model-level would not be currently achievable. The various representations at the feature-level are more limited (MFCC, LFCC, LPC, power spectrum), but the various types of signal enhancement and pre-processing in use make even feature-level standardization unlikely. However, standardization at the level of collection, representation, and storage of speech data could be achievable.

To this end, there are at least two current efforts. The first is that of ISO/IEC JTC1 SC37, 19794: Biometric Data Interchange Format-Part 13: Voice Data. This project is at the "Working Draft" stage, but is not progressing rapidly. The standard allows for storage of speech data in the

following formats: Pulse Code Modulation (PCM), Adaptive Differential PCM (ADPCM), Global System for Mobile communication (GSM), Adaptive Multi-Rate (AMR), G.711 A-law and mu-law, G.722.1, and G.722.2 AMR-Wideband (WB), G. 723, G. 728, and G. 729. The header in the data record block can accommodate some description of the collection conditions, such as “office environment,” “street,” “crowded,” “unknown,” “pin-drop silence,” “silence,” “quiet,” “undefined,” collection device, such as “handheld,” “telephone,” “mobile phone,” “others,” and microphone, such as “carbon,” “electret,” and “unknown.” The working draft currently contains no “best practices” or guidance on data collection.

The US DoD has developed a different data collection standard to support the CORVET (Coordinate Operational Resources for the Voice Exploitation Technology) data collection effort. These informal standards advocate use of prompted-text speech and have been highly controversial within the speaker recognition community. Voice samples from about 9,000 persons have been collected in Iraq using this standard. Because of the uncertainty of the utility of prompted-text speech, as opposed to free-style conversational speech, these collection and storage standards must be carefully evaluated.

The Linguistic Data Consortium collects data in the Speech File Manipulation Software (SPHERE) format developed by NIST IAD. The metadata for the collected speech is not stored with the data file, but rather in a separate file which gives all the metadata for the entire collection effort. The LDC, however, notes that the Microsoft “wav” standard has become universally popular and can support transfer of raw data.

The current standard for exchange of forensic biometric information is ANSI/NIST ITL 1-2000/2006. There is no provision in this standard, which has de facto international recognition, for the inclusion of speech data. The standard allows a “Type 99” record for any data collected using an international standard, but there is currently no international standard for voice data. Consequently, there is currently no way to formally exchange speaker data, except on an ad hoc basis.

#### **4.8.1 Forensic Capabilities**

The forensic applicability of speaker recognition has long been debated, in various books, journal articles, and a study by the National Research Council on this subject. Clearly, there are investigative uses for automated speaker recognition, but forensic admissibility under Daubert for results of ASR systems given the current state of the art is highly questionable.

In the realm of law enforcement, typical applications include wiretapping, body wiring techniques, surveillance audio, and prisoner monitoring. Because the recording environment is not typically controlled, a major challenge to verification and identification is ambient noise. Consequently, researchers are seeking ways to increase the robustness of speaker recognition in environments by using realistic noise (conversational babble, and city street noises) as part of their tests.

Other researchers have investigated the effect of disguised speech on speaker recognition systems. Zhang (2008) discusses a newly developed FASRS on which ten types of voice disguises common to forensic casework were tested. Researchers at the University of Avignon used a simple imposter voice transformation method, which transforms a speaker's speech signal to increase its likelihood to resemble the GMM corresponding to another speaker. "The results show that this simple voice transformation allows a drastic increase of the false acceptance rate, without a degradation of the natural aspect of the voice" (Matrouf and Bonastre, 2005).

To answer the courts' demand (in 1993 the U.S. Supreme Court decided in *Daubert v. Merrell Dow Pharmaceuticals* that, in order for scientific evidence to be admissible, it should be known to what extent the method has been and can be tested) for rigorous evaluation of forensic evidence, speaker recognition researchers began to examine ways in which Bayesian and Bayesian-like logic could be used to calculate the probability that a speaker's utterance matches a speech sample.

Rose describes the likelihood ratio (LR) as follows:

*Likelihood ratio is the ratio of the probability of the evidence assuming samples have come from same speaker to the probability of the evidence assuming the samples have come from different speakers [Rose, 2004].*

This is different from the likelihood or probability that a speaker's utterances are those of a sample.

#### **4.8.2 Vulnerabilities**

Training data should be very similar to the testing data. Some factors that may affect performance include:

- Speaker characteristics
- Vocal effort—whisper to scream
- Speech style—read, extemporaneous or drunk
- Speech rate—slow to fast
- Speech length (very short vs. very long utterances)
- Aging—time between sessions 1 hour – 10+ years
- Speaker health
- Native language
- Number of target speakers
- Channel and recording characteristics
- Signal to noise ratio
- Type of noise—white, babble, music, other speakers (crosstalk)
- Room acoustics—reverberation and echoes
- Sensor type—high quality microphone, telephone microphone...
- Distance of sensor from speaker



- Bit-error rate in compression and transmission
- Single speaker vs. multiple speakers at the same time.

Speaker recognition systems are vulnerable to impersonation and concealment: impersonation being done through “replay attacks” or electronic voice alteration, and concealment through electronic toys or more simple means to alter voice frequency characteristics. Spoofing speaker recognition technologies through impersonation was the topic of a special session at the Acoustic Society of America meeting in 1972. Journal papers on spoofing through voice impersonation have published regularly since then.

The “Speaker-Key” system developed by ITT industries in the 1990s had a novel approach to prevent replay attacks. The system was designed for monitoring offenders in home incarceration programs. Data subjects enrolled nine digits, “one, two, three....” and the numbers “twenty, thirty....ninety.” When called at home by the computerized system containing speaker recognition technology, they were asked to repeat randomly selected “combination lock” numbers, such a “39, 21, 43.” Under the simplified assumption of no co-articulation effects (that is, that current voice response does not depend upon previous or future responses), the response to “39” should be similar to the concatenation of the enrolled patterns “thirty” and “nine.”

It is commonly held that current speaker recognition technologies are not subject to impersonation attacks by mimics, but electronic voice mimicking can be a threat. In a 2008 International Conference on Acoustics, Speech and Signal Processing (ICASSP) paper, three Carnegie Mellon University (CMU) professors test the degree to which voice transformation is a threat to SID. They conclude that GMM methods are vulnerable, but that techniques based on other technologies (phonetics, prosody) are not.

## 4.9 Technology Objectives

The following are some recommended technology objectives for the field of voice biometrics:

- Stream-line the enrollment and detection in real time and provide a standard confidence ratio.
- Fuse multiple biometric matching such as record spoken signature (Humm, 2008) while taking other biometric data to prevent fraud at booking sessions.
- Build robust forensic voice databases and train robust speaker ID models.

## 4.10 Addressing Daubert Admissibility for Voice

The fundamental objective for the FBI with regard to biometric technologies must be in addressing the current Daubert admissibility gap [Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993)]. As with other technologies presented as “scientific,” automated speaker recognition must be:

- Testable, and been previously tested
- Subjected to peer review and publication

- Have a known or potential error rate
- Subject to existence and maintenance of standards controlling its operation
- Have widespread acceptance within a relevant scientific community.

