WVU Biometric Data Collection Projects

Jeremy M. Dawson
West Virginia University
Statler College of Engineering and Mineral Resources
Lane Dept. of Computer Science and Electrical Engineering

The results presented herein were generated by work performed under **FBI** contract numbers POA8A806585, POA9A906229, POA2A201589, DJF-13-1200-A-0000651, POA1A103721, POA2A201564, DJF-13-1200-A-0000625, and DJF-14-1200-A-1115904, **ONR** contract numbers N00014-12-1-0931 and N00014-08-1-0895, **ManTech** contract numbers 25922 (2010-IJ-CX-K024) and MASI-14-WVURC-F-828-29156, **DHS** contract number IIP-0641331 **DOJ** contract number 2010-DD-BX-0161, as well as awards from the **Center for Identification Technology Research (CITeR)**.

Outline

- **Why do we collect biometric data?**
  - **Create test dataset**
  - **Sensor evaluation/ Interoperability**
  - **Human Factors**
  - **Explore new modalities**
- Common items associated with collection preparation
  - IRB: things-to-know
  - Worker training
- Dataset utility & longevity
  - Lab vs. operational environment
  - Sensor suite selection
  - Data storage/management

Overview

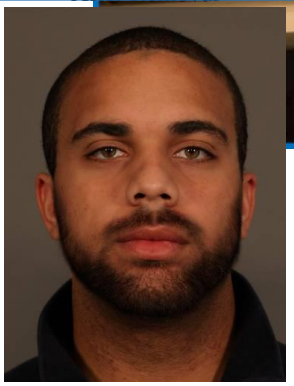
Since 2008, WVU has performed large, medium, and small scale biometric data collection projects to accomplish the following goals:

- Build research datasets to train humans, algorithms, and systems
- Evaluate prototype sensor operation
  - Data interoperability (e.g. contactless vs. contact-based fingerprint sensors)
  - Human factors
- Explore the application of new modalities/methods
  - Short-wave infrared (SWIR) imagers for cross-spectral facial identification
  - Biometrics in difficult environments
  - Bimolecular biometrics

FBI Collections – Test Datasets

Lab Collections:

- 2008-Present; collected to:
  - "Build robust dataset for future applied research efforts, including prototype device and algorithm development"
  - "Develop training materials, and in proficiency testing and competency testing"
- Primarily face (stills & video), fingerprint, and iris
  - *2008:* large latent collection (10-print, palm, major case, latent impressions)
  - *2009:* added non-ideal face (expressions, digital disguise), archival photos
  - *2012:* added hand geometry, 'eyes closed' face images, emphasis on repeat visit 1-2 months later
  - *2013:* added unscripted voice, audio booth, SWIR face
- **Total of 4532 datasets, 550 repeat visits and counting**



Facial Hair Before

After Addition

FBI Collections – Test Datasets

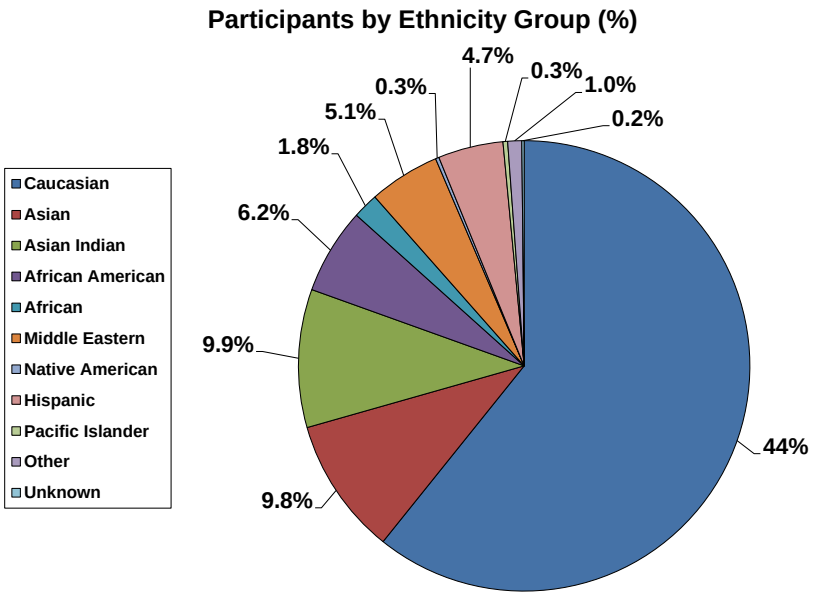
‘Twins Days’ Collections:

- 2010-Present; collected to:
  - “Build robust dataset for future applied research efforts, including prototype device and algorithm development”
  - “Develop training materials, and in proficiency testing and competency testing”
- Limited area (10’x10’ tents), limited power
- Environmental factors: heat, rain, sun angle
- Primarily face (stills & video), fingerprint, and iris
  - *2013* added twin audio collection
- **Total of 1736 datasets, 197 repeat visits and counting**

