Biometrics: Best Practices and Applications

Dr. James L. Wayman
jlwayman@aol.com

Outline

- Scientific Practices
  - What is “science”?
  - What is “biometrics”?
    - Recognition without identity
    - Taxonomy of applications
    - Example operational systems
  - New conceptualizations
- Testing Best Practices
  - ISO/IEC 19795-2
  - ISO/IEC 19795-6
- Estimating Performance of Large-Scale Systems
- Uncertainty Assessment
  - ISO Guide 98

Scientific Best Practices in Biometrics

- So what is "science"?
  - From Latin "scientia": to know
  - But knowledge can be of many types, so a narrower definition from Oxford English Dictionary:

  "the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment"

Scientific Best Practices in Biometrics

- We observe the real world
  - No “arm chair” experiments
- We reason inductively, then limit our conclusions to the scope of the observations
  - Results under one set of collection conditions do not translate to any other set of conditions
- We question all results
  - Question this talk.
  - Are results confirmed within the current social structure?

Scientific Values Informing Best Practices

T.Kuhn, “Objectivity, Value Judgment, and Theory Choice” (1972)

1. External consistency (accuracy) with regard to experimental results
2. Internally consistent (with itself and other theories)
3. Broad scope
4. Simplicity
5. Fruitfulness

We should practice techniques that support our values.

What is “Biometrics”?

- “Biometrics” – the application of statistical methods to biological data – Oxford English Dictionary, 10th Edition, 2002
- “The active pursuit of biological knowledge by quantitative methods” -- R.A. Fisher (1948)

S. Stigler, "The Problematic Unity of Biometrics", *Biometrics* (56)3, 2000

- "What is biometry? Our modern subject of biometry is amazingly diverse; so much so that the question could be raised as to whether or not it has sufficient unity to constitute a single discipline."
- William Whewell (1794-1866)
  - …there is a problem in Biometry (if you choose to call your calculations on lives by a Greek name) which may perhaps be included in something you have done.... It is this: "It is said to be ascertained that to put off to a later period of life the average age of marriage, does not diminish the average number of children to a marriage. This being assumed, to find the effect on the increase of the population produced by a given retardation of the average age of marriage."

Our Definition

"Biometrics" -- the automated recognition of individuals based on their biological and behavioral characteristics --- ISO/IEC JTC1 SC37 Working Group 1

More definitions in upcoming ISO/IEC 2382 – Part 37

Some Consequences of Our Definition

1. Biometrics without identity
2. Recognition, not "verification" or "identification"

   Old concepts!
3. No taxonomy of "behavioural" and "biological"
4. Non-recognition can be as important as recognition – The Zen of Biometrics
5. Non-automated approaches out of scope
6. Biometrics without enrolment

Some Interesting Examples

Disney World Theme Park

Some Interesting Examples

- FBI
  - Linking cases through recognizing biometric characteristics from unknown individuals
  - Counting number of speakers in a conversation
  - ANSI/NIST ITL Type-11 “Voice Signal Record”
  - Recognizing unknown person as seen before
- Health Care
  - Anonymous health screening using iris recognition

Some Interesting Examples

- Customization

Conceptualizing Biometrics in a New Way

ISO/IEC 2382-37, "Biometric Vocabulary", 2013

"Who Goes There? Authentication Systems through the Lens of Privacy", L.Millett and S. Kent (eds.), National Academies of Science Press, 2003

"Biometric Recognition: Challenges and Opportunities", J. Pato and L. Millett (eds.), National Academies Press, 2010

Biometric Characteristics are Not "a Replacement for PINs and Passwords"

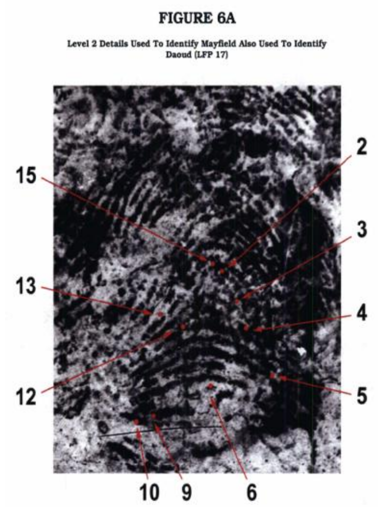
- Accessible without knowledge/consent
- Not exactly repeatable
- Enduring
  - Blanchette and Johnson, "Data retention and the panoptic society: The social benefits of forgetfulness", (2002)
- Not application specific
- "Sticky"
- Allow record linkage across systems
- Require specialized collection hardware
- Can establish subject is not known to the system
- Universal "accessibility" issues not well explored
- Different security issues and evaluation methodology

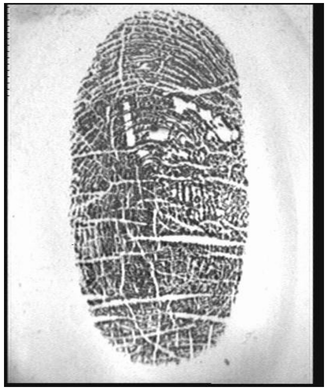
Inherent Challenges of Biometric Recognition

- Availability
  Acquisition of biometric reference
- Distinctiveness
  Between-class variation
  Mayfield – Doude confusion

www.usdoj.gov/oig/special/s0601/PDF_list.htm

- Stability
  Within-class variation





Beyond Access Control Apps

- Two applications
  - Establishing a person is known (recognized)
    - with identifier
    - without identifier
  - Establishing a person is not known (recognized)
- No biometric method guarantees the validity of the non-biometric data in the database
- Our thinking and terminology remain biased towards access control apps

Positive Claims

- To prove I am known to the system
- Prevent multiple users of a single identity
- Comparing sample set against single stored reference set
- False match allows fraud
- False non-match is inconvenient
- Multiple alternatives
- Can be voluntary
- At large scale, most claims true, so false negatives will be the problem

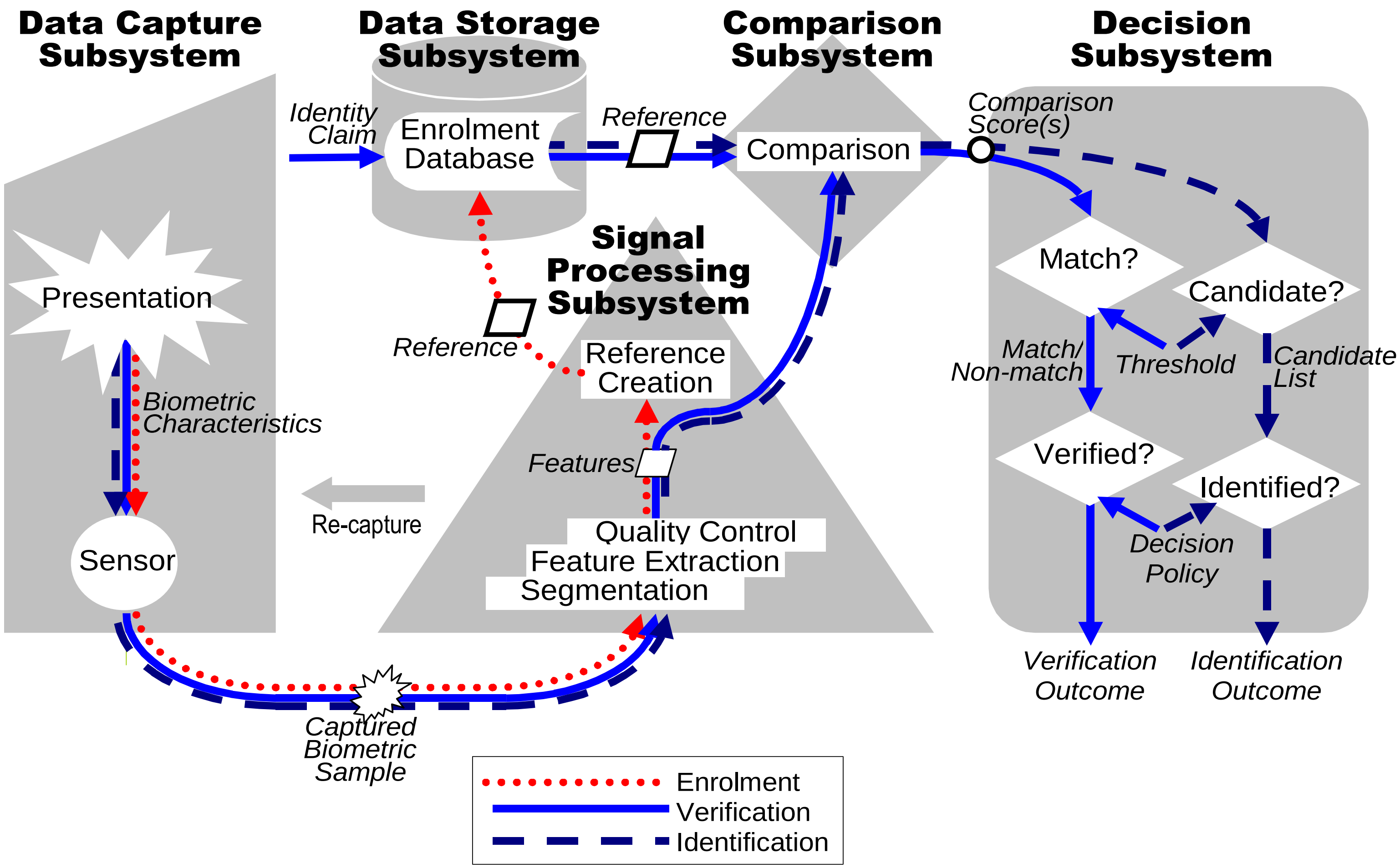
Negative Claims

- To prove I am not known to the system
- Prevent multiple identities of a single user
- Comparing sample set to all stored reference sets
- False non-match allows fraud
- False match is inconvenient
- No alternatives
- Mandatory for all users
- At large scale, most claims true, so false positives will be the problem

