Lessons from Collecting a Million Biometric Samples

Patrick J. Flynn
Kevin W. Bowyer
University of Notre Dame

P. Jonathon Phillips
National Institute of Standards and Technology

Setting the Stage

- Largest face dataset 2000
  - FERET 14,126 images; 1199 people

- FRVT 2002 Performance
  - Highly controlled frontal face images
 - FRR = 0.20 at FAR = 1 in 1000

Motivation

Face Recognition

- Effects of 1-2 year face aging
- Fat dataset design
  - Large number of images of each person

- Improve performance by order of magnitude
  - FRR = 0.02 @ FAR = 1 in 1000
  - Controlled frontal face recognition
 - Face Recognition Grand Challenge (FRGC)

A Decade Long 2002-2012

- **986,246** enrolled biometric samples plus boutique collections

- 2002-2006
  - Improving Face and Benchmarking Iris
  - FRGC, FRVT 2006, ICE 2005 and 2006
- 2007-2010
  - Face and Iris Portals, and Video
  - MBGC and MBE 2010
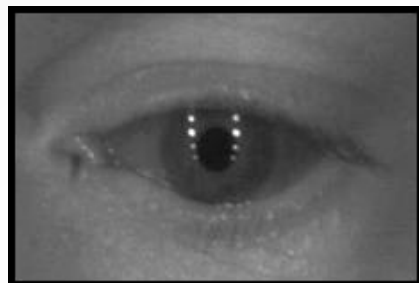- 2011-2012
  - The Point and Shoot Face Recognition Challenge
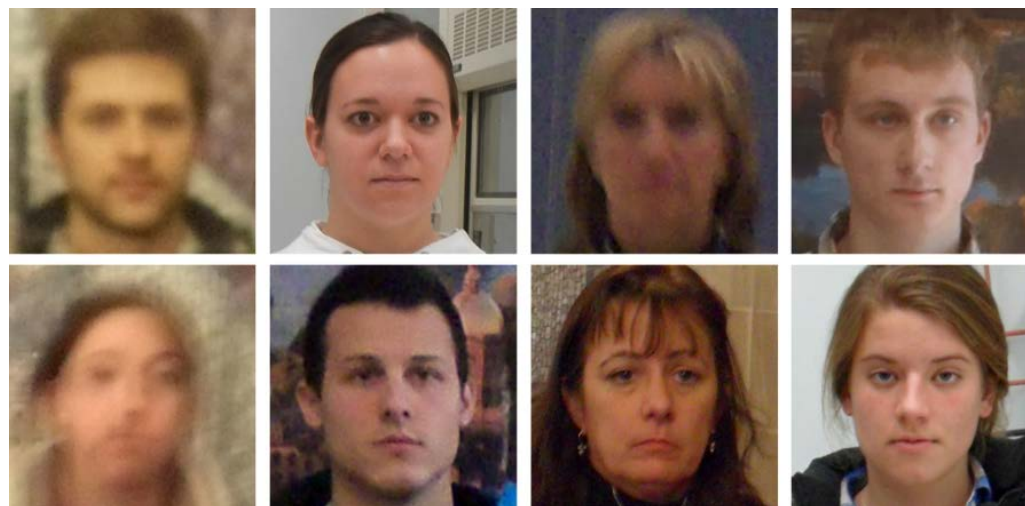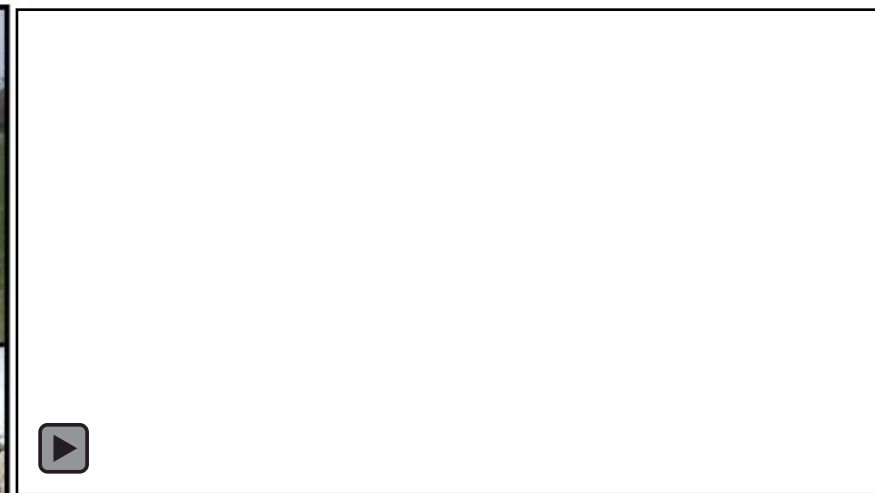
Samples Collect by Academic Year
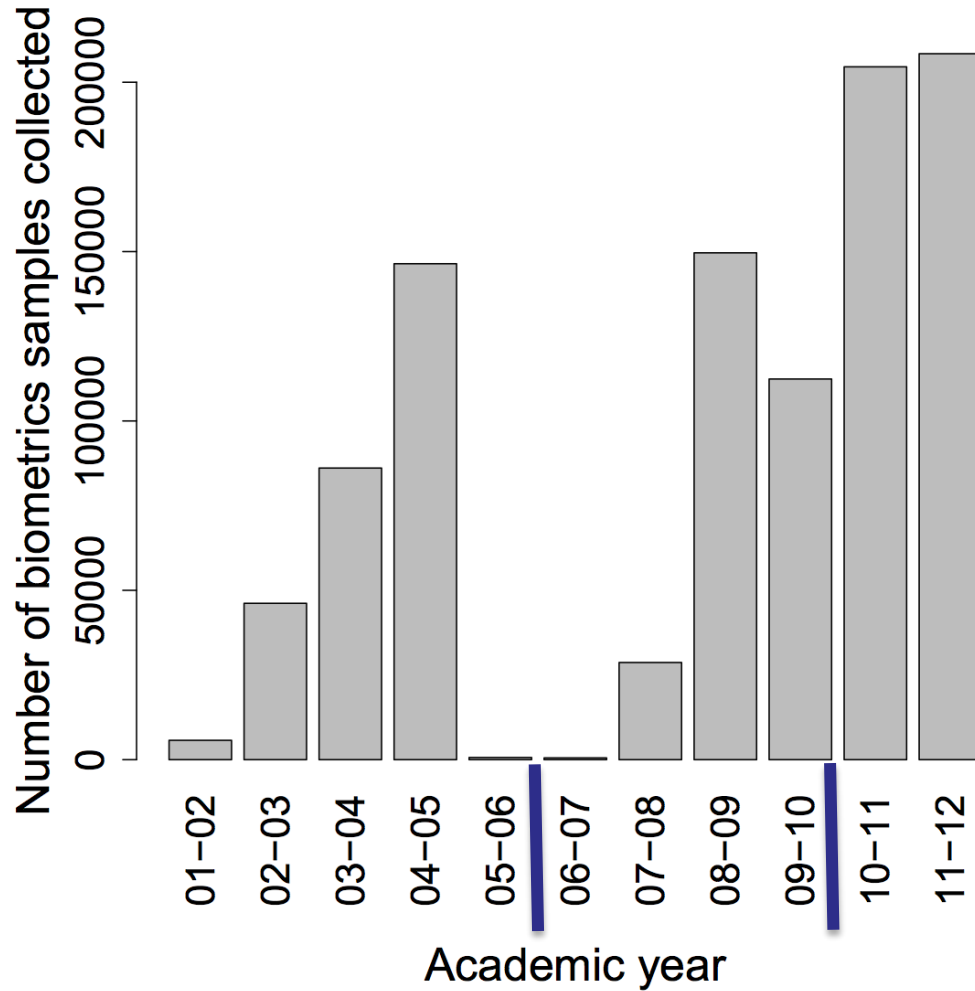
Examples 2002—2006
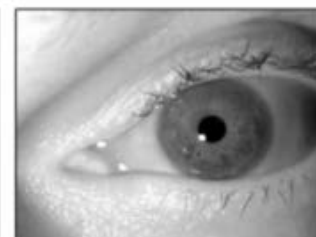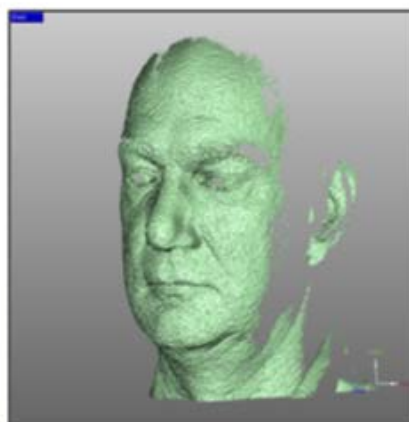
Examples 2007—2010

Examples 2011—2012

"You want to do WHAT?"

- Managing and distributing identifying data without running afoul of your research sponsor, your IRB, or your Office of Counsel

- Collect, groundtruth, annotate, organize, and distribute without running afoul of…

- Quickly access basic information about the dataset without…

General Approach to Sampling

- "Subject session": all data collected from a subject when they show up in the lab during a week of acquisitions

- Size of "subject session" has increased DRAMATICALLY with time...