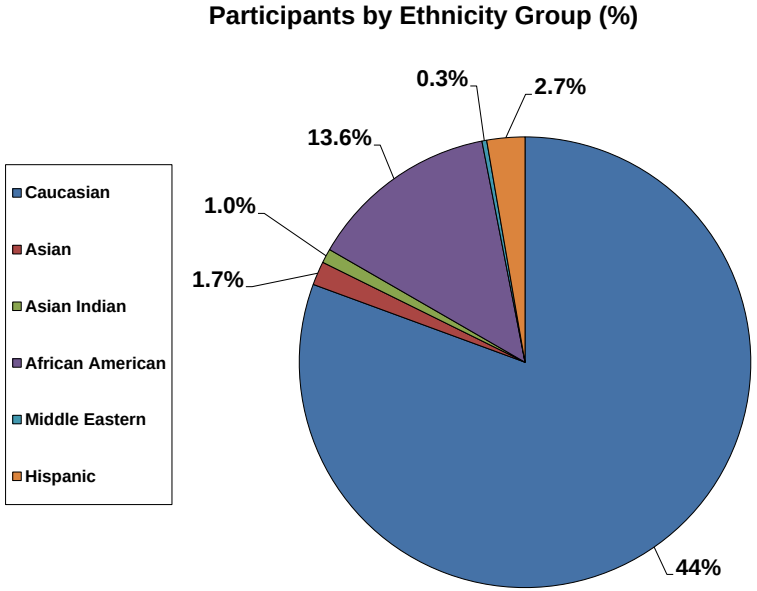
FBI Collections – Test Datasets

Demographic Variance:

2012 Lab



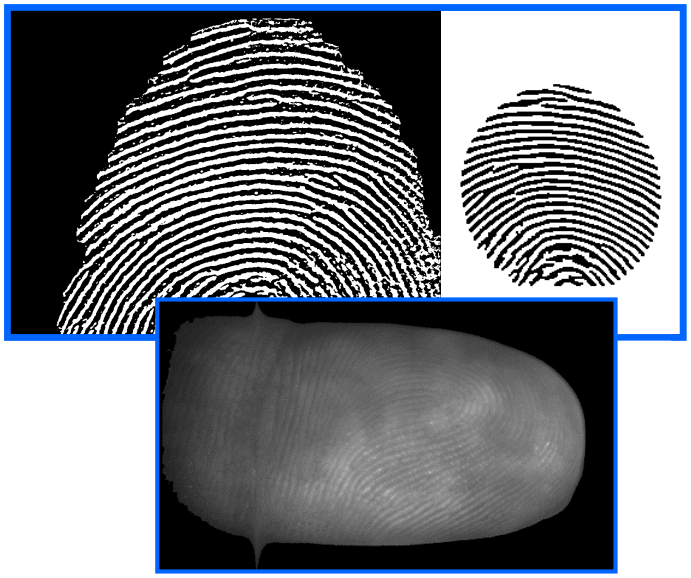
2014 Twins



DOJ & DHS Collections – Sensor Interoperability & Human Factors

3D & Contactless Fingerprints:

- 2010 DHS Collection – Goal: Evaluate data collected from two prototype non-contact fingerprint capture systems
- *Sensors:* Flashscan3D single finger and GE 4-finger phase I prototypes
- Ground truth: Crossmatch Guardian, 10-print cards
- **122 participants, 19 repeats**



- 2012 & 2015 ManTech/DOJ Collection – Goal: Evaluate data interoperability and perform qualitative assessment of operation
- *Sensors 2012:* Crossmatch Guardian R2, Crossmatch SEEK II, i3 DigID Mini, L1 Touchprint 5300, TBS 3D-Enroll (commercial; Series 11), FlashScan3D D1 single-finger (V2), FlashScan3D D4 four-finger (V1)
- *Sensors 2015:* Crossmatch Guardian R2, Crossmatch SEEK Avenger, NG BioSled, Moprho Ident, Morpho Finger-on-the-Fly, ANDI On-the-Go, Flashscan D1 (production), IDAir InnerID (iPhone app)
- Ground truth: 10-print cards (scanned)
- **500 participants for 2012, 400 planned for 2015**

L. Lugini, E. Marasco, B. Cukic, and J. Dawson, "Removing Gender Signature from Fingerprints," in Proc. Biometrics & Forensics & De-identification and Privacy Protection (BiForD), May 2014, Croatia.

