FR 2002
VT 2002

FACE RECOGNITION

VENDOR TEST 2002

New Evaluation Statistics for Measuring the Performance of Biometric Technologies

Patrick Grother,

P. Jonathon Phillips, Ross Michaels

NIST
National Institute of Standards and Technology

FR VT 2002

Goals

- Assess performance on large real-world data sets
- Identify new promising approaches
- Measuring progress on hard problems
  - Pose variation
  - Images taken months/years apart
  - Video sequences

NIST National Institute of Standards and Technology

Team

- Conducted by NIST
- Principals
  - Dr. Jonathon Phillips, Test Director
  - Duane Blackburn
  - Mike Bone
  - Patrick Grother
  - Ross Micheals
- Principals from FERET and FRVT 2000
- Twenty people supporting






National Institute of Standards and Technology

Team

- **Supporters**
  - Counterdrug Technology Assessment Center
  - U.S. Customs Service
  - Department of Energy
  - Drug Enforcement Administration
  - U.S. Secret Service
  - Technical Support Working Group
  - Immigration and Naturalization Service
  - Canadian Passport Office
  - UK Biometrics Working Group
  - Australian Customs
- **Sponsors**
  - National Institute of Standards and Technology
  - Defense Advanced Research Projects Agency
  - Department of State
  - National Institute of Justice
  - Transportation Security Administration
  - Federal Bureau of Investigation

National Institute of Standards and Technology

Participants

- AcSys Biometrics
- Cognitec Systems GmbH
- C-VIS
- Dream Mirh Co.
- Electronic System Products
- Eyematic Interfaces
- Iconquest
- Imagis Technologies
- IRID
- Phoenix Systems Corporation
- VicarVision
- Viisage
- Visionics
- VisionSphere

Terminology

Rank	Value	
1	6.6	
2	5.3	
3	5.2	
4	4.8	
5	3.1	
6	2.7	
7	2.0	
8	1.6	

Probe (unknown image)

Match (same ID as probe)

Gallery (known images)

National Institute of Standards and Technology

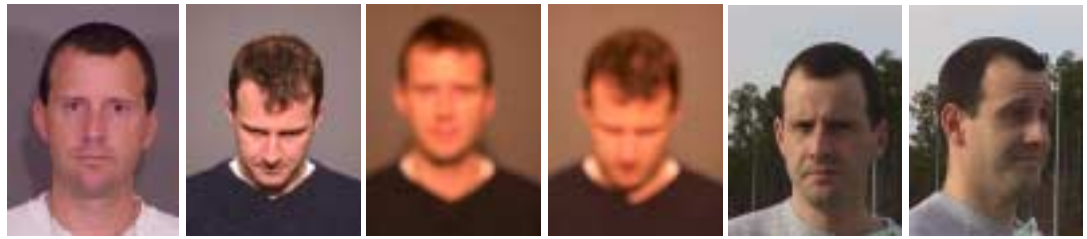
FR 2002
VT 2002

Human ID Evaluation Framework

- HEF allows measurement of biometric performance **empirically**:
  - Available to other biometrics (face, fingerprint, iris …)
  - Composite biometrics (face + fingerprint, face + gait …)
  - Off-line (live human subjects not required)
  - Repeatable (across many vendors)
  - Proctored (real ids hidden from vendors)
  - Very large numbers of subjects
  - FERET Protocol as previously in FERET, FRVT 2000
  - Suitable for
    - Identification, Verification, Watch List
    - Different scenarios