- … hence, overall data "burden" has increased (collecting >100GB per acquisition week, 10 weeks/year)

100 GB per week... what?

- We collected images and videos of people's faces and irises and scripted "activities" using many consumer and specialized cameras
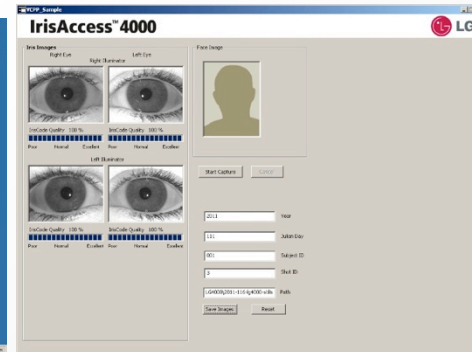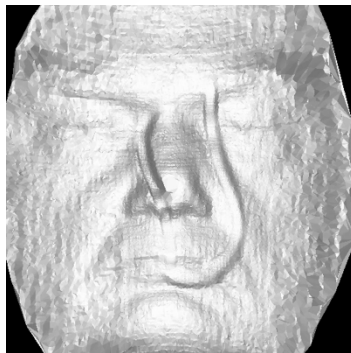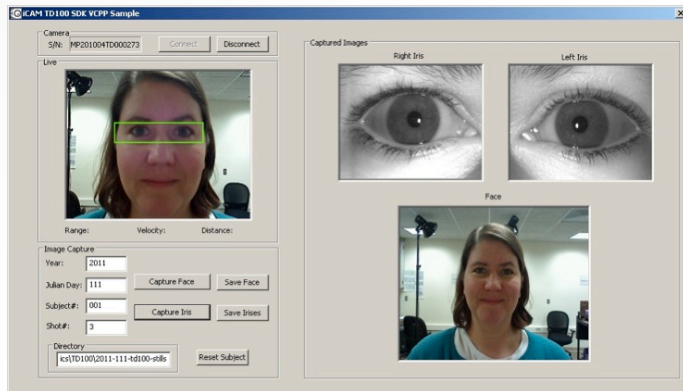
An acquisition week

- Collected data for 18 hours spread over Tuesday-Thursday

- Three locations (studio and two other locations with uncontrolled illumination)

- 50 undergraduate operators, 15 GRA operators, etc.

- End-of-day backup of data from cameras and host computers to secure storage

After an acquisition week... enrollment.

- Ingested data into "BXgrid" portal/database/distributed file system
- Dozens of "validation" assignments to operators

Ground Truth

- Subject
  - Identity
  - Meta-data: Sex, race, YOB, eye color…

- Sample
  - Identity
  - Modality
  - Date
  - Sensor
  - Location
  - Eye coordinates
  - …

Bxgrid

- Need infrastructure to keep track and organize 1 million items

- Bxgrid: Database-backed, distributed-file-storage-using, Web-front-ended biometric data management system

Browse Source Datasets:

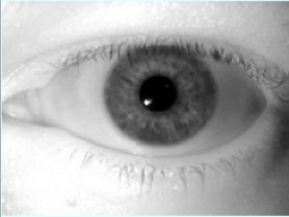|	
1. Face Images (bxgrid)	254642 records
2. Face Videos (bxgrid)	33808 records
3. Iris Images (bxgrid)	350201 records
4. Iris Videos (bxgrid)	11753 records
5. 3D Face Scans (bxgrid)	13210 records
6. Environments (bxgrid)	63 records
7. Illuminants (bxgrid)	27 records
8. Sensors (bxgrid)	49 records
9. Stages (bxgrid)	49 records
10. Subjects (bxgrid)	2922 records
11. Collections (bxgrid)	54 records
12. Metadata Log (bxgrid)	363597 records

BXgrid

Successes

- Deployable face recognition algorithms
  - FRGC->FRVT 2006->MBE 2010-> FBI Next Generation Identification
- First independent benchmarking of iris
  - ICE 2005 and 2006
- Characterization of 3D face performance
  - FRGC and FRVT 2006
- Fat datasets
  - Good, bad, and ugly
  - Covariate analysis

Successes

- Understanding Long-wave infrared
- Twins—face and iris
- Iris template aging
- Understanding properties of iris
  - Contacts, dilation, forensic comparisons
- Understanding performance on faces
  - Factors, covariates, and quality
- Fusing humans and algorithms
- Non-face identity cues
  - Still and video

FG2015: Successes and Future

- Wednesday 11:00-11:15

  Comparison of Human and Computer Performance Across Face Recognition Experiments

- Wednesday 16:15-17:15

  Report on the FG2015 Video Person Recognition Evaluation

- Thursday 11:00-12:20

  Panel "Promise and Perils of Found Data"

Conclusions

- 1 million biometric samples collected
  - Still, 3D and video face/person
  - Classical and at a distance iris
  - Long and near infrared face
  - Multi-modal
- Developed and assessed algorithms
- Basic scientific investigation of face and biometrics

Questions