Justice Blackman, in writing the above in *Daubert*, specifically called out speaker recognition techniques, citing *United States v. Smith*, 869 F. 2-D 348, 353-354 (CA7 1989) in referencing a “known or potential error rate.” Although laboratory error rates for the past NIST SRE tests are known, this may not be an adequate estimate of forensic error rates, where speech exemplars can be obtained in office conditions. It is highly significant that current and future NIST SRE tests may contain target speakers recorded under office conditions, supplying better data by which error rates can be estimated.

But for such error rate estimations to have operational significance in forensic applications, operational and test data collection protocols must be commensurate. Currently, test data protocols are controlled by NIST, LDC, and NSA. Operational protocols are controlled by the DoD, local police agencies, and, perhaps, the FBI. As previously discussed, there are no universal standards for any of these protocols. Consequently, the main challenge for the FBI and the U.S. government, in developing forensically admissible speaker recognition technologies, is to create uniform data collection protocols and data formats (including both speech waveforms and meta-data).

Because use of these standards will be by communities outside the direct control of the FBI (e.g., DHS, DoD, and foreign organizations), the standards must be created with the consensus of all stake holders. Here, the FBI has two potential paths: 1) pursuing the ISO/IEC JTC1 SC37 route; or 2) adding speech data formats and collection protocols to the ANSI/NIST ITL 1-2000/2006 documents. Because of the previous success of ANSI/NIST standards development in the area of fingerprinting and face recognition, it is our recommendation that the FBI follow the latter path. We recommend that the FBI work with NIST IAD to develop a speech data format for inclusion in ANSI/NIST ITL 1-2000/2006 and a “best practice” collection guideline for forensic applications. This format could be known as a “Type 19” record within the format.

The Daubert criteria of “widespread acceptance with a relevant scientific community” can only follow the development of such standards and will require FBI funding of the necessary scientific testing and background literature to support this acceptance. The infrastructure for this development already exists through the NIST/NSA speaker recognition community.

#### **4.11 Recommendations**

Recommend that the FBI support the following:

- Direct funding to NIST IAD for broadening SRE to include test protocols of operational interest to the FBI.
- Direct funding to LDC to establish test and development databases supporting forensic applications of speaker recognition technology.

- Fund industrial and academic groups already actively involved in the NIST SRE to continue their involvement. Such groups have been working without U.S. government funding, but cannot be expected to increase their output or performance without some promise of gold at the end of the rainbow.
- Create robust data collection protocols and “best practices” involving both telephone and office environment speech aimed at lowering error rates.
- Leverage the relevant international work to support scientific acceptance of forensic speaker recognition technologies, such as that by the Forensic Science Service, the University of Lausanne, and the Netherland Forensic Institute.
- Develop a plan for integrating speaker data with other modalities.

Having developed Daubert acceptability, pragmatic issues must also be resolved:

- Developing rapid hardware/software systems for “real time” processing of speech data against a large number of recognized target speakers. Commodity hardware such as GPUs or multi-core processors hold promise for making high performance processing cost effective.
- Developing additional “chain of custody” protocols and standards applicable to speech data collected by a variety of agencies, most outside the FBI.
- Developing forensically acceptable pre-processing algorithms for enhancing speech data, including robust activity detection and noise suppression.
- Working with the DoD, DHS and other agencies with a mission of combating terrorism to develop policies and procedures for implementing data collection protocols.
- Developing in-house capability for expert testimony at trial regarding the results of speaker recognition technologies.

It is recommended that the FBI begin a series of workshops with relevant stakeholders (DoD, DHS, DNI, NIST, NSA, LDC, and foreign allies) to outline a specific path forward and develop a timeline and a budget for this work.

This page intentionally left blank.

## 5 Handwriter Recognition

There are two primary approaches to automated recognition of persons using handwriting within the field of biometrics:

- 1) Dynamic: this is a multi-dimensional signal processing approach based on the movement of the pen during the writing process. The position and sometimes orientation of the pen is sampled thousands of times per second to create a time record of the pen's movement. This approach is sometimes called "on-line" recognition and is "text-dependent," meaning that the content of the writing must be predictable, limited usually to the writer's signature. Depending on the pen and tablet used, dynamic writing samples can consider pen pressure, 3-D pen orientation, pen up and down events, and high speed tracking of the pen for each stroke, character, word, or "gesture."
- 2) Static recognition: an image processing approach to recognizing the writer of a document through the shape of characters written. The entire written document is sampled once as a 2-D image. This approach is sometimes called "off-line" recognition and is "text-independent," in that the goal is to recognize the writer regardless of what is written. The analysis of questioned documents is a static recognition problem. There is no information on how the sample was produced other than what is physically presented. The timing components of starts, stops, and velocity estimation can at best only be inferred through second order indicators.

Neither of these approaches is to be confused with "digital signatures," an encryption technique associated with Public Key Infrastructures, designed to demonstrate both that an electronic document can be attributed to an encryption key holder and that the document has not been altered. Nor should these techniques be confused with automated handwriting recognition, which seeks to recognize handwritten characters, regardless of who wrote them. While both digital signatures and automated handwriting recognition are of great potential interest to the FBI, they are both outside the immediate scope of biometrics and will not be discussed in any detail. Questioned document examinations may also involve analysis of writing instruments, inks, papers, and other analysis techniques, which are outside of the scope of biometrics and are not discussed in any detail. This section will be focused on the recognition of the writers of handwritten text. Although this field is sometimes called "Handwriting Recognition," it is more aptly named "Handwriter Recognition." We will differentiate these terms here, using "handwriting" recognition to refer to the automated recognition of the content of a handwritten sample (as in Optical Character Recognition) and "handwriter" recognition to refer to recognition of the person who wrote the sample.

Development of dynamic handwriter, or signature, recognition dates to the mid-1960s. There have been several attempts at commercial applications, with a high profile, customer-oriented, pilot project recently completed by the Nationwide Building Society, a financial institution in the

United Kingdom. Dynamic signature recognition techniques are often discussed at biometric conferences and in the signal processing literature. There is one completed international standard in this field (ISO/IEC 19794-7:2007 “Biometric data interchange formats – Part 7: Signature/sign time series data”) and a second under development (ISO/IEC 19794-11, “Biometric data interchange formats – Part 11: Signature/sign Processed Dynamic Data.” Although it is possible to create a hypothetical situation in which an attempt to forge a dynamic signature becomes of forensic interest, there are insufficient applications that collect dynamic signature to make this technology of primary interest to the FBI. Consequently, dynamic signature recognition will not be discussed further in this document.