But not all applications of biometrics involve claims

Testing Best Practices

System Diagram from SC37 Standing Document 11



Why Performance Evaluations are Conducted

- Demonstration of capabilities
  - Is this suitable for my application / environment
- Procurement
  - Does system meet specification ?
  - Does system A outperform system B?
    - Will it be significantly cheaper to run?
- Performance prediction
  - What performance will be achieved ?
- Performance monitoring
  - What performance are we achieving?
  - Where are problems arising?
- Optimisation / algorithm improvement
  - Tuning of algorithm
  - Development/proof of new techniques

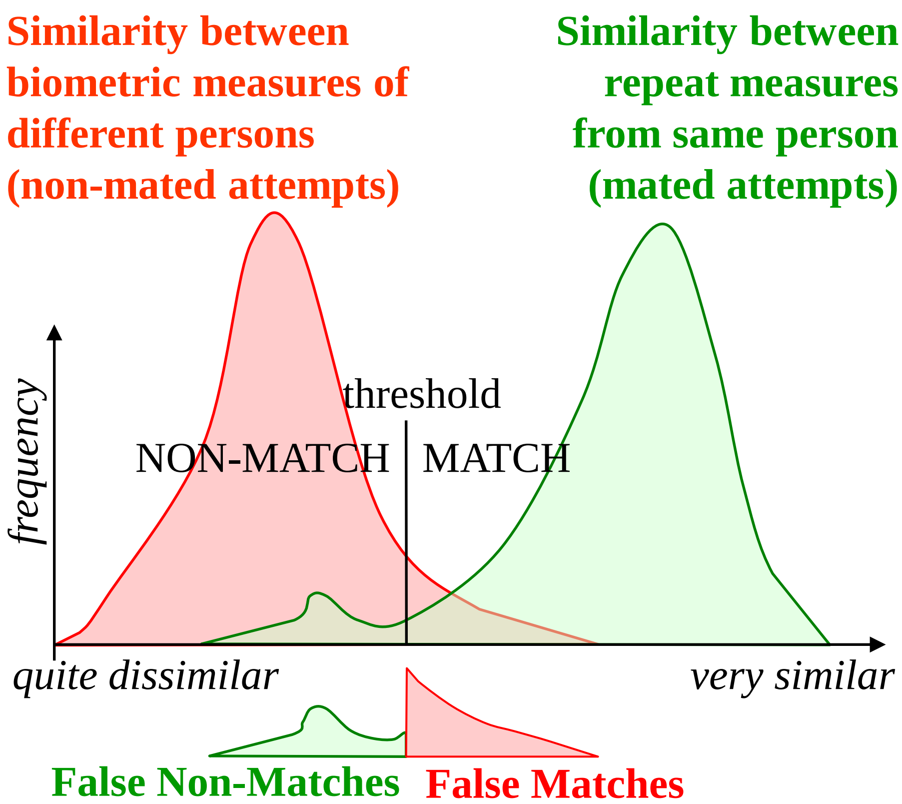
Aspects of Performance

- Things to measure
  - Technical performance/accuracy
  - Throughput
  - Interoperability
  - Conformance
  - Reliability, availability, maintainability
  - Security & vulnerability analysis
  - Safety
  - Usability
  - Public perceptions/acceptance
- Components
  - Sensors
  - Algorithms
    - Feature extraction
    - Comparison
    - Quality assessment
  - Interfaces
  - Processes
  - People
  - Environments

Fundamental Technical Performance Metrics for Biometric Systems

- Universal – *Biometric is possessed and measurable on all people*
  - **Failure-to-enrol rate**: Proportion of people unable to enrol
- Distinctive – *Biometric measure different for different users*
  - **False match rate**: Proportion of "impostor" comparisons (i.e. between measures from different people) that are deemed to match
- Repeatable – *Biometric measure similar across time for each user*
  - **False non-match rate**: Proportion of "genuine" comparisons (I.e. repeat measures from the same person) that fail to match
- Accessible – *Biometric measure easily acquired by the sensor*
  - **Failure-to-acquire rate**: Proportion of cases measure can't be acquired
  - **Throughput rate**: Times taken to enrol & to be verified

Performance Metrics are Inter-dependent



- Relax threshold for deciding if biometric measures match
  - \+ Fewer false non-matches
  - – More false matches
- Allow enrolment of poor quality biometric measures
  - \+ Fewer failed enrolments
  - – More matching errors
- Spend less time collecting biometric measure
  - \+ Faster throughput
  - – More failed enrolments & More matching errors

False non-match rate is quoted with corresponding false match rate

What is the Shape of These Distributions?

- Wu and Wilson, "Nonparametric Analysis of Fingerprint Data", NISTIR 7226, May 2005, http://www.itl.nist.gov/iad/894.03/pact/ir_7226.pdf
- "This paper demonstrates that, for large-scale tests, the match and non-match similarity scores have no specific underlying distribution function. The forms of these distribution functions require a nonparametric approach for the analysis of the fingerprint similarity scores.

