Biometric Performance
Biometric Systems (DTU 02238)

Christoph Busch

Session 7 and 8

DTU

Overview Biometric Performance

Structure of this session

- Issues in Performance Tests

- Fundamental Metrics

- Reporting

- Standards

Categorization of Biometric Systems

Costs

- Installation- and maintenance costs

Operating expense

- Duration (transaction time) and complexity of operation
- Adaptation time

Biometric performance

- How precise does the system recognize individuals?
- What are the error rates?

Presentation Attack Detection (PAD)

- Does the system detect artefacts
  of biometric characteristics (a.k.a fakes)

Issues in Performance Tests
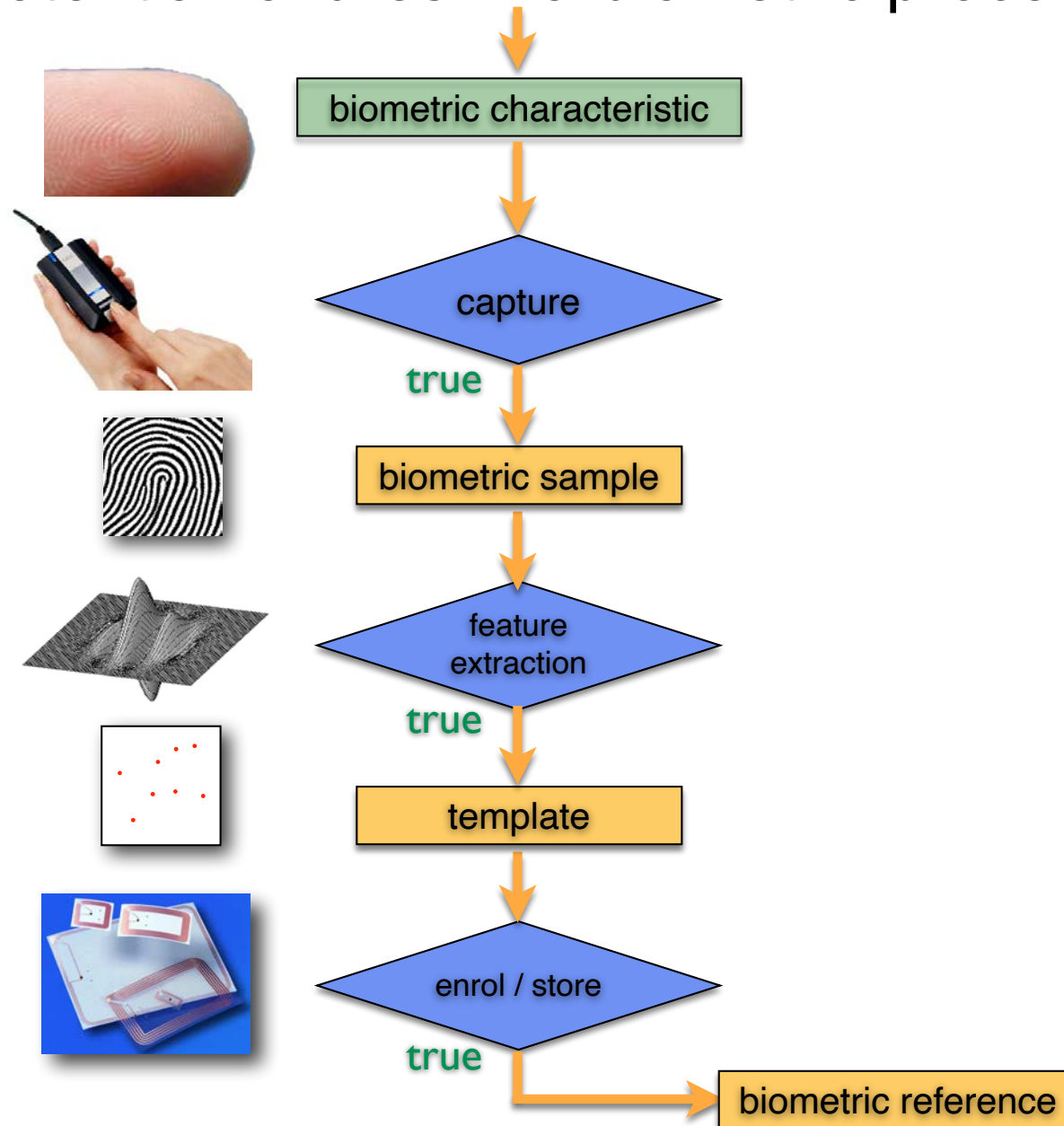
Performance Test and Errors

Why measure failures?