What is of great interest, however, is static handwriter recognition -- the automation of the expert handwriting analysis process well known in forensic sciences and the field of questioned documents. Handwriting analysis by expert witnesses was thoroughly analyzed against the *Daubert* requirements in *U.S. v. Prime*<sup>54</sup> and *U.S. v. Crisp*<sup>55</sup>. It was determined that handwriting analysis by expert witness meets the *Daubert* criteria for admissibility as scientific evidence. Today’s technologies seek to automate many of the techniques used by these handwriting analysis experts, if not to give positive identification, at least to supply a candidate list of possible writers from a database of known writers, or to supply possible linkages between unknown writers of multiple documents. Like all biometric methods, automated handwriter recognition must be robust against natural variations across multiple samples from a single writer taken over time (within-class variation), while detecting the variations between writers (between-class variation).

## 5.1 Technology Background

Handwriter recognition adheres to approaches similar to other biometric modalities with four major phases in the process: data collection, segmentation, feature extraction (template generation), and statistical analysis (feature comparison).

Data collection is the means by which a machine acquires the data to process. In static handwriter recognition, the collection is done with a digital, “flat bed” scanner. Documents, letters, and other media are scanned at a high, archival resolution and typically also at a lower resolution of 300 ppi for automated methods. It is unclear what the minimum resolution should be to ensure reliable capture of all possible features for automated recognition. The current FBI prototype system, FLASH ID, uses 300 ppi as a minimum and requires that document images in the repository be at the same resolution as the search document image.

Segmentation is the removal of the information of interest from the background. In handwriter recognition, this includes segmenting the letters and words from each other.

---

<sup>54</sup> *United States v. Michael Stefan Prime*, 220 F. Supp.2d 1203 (W.D.Wash. 2002).

<sup>55</sup> *United States v. Patrick Leroy Crisp*, 324 F.3d 261 (4th Cir. Court of Appeals, 2003).

Feature extraction is identifying the quantitative measurements of a sample in order to establish the characterization of a specific writing style. There are features present at all levels of a document, in descending order of scope: global features, paragraph features, sentence features, word features, and character or stroke features. For manual (non-automated) handwriter recognition, these features have been compiled<sup>56</sup> into a list of twenty-one discriminating elements:

- A. Elements of Style
  - 1. Arrangement
  - 2. Class of allograph
  - 3. Connections
  - 4. Design of allographs and their construction
  - 5. Dimensions
  - 6. Slant or slope
  - 7. Spacings
- B. Elements of Execution
  - 8. Abbreviations
  - 9. Alignment
  - 10. Commencements and terminations
  - 11. Diacritics and punctuation
  - 12. Embellishments
  - 13. Legibility or writing quality
  - 14. Line continuity
  - 15. Line quality
  - 16. Pen control
  - 17. Writing movement
- C. Attributes of all writing habits
  - 18. Natural variations or consistency
  - 19. Persistency
- D. Combinations of writing habits
  - 20. Lateral expansion
  - 21. Word proportions.

Automated handwriter recognition could, in theory, use any or all of these discriminating elements or “features.” In the studies examined herein, the basis for analysis was the shape of individual letters manually segmented from larger documents.

---

<sup>56</sup> R. A. Huber and A. M. Headrick, *Handwriting Identification: Facts and Fundamentals*, Boca Raton, FL: CRC Press, 1999.

### **5.1.1 Individuality of Handwriting**

In 2001, the National Institute of Justice sponsored research working to validate the hypothesis that handwriting is individuating.<sup>57</sup> The paper, later appearing in a 2002 issue of the *Journal of Forensic Sciences*, established the ability to determine the writer of a document with a high degree of confidence by analyzing handwriting samples from 1,500 individuals whose diversity was representative of the U.S. population. The research further supported handwriting as viable courtroom evidence. The variation of human writing between subjects is so complex that it rivals that of other biometric approaches such as DNA or iris identification analysis.

## **5.2 State of the Industry**

### **5.2.1 Data Collection**

Although data collection for handwriter recognition is generally done with a commercially available “flat bed” scanner, advances in digital cameras with optical zoom magnification that can reveal and record details will no doubt benefit this field. These instruments can help document examiners detect indented impressions in documents and allow additional handwriting data to be collected and stored digitally. Two vendors in particular are mentioned later whom provide such instruments.

### **5.2.2 Automated Handwriter and Handwriting Recognition (OCR) Technologies**

Automated handwriting recognition has been present for a number of years. The handwriting analyzers are capable of recognizing letters, characters, and words across wide variations in individual handwriting patterns. The problem of writer identification, however, requires specific enhancement of these variations that are characteristic to a unique writer’s hand. Therefore, handwriting and handwriter recognition present two opposing facets of handwriting analysis. The traditional character based recognition algorithms are concerned with the evolving pattern of combining the letters of the alphabet, wherein writer identification the focus lies on the shape and contours of characters as written by an individual. Lambert Schomaker believes that the advances in writer identification could aid the recognition process if information on the writer’s general writing habits and idiosyncrasies are available to the handwriting recognition system.<sup>58</sup>

In recent years writer identification and verification have received significant attention due to its forensic applicability (e.g., anthrax letters, robbery notes, the Ramsey case). A writer identification process involves a one-to-many (1:N) search of a large database of handwriting samples of known authorship and it produces a statistically determined list of candidates. This

---

<sup>57</sup> S. N. Srihari, S. Cha, H. Arora, S. Lee, "Individuality of Handwriting: A Validation Study," *ICDAR*, p. 0106, Sixth International Conference on Document Analysis and Recognition (ICDAR'01), 2001

<sup>58</sup> Text-Independent Writer Identification and Verification Using Textural and Allographic Feature



candidate list is further scrutinized by the forensic expert who makes the final decision regarding the identity of the questioned sample's author. The identification problem is that previous handwriting samples of enrolled individuals must exist. Writer verification involves a one-to-one (1:1) comparison of an individual's handwriting against known handwriting samples. The writer's handwriting must be automatically detected in a stream of handwritten documents.<sup>59</sup>

### **5.3 Growth and Markets**

There are many tools and systems currently in use today for writer recognition. Most are controlled by governments but some commercial and public technologies are available. Banking and commerce applications are candidate markets to adopt some of these technologies for point of sale authentication, but have not done so yet in any significant way. The tablet PC market offers a unique application platform for dynamic signature verification. However, these products tend to use stylus information more for recognizing characters and gestures (for application control). Signature or writing recognition applications are not widely used.

#### **5.3.1 Data Collection Tools**

In addition to traditional "flat bed" scanners, there are a variety of alternate light sources and imaging techniques that may assist with the identification of particular papers and inks. There are also a number of new tools commercially available for enhancing the process of digitizing handwritten marks left on papers and other surfaces. Some representative examples follow:

##### **5.3.1.1 ForensicXP**

ForensicXP is a digital forensic imaging spectrograph on the market. While similar in principle with the traditional video spectral comparator, the instrument is based on the latest 4-D hyper-spectrum digital technology and implements a fully computerized operation for questioned document processing. The principle of operation is automatic hyperspectrum measurement and processing. The instrument was created as a result of five years of research and development and has drawn the attention of the forensic experts by its exceptional sensitivity and resolution. Due to new hyper-spectral processing, the instrument can reveal some of the most difficult cases of obliterated writings including graphite, printer, and ink.<sup>60</sup>

##### **5.3.1.2 Foster and Freeman**

Foster and Freeman of the United Kingdom<sup>61</sup> provide several instruments for questioned documents examinations.