Distributions

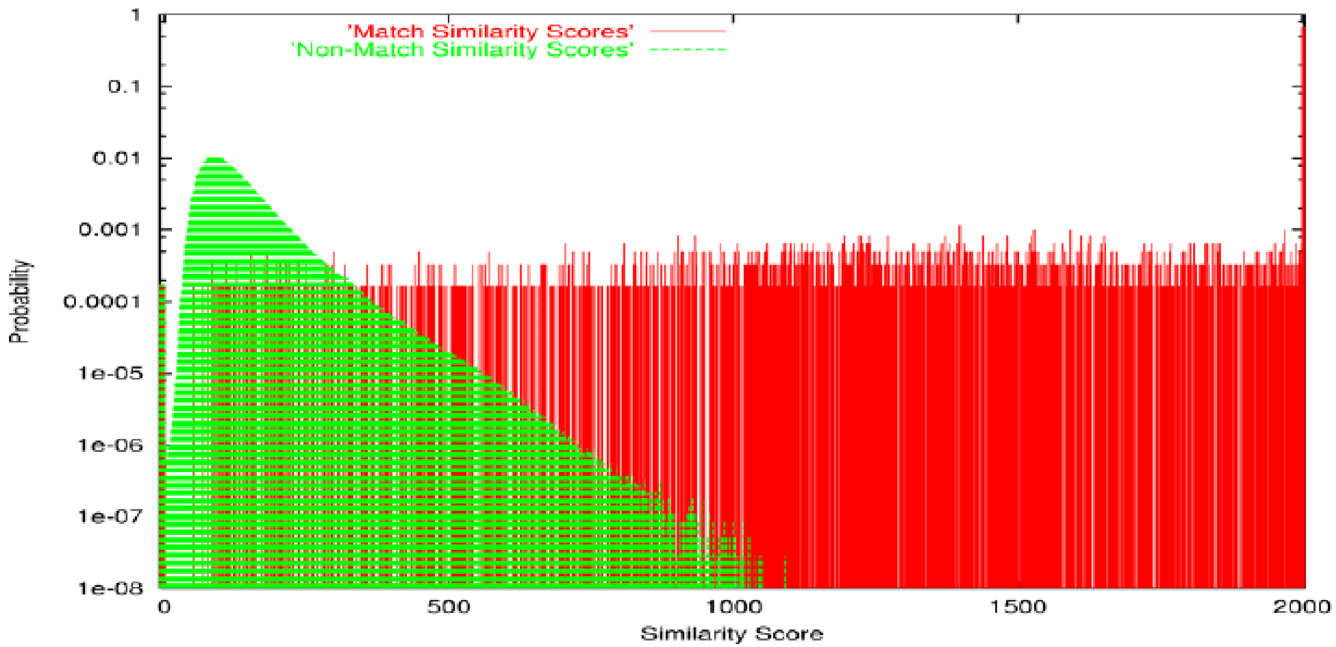
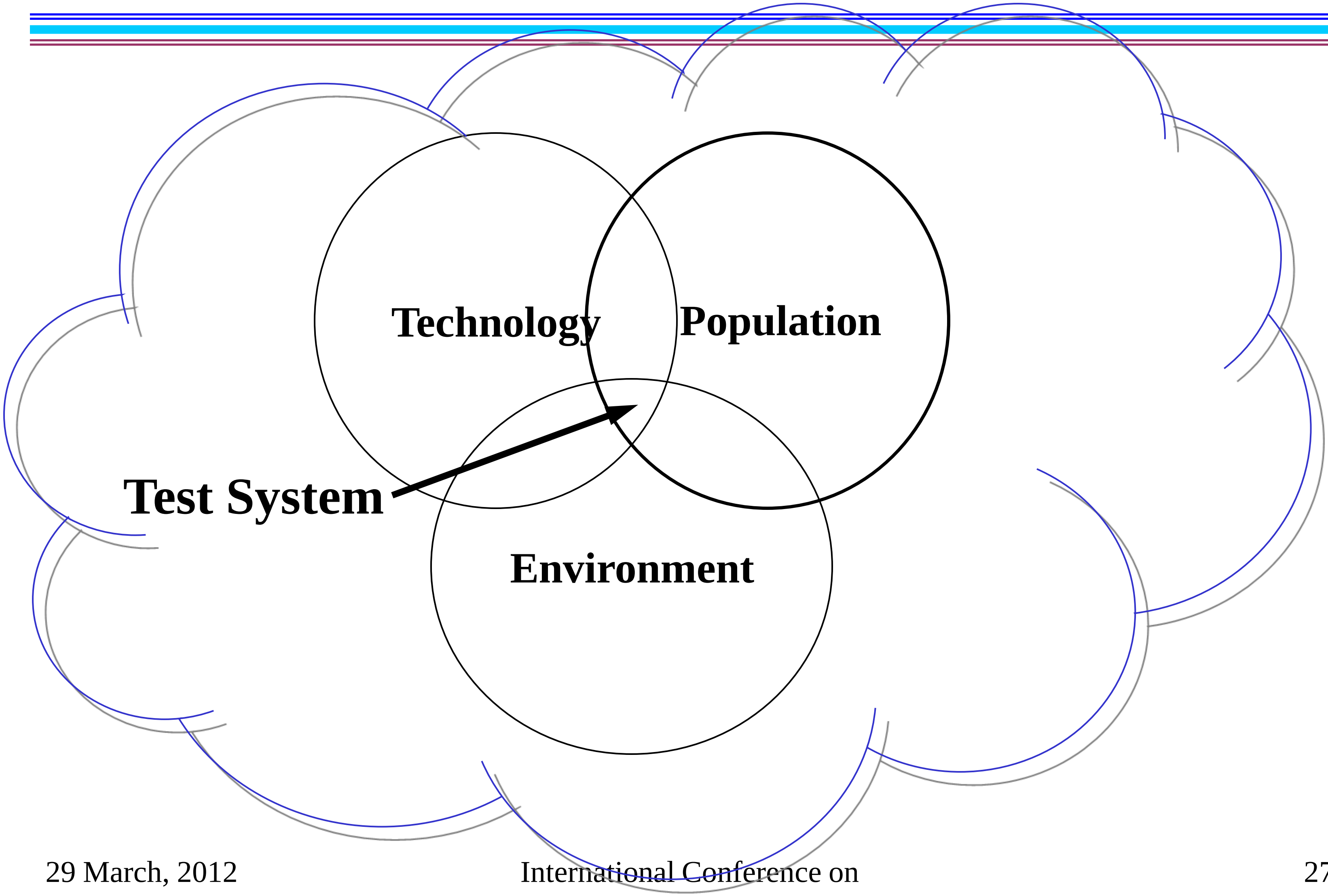
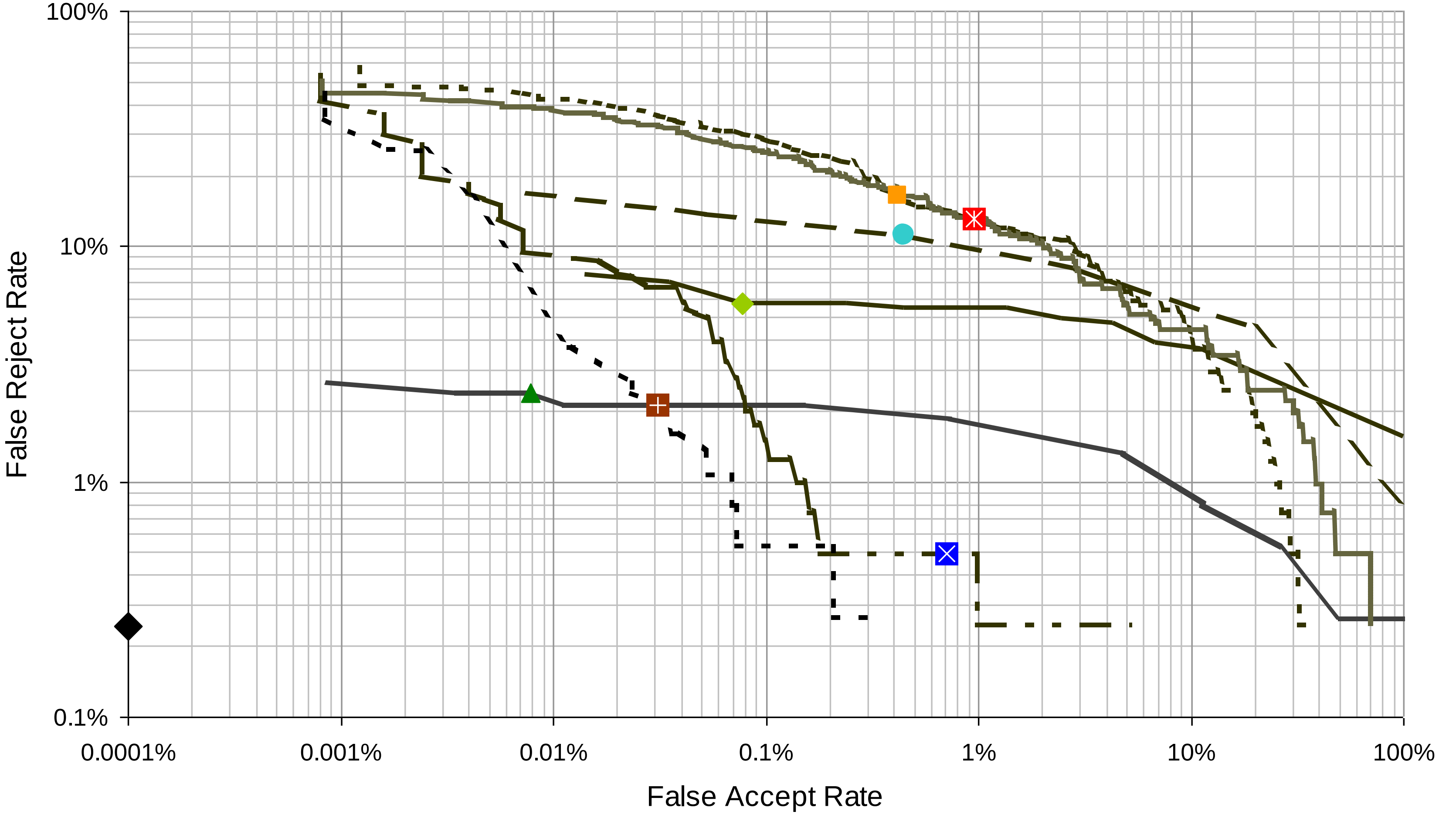


Figure 1 The discrete probability distribution functions of the match and non-match similarity scores generated by using the fingerprint-image matching Algorithm 1. The integral similarity scores run from 0 to 2000. The widths of peaks at the highest score and at the lowest score are enlarged to show the characteristics of the distributions.

Determining Distributions



Detection Error Trade-Off (DET) Curves
(Thank you, Tony Mansfield)



An Alternative Approach

- “Closed set” testing
  - I know you’re enrolled
  - Which one are you?
- Non-parametric reporting
  - Rank-order statistics
  - “Accuracy”: Probability that the “true” match is within the top m matches
  - “False matches” not defined
  - “Impostors” don’t exist

Cumulative Match Scores

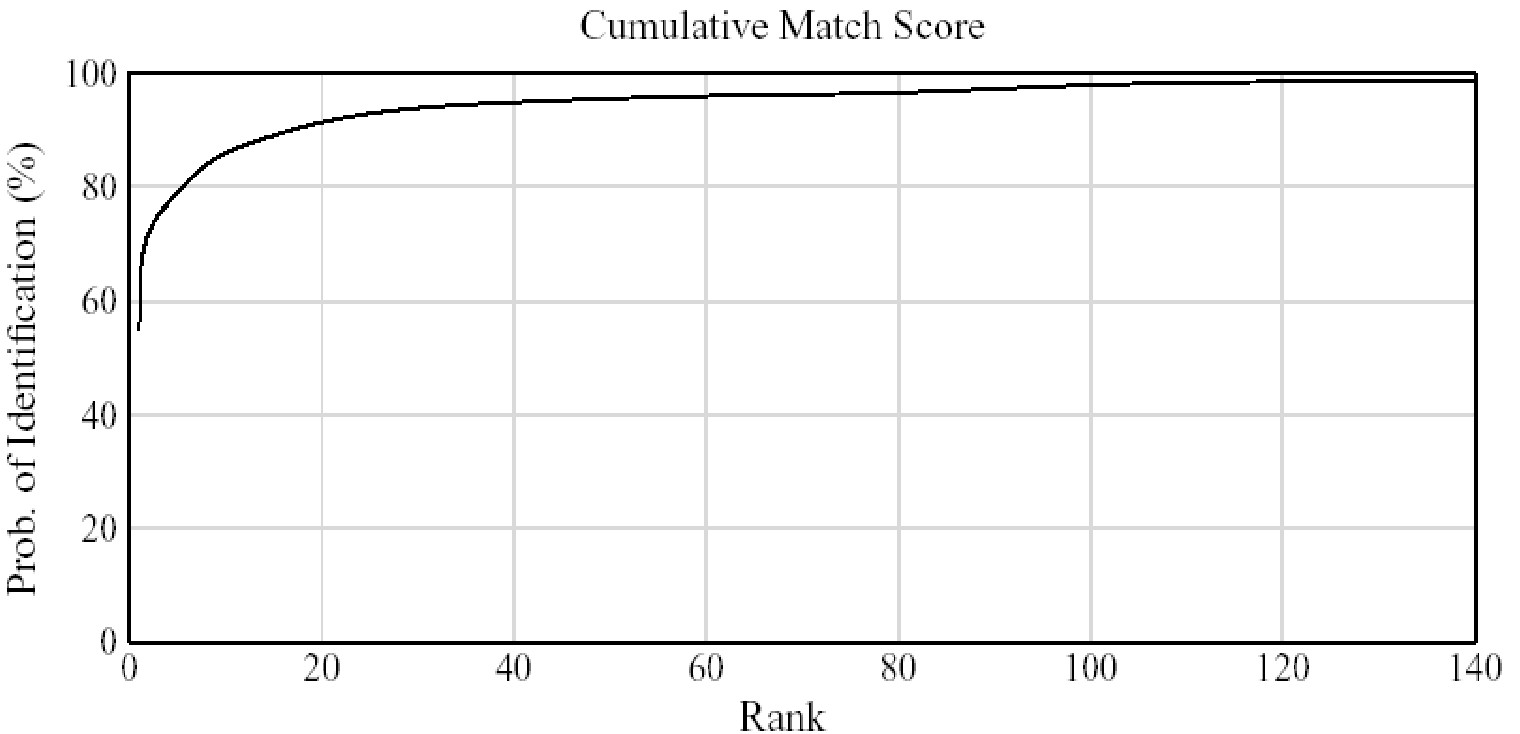


Figure D-1: *Sample cumulative match score (CMS).*

- From M. Bone and D. Blackburn (2002)