- Murphy' Law teaches us:
  „Whatever can go wrong, will go wrong.“

  ▸ It is just a matter of the point in time ...

  ▸ ... and the likelihood
  that a failure happens

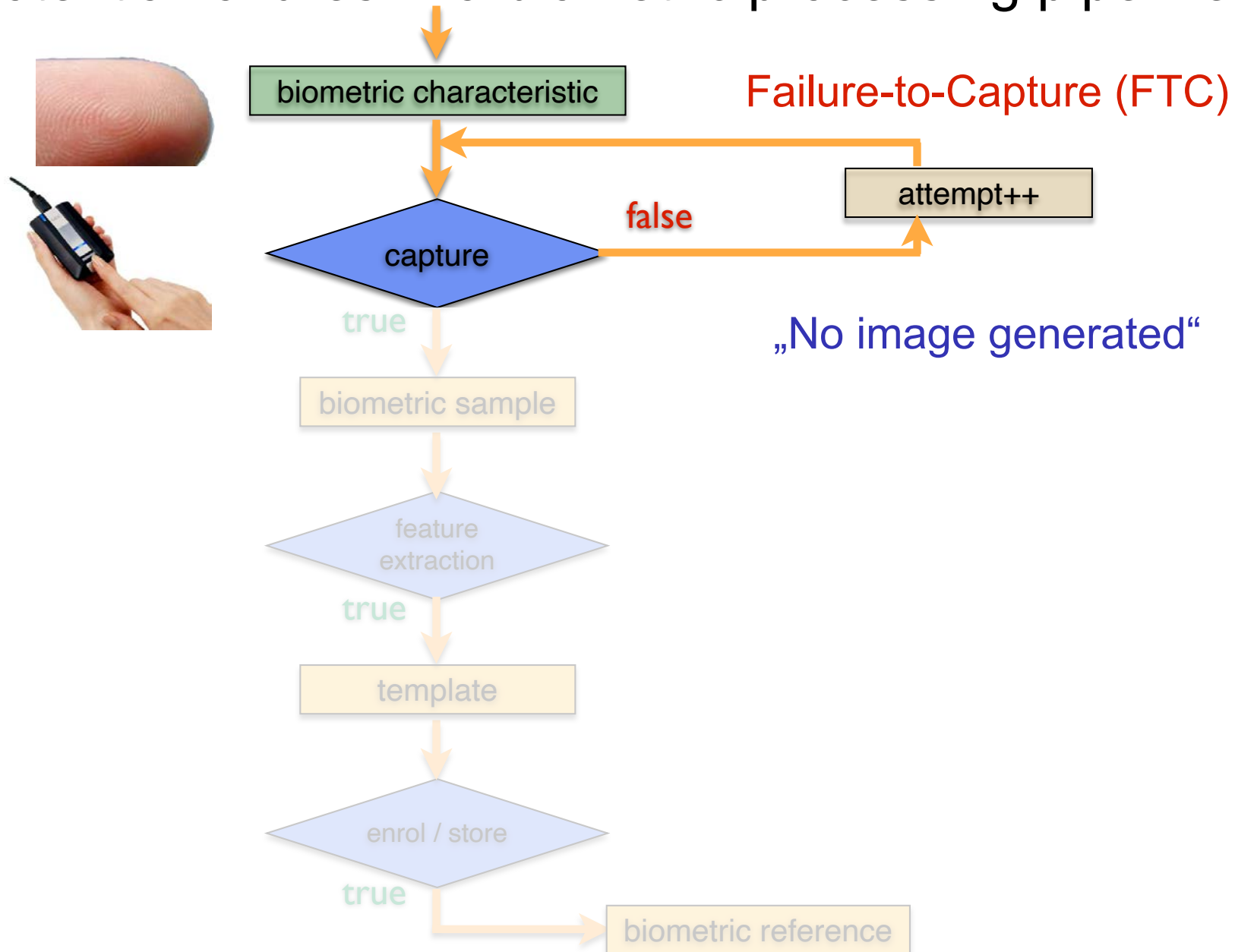 ▸ There are small failures ...