---

<sup>59</sup> Text-Independent Writer Identification and Verification Using Textural and Allographic Features

<sup>60</sup> [http://www.sciencedgl.com/Forensic/Forensic\\_hardware.html](http://www.sciencedgl.com/Forensic/Forensic_hardware.html)

<sup>61</sup> <http://www.fosterfreeman.com/products/documents/>

An Electrostatic Detection Apparatus (ESDA) detects indented writing on questioned documents by creating an invisible electrostatic image of the indented writing which is then visualized by the application of charge sensitive toners. The sensitive imaging process reacts to sites of microscopic damage to fibers at the surface of a document, which have been created by abrasive interaction with overlying surfaces during the act of handwriting.

The Video Spectral Comparator (VSC) 6000 digital imaging system provides document examiners a range of capabilities for detecting irregularities on altered and counterfeit documents. The applications of such a system include:

- Revealing concealed or obliterated information
- Examining watermarks
- Examining UV activated security features
- Detecting security features printed with anti-stokes inks
- Examining invisibly embedded personal information
- Detecting alterations by revealing the presence of chemically different inks
- Side by side document examination
- Casework reporting facilities.

Foram is a range of Raman Spectrometers for the examination of questioned documents and other forensic material. The instrument offers features specific to molecular structure and can provide valuable reference points for comparing and differentiating materials. It is non-destructive, provides up to 500x magnification, and can be operated on samples as small as 5 microns in diameter. It has data archiving with search and match facilities.

### **5.3.2 Handwriter Identification Tools**

#### **5.3.2.1 European FISH**

In 1977 the Germany's Bundeskriminalamt (BKA) funded a research project called Forensic Information System for Handwriting (FISH). A prototype was completed in 1989, and after several modifications and improvements, FISH proved capable of classifying handwriters by either using text-independent features or relying on a series of interactive, computer-aided measurements performed by an operator. The text-independent component is based on pattern recognition, where a handwritten document is evaluated and compared to a standard before storing each measurement as a pattern difference. The interactive approach looks at writing in a more traditional sense by considering such qualities as slope, height, width, upper and lower extensions, and distance between baselines, oval height, and the shape of loops. FISH is presently used by German authorities as well as certain federal agencies in the U.S. to consolidate and

associate unsolved cases, link handwritten documents, and assist investigations involving missing and exploited children.<sup>62</sup>

In 1991, FISH was comprised of 55,000 handwriting samples from 25,000 individuals. Thus, much effort has been devoted to compiling this database of handwriting.

#### **5.3.2.2 U.S. Secret Service FISH**

Maintained by the U.S. Secret Service (USSS), Forensic Information System for Handwriting (FISH) enables document examiners to scan and digitize text writings such as threatening correspondence to elected officials. A document examiner scans and digitizes an extended body of handwriting, which is then plotted as arithmetic and geometric values. Searches are made on images in an offline database, producing a list of probable candidate matches. The questioned writings, along with the closest matches, are then submitted to the Document Examination Section for confirmation.<sup>63</sup>

An attempt was made to contact a FISH expert inside the USSS in order to discuss the system in greater detail; however the request for a meeting was refused.

In addition to the capability of recognizing handwriters, the USSS has related technologies available to help link documents through paper and ink comparisons. They maintain a database of paper specimens that can determine how the paper was processed, what type of tree the paper came from, and where and when the paper was made and possibly sold. The Instrument Analysis Services Section houses the International Ink Library. It contains over 8,000 samples. This collection is used to identify the source of suspect writing by not only providing the type and brand of writing instrument, but the earliest possible date that the document could have been produced. This section also maintains a watermark collection of over 22,000 images as well as collections of plastics, toners, and computer printer inks.

#### **5.3.2.3 CEDAR-FOX**

CEDAR-FOX is a document analysis system developed by the Center of Excellence for Document Analysis and Recognition (CEDAR) at the University of Buffalo. The system has a variety of functions geared toward the Questioned Document Examiner (QDE):

- Handwriting recognition
- Writer identification and verification
- Signature verification
- Image processing

---

<sup>62</sup> Jan Seaman Kelly and Brian S. Lindblom, *Scientific Examination of Questioned Documents*, Second Edition, CRC Press 2006.

<sup>63</sup> <http://www.ojp.usdoj.gov/nij/journals/258/forensic-databases.html>

- Handwriting segmentation
- Several search modalities.

As a document management system for forensic analysis, CEDAR-FOX provides users with three major functionalities. First it can be used as a document analysis system. Second it can be used for creating a digital library for forensic handwritten documents, and third, it can be used as a database management system for document retrieval and writer identification. As an interactive document analysis system, a graphic interface is provided which can scan or load a handwritten document image. The system will first automatically extract features based on document image processing and recognition. The user can then use tools provided to perform document examination and extract document metrics. These tools include capabilities such as image selection, image enhancement, and contour display.

During writer verification, when a known document is compared to a questioned document, CEDAR-FOX analyzes the questioned document and calculates a similarity score they call the Log Likelihood Ratio (LLR) that the two documents are from the same writer as opposed to different writers. Additionally, the system has the ability to learn from known documents for writer verification. The system requires a minimum of four samples from the same writer in order for the system to be trained on the writer's handwriting. Following system learning/training, questioned documents can then be compared against the known samples from the writer for verification purposes. CEDAR-FOX uses a batch processing feature for writer identification. From a set of known documents, the system can find the closest documents to a questioned document.

CEDAR-FOX has two signature verification modes. In the first signature verification mode, a known and a questioned signature are compared; the system then generates a score indicating whether the questioned signature is genuine or not. The second signature verification mode has the ability to learn from known signature samples. The system recommends that a minimum of four known samples be used. Following system training on the known writer's signature, questioned signatures can be compared and a probability score is generated which indicates whether the questioned signatures are likely to be genuine or not.

The current system is intended to further enhance and expedite the QDE's analysis. CEDAR-FOX has been tested by the Canada Border Agency, the U.S. Secret Service, the Federal Bureau of Investigation, and is currently being evaluated by the San Diego Police Department and Minnesota State Patrol. In addition, the system has been licensed by the Netherlands Forensic Institute and the Canada Border Services Agency (CBSA). A trial version of CEDAR-FOX is

available for download.<sup>64</sup> An Arabic version of CEDAR-FOX, known as CEDARABIC, is also available.

#### **5.3.2.4 FLEX-Tracker and FLEX-Miner**

The FLEX-Tracker, a handwriting biometric product developed by Gannon Technology,<sup>65</sup> utilizes a technique called graph-based pattern matching, which statistically compares measurements on similar objects across different handwriting samples to identify characteristics unique to an individual writer. These characteristics include pen strokes, loops, crossed lines, and size which can be analyzed and then translated into a mathematical characterization or “identifier” for each writer.

The FLEX-Miner indexes and facilitates in-language key word search across massive repositories of multi-language handwritten documents. Directed Workflow, Gannon Technology’s proprietary document screening process, enables humans to make final checks on large quantities of pictographically matched results.