Why I Object to Closed-Set Testing

- Apparent relationship:

"Rank k accuracy" = f (N, M, k, algorithm, database)

- Actual relationship:

N,M,k, database = f (desired accuracy claim, algorithm)

- "Accuracy" allows for comparison of algorithms only if N, M, k, and database are the same for each
- So "accuracy A = 90%" might be better than "accuracy B = 95%"
- Results from different tests are incommensurate.

Technical Test Types (NIST, 1999)

- Technology, Scenario, Operational
  - Technology: testing the algorithms
  - Scenario: testing the human-machine interface
  - Operational: testing mob behavior

ISO/IEC 19795: Biometric Performance Testing and Reporting

Multipart Standard

1. Principles & framework
2. Methodologies for technology/scenario evaluation
3. Modality specific testing (Technical Report)
4. Performance & interoperability performance testing
5. Access control test scenario
6. Methodologies for operational evaluation
7. Testing of match-on-card biometrics

ISO/IEC 19795-1 – Contents

1. Scope
2. Conformance
3. Normative References
4. Terms & Definitions
5. General biometric system
6. Planning the evaluation
7. Data collection
8. Analyses
9. Record keeping
10. Reporting results

A. Difference between types of evaluation
B. Test size and uncertainty
C. Factors influencing performance
D. Pre-selection
E. Identification performance as function of database size
F. Algorithms for DET and CMC curve generation

Introductory elements

Normative part

Informative Annexes

Data Collection

- Matching errors (FMR, FNMR) can be smaller than errors occurring in operational procedures
- Avoidance of data collection errors
  - Metadata errors
    - Wrong PIN, user ID
    - Wrong body part
  - Corpus errors
    - Blank corrupt images
    - Test subjects/operators using system incorrectly
- Correct usage needs to be defined in advance
- Procedures needed to prevent/detect/correct errors

Volunteer “Crew”

- Demographically similar to target population
- Use of volunteers will unavoidably bias results
- Standard human subject protections

Mated Transactions

- If external consistency is required, replicate target environment as closely as possible.
- Time delay from enrollment to replicate target "Template Aging"
- Good faith user attempts to replicate enrollment pattern

Mated Distributions

- One sample from multiple individuals
  - BEST, but expensive
- Multiple samples from one individual
  - WORST, but cheap
- “Balanced” sampling

Non-mated Transactions

- Unknown impostors required
  - “Jackknife”
- Good faith user attempts to replicate own enrollment pattern
- “Zero effort”

Some Fundamental (but often violated) Principles

- Separate "system training" and testing databases
  - i.e. Genuines and impostors cannot be from set used to create basis vectors
- Artificial images are phony
  - We are not God and don't know how people are made
  - Unfair positive bias to systems making similar assumptions

Some Fundamental (but often violated) Principles

- Test data must be “unseen” (sequestered)
  - What to do about overtraining on non-sequestered elements?
  - How do we assess performance improvement over time?

ISO/IEC 19795-6: 2012

Testing methodologies for operational evaluation

Purposes of Operational Tests

- determine if performance meets the requirements for a particular application or the claims asserted by the supplier;
- determine how to adjust system to improve performance;
- predict performance with increase in subjects, locations, or devices;
- obtain information on the target population and environmental parameters found to affect system performance;
- obtain performance data from a pilot implementation;
- obtain performance data to benchmark future systems

Hidden Factors Impacting Performance Measures

- performance of the system might improve as subjects habituate or degrade as subjects' biometric characteristics age over time away from their enrolled references.
- The performance observed in testing can depend on the operational personnel, such as attendants or biometric examiners, as well as the biometric subjects.

Operational Metrics

- throughput for enrolment and recognition transactions,
- failure-to-enrol rate,
- system rejection rate (in verification systems),
- system identification rate (in identification systems)
- false accept rate and false reject rate (in verification systems when the evaluation can establish ground truth),
- false-positive identification error rate and false-negative identification error rate (in identification systems when the evaluation can establish ground truth).

Some Examples of Operational Tests

- SmartGate
- EasyPass

Australian SmartGate

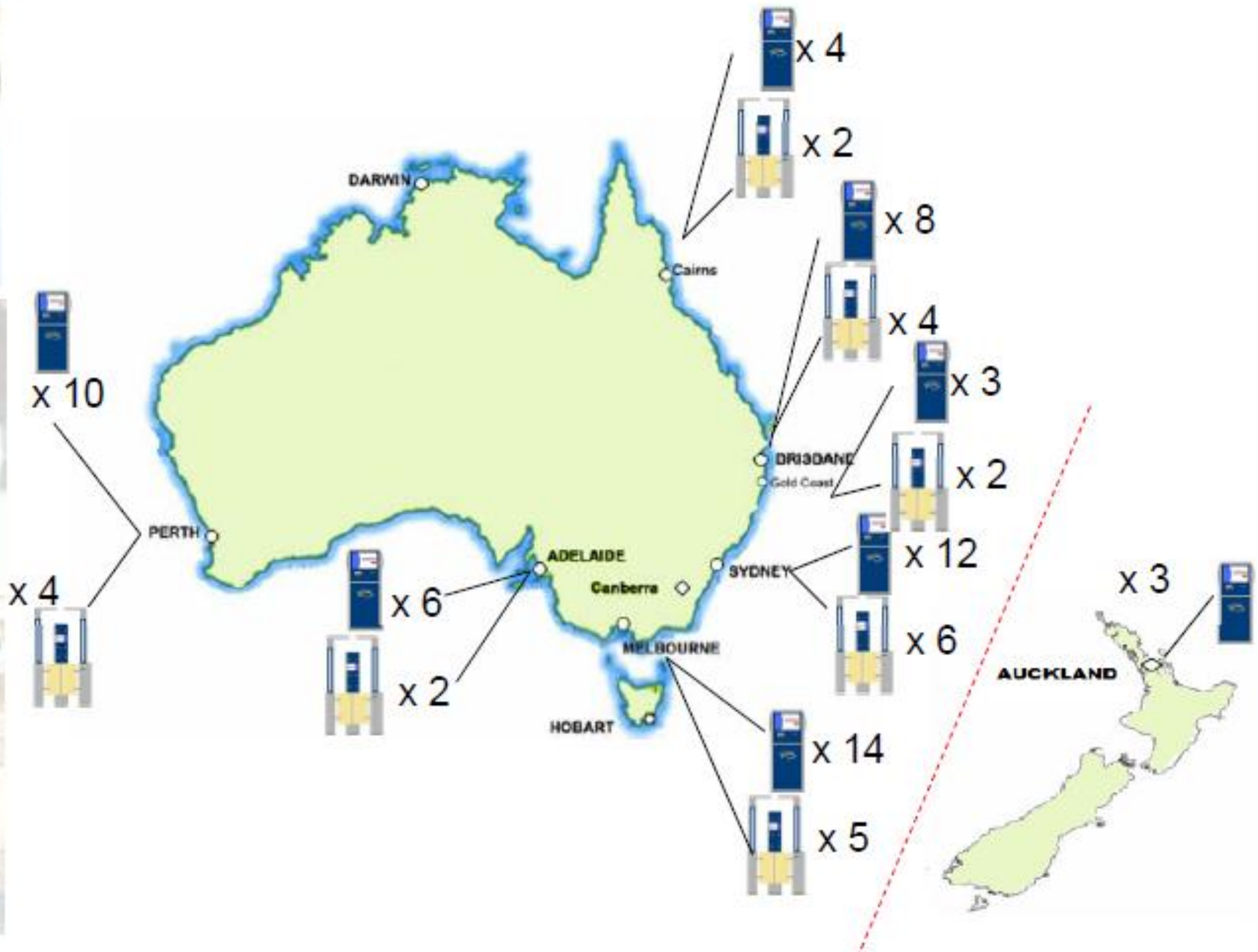
- Can the primary line immigration processing be replaced by automated system?
  - Security features on passport
  - Check for persons of interest
  - Match passenger to passport
  - Passenger clearance recorded
  - Can refer passenger to immigration or call for more checks
- But automated system must be:
  - Voluntary
  - Quick and accurate
  - Preserve all other parts of security system

Face Recognition as One Element of SmartGate

- To match passenger to passport under harsh lighting
  - In Australia, duty free shops are located in the entry control area
  - Lighting, windows, window treatments and screens in entry control area are owned by airport authority
- If face recognition from the image on the e-passport is possible, no specialized enrolment would be required
  - This differs from UK IRIS, CANSPASS, Ben Gurion, US Global Traveler, (former) US-INSPASS
- But SmartGate would have no control over enrollment conditions

SmartGate evolution

Where is it?