  ▸ ... and larger disasters
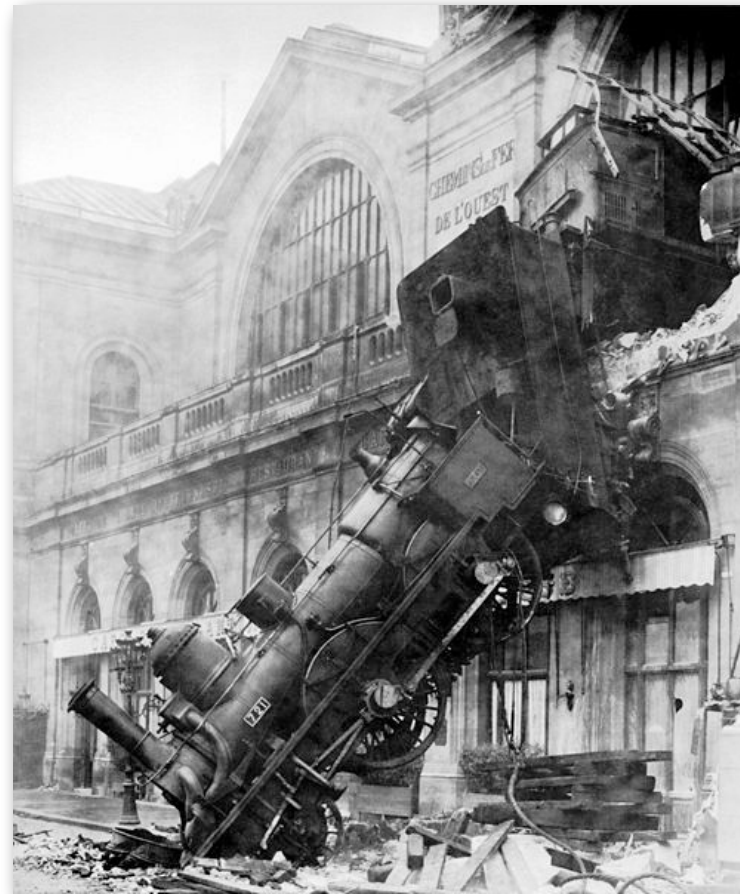
Potential failures in a biometric processing pipeline
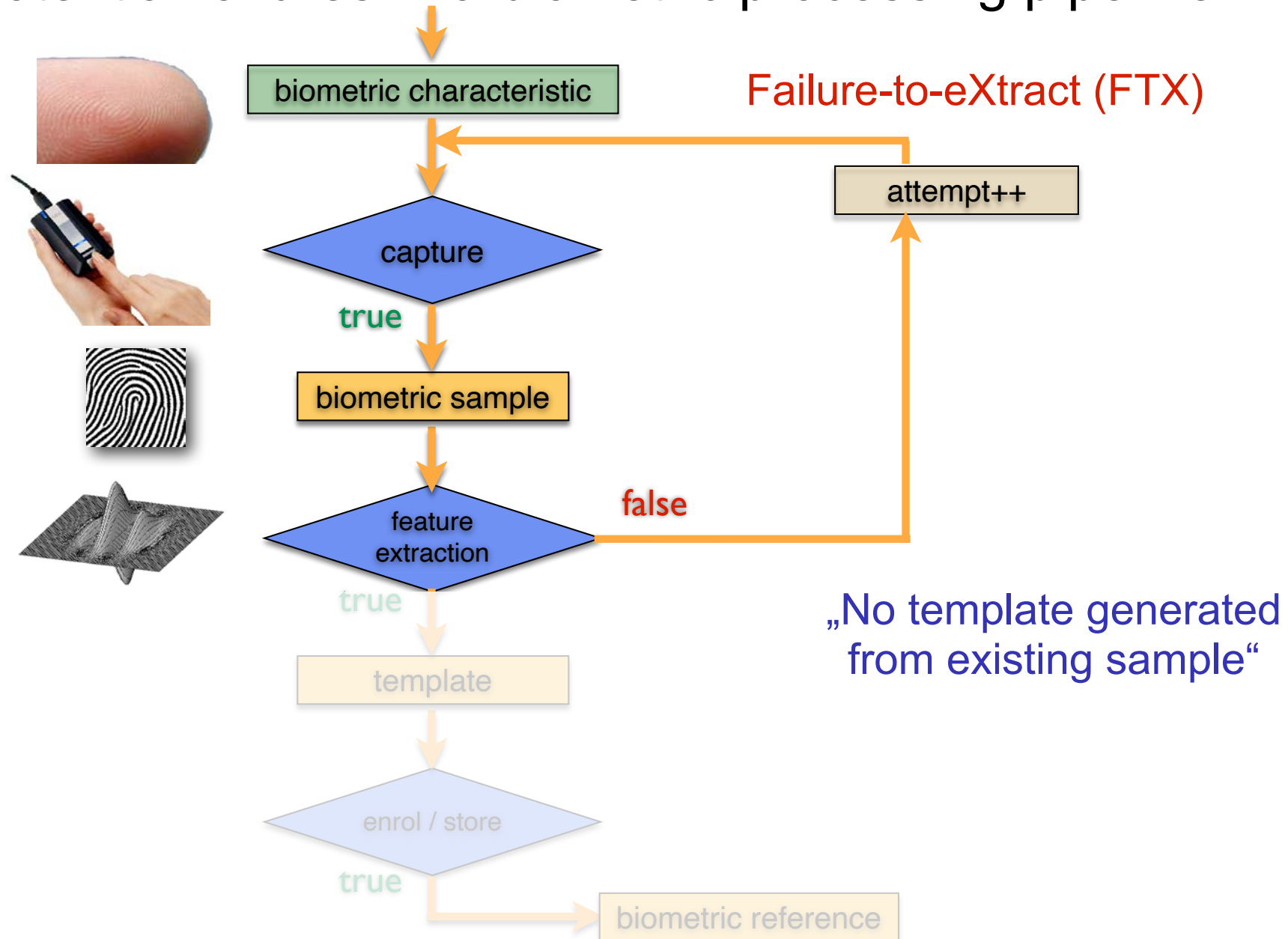
Failures