#### **5.3.2.5 FLASH ID**

FLASH ID is an FBI handwriting biometric product. FLASH ID represents an automated process for extracting graphical data from handwritten documents, analyzing this data using established statistical methods and matching documents based on similarity of the captured writing.<sup>66</sup>

FLASH-ID extracts features graphically from written documents at the level of individual letters and characters. Extracted features can be writer-dependent – that is, different features can be deemed individuating for different persons. The FLASH ID system was designed to use features not tied to a particular alphabet and thus intended and believed to be language independent. The property of language independence has not yet been fully validated.

#### **5.3.2.6 European Script**

The Script system has been in use in the Netherlands since 1996. The handwriting processing procedure is much more automated than the European FISH system. Script is designed to perform quantitative information analysis by providing information about the frequency of certain writing features that are encountered in the data set, as well as studying the effects of natural variation.

A problem with the FISH and Script systems is that large amounts of handwritten data are needed to train the systems on each writer and large amounts of sample data are needed to identify the writer from those in the training database. Nowadays, handwriting is found less often at crime scenes attributed primarily to increased use of computers; hence, the number of critical cases

---

<sup>64</sup> [http://www.cedartech.com/products\\_cedarfox.html](http://www.cedartech.com/products_cedarfox.html)

<sup>65</sup> <http://www.gannontech.com/index.html>

<sup>66</sup> <http://www.gannontech.com/index.html>

involving handwriting has decreased. From the limited results in practice, it can be concluded that the use of these systems is not economically feasible in regular casework.<sup>67</sup>

### **5.3.2.7 Arkansas State Crime Laboratory**

The Arkansas State Crime Laboratory has developed a non-automated system based on eight handwritten letters: *a, d, f, g, i, k, r,* and *t*. Handwritten documents submitted to the lab are classified based on these features and compared to other samples filed in a database. Each candidate match is compared to the questioned document by a handwriting expert to determine if a match exists.<sup>68</sup>

## **5.4 Performance and Accuracy**

The National Institutes of Standards and Technology Information Technology Laboratory, Information Access Division, regularly conducts tests of fingerprint, face, speaker and iris recognition technologies<sup>69</sup>, as well as “text retrieval” systems. Optical character recognition systems have also previously been tested, however, there has been no test program for handwriter recognition. Consequently, the only performance results available are those self-reported by system developers, each using their own handwriting database and test protocols. Several documents available on self-reported performance tests of handwriter recognition systems were reviewed and found to be inconsistent with published international standards for biometric testing and reporting (ISO/IEC 19795-1:2006: “Information technology – Biometric performance testing and reporting – Part 1: Principles and framework”). Consequently, it is not possible to interpret results across studies or to place these results within a scientific framework. Nonetheless, the studies are summarized here.

It was discovered that USSS FISH and CEDAR-FOX are the two systems that have been established the longest and have been at least basically evaluated. Unfortunately the performance numbers for the USSS FISH system were not available, however according to some European studies, the target performance for the European FISH forensic writer identification system is a near 100 percent recall of the correct writer in a closed-search of 100 writers, computed from a database on the order of 10,000 samples (the size of the current European forensic database). As closed-searches do not allow for the possibility that the writer of the sample is actually unknown and the performance figures decrease rapidly with increasing number of writers known to the

---

<sup>67</sup> <http://forensic.to/Dissertation.pdf>

<sup>68</sup> Jan Seaman Kelly and Brian S. Lindblom, *Scientific Examination of Questioned Documents*, Second Edition, CRC Press 2006

<sup>69</sup> <http://www.itl.nist.gov/iad/894.03/>

system, it is not clear how this system would perform in a “real world” application. Therefore, the target performance still remains an ambitious goal.<sup>70</sup>

In one current CEDAR-FOX case study, *On the Discriminability of the Handwriting of Twins*, undertaken by the University of Buffalo’s CEDAR, handwriting of 206 pairs of twins (412 people) was examined. The study also included 1,648 non-twins (members of the general population). It appears that the false negative/false positive error rates for twins was 12.6 and 13.2 percent, and for non-twins was 3.2 and 4.2 percent.<sup>71</sup>

Another case study<sup>72</sup> by the same authors was performed on CEDAR-FOX analyzing offline signature performance where 55 individuals contributed 24 signatures, creating 1,320 genuine signatures. Some were asked to forge three other writers’ signatures, eight times per subject, thus creating 1,320 forgeries. Each signature was scanned as a 300 dpi grey-scale image which was then converted to binary and subjected to noise removal and slant normalization. Three unique features were extracted from each signature: 1) image gradient analysis of all pixels in the rectangle, 2) distribution of pixels of signature, and 3) geometrical and topological features corresponding to stroke segments. Each image was normalized to fit in a 4x8 grid before features were extracted. This study differentiated between false negatives (false rejection rate) and false positives (false acceptance rate), but also included a non-standard metric called an “average error rate,” where  $AER = [(FRR + FAR) / 2]$ . Those rates, calculated using a variety of approaches over the same data, are given in Table 5-1. The data is inconclusive as to which is the best analysis approach due to the dataset being rather small. However, the researchers on this study claim that as the size of the training samples increases, the performance is improved.

**Table 5-1. Writer-Independent Methods with 1 and 16 Training Samples**

Methods	Samples	FRR(%)	FAR(%)	AER(%)
Distance Stats	1	27.6	27.8	27.7
Naive Bayes	1	27.2	26.0	26.6
Distance Stats	16	21.3	22.1	21.7
Naive Bayes	16	22.9	24.1	23.5

<sup>70</sup> Marius Bulacu and Lamber Schomaker, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 29, NO. 4, April 2007, Text-Independent Writer Identification and Verification Using Textural and Allographic Features

<sup>71</sup> On the Discriminability of the Handwriting of Twins, Sargur Srihari, Chen Huang and Harish Srinivasan, June 2007, <http://www.cedar.buffalo.edu/~srihari/papers/TR-04-07.pdf>

<sup>72</sup> <http://www.cedar.buffalo.edu/~srihari/papers/BBC-IAPR-Springer.pdf>

CEDAR-FOX claims a 97.94 percent “writer identification rate,” again a non-standard metric not comparable to metrics reported in the other studies, based on a study of 1,500 writers.<sup>73</sup> In this study the researchers collected 3 documents for each writer. For the same-writer category the writer set has been divided into two sets: 500 for training and 500 for testing. Therefore, for each character there are 3x500 distances between samples belonging to the same writer in each of the training and testing sets. For the different-writer set 1,500 pairs of documents (from different writers) were randomly chosen from the first 500 writers for each of the training and test sets. Two types of features are extracted from the handwritten documents for classification. The first type of feature known as a micro-feature is a binary string extracted from identified and recognized character components. The second type of feature, known as a macro-feature, is extracted from processing of the document images for global characteristics of a writer’s individualities. The final discrimination between two documents is done based on the combination of modified likelihood ratios of both micro and macro features. According to this study, writer discriminability is highest when all of the macro-features are used. It is least when only ten numerals (digits) in the handwritten document are used. Each character has certain discriminability value associated with it; therefore, the more characters a document contains the greater the system’s performance accuracy. However, the more features that are extracted the slower the document processing time. The final result in this study showed that the performance for writer identification using a document set from 975 writers resulted in an identification rate around 60 percent using only macro features. In using macro features plus micro features extracted only from 10 characters, the identification rate was 89.1 percent, and in using macro features and micro features extracted from 62 characters, the identification rate was 97.94 percent.