DOJ & DHS Collections – Sensor Interoperability & Human Factors

3D & Contactless Fingerprints:

DOJ & DHS Collections – Sensor Interoperability & Human Factors

Long-Range 3D Face:

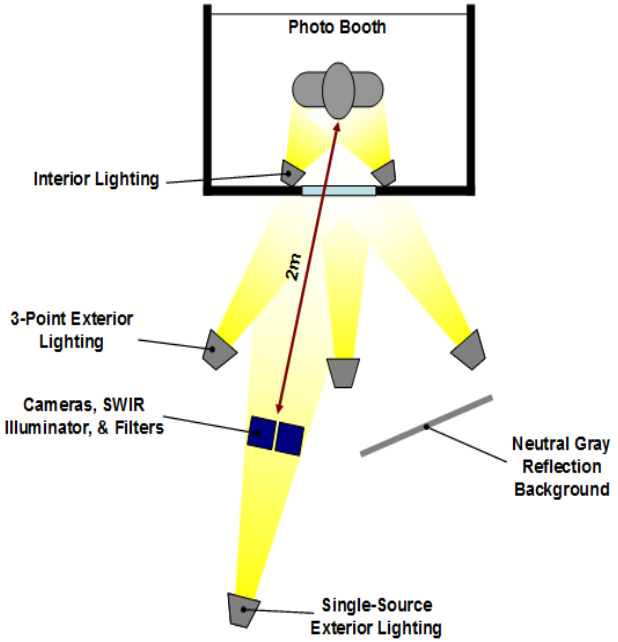
- 2012 & 2013 ManTech/DOJ Collection – Goal: Evaluate data interoperability and perform qualitative assessment of operation
- 2012 Sensors: Stereovision binoculars prototype (V1), Sony DEV 5 digital recording binoculars
- 2013 Stereovision binoculars prototype (V2)
- Ground Truth: Digital SLR camera
  - Outdoors: Canon 5D MkII digital SLR camera with a Canon EF 800mm f/5.6L IS USM Autofocus Lens
  - Indoors: Canon 5D Mk II digital SLR camera with a Canon EF 70-200mm (f/2.8, image stabilized) lens, standard 5-pose mugshots
- **100 participants each, 2012 & 2013**

ONR Collections – SWIR Biometrics

2011-2013 Face in Challenging Environments

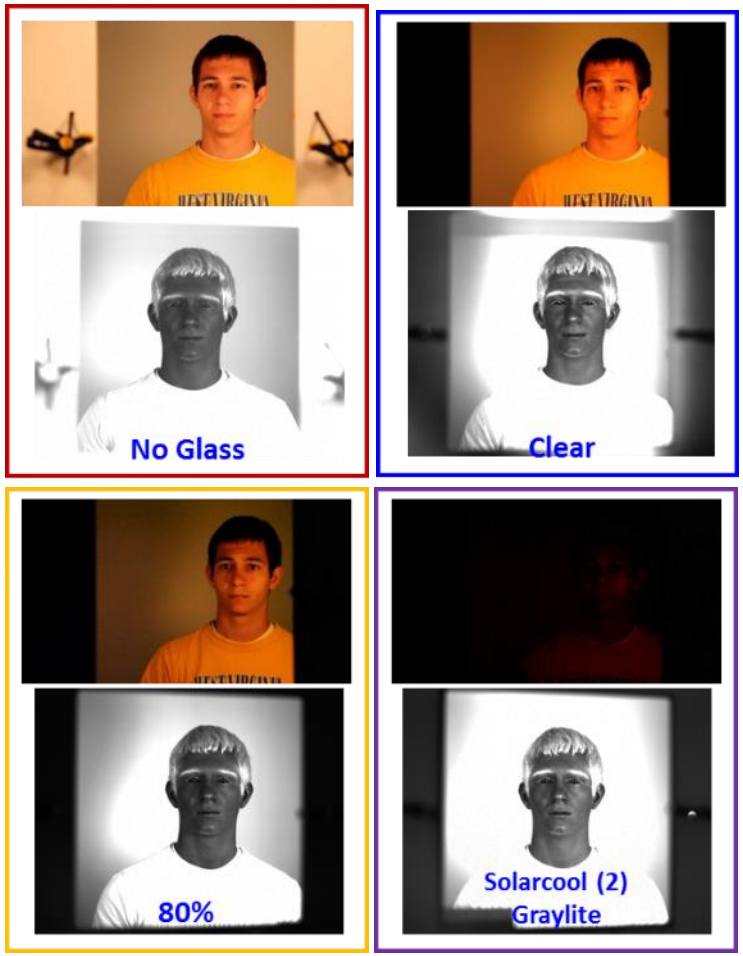
- Goal: Develop algorithms for cross-spectral face matching at night and obstructed by tinted materials
- SWIR imager, active (1150nm laser source), tungsten, and natural illumination
  - 1050-1650nm wavelengths, filtered at 100nm bands
- Phase I: Indoor collection under varying lighting conditions
  - **138 participants**
- Phase II: Outdoors collection under environmental lighting, both day and night
  - **200 participants**

J. Ice, N. Narang, C. Whitelam, N. Kalka, L. Hornak, J. Dawson, and T. Bourlai, "SWIR Imaging for Facial Image Capture Through Tinted Materials," Proc. SPIE, 8353, p. 83530S, 2012.



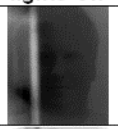
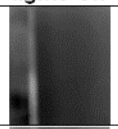
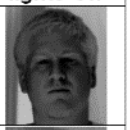
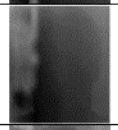
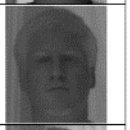
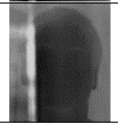
ONR Collections – SWIR Biometrics

2011-2013 Face in Challenging Environments

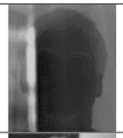
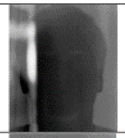
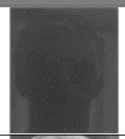
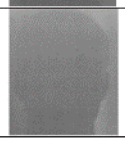


Sample Indoor Images

Sample Daytime Images

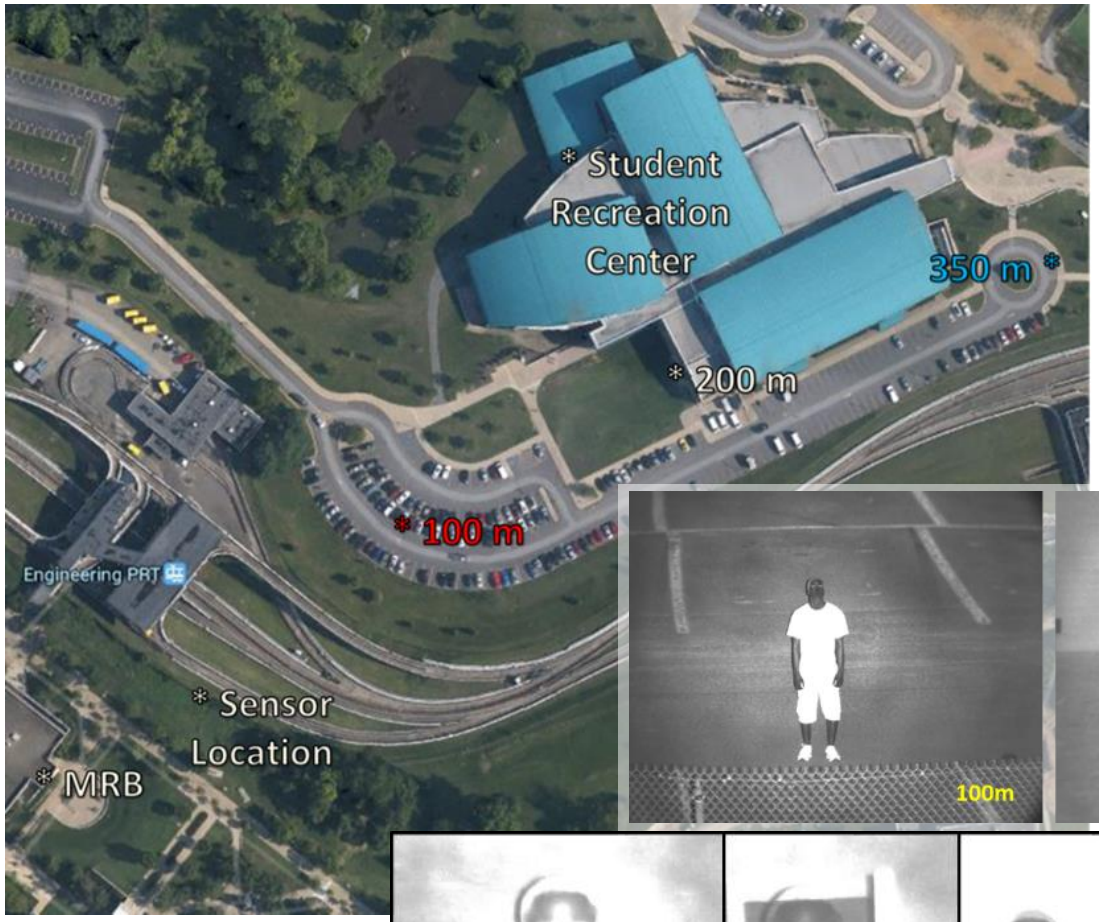
	Ray Ban Sunglasses		Tinted Glass Panel On		Tinted Glass Panel Off	
	Participant facing sun	Participant facing sway From Sun	Interior booth lights On	Interior booth lights Off	Interior booth lights On	Interior booth lights Off
1150nm						
1350nm						
1550nm						
SWIR Illuminator						

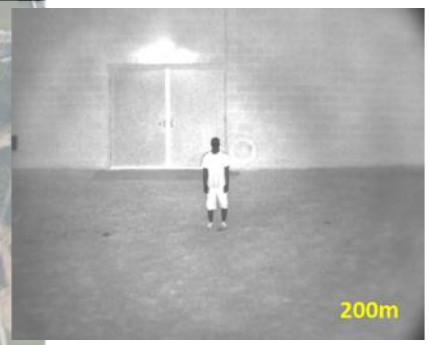
Variations in Image Quality with Varying Collection Conditions (all images @ 1550nm)

	Glass On, Interior Lights Off, SWIR Illuminator On	Glass On, Interior Lights Off, No SWIR Illuminator	Ray Ban Sunglasses, Facing Away From Sun	Ray Ban Sunglasses, Facing Toward Sun	Ray Ban Sunglasses, No SWIR Illuminator	Ray Ban Sunglasses, SWIR Illuminator
Day						
Overcast						
Night (street lamp illumination)						

ONR Collections – SWIR Biometrics

2013 Long-Range SWIR Face





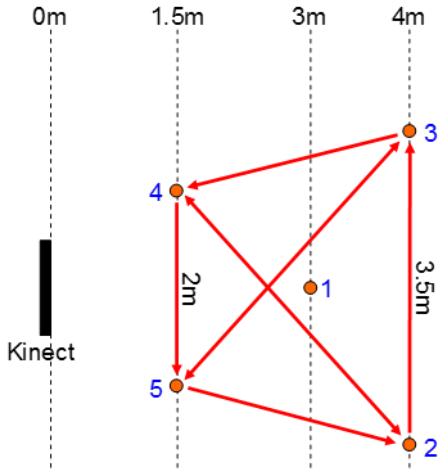
- Performed in partnership with WVHTCF (Fairmont, WV) using TINDERS imager
- SWIR images captured at 100, 200, & 350 meters
- Faces captured behind tinted glass at each location
- **104 participants**

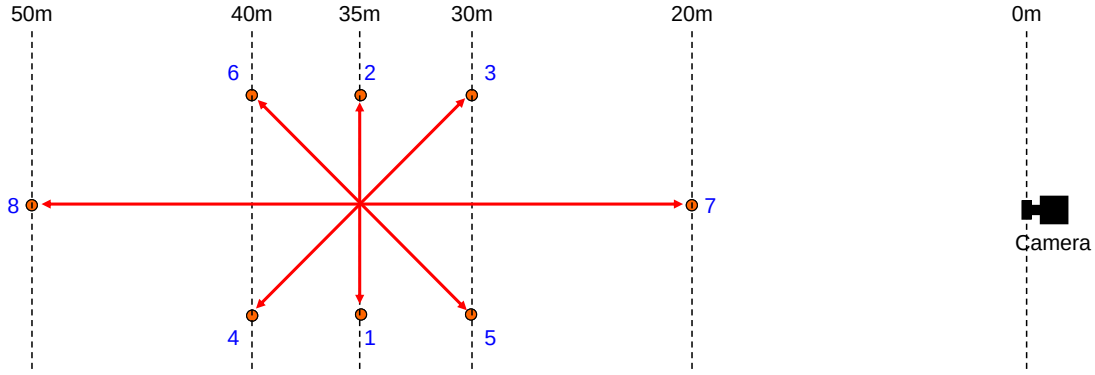
ONR Collections – SWIR Biometrics

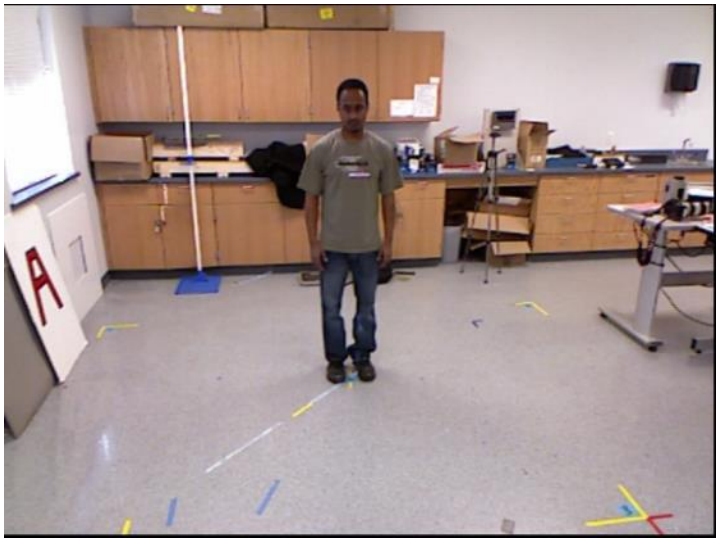
2011 Gait & Body Measurements

B. DeCann, A. Ross, and J.M. Dawson, "Investigating gait recognition in the short-wave infrared (SWIR) spectrum: dataset and challenges," Proc. SPIE 8712, Biometric and Surveillance Technology for Human and Activity Identification X, 87120J, May 31, 2013.

- Gait video captured with MS Kinect (indoors, short range) and SWIR camera (outdoors, long range)
- Body measurements recorded as well
- **157 participants**





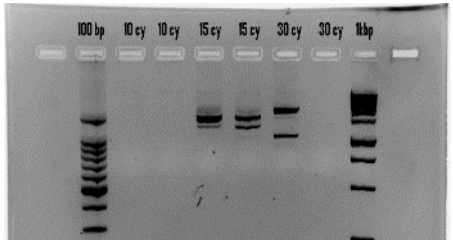


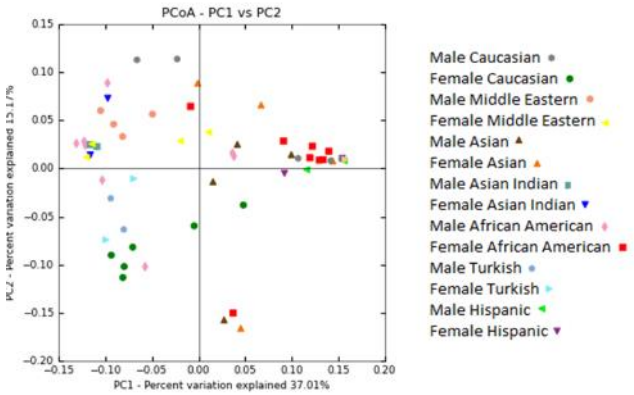
CITeR/DOJ Bimolecular Biometrics

DNA & Face Images

- 5-pose face images and blood samples
- **250 participants**
- 20 sequenced genomes

A.B. Holbert, H.P. Whitelam, L.J. Sooter, J.M. Dawson, and L.A. Hornak, "Evaluation of Hand Bacteria as a Human Biometric Identifier," in Proc. IEEE 14th International Conference on BioInformatics and BioEngineering, pp. 83-89, November 10-12, Boca Raton, FL (2014).





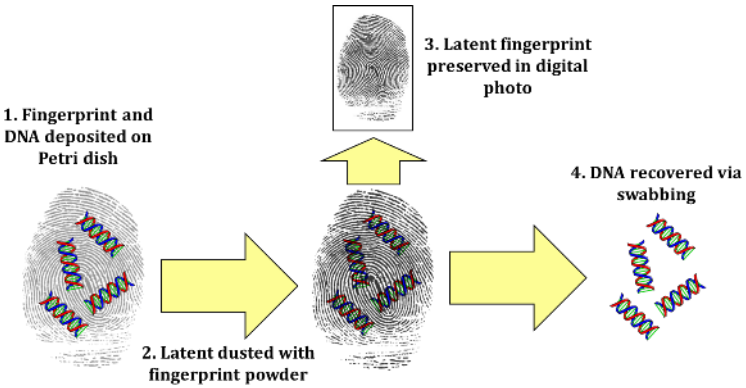
56 Samples compared by Gender and Ethnicity

Hand Bacteria

- Hand swabs from right/left hands
- **250 participants**
- 56 samples isolated and sequenced (16s rRNA)

Touch DNA & Latent Fingerprints

- Latent impression on plastic
- Touch DNA recovered from fingerprints
- **35 participants**



NetBio Instrument Validation

- 5 minute buccal swab; performed in high-traffic areas on campus
- Two 2-day collections; **600 collected first collection, 200 second collection**

Outline

- Why do we collect biometric data?
  - Create test dataset
  - Explore new modalities
  - Sensor evaluation/ Interoperability
  - Human Factors
- **Common items associated with collection preparation**
  - **IRB: things-to-know**
  - **Worker training**
- Dataset utility & longevity
  - Lab vs. operational environment
  - Sensor suite selection
  - Data storage/management

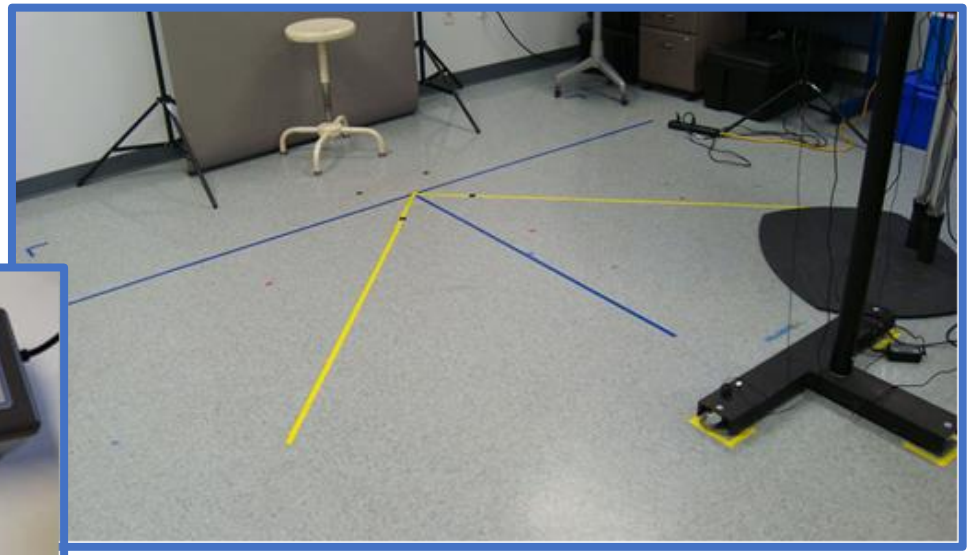
IRB Protocol Review – Things to Know

- Most biometric collections are considered minimal risk studies, however…
  - Prototype ‘devices’ necessitate full-board review, inclusion of safety documentation in protocol (typically exempt from FDA certification since assembly of COTS components)
  - Human DNA collection may require additional biosafety protocol(s), necessitate full board review
- If planned, data release or sharing needs to be explained clearly in consent form
- Collection of physical metadata (height, weight, etc.) does not require HIPAA forms if not correlating to participant health

Worker Training

- Easy-to-use sensor interfaces crucial to data consistency

- Standard operating procedures essential

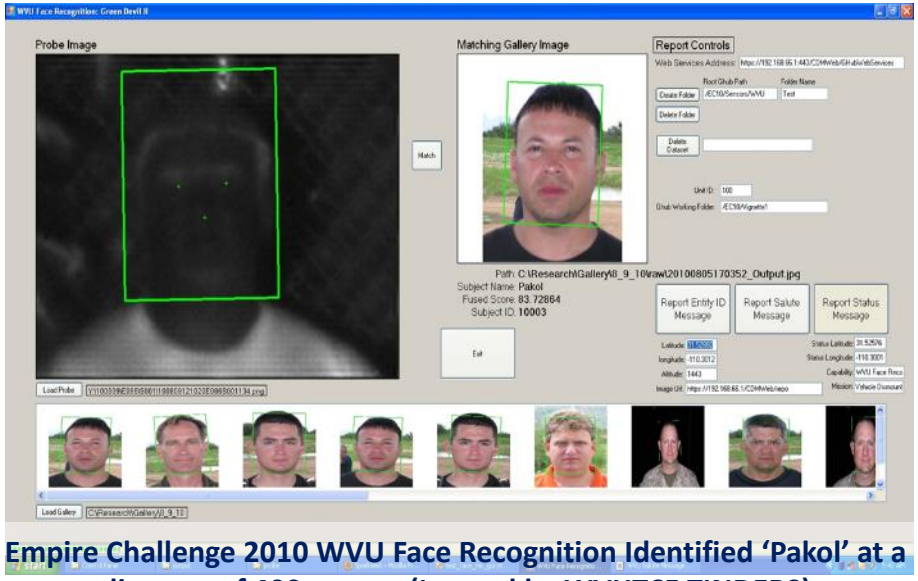


Outline

- Why do we collect biometric data?
  - Create test dataset
  - Explore new modalities
  - Sensor evaluation/ Interoperability
  - Human Factors
- Common items associated with collection preparation
  - IRB: things-to-know
  - Worker training
- **Dataset utility & longevity**
  - **Lab vs. operational environment**
  - **Sensor suite selection**
  - **Data storage/management**

Lab vs. Operational Conditions

- Laboratory settings allow for control over common variables impacting data quality (lighting, presentation, etc.)
  - Sometimes results in data that is “too good”
  - Some quality variance due to sensor variance, operator habits; helpful to track operator/station IDs
- Operational conditions pose challenges to algorithms developed solely on lab data
  - Distance, environmental factors (darkness, weather), lack of enrollment opportunity, subject cooperation, collection speed, etc.



Empire Challenge 2010 WVU Face Recognition Identified ‘Pakol’ at a distance of 400 meters (Imaged by WVHTCF TINDERS)

Drastic change in grass height over course of exercise; occluded key features needed for gait recognition

Sensor Selection

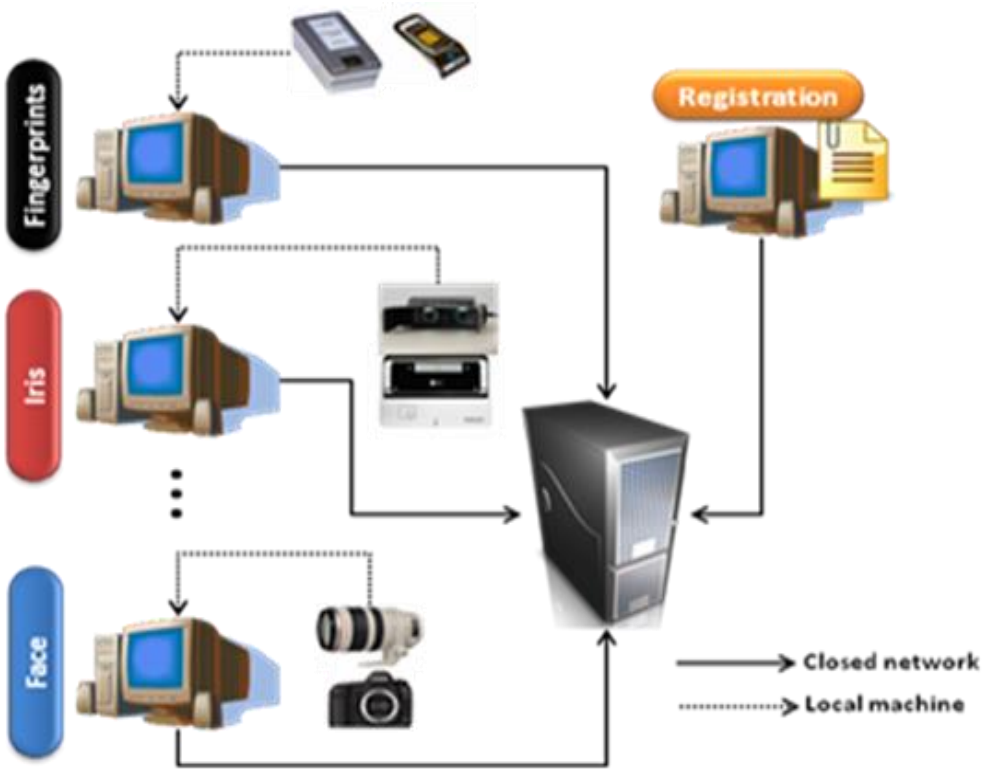
- Sensor technology continually improving/updated
- Legacy data may still see widespread use
- Necessitates co-collection of data from new and old devices

Integrated Biometrics Sherlock – TFT technology

Sarnoff IOM – High-res Face + Iris

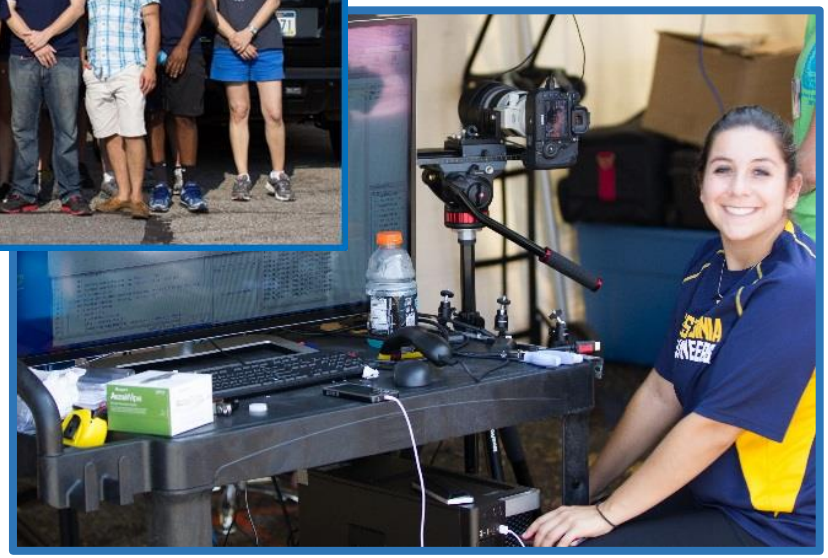
Data Storage & Management

- Data storage needs can grow quickly
  - 2009 FBI collection – **1.2TB**
  - 2012 FBI collection – **3.5TB**
  - 2013 FBI Collection - ??? Audio files are **5GB per participant**



- Should data be kept indefinitely?
  - Sensors may no longer be relevant
  - IRB may require limits on longevity
- Does your data require a release policy?
  - IRB may require release plan if data will be shared
  - Staff may be needed to maintain release requests

Thank You!



QUESTIONS?

- FBI datasets are available upon request; contact Joey Newell (Joey.Newell@ic.fbi.gov)
- ONR SWIR data availability is contingent upon sponsor approval after project conclusion (2015)
- Other dataset inquiries can be directed to Jeremy.Dawson@mail.wvu.edu