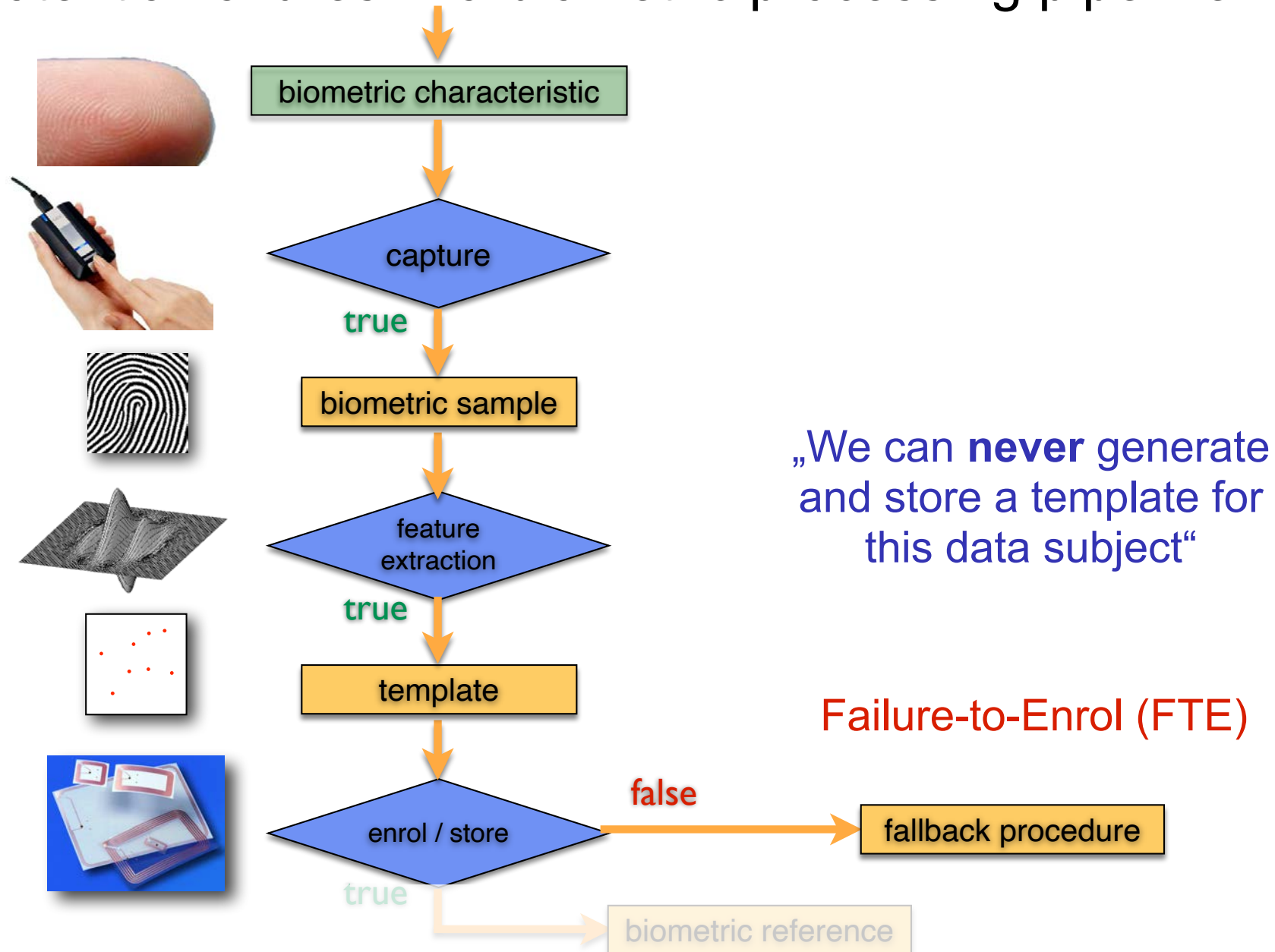Potential failures in a biometric processing pipeline