## **5.5 Standardization and Interoperability**

### **5.5.1 ANSI/NIST ITL 1-2007**

The de-facto standard for international forensic data exchange is the ANSI/NIST ITL 1-2007, which now allows exchange of fingerprint, mug shots, iris images and scar/mark/tattoo data. This standard also accommodates a “Type 99” record, which is any type of data (not addressed elsewhere in the standard) for which an international standard exists. There is no provision in the current ANSI/NIST ITL 1-2007 standard for the exchange of handwriter data.

---

<sup>73</sup> Sargur N. Srihari and Zhixin SHI, Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL’04), Forensic Handwritten Document Retrieval System, Center of Excellence for Document Analysis and Recognition (CEDAR), University at Buffalo, State University of New York, Buffalo, USA Srihari@cedar.buffalo.edu



## **5.5.2 WANDAML**

WandaML is an XML-based markup language for the annotation and filter journaling of digital documents. It addresses the needs of forensic handwriter data examination by allowing experts to enter information about writer, material (pen, paper), script, and content, and to record chains of image filtering and feature extraction operations applied to the data. Annotations may be organized in a structure that reflects the document layout via a hierarchy of document regions. WANDAML lends itself to a variety of applications, including the annotation of multiple types of handwriter documents (online or offline), images of printed text, medical images, and satellite images. FISH uses WANDAML for data export.

## **5.6 Data**

There are many handwriter based databases that have been used to generate numerous academic papers and performance statistics. Unfortunately, most academic datasets are quite small which reduce the research and scientific capabilities of a system assessment. A few of the larger and most popular databases are described here.

### **5.6.1 CEDAR**

CEDAR is an online handwriter database of 200 subjects for a total of 105,573 words. The database contains both printed and cursive writings. A digital tablet with a stylus writing device was used to capture sample handwritings. Twelve passages were used from different types of English usage (i.e., business, legal, scientific, informational).

### **5.6.2 The U.S. Government Agencies**

#### **5.6.2.1 National Institute of Standards and Technology (NIST)**

NIST Special Database 19 contains the entire NIST corpus of training materials for hand printed document and character recognition. It publishes Handprinted Sample Forms from 3,600 writers, 810,000 character images isolated from their forms, ground truth classifications for those images, reference forms for further data collection, and software utilities for image management and handling.<sup>74</sup> This database, although extensive, was collected for handwriting, not handwriter, recognition. NIST has historically conducted tests and encouraged industrial and academic development of biometric systems, such as face, iris, speaker, fingerprint, and optical character recognition systems, and has collected extensive test and development databases for each of these modalities. NIST has not, however, conducted tests of automated handwriter recognition.

---

<sup>74</sup> <http://www.nist.gov/srd/nistsd19.htm>

### **5.6.2.2 The United States Secret Service (USSS)**

The USSS is in possession of two databases of importance to handwriter recognition: 1) the FISH system and 2) the International Ink Library. The International Ink Library, maintained jointly by the USSS and the Internal Revenue Service, includes more than 9,500 inks, dating as far back as the 1920s. Every year, pen and ink manufacturers are asked to submit their new ink formulations, which are chemically tested and added to the reference collection. In a questioned document analysis, linking of inks is evidence that two documents were written with the same class writing utensil. Open-market purchases of pens and inks ensure that the library is as comprehensive as possible. The USSS generally provides assistance to law enforcement on a case-by-case basis.<sup>75</sup>

### **5.6.2.3 Federal Bureau of Investigation (FBI) Files**

The FBI Questioned Documents Unit maintains two types of files for their investigations: reference files, and standard files. The reference files are information drawn from casework which is used to relate incoming data to previously examined material, for example, to make an association between two threatening notes. The standard files are repositories for manufacturer's and similar primary-source data which are used to determine the source of an item of evidence (e.g., the maker of a style of typeface). FBI reference and standards files with handwriting or signature data include:

- Anonymous Letter File and Bank Robbery Note File
- National Fraudulent Check File
- National Motor Vehicle Certificate of Title File
- Government Issued Documents (e.g. drivers licenses, social security cards).

### **5.6.3 Requested Writings**

In the 1920s, Albert S. Osborn composed standard texts,<sup>76</sup> containing examples of each letter, in both uppercase and lowercase, all numerals, and various punctuation. This "London Letter" and another similar text from the same publication, known as the "Dear Sam Letter." is widely used as requested writings or source documents when collecting handwriting samples.

---

<sup>75</sup> <http://www.ojp.usdoj.gov/nij/journals/258/forensic-databases.html>. For additional information, contact 202-406-5708.

<sup>76</sup> A.S. Osborn, *Questioned Documents*. 2<sup>nd</sup> edition, Albany, NY: Boyd Printing Company. Reprinted, Chicago, IL: Nelson Hall Co., 1929.

Our London business if good, but Vienna and Berlin are quiet. Mr. D. Lloyd has gone to Switzerland and I hope for good news. He will be there for a week at 1496 Zermatt St. and then goes to Turin and Rome and will join Col. Parry and arrive at Athens, Greece, Nov. 27th or Dec. 2nd. Letters there should be addressed: King James Blvd. 3580. We expect Charles E. Fuller Tuesday. Dr. L. McQuaid and Robt. Unger, Esq., left of the 'Y.X.' Express tonight.

**Figure 5-1. “London Letter”**

Modern research has improved upon the “London Letter” text to provide for greater variation of character usage and their placement within words and sentences. An example of a modern requested writing is known as the “CEDAR Letter.” This sample is designed to present each letter in the initial, capitalized position, and each lower case letter in lead, internal and terminal positions within words.<sup>77</sup> In all but one case, a terminal lower-case “j,” the CEDAR Letter provides for all variations.

---

<sup>77</sup> Cha, S.H., Srihari, S.N., *Writer Identification: Statistical Analysis and Dichotomizer*, Lecture Notes in Computer Science, Volume 1876, Jan 2000, Page 123

From  
Jim Elder  
829 Loop Street, Apt 300  
Allentown, New York 14707

Nov 10, 1999

To  
Dr. Bob Grant  
602 Queensberry Parkway  
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later; were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!  
Jim

**Figure 5-2. "CEDAR Letter"**

## **5.7 Technology gaps and Challenges**

What follows are the general recommendations, categorized by reasonable timeframes, that should be considered when evaluating handwriter recognition applications, research and development.