NIST
National Institute of Standards and Technology

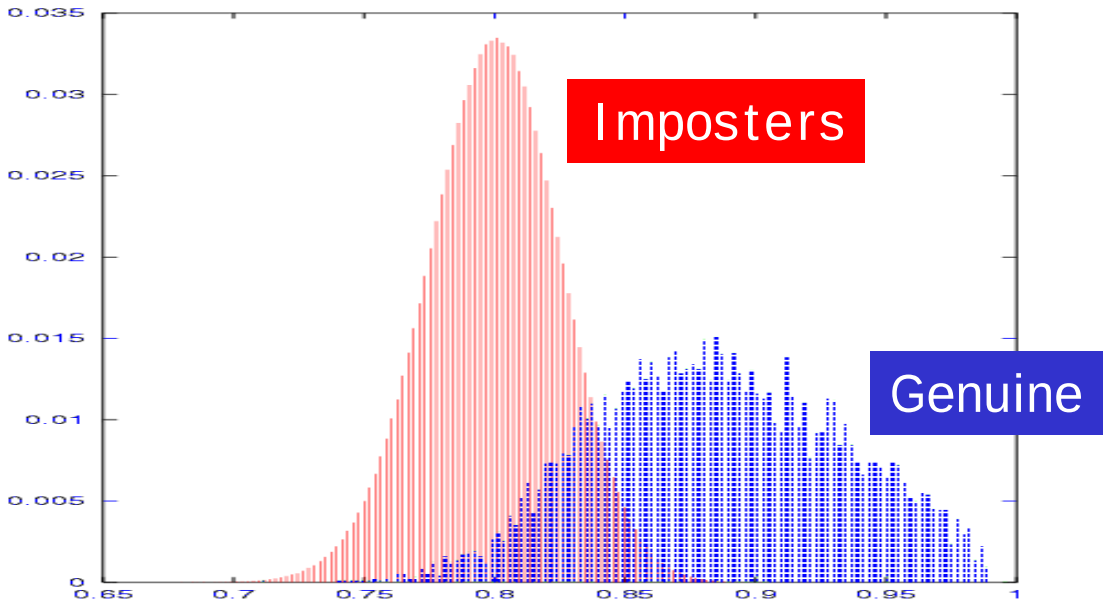
Human ID Evaluation Framework

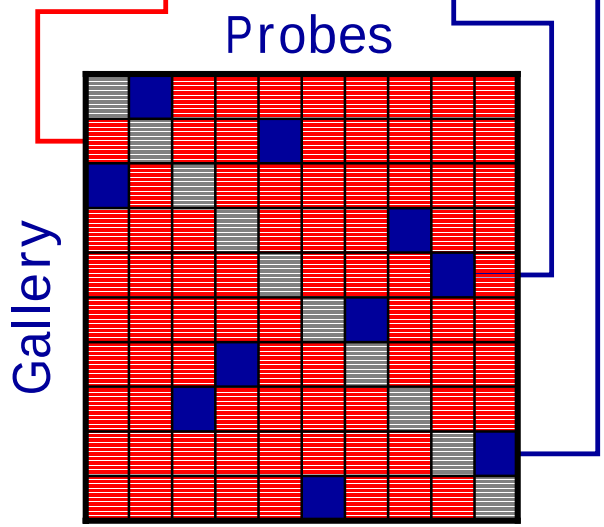


0.10	0.05	0.19	0.01	0.03	0.01
0.22	0.09	0.11	0.02	0.15	0.02
0.98	0.44	0.88	0.29	0.77	0.21
0.39	0.19	0.34	0.09	0.31	0.02

National Institute of Standards and Technology

Biometrics : The Problem





National Institute of Standards and Technology

FR VT 2002

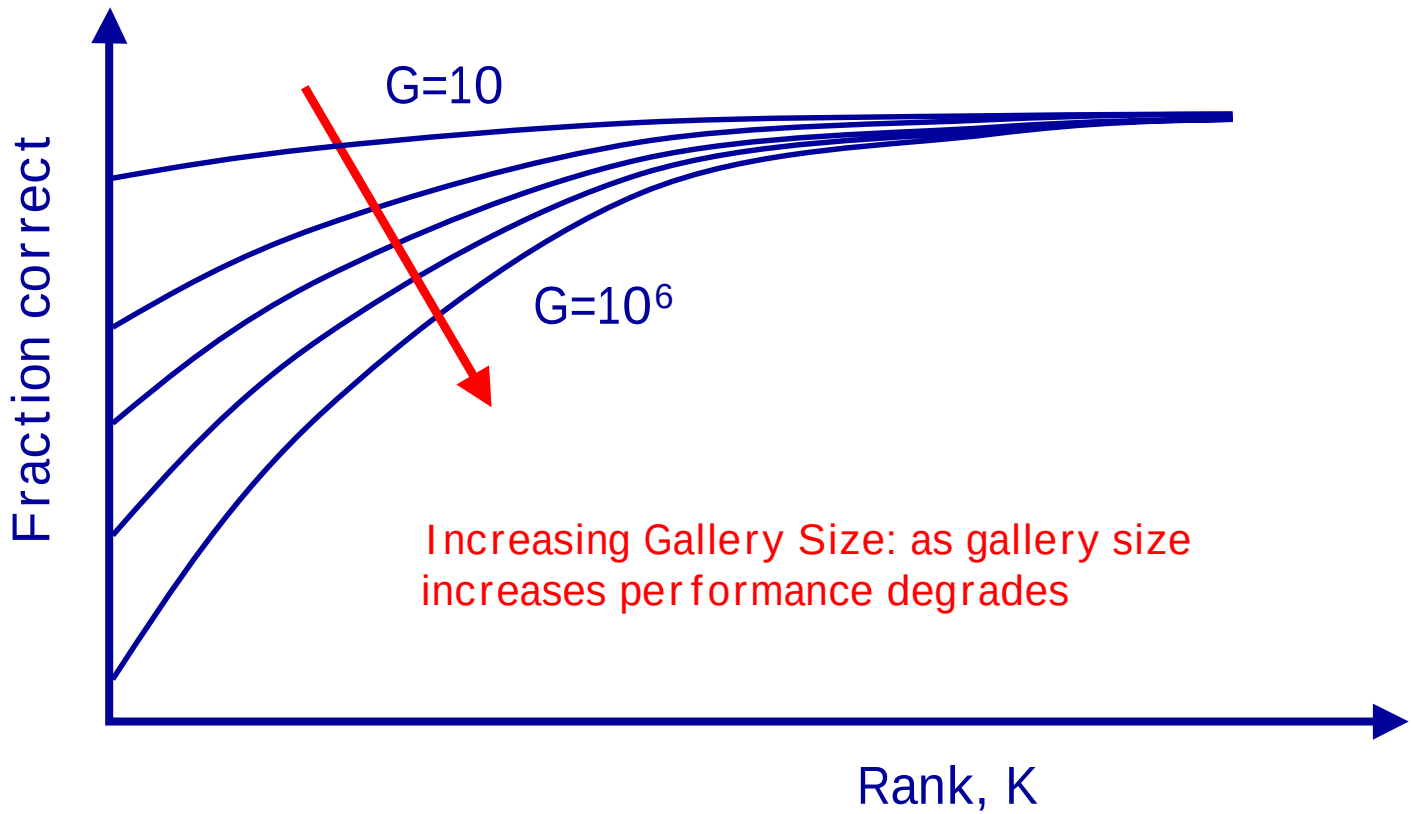
Biometric Uses

- Identification (1-N)
  - Who is that?
- Verification (1-1)
  - Is he who he claims to be?
- Watch list (1-N)
  - Is he on the watch list, and if so, who is he?

NIST National Institute of Standards and Technology

Identification (1-N)

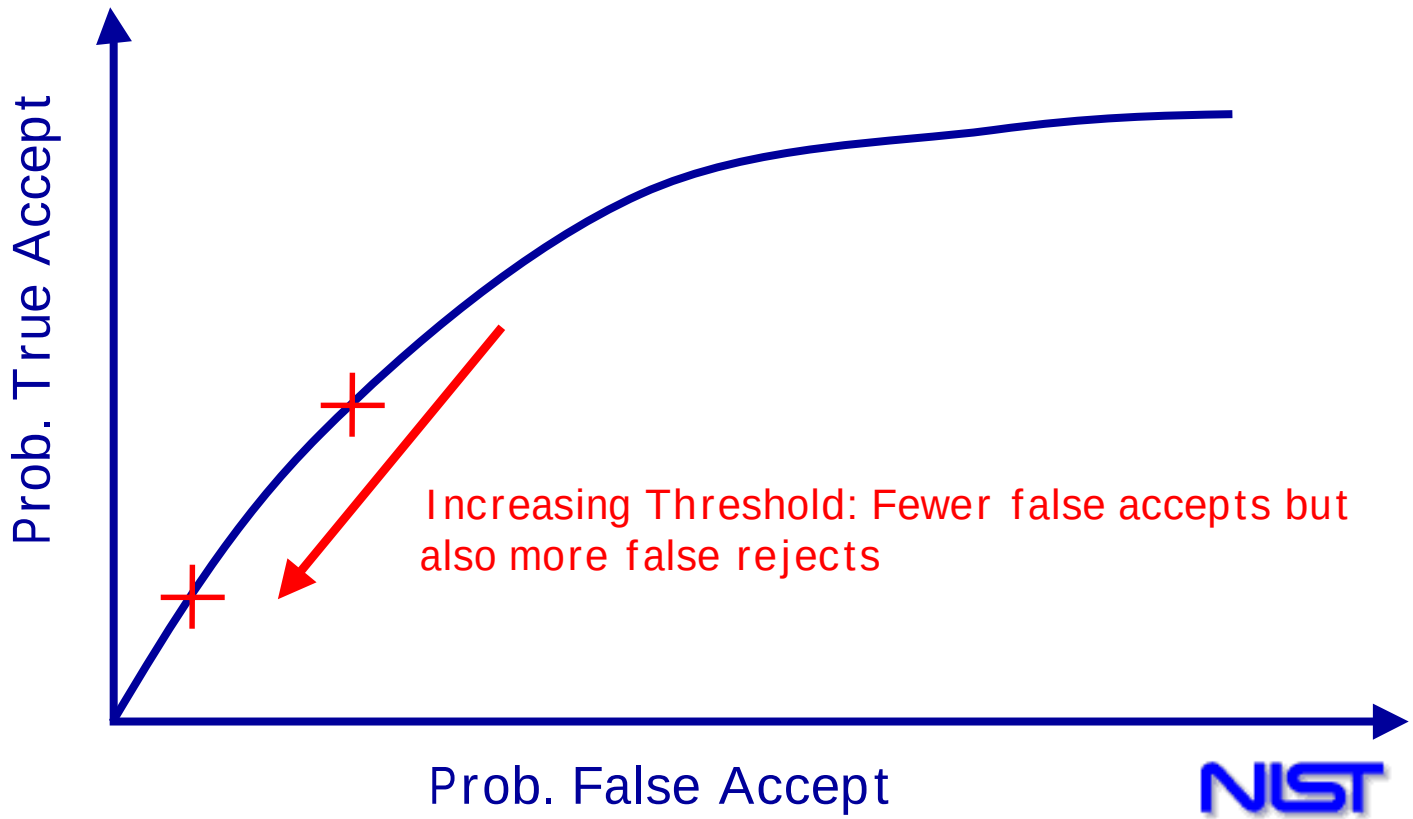
Operationally: Compare current image with all images.
How many are matched within top K guesses



National Institute of Standards and Technology

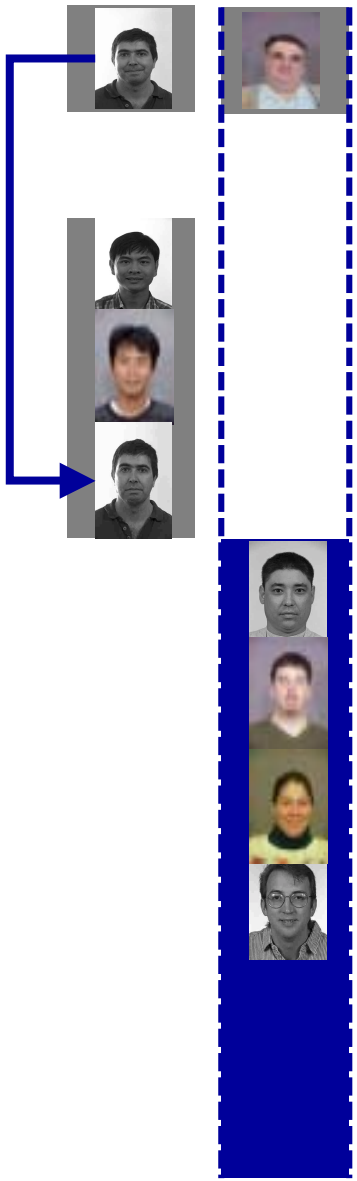
Verification (1-1)

Operationally: Compare current image with previous image.
Decision: Accept or reject a claim?
Vary Threshold: Tradeoff bad guys in vs good guys out

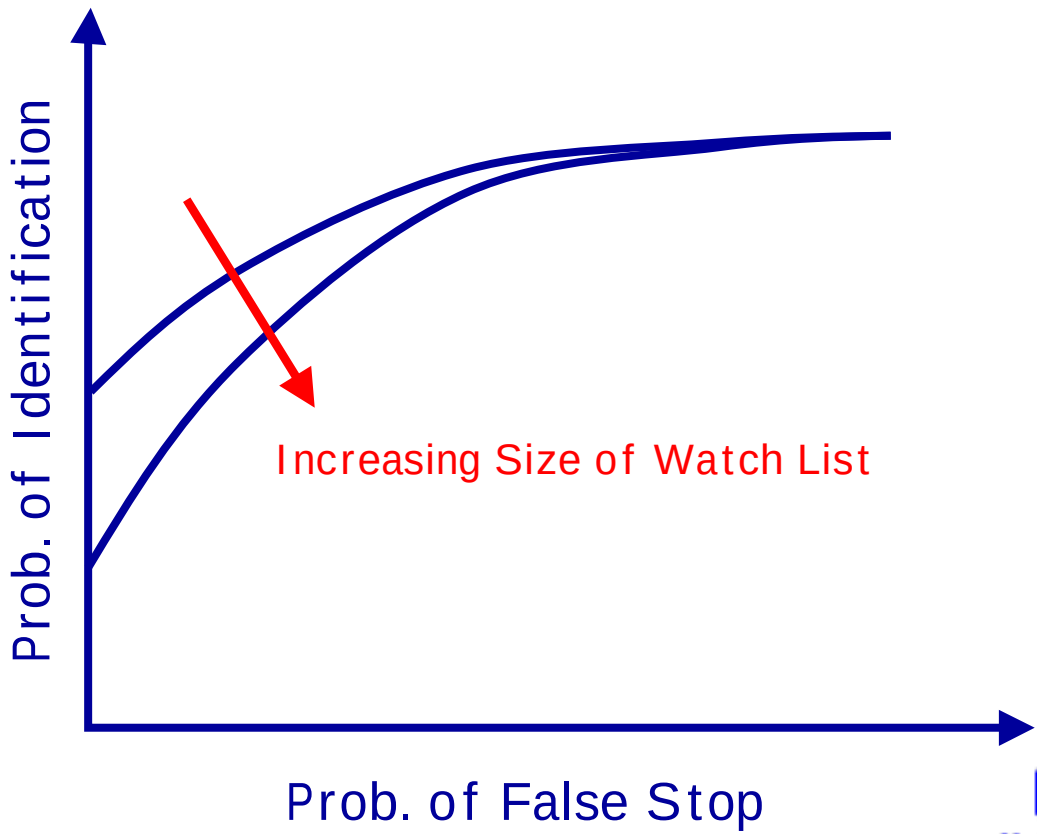


National Institute of Standards and Technology

Watch List (1-M < N)



Operationally: Compare current image all in watch list
Decision: Accept or reject a claim? If rejected, then who is it?
Threshold: Tradeoff stopping friends vs missing bad guys

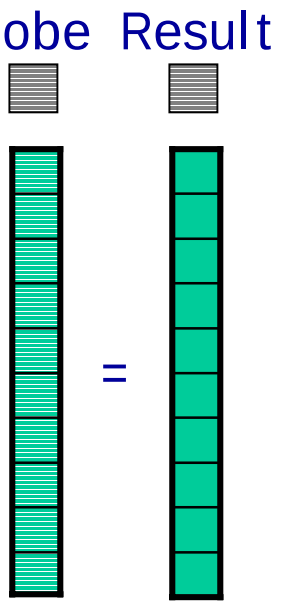


National Institute of Standards and Technology

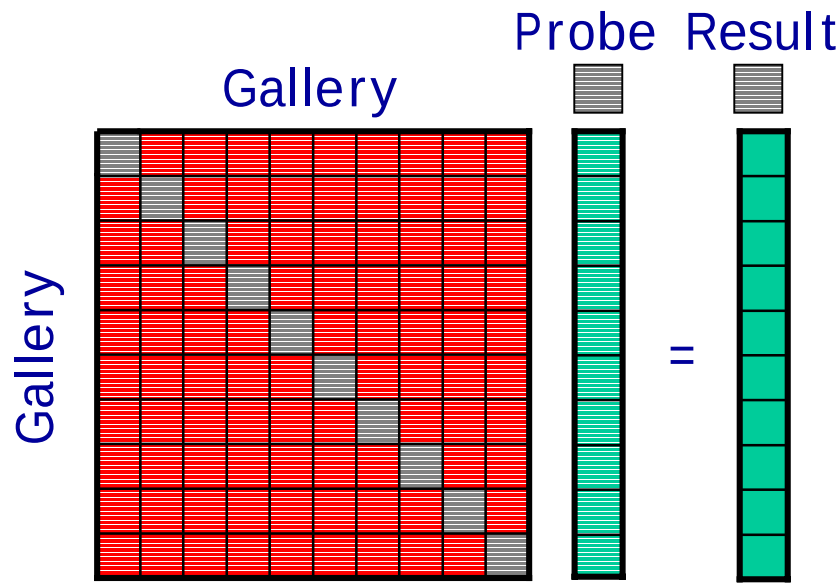
Normalization

Similarity element $s_{ij} = f(x_i, x_j)$ i.e. just a function of two images. But there is information in the other scores.

Option 1: Use information in the probe to gallery elements



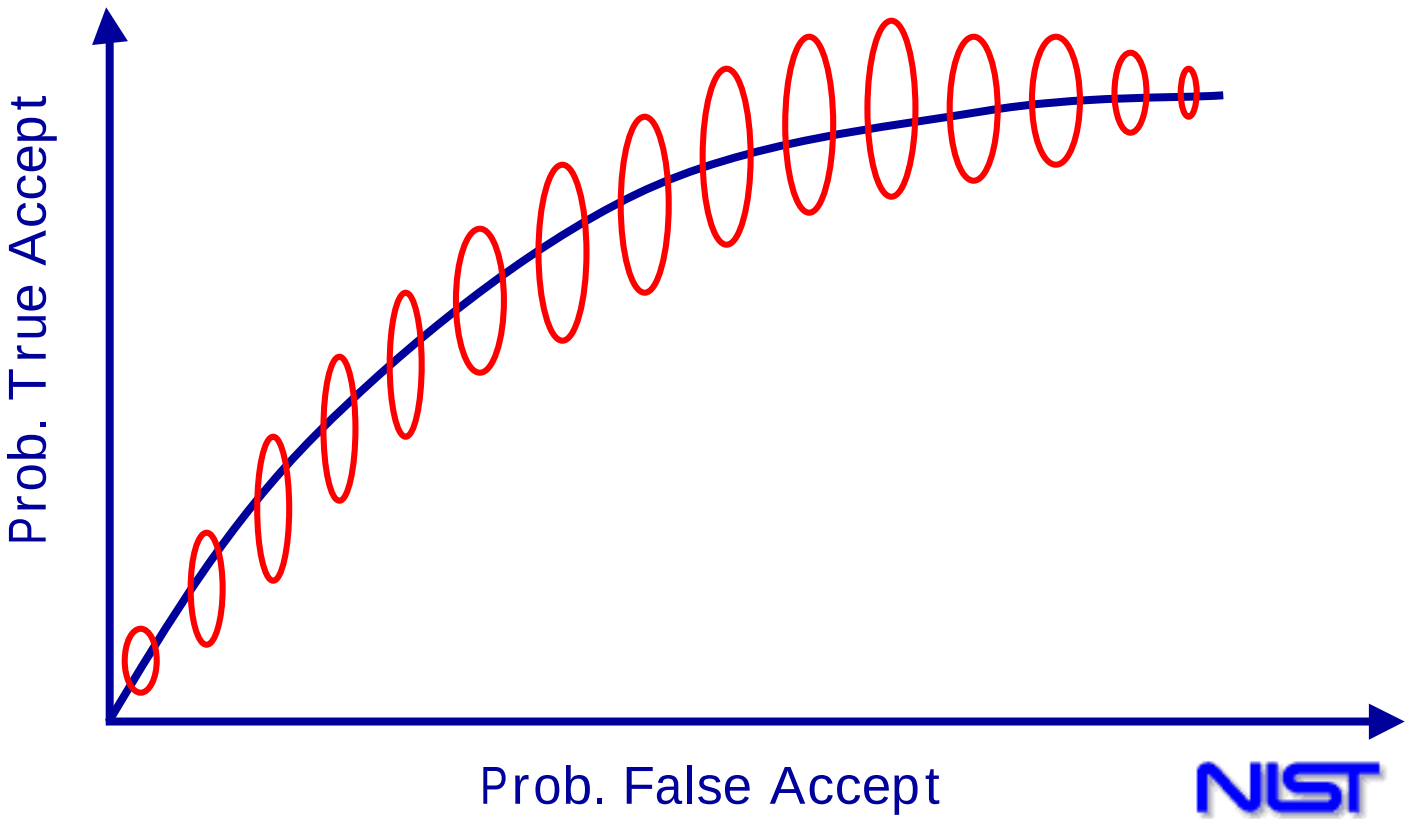
Option 2: Use information in both probe to gallery and gallery x gallery elements



National Institute of Standards and Technology

Standard Error

Problem: How Stable are the Estimates of Performance
Method: Divide the data; make multiple estimates.



National Institute of Standards and Technology

Standard Error & Confidence

- Observed Statistical Differences Between Standard Error & Confidence

Standard Error

- measurable observed variation

Confidential Interval requires

- distributional assumptions
- no accepted distributional assumptions

FR 2002
VT

Conclusion

- Thorough measures of performance
  - Identification
  - Verification
  - Watch List
  - Normalization
  - Standard Error Estimation
- Applicable to other biometric technology
- Basis for evaluation protocol standard

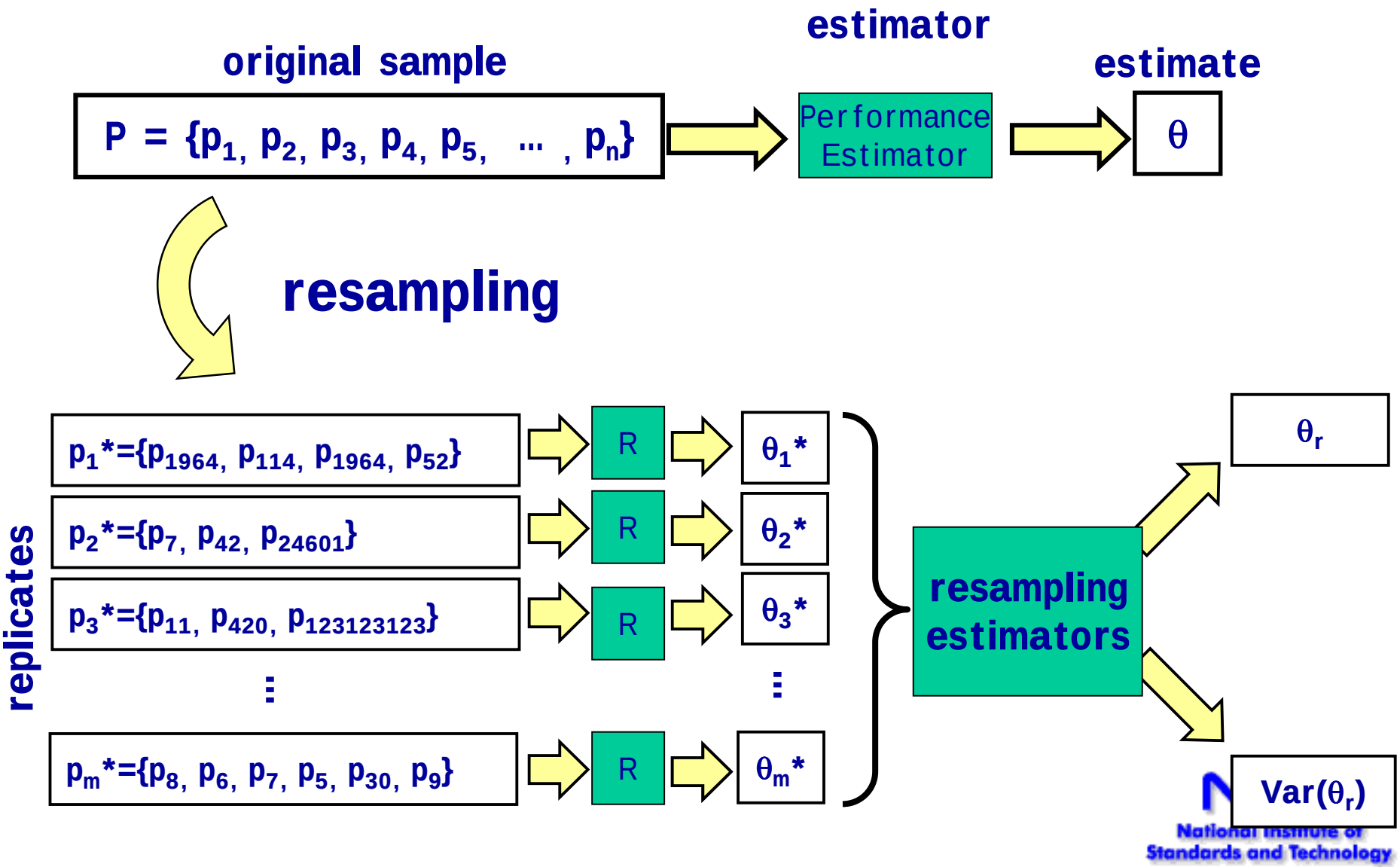
NIST
National Institute of Standards and Technology

FR VT 2002

Back-Up Slides

National Institute of Standards and Technology

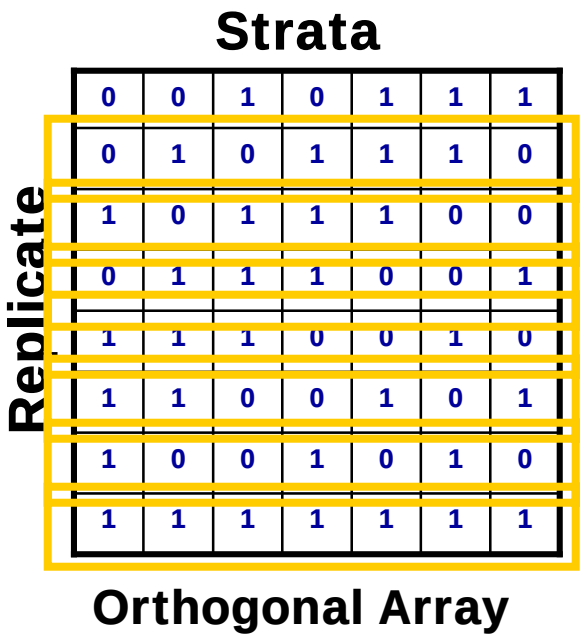
Resampling Techniques

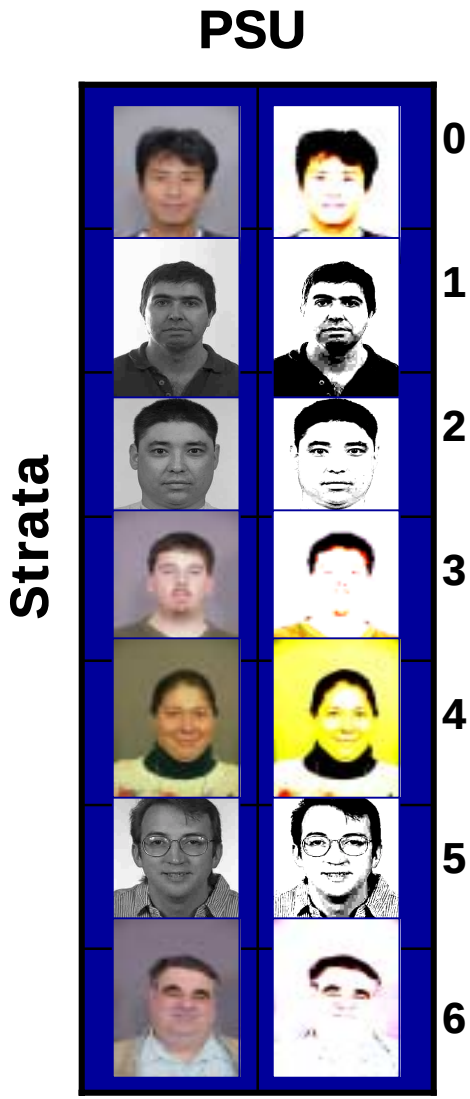


Resampling Techniques

- **Resampling** techniques examine statistics of subsets, combinations, and permutations of the original data.
- Resampling techniques are almost always **computer intensive**.
- The most common resampling methods are the **bootstrap** and the **jackknife**.
- However, these require **iid**.

Balanced Repeated Replication (BRR)





$\alpha_0 = \{$  $\}$

$\alpha_1 = \{\ldots\}$

$\alpha_2 = \{\ldots\}$

$\alpha_3 = \{\ldots\}$

$\alpha_4 = \{\ldots\}$

$\alpha_5 = \{\ldots\}$

$\alpha_6 = \{\ldots\}$

$\alpha_7 = \{\ldots\}$

NIST National Institute of Standards and Technology

FR 2002
VT 2002

Alternative Titles Perhaps

Biometric Systems: Do they work? How do you find out?

Beyond The Hype:
Measuring Performance of Biometric Systems

NIST
National Institute of Standards and Technology

Balanced Repeated Replication (BRR)

Strata

Replicate							
0	0	0	1	0	1	1	1
1	0	1	0	1	1	1	0
2	1	0	1	1	1	0	0
3	0	1	1	1	0	0	1
4	1	1	1	0	0	1	0
5	1	1	0	0	1	0	1
6	1	0	0	1	0	1	0
	1	1	1	1	1	1	1

Orthogonal Array