biometric characteristic

Failure-to-Capture (FTC)

capture

false

attempt++

true

„No image generated"

biometric sample

feature extraction

true

template

enrol / store

true

biometric reference

Failures

Potential failures in a biometric processing pipeline

biometric characteristic

Failure-to-eXtract (FTX)

capture

true

biometric sample

attempt++

feature extraction

false

true

template

enrol / store

true

biometric reference

„No template generated from existing sample"

Failures

Potential failures in a biometric processing pipeline

biometric characteristic

capture

true

biometric sample

feature extraction

true

template

enrol / store

false → fallback procedure

true

biometric reference

„We can **never** generate and store a template for this data subject"

Failure-to-Enrol (FTE)

Performance Metrics

Failure-to-Capture Rate (FTCR)

aligned to ISO-HBV: *proportion of failures of the biometric capture process to produce a captured biometric sample of the biometric characteristic of interest*

$$FTCR = \frac{N_{tca} + N_{nsq}}{N_{tot}}$$

where

N_{tca} is the number of terminated capture attempts

N_{nsq} is the number of images created
with insufficient sample quality

N_{tot} is the total number of capture attempts

Definition of FTC: https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-37:ed-2:v1:en:term:3.9.5

Performance Metrics

Failure-to-eXtract Rate (FTXR)

Def: *proportion of failures of the feature extraction process to generate a template from the captured biometric sample*

$$FTXR = \frac{N_{ngt}}{N_{sub}}$$

where

N_{ngt} is the number of cases where no template was generated

N_{sub} is the total number of biometric samples
being submitted to the feature extraction

Performance Metrics

Failure-to-Enrol Rate (FTER)

aligned to ISO-HBV: *proportion of a specified set of biometric enrolment transactions that resulted in a failure to create and store a biometric enrolment data record*

$$FTER = \frac{N_{nec}}{N}$$

where

N_{nec} is the number of cases where the biometric characteristic of the subject cannot be captured at all