### **0 to 2 Years**

- Baseline handwriter recognition performance for questioned documents by conducting comparative analysis between systems; fund and leverage the experience of NIST for evaluating recognition performance of handwriter recognition as well as the underlying feature extraction processes.
- Propose standard feature representations derived from leading research and current prototype systems, and advance these through NIST.

- Collect progressively larger known test sets for training, development, and testing of existing and future systems.
- Request case feedback to better establish ground truth and performance metrics (human and automated).
- Refine support tools for human visualization, mark up, and verification of features.

## **2 to 5 Years**

- Integrate writer recognition with character, text, and language recognition.
- As non-handwritten communications become more prevalent, such as blogging, text messaging and emails, there is a growing need to identify writers not by their written script, but by analysis of the typed content. Currently, there are some studies in the area of writer's colloquial analysis that may lead to the emerging technology of writer identification in the "blogosphere." These technologies could possibly create a profile and even identify a writer's identity. Similar to colloquial speech analysis, studies have shown that bloggers and chatters use colloquial forms of writing instead of standard forms when blogging, chatting, or text messaging. Recommend investment in scientifically-based text-independent e-mail and blog writer identification and document linking.

## **5 to 10 Years**

- Consider, for investigative use, integrating automated services in Next Generation IAFIS for handwriter recognition. An initial form of integration could be the cross referencing of confirmed samples (solved questioned documents) to their corresponding criminal files.

## **5.8 Conclusions**

Handwriting is a biometric characteristic subject to all manner of natural and random variation. Some of these variations can be due principally to the lack of machinelike precision in the human body, but can also be accentuated by external factors, such as writing position, writing instrument, and care of execution. Writing variation is also influenced by physical and mental conditions such as fatigue, intoxication, drug use, illness, advanced age, nervousness, and the writer's urgency or emotional state. These factors make handwriter recognition a challenge for the scientists and researchers to overcome. Tools such as FLASH ID and CEDAR-FOX are extracting several features from handwritten texts useful for the recognition of writers and for linking unknown writers of available samples. However, a large number of sample writings per subject over time are required to accurately capture and understand fluctuations due to natural variation. Large datasets useful to the robust development and testing of automated handwriter analysis are not currently available to the government, and the science of automated handwriter recognition remains in its infancy. Although the U.S. government has for the last decade sponsored tests of fingerprint, iris, speaker, and face recognition technologies, no tests of handwriter recognition

have been conducted. On the other hand, increased use of computers for communication is diminishing the prevalence and necessity of handwritten notes, just as electronic signatures are becoming substitute for handwritten signatures, perhaps making forensic handwriter analysis less frequent in criminal investigations.<sup>78</sup>

As handwriting can be most accurately compared with automated tools when the questioned specimens and the sample set are written under comparable conditions, these automated handwriter recognition methods only serve as a tool in the toolboxes of the Forensic Document Examiners. Automated writer recognition cannot replace Forensic Document Examiners. Only through logical reasoning and the application of scientific principles by a qualified expert can the writer of a contested handwritten document be accurately established in a court of law.

## **5.9 Addenda**

### **5.9.1 Applicable Organizations and Conferences**

The following is a list of relevant organizations and conferences with interest and expertise in handwriter recognition.

- International Conference on Document Analysis and Recognition (ICDAR), held on odd years (<http://www.icdar2007.org/> & <http://www.icdar2009.org/>)
- International Conference on Frontiers in Handwriting Recognition (ICFHR), held on even years (<http://www.cenparmi.concordia.ca/ICFHR2008/index.html>)
- American Society of Questioned Document Examiners (ASQDE) (<http://www.asqde.org/>)
- American Academy of Forensic Sciences, Questioned Documents Section (<http://www.aafs.org/>)
- American Board of Forensic Document Examiners (ABFDE) (<http://www.abfde.org/>).

---

<sup>78</sup> Keystroke dynamics is the electronic counterpart to handwriter recognition. In contrast to handwritten notes that result in and emphasize static analysis, keystroke recognition is based on dynamic capture and requires special software to capture keystroke timing events.

## Appendix A Speaker References

- Alexander, A. et al, 2004, "Handling Mismatch in Corpus-Based Forensic Speaker Recognition," Odyssey 2004, The Speaker and Language Recognition Workshop, Toledo, Spain, pp. 69–74.
- Beck, S. D. et al, 2004, "A Bilingual Multimodal Voice Corpus for Language and Speaker Recognition (LASR) Services," Odyssey 2004, The Speaker and Language Recognition Workshop, Toledo, Spain.
- Bonastre, J. F. et al, 2003, "Person Authentication by Voice: A Need for Caution," *EUROSPEECH-2003*, ISCA, Geneva, Switzerland.
- Botti, F. et al, 2004, "On Compensation of Mismatched Recording Conditions in the Bayesian Approach for Forensic Automatic Speaker Recognition," *Forensic Science International*, Vol. 146: Supplement 1, 2 December 2004, pp. S101-S106.
- Byrne, C., P. Foulkes, 2004, "The 'Mobile Phone Effect' on Vowel Formants," *Speech, Language, and the Law*, Vol. 11: 1. pp. 1350-1771.
- Campbell, J. P. et al, 2004, "The MMSR Bilingual and Cross-channel Corpora for Speaker Recognition Research and Evaluation," ODYSSEY 2004-The Speaker and Language Recognition Workshop, May 31–June 3, 2004, Toledo, Spain.
- Campbell, J. P. et al, 2003, "Fusing High- and Low-Level Features for Speaker Recognition," *EUROSPEECH-2003*.
- Campbell, J.P., D.A. Reynolds, 1999, "Corpora for the Evaluation of Speaker Recognition Systems," <http://www.apl.jhu.edu/Classes/Notes/Campbell/SpkrRec/icassp99.htm> (last access 28 March 2008).
- Campbell, W. M., D.E. Sturim, D. A. Reynolds, May 2006, "Support Vector Machines Using GMM Supervectors for Speaker Verification," *IEEE Signal Processing Letters*, 13(5), pp. 308-311.
- Campbell, W. M. et al, 2004, "Language Recognition with Support Vector Machines," ODYSSEY 2004-The Speaker and Language Recognition Workshop, May 31–June 3, 2004, Toledo, Spain.
- Cieri, C. et al, 2007, "Resources for New Research Directions in Speaker Recognition: The Mixer 3, 4, and 5 Corpora," Philadelphia: University of Pennsylvania.
- Doddington, G., et al, 1998, "Sheeps, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation," *Proceedings of the ICSLP*.
- Drygajlo, A. et al, 2007, March 2007, "Forensic Automatic Speaker Recognition," *IEEE Signal Processing Magazine*, pp. 132-135.