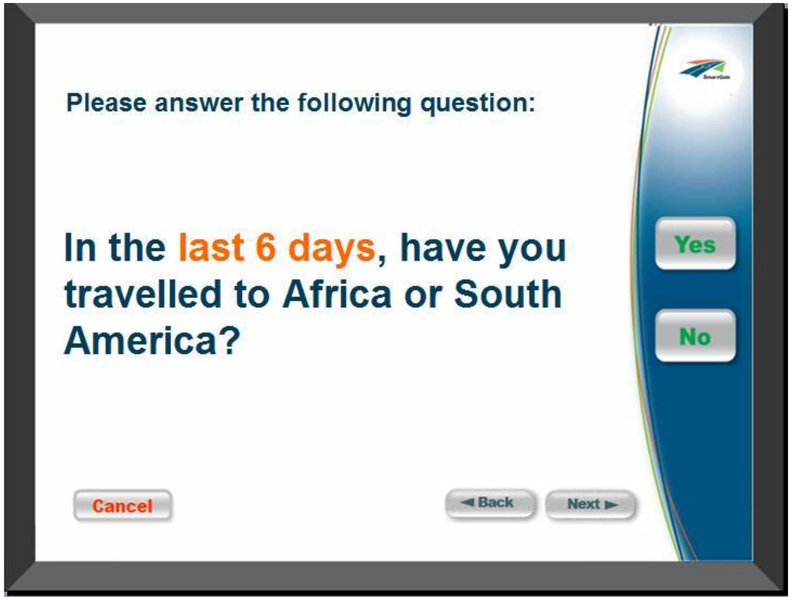

x 4
x 2
x 8
x 4
x 3
x 2
x 10
x 4
DARWIN
Cairns
PERTH
ADELAIDE
Canberra
SYDNEY
MELBOURNE
HOBART
x 6
x 2
x 12
x 6
x 14
x 5
x 3
AUCKLAND

A few facts and figures

- 55% of eligible travellers choosing to use SmartGate currently
- Over 2.3 million users since initial implementation in Brisbane 2007
- Since July 2009: 1.3 million users
- 60 kiosks and 25 gates in operation around the country

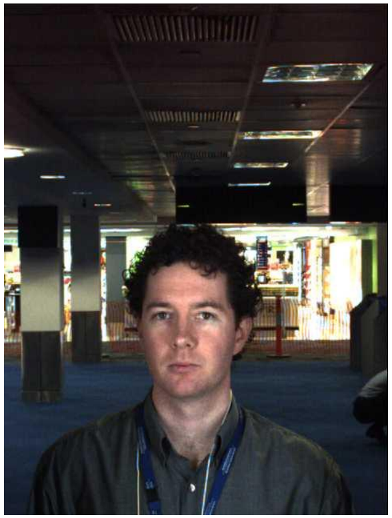
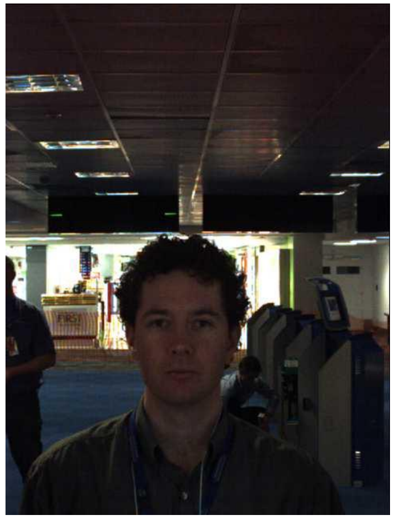
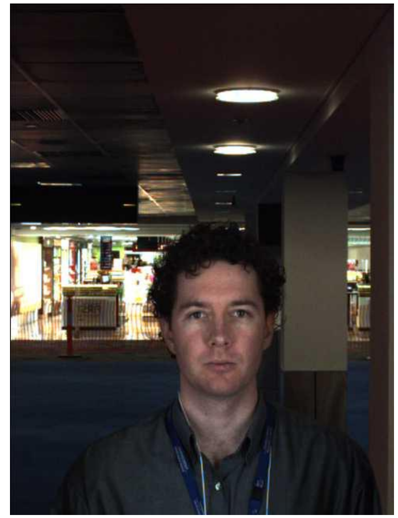
Two Stage Process

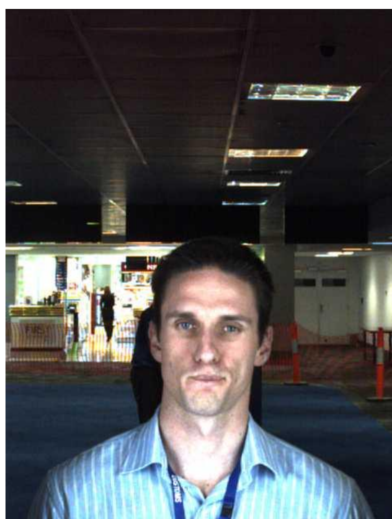
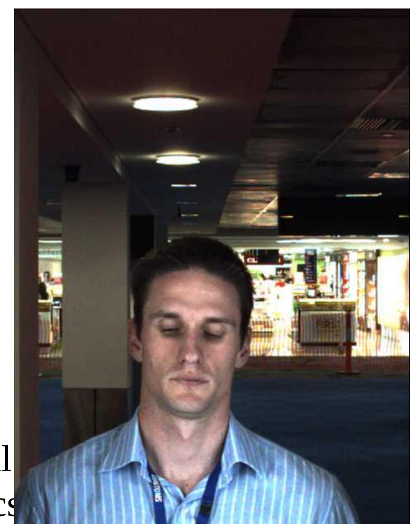
- Kiosk
  - e-Passport read
  - Passport validity/eligibility verified
  - Health and character questions
  - Ticket issued
- Gate
  - Face recognition
  - 3 cameras at different heights
  - Final clearance check
  - Passenger entry recorded



CUSTOMS
WAIT HERE
WAIT HERE
WAIT HERE
WAIT HERE
CUSTOMS

SmartGate Images

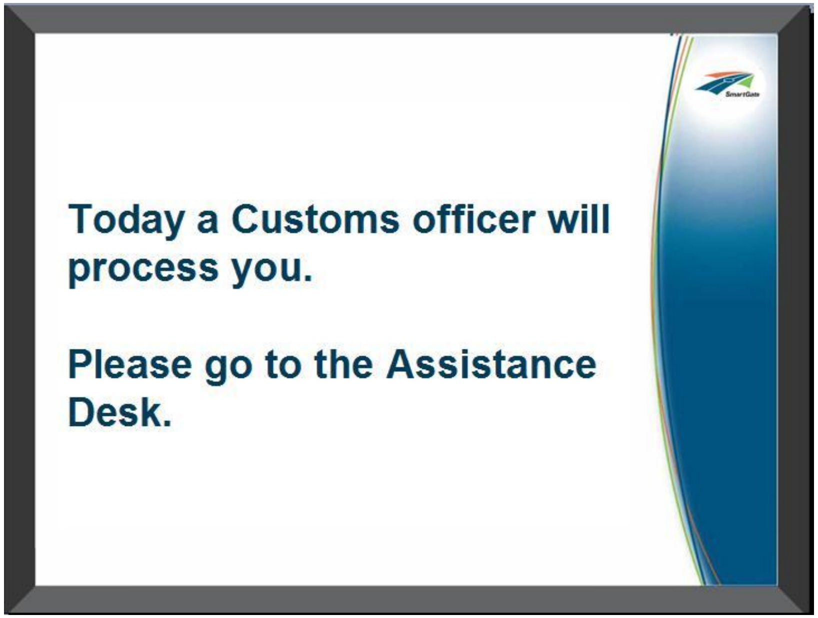

Challenges

- ISO/IEC 19794-5 passport photo compliance is necessary but not sufficient
- People have been trained on different processes
- No international standards for signs and symbols relating to biometrics
  - ISO/IEC 24779-1 in working draft stage

Referred from Kiosk

Referrals include:

- Under 18
- Passport cannot be read
  - Damaged pages
  - Damaged chip
  - Misplaced in reader
- Abandon process
- Response to health and character questions



Referred from Gate

Referrals include:

- Not looking at camera
- Passport photo issues
- Not PIE compliant at gate
- Wrong camera automatically selected
- Usually multiple issues

Federal Office
for Information Security

The EasyPASS experience at Frankfurt Airport

Markus Nuppeney – Federal Office for Information Security (BSI)

Markus Nuppeney London - October 20, 2010 1

Federal Office
for Information Security

BSI

Operational figures –
Rejections at EasyPASS entrance
(Oct. 2009 – March 2010)

- ≈ 38.500 documents presented on DocReader
- ≈ 17.500 users passing through EasyPASS
- 55% rejection rate at the EasyPASS entrance
  - **28%** : Document could not be read optically (wrong presentation of the document on the DocReader)
  - **20%** : Rejected based on the EasyPASS policy (no ePassport, underage, no EU/EEA/CH citizen)
  - **7%** : ePassport RF reading aborted
- 45% of the ePassport reading attempts result in a border passage via EasyPASS

Markus Nuppeney   London - October 20, 2010   11

Federal Office
for Information Security

Operational figures –
Rejections by EasyPASS control process (Oct. 2009 – March 2010)

- ≈ 17.500 users passing through EasyPASS
- ≈ 15.000 users passing EasyPASS automatically
- 85,7% success rate
  - border crossing without manual interaction
- 14,3% rejection rate
  - additional manual inspection by border guard

Markus Nuppeney London - October 20, 2010 13

Federal Office
for Information Security

Operational figures –
Rejections by EasyPASS control process
(Oct. 2009 – March 2010)

Decomposition of the **14,3%** rejection rate

- **5,6%** rejected due to face verification failed **@ ≈ 0,1% FAR**
  - **2,2%** Failure-to-Capture
    (no image(s) containing a face delivered by the camera system)
  - **3,4%** Failure-to-Match
    (comparison score below threshold or template generation on DG2 or live image(s) failed)
- **8,7%** rejected by the system due to other reasons
  (e.g. non compliant user behaviour, document check failed, hits from background database checks)

Markus Nuppeney London - October 20, 2010 14

Federal Office
for Information Security

Operational figures –
Process time (Oct. 2009 – March 2010)

- ≈ **15.000** users passing EasyPASS automatically
- ≈ **18 sec.** average time period to pass the eGates
  - Time from presenting the ePassport on the DocReader until the system is ready to process next traveller
- Average time periods for main sub-processes
  - **5 - 6 sec.** for Reading and checking ePassport data (optical and electronic checks)
  - **5 - 6 sec.** for the traveller to enter the eGate
  - **1 sec.** for biometrics (face capture and comparison)
  - **5 - 6 sec.** for the traveller to leave the eGate

Markus Nuppeney London - October 20, 2010 16

Best Practices for Estimating Performance of Large-scale Systems

A reductionist model

"Under the simplifying, but approximate, assumption of statistical independence of all errors, (the) independent variables are bin error rate, penetration rate, sample-template ('genuine') and 'impostor' distance distributions, number of active templates or user models in the database, N, and the number of samples submitted for each transaction, M"

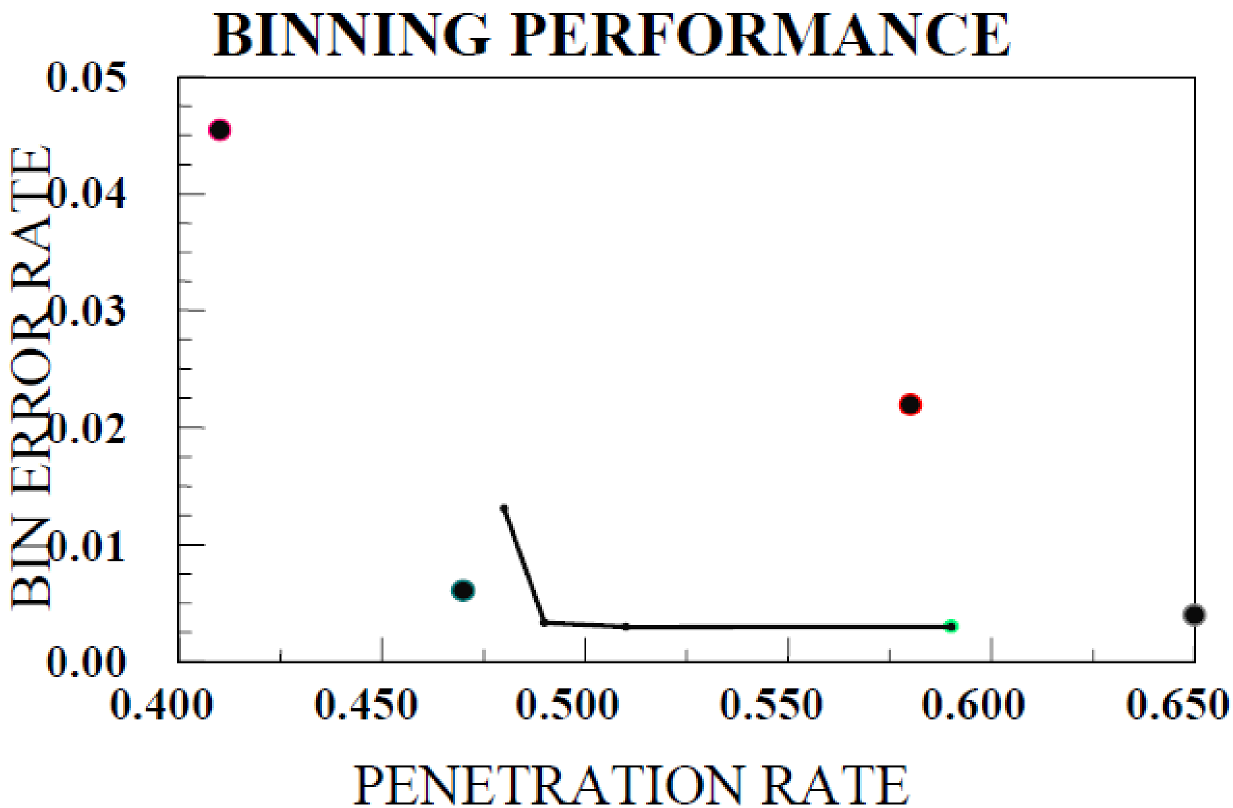
- When N = 1, equations must degenerate to "verification" system.

Bernoulli Assumptions, Binomial Results

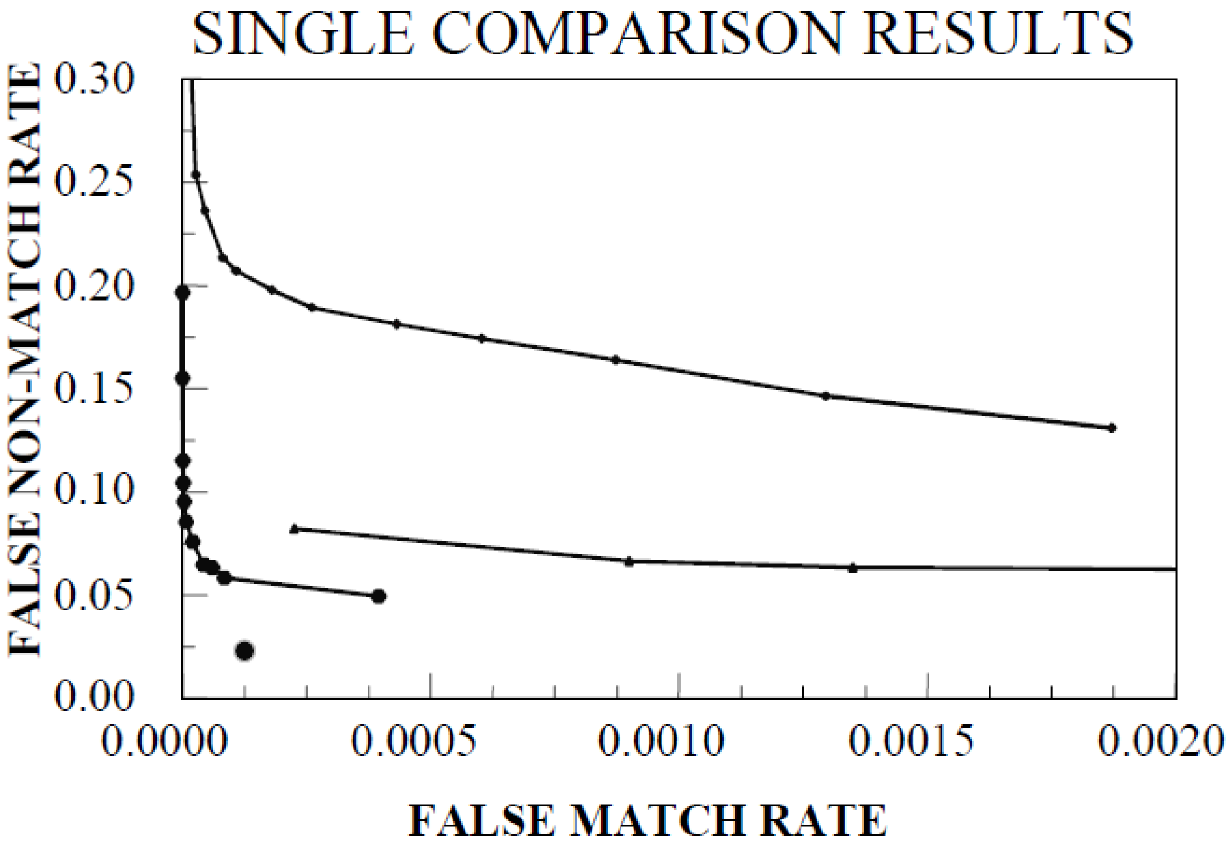
$$FNM_{sys} = \varepsilon_{ensemble} + [1 - \varepsilon_{ensemble}] \prod_{i=1}^{m} \left[1 - (1 - FNM_i) \sum_{j=Q-1}^{T-i} \binom{T-i}{j} (1 - FNM_U)^j (FNM_U)^{T-i-j} \right]$$

$$FMR_{sys} = 1 - \prod_{i=1}^{m} \left[1 - FMR_i * \sum_{j=Q-1}^{T-i} \binom{T-i}{j} FMR_U{}^j (1 - FMR_U)^{T-i-j} \right]^{N*P_i}$$

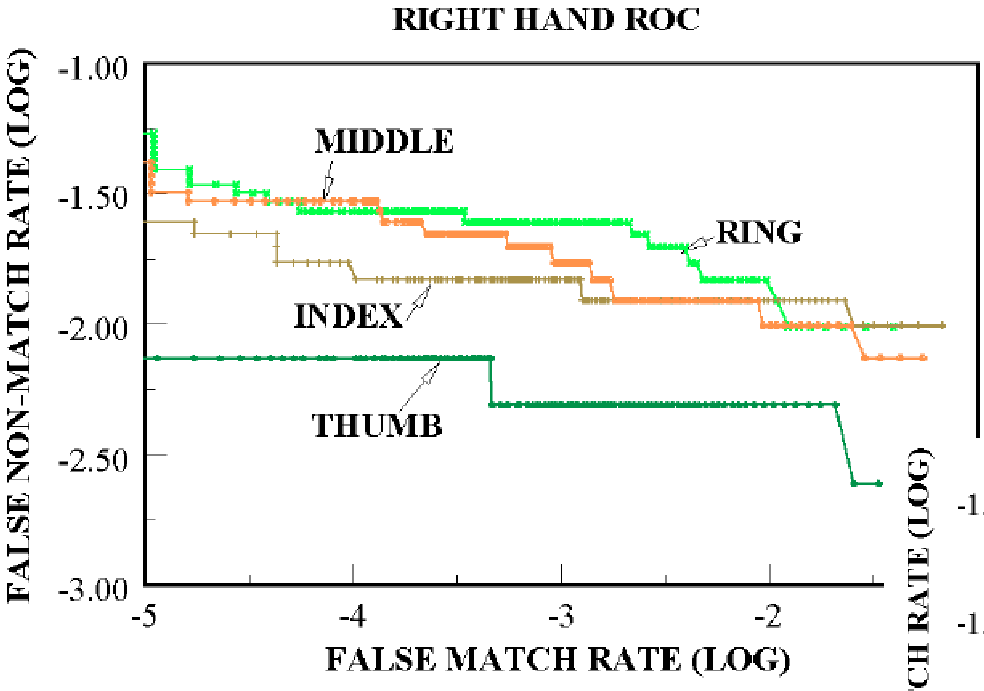
Estimating the Parameters

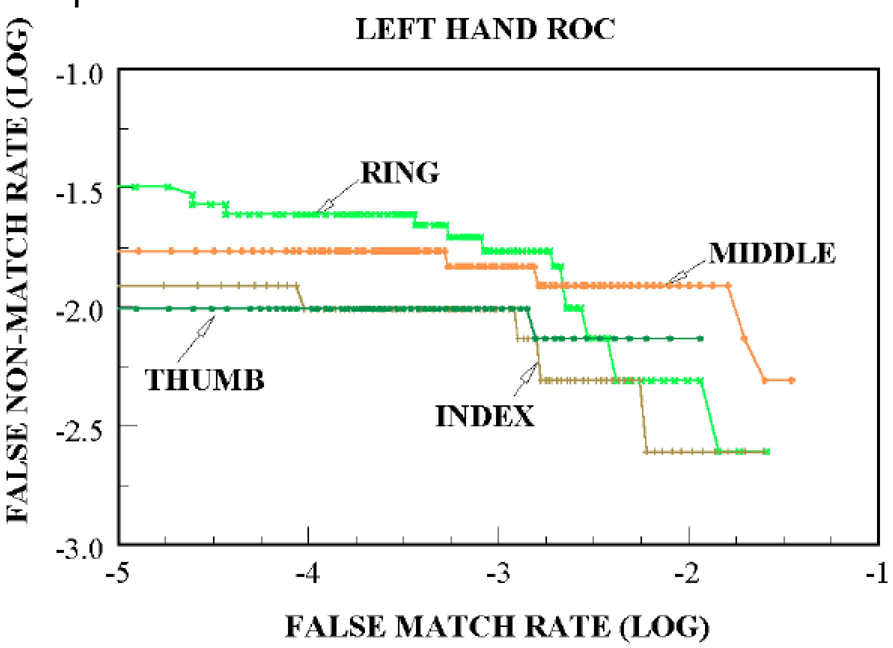


Estimating the Parameters



Finger Variability





Penetration Rate Correlations

TABLE 4: TWO-FINGER BINNING STATISTICS

Finger	Error Rate	Error if independent	Penetration Rate	Penetration if independent	
				FBI Data	Test Data
Thumb	0.005	0.005	0.52	0.30	0.47
Index	0.007	0.007	0.25	0.19	0.20
Middle	0.015	0.019	0.55	0.71	0.49
Ring	0.017	0.017	0.55	0.44	0.49

TABLE 5: MULTIPLE-FINGER BINNING STATISTICS

Fingers	Error Rate	Error if independent	Penetration Rate	Penetration if independent	
				FBI Data	Test Data
Four: Thumb and index	0.012	0.012	0.15	0.059	0.093
Eight: Thumb index, middle, ring	0.040	0.048	0.08	0.018	0.022

Further Reduction

- M. E. Schuckers, "Using the beta-binomial distribution to assess performance of a biometric identification device", *2003*.
- Because ….each individual will have their own probability of success, then p, the… probability of success, is not the same for each user. Thus, the binomial is not appropriate for assessing the performance… when combining outcomes from multiple users. Consequently, we need a model that allows for variability in the probability of success among individuals and that allows for the possibility that trials by a given individual are not independent. One such model is the Beta-binomial model or, more formally, the product Beta binomial.

$$f\left(\vec{x}|\alpha,\beta,\vec{n}\right)=\int f\left(\vec{x},\vec{p}|\alpha,\beta,\vec{n}\right)dp=\int f\left(\vec{x}|\vec{p},\vec{n}\right)f\left(\vec{p}|\alpha,\beta\right)dp$$

$$=\prod_{i=1}^{m}\binom{n_i}{x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+x_i)\Gamma(\beta+n_i-x_i)}{\Gamma(\alpha+\beta+n_i)}$$

- Where there are n individuals tested m times and α,β are parameters of the Beta distribution of p among the individuals

Alternative Approaches to Estimating FPIR in Large, Negative Claim Systems

- Jarosz, Fondeur, Dupré, “Large-scale Identification System Design” (2005)
  1. Extrapolation from experience
  2. Identification as succession of N verifications
  3. Extrapolation from extreme value
  4. Extrapolation when distance can be modeled

The influence of classification on reductionist models:

$$\text{FMR} = f(\text{binning})$$

Extrapolation

- No false positives implies that1st non-mated comparison is not a false positive & 2nd non-mated comparison is not a false positive & ….
- 1- False positive identification rate = $(1\text{-FMR})^N$
- ln (1-FPIR) = N ln (1-FMR)

  If $x \ll 1$, then $\ln(1-x) \approx -x$

So FPIR ≈ N (FMR) if FMR<<<1

So we confirm this experimentally by observing how FPIR increases with N for small N, then extrapolate as N increases

Extreme Value

- Compare each of M independent biometric probes against N independent references
- For each probe, record the best score
- Regardless of how all the NxM scores are distributed, the M best scores will be distributed in one of only three possible distributions (Gumbel, Fréchet, Weibull)
- From the M best scores, estimate the distribution
- Using the estimated distribution, determine probability that a best score will be greater than any threshold

Best Practices in Uncertainty Estimation

- Youden, "Enduring Values", Technometrics, 1972

Table 1. Different values reported for the Astronomical Unit (from Youden, 1972)

Number	Source of measurement and date	A.U. in millions of miles	Experimenter's estimate of spread
1	Newcomb, 1895	93.28	93.20–93.35
2	Hinks, 1901	92.83	92.79–92.87
3	Noteboom, 1921	92.91	92.90–92.92
4	Spencer Jones, 1928	92.87	92.82–92.91
5	Spencer Jones, 1931	93.00	92.99–93.01
6	Witt, 1933	92.91	92.90–92.92
7	Adams, 1941	92.84	92.77–92.92
8	Brouwer, 1950	92.977	92.945–93.008
9	Rabe, 1950	92.9148	92.9107–92.9190
10	Millstone Hill, 1958	92.874	92.873–92.875
11	Jodrell Bank, 1959	92.876	92.871–92.882
12	S. T. L., 1960	92.9251	92.9166–92.9335
13	Jodrell Bank, 1961	92.960	92.958–92.962
14	Cal. Tech., 1961	92.956	92.955–92.957
15	Soviets, 1961	92.813	92.810–92.816

Duhem-Quine Thesis and Testing Holism

- "…the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses"" – Duhem, 1906
- the results of any scientific test reflect the totality of conditions of the test ("the unit of empirical significance"), including instrumentation, background assumptions, auxiliary hypotheses, and even the theories being tested themselves. So what we measure in any experiment is the totality of all the elements existing in both the physical and intellectual environment of the test and, further, the measurements must be expressed using words and concepts that themselves may be subject to change as our understanding progresses.

An Example of Statistical Control of the Unit of Empirical Significance

- What is the “speed of sound”?
  - What medium?
    - Air?
      - At what precision do we need the results?
        - » High?
          - » What temperature?
          - » What pressure?
          - » What molecular composition?
          - » What unknown influence variables? (humidity, salt content, moon phase…)?

ISO Guide 98, “Guide to Expression of Uncertainty in Measurement”

- “… in principle, a measurand cannot be *completely* described without an infinite amount of information. Thus, to the extent that it leaves room for interpretation, incomplete definition of the measurand introduces into the uncertainty of the result of a measurement a component of uncertainty that may or may not be significant relative to the accuracy required of the measurement’”
- “..when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured”

The ISO Concept of Uncertainty Estimation Techniques

- Type A: Evaluation by statistical methods means estimation of a component of uncertainty using statistical methods applied to **replicated indications** obtained during measurement.
- Type B: Other means of evaluation include information derived from authoritative publications, for example in the certificate of a certified reference material, or based on expert opinion.
- Appears to combine frequentist and subjective measures in a way that neither frequentists nor Bayesians can endorse

The ISO Concept of Uncertainty

	Type A Statistical	**Type B Other**
Random	Classic "confidence intervals"	
Systematic		

Which type of error (random, systematic) dominates and how should it be estimated?

Neyman's "Confidence Intervals"

- Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, Vol. 236, No. 767 (Aug. 30, 1937), pp. 333-380

X—Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability

By J. NEYMAN

Reader in Statistics, University College, London

(Communicated by H. JEFFREYS, F.R.S.—Received 20 November, 1936—Read 17 June, 1937)

CONTENTS

I—INTRODUCTORY

(*a*) *General Remarks, Notation, and Definitions*

We shall distinguish two aspects of the problems of estimation : (i) the practical and (ii) the theoretical. The practical aspect may be described as follows :

(*ia*) The statistician is concerned with a population, π, which for some reason or other cannot be studied exhaustively. It is only possible to draw a sample from this population which may be studied in detail and used to form an opinion as to the values of certain constants describing the properties of the population π. For example, it may be desired to calculate approximately the mean of a certain character possessed by the individuals forming the population π, etc.

(*ib*) Alternatively, the statistician may be concerned with certain experiments which, if repeated under apparently identical conditions, yield varying results. Such experiments are called random experiments, (*see* p. 338). To explain or describe

VOL. CCXXXVI.—A 767 (Price 6*s*.) 2 Z [Published August 30, 1937

Neyman's Applications of "Confidence Intervals"

- "(ia) The statistician is concerned with a population, π, which for some reason or other cannot be studied exhaustively. It is only possible to draw a sample from this population which may be studied in detail and used to form an opinion as to the values of certain constants describing the properties of the population ,π…..
- (ib) Alternatively, the statistician may be concerned with certain experiments which, if repeated under apparently identical conditions, yield varying results."

A Different Approach by GUM

- Subsumes Neyman “confidence intervals” , but covers a much broader range of conditions, including experiments which cannot be repeated under identical conditions, as in biometrics
- “interval”: possible values of the measurand given combined random/systematic uncertainty evaluated by Type A and Type B methods
- “level of confidence” to describe the estimated probability that the measurand lies within that interval

Technology Tests

- Model: NIST "Proprietary Fingerprint Template Testing"

	DHS2	DOS	POE	POEBVA
D	0.9917	0.9845	0.9955	0.9932
F	0.9893	0.9944	0.9979	0.9979
H	0.9870	0.9978	0.9993	0.9994
I	0.9904	0.9978	0.9992	0.9992

- Measurand: TAR=(1-FNMR) at FMR=0.0001 for algorithm X against database Y
- Completely repeatable and reproduceable within hardware truncation and memory leakage limits
- Systematic uncertainty: Actual measurand (error rate against test key) is a proxy for stated measurand
- No "confidence intervals" because nothing is repeated under identical conditions and no data is random sample of a larger population.

Concluding Remarks on Uncertainty

1. "Uncertainty" is a broader concept than "error"; it is the doubt about how well the test result represents the quantity measured (or being said to be measured).
2. A central source of uncertainty is definitional incompleteness in specifying the "unit of empirical significance" for the measurand – full specification of which would require "infinite amount of information".
3. What we are measuring is often only a proxy for the measurand of real interest, even if fully defined, which adds yet another source of uncertainty in our measurement.
4. How we control, measure and report the values in a test must reflect how we expect those values to be used by others

Concluding Remarks on Today's Talk

- Three characteristics of science are:
  1. Reliance on real-world data
  2. Inductive generalities from specific observations
  3. A critical social structure
- Science values:
  1. Accuracy
  2. Consistency
  3. Broad scope
  4. Simplicity
  5. Fruitfulness
- We seek to take a scientific approach in developing best practices in biometrics

Concluding Remarks

- The term “biometrics” has been used historically to mean many different things.
- Our meaning is automated human recognition using behaviours and biology
- Our field has re-evaluated basic concepts in the last decade.
  - “Identity” is outside our scope
- Our form of “biometrics” can be used to verify that someone is recognized or that someone is not recognized.
- We can recognize people without knowing “who” they are

Concluding Remarks

- Biometric applications are much broader than access control.
- Biometric systems do not generally compete with PINs and passwords
- Fundamental challenges are within- and between-class variation (a “class” is an individual)
- There are now international standards for testing biometric systems.
- Nonetheless, tests differ because motivations for testing differ

Concluding Remarks

- Technical performance metrics are
  1. False match
  2. False non-match
  3. Throughput rate
  4. Failure to enrol
  5. Failure to acquire
- Performance metrics are inter-related and cannot be changed independently
- Tests can be classified as
  1. Technology
  2. Scenario
  3. Operational

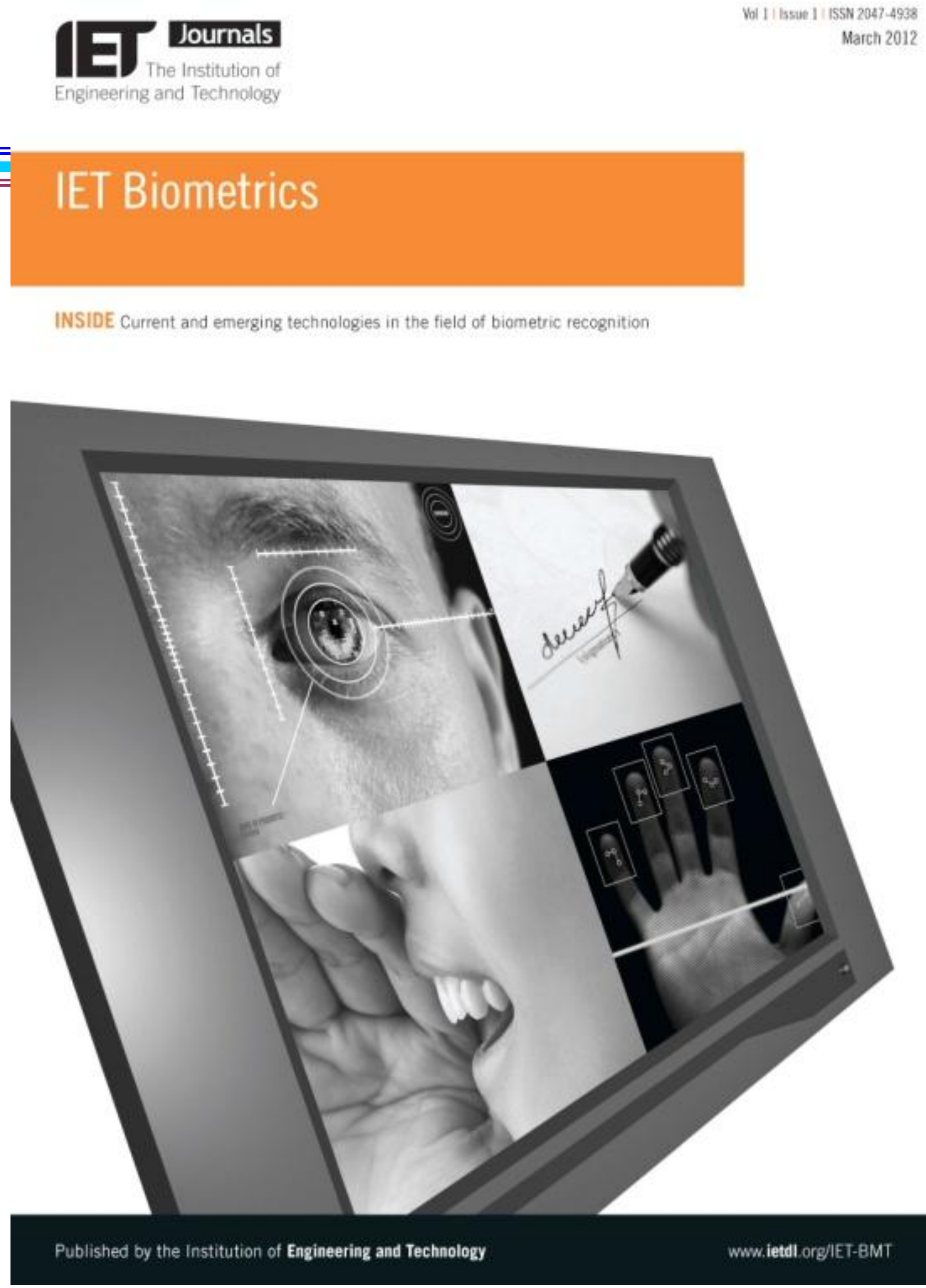
Concluding Remarks

- There is a test standard document for each type of test.
- Operational tests present special challenges with regard to "ground truth"
- Operational tests on two border control systems were discussed.
- Four different approaches have been proposed for estimating performance of large-scale systems
  1. Extrapolation from experience
  2. Identification as succession of N verifications
  3. Extrapolation from extreme value
  4. Extrapolation when distance can be modeled

Concluding Remarks

- The current standard for estimating uncertainty in laboratory measurements is the ISO/IEC Guide 98
- “Confidence intervals” have been replaced in our thinking with the broader concept of “coverage intervals”
  - Coverage intervals include both systematic and random error
  - Coverage intervals are estimated using both mathematical and expert techniques
- “Biometrics” in the 21st Century has been characterized by fundamental change and advancement.

A new journal for the biometrics community



Latest news ....

The first issue is complete and will appear end March/early April 2012 (ie. is imminent!)

Papers are now being accepted for future issues in Volume 1