N is the total number of subjects intended to be enroled in the biometric application

Definition of FTER https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-37:ed-2:v1:en:term:3.9.7

Performance Metrics

Failure-to-Acquire Rate (FTAR)

aligned to ISO-HBV: *proportion of a specified set of biometric acquisition processes that were failure to accept for subsequent comparison the output of a data capture process*

Note: This is caused by either a FTCR or a FTXR in the
        in the verification process. No probe feature vector.

$$FTAR = FTCR + FTXR * (1 - FTCR)$$

Definition of FTAR https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-37:ed-2:v1:en:term:3.9.4

Issues in Performance Tests

Biometric performance for a failure to occur

- specified in proportions (error rates)
- assumes random nature. Correct?

What else can cause failures?

- Failures / errors caused by:
  - ‣ Violation of persistence property
    - variations of the observed biometric characteristic
    - ageing, spontaneous variations
  - ‣ Environmental factors
    - illumination, background noise, …
  - ‣ Quality of the samples

Issues in Performance Tests

Validate the test data - before you fix the corpus

- Control the quality of the corpus
  - „Ground truth" could be error prone
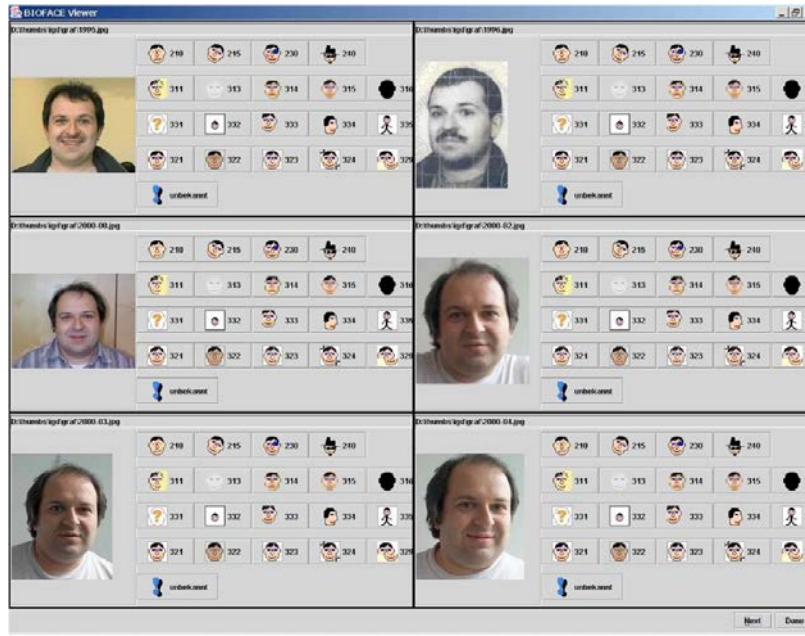
  - empty chair
- Examples of typical testing mistakes
  - Fingerprint images assigned to the wrong subject ID
  - Volunteers using multiple identities

Issues in Performance Tests

Scientific approach

- Separation of training/development data versus test data

- Avoid to use artificially generated data (synthetic data)

- Size of data base must be sufficient
  for the claimed error rates
  - ‣ rule of 3

Test Size versus Uncertainty

Rule of 3

- The Rule of 3 addresses the question
  "*What is the lowest error rate that can
  be statistically established with a given number N
  of independent identically distributed comparisons?*"

- This value is the error rate *p* for which
  the probability of zero errors in *N* trials, purely by chance,
  is (for example) 5%.

- This gives:

$$p \approx \frac{3}{N}$$

 for a 95% confidence level

Fundamental Metrics
in Technology Testing

Basic Metrics

Comparison scores

- **comparison score** $c(Q,R)$**:**
  - ▸ *numerical value (or set of values) resulting from a comparison*
    NOTE: the term „~~matching score~~" is deprecated by ISO

- **similarity score** $s(Q,R)$ **:**
  - ▸ *comparison score that increases with similarity*

- **distance score / dissimilarity score** $d(Q,R)$**:**
  - ▸ *comparison score that decreases with similarity*
  - ▸ Some distance measure
  - ▸ Often $d(Q, R) \geq 0$