- Gonzalez-Rodriguez, J. et al, 2006, "On Robust Estimation, Interpretation, and Assessment of Likelihood Ratios in Forensic Speaker Recognition," *Computer Speech and Language*, Vol. 20: 2-3, April-July 2006, pp. 331-55.
- Humm, A. et al, 2008, "Spoken Signature for User Authentication," *Journal of Electronic Imaging*, Vol. 17: 1, pp. 1-11.
- Jin, Q., A. Toth, A. Black, T. Schultz, 2008, "Is voice transformation a threat to speaker identification?" ICASSP2008, Las Vegas, NV.
- Kajarekar, S.S., et al, 2006, "A Study of Intentional Voice Modifications for Evading Automatic Speaker Recognition," *Proceedings of the IEEE Odyssey 2006 Speaker and Language Recognition Workshop*, San Juan, Puerto Rico.
- Kajarekar, S.S. et al, 2004, "Modeling NERFs for Speaker Recognition," *Odyssey 2004, The Speaker and Language Recognition Workshop*, Toledo, Spain.
- Kenny, P., P. Dumouchel, 2004, "Experiments in Speaker Verification Using Factor Analysis Likelihood Ratios," *Odyssey 2004, The Speaker and Language Recognition Workshop*, Toledo, Spain.
- Kersta, L., 1962, "Voiceprint Identification," *Nature* 196, pp 1253-1275.
- Nakasone, H. et al, 2004, "Pitch Synchronized Speech Processing (PSSP) for Speaker Recognition," *Odyssey 2004, The Speaker and Language Recognition Workshop*, Toledo, Spain.
- Nakasone, H., S. Beck, June 2001, "Forensic Automatic Speaker Recognition," *Proc. 2001 A Speech Odyssey*, ISCA, pp. 18-22.
- Nakasone, H., S.D. Beck, 2001, "Forensic Automatic Speaker Recognition," *2001: A Speaker Odyssey*, pp. 139-144, Crete, Greece.
- Nakasone, H., 2003, "Automated Speaker Recognition in Real World Conditions: Controlling the Uncontrollable," *EUROSPEECH-2003*, pp. 697-700.
- Przybocki, M.A., A. Martin, 2004, "NIST Speaker Recognition Evaluation Chronicles," *ODYSSEY 2004-The Speaker and Language Recognition Workshop*, May 31–June 3, 2004, Toledo, Spain.
- Reynolds, Douglas A., et al, 2000, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing* 10: pp. 19–41.
- Reynolds, D.A., R. C. Rose, 1995, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Transactions on Speech and Audio Processing*, 3: 1, pp. 72-83.
- Romito, L., V. Galata, 2004, "Toward a Protocol in Speaker Recognition Analysis," *Forensic Science International*, Vol. 146, Supplement 1, 2 December 2004, pp. S107-S111.



- Rose, P., 2006, "Technical Forensic Speaker Recognition: Evaluation, Types and Testing of Evidence," *Computer Speech and Language*, Vol 20: 2-3, April-July 2006, pp. 159-91.
- Rose, P., 2004, "Technical Forensic Speaker Identification from a Bayesian Linguist's Perspective," in Javier Ortega-García, et. al., *Odyssey 2004: The Speaker and Language Recognition Workshop*, Toledo, Spain, May 31-June 3, 2004, pp. 3-10.
- Shriberg, E., 2007, "Detecting Nonnative Speech Using Speaker Recognition Approaches," *Proc. Odyssey Speaker and Language Recognition Workshop*, n.p. Stellenbosch, South Africa, January 2008.
- Solomonoff, A. et al, 2004, "Channel Compensation for SVM Speaker Recognition," *Odyssey 2004, The Speaker and Language Recognition Workshop*, Toledo, Spain.
- Spaun, N., 2007, "IEEE Biometric Theory, Applications, and Systems," *Forensic Biometrics from Images and video at the Federal Bureau of Investigation*.
- Stolcke, A., S. Kajarekar, 2008, "Recognizing Arabic Speakers with English Phones," *Proceedings of the Odyssey Speaker and Language Recognition Workshop*, Stellenbosch, South Africa.
- Reynolds, Douglas, et al, 2002, "SuperSID: Exploiting High-Level Information for High-Performance Speaker Recognition," *SuperSID Project Final Report*, Johns Hopkins University, Center for Language and Speech Processing.
- Zhang, C., T. Tan, 2008, "Voice Disguise and Automatic Speaker Recognition," *Forensic Sciences International* Vol. 175: 2-3, March 2008, pp. 118-122.

This page intentionally left blank.

## Appendix B Acronyms

ANSI	American National Standards Institute
API	Application Programming Interface
ASR	Automatic Speech Recognition
BAT	Biometrics Automated Toolkit
BISA	Biometrics Identification for Secure Access
CESG	Computer Electronic Security Group
CLSP	Center for Language and Speech Processing
CORVET	Coordinate Operational Resources for the Voice Exploration Technology
DET	Detection Error Tradeoff
DHS	Department of Homeland Security
DoD	Department of Defense
EBGM	Elastic Bunch Graph Matching
FAR	False Acceptance Rate
FASR	Forensic Automatic Speaker Recognition
FBI	Federal Bureau of Investigation
FEARID	Forensic Ear ID
FMR	False Match Rate
FNMR	False Non Match Rates
FRGC	Face Recognition Grand Challenge
FRR	False Reject Rate
FRVT	Face Recognition Vendor Tests
FTA	Failure to Acquire
FTE	Failure to Enroll
GMM	Gaussian Mixture Model
HD	Hamming Distance
HIIDE	Handheld Interagency Identification Device
HMM	Hidden Markov Model
IBG	International Biometrics Group
IC	Intelligence Community
ICAO	International Civil Aviation Organization
IREX	Iris Recognition Exchange
IRIS	Iris Recognition Immigration System
ITIRT	Independent Testing of Iris Recognition Technology
JHU	Johns Hopkins University
LDC	Linguistic Data Consortium
LED	Light Emitting Diodes
LID	Language Identification
LR	Likelihood Ratio

MBARK	Multimodal Biometric Acquisition Research Kiosk/Multimodal Biometric Application Research Kit
NIR	Near Infrared
NIST	National Institute of Standards and Technology
NSTC	National Science and Technology Council
OEM	Original Equipment Manufacturer
ONR	Office of Naval Research
PCA	Principal Component Analysis
PPI	Pixels Per Inch
ROC	Receiver Operating Characteristic
ROI	Regions of Interest
SABER	State of the Art Biometrics Excellence Roadmap
SVM	Support Vector Machine
TSWG	Technical Support Working Group
UBM	Universal Background Model
U.S.	United States
UAE	United Arab Emirates
UK	United Kingdom

## Appendix C Revisions

Date	Revision	Notes
June 30, 2008	v1.1	Document completion
October, 2008	v1.2	Additional review and editing for public release
March 27, 2009	v1.3	<p>Corrected formatting errors to section 2 figures 2-3. “IRIS06” changed to “IRIS06” (page 2-24). Corrected section headers for R&amp;D (2.6) and Human Issues (2.7).</p> <p>Modified two paragraphs on page 2-2 to clarify current availability of multi-vendor products.</p> <p>Clarified iris only polar format (page 2-28).</p> <p>Table 2-1, added annotation that some stand-off iris provide face image(s).</p> <p>Incorporated Appendix A into body of iris chapter. (page 5-11) Corrected description for intended use of “Type 99” records.</p>

This page intentionally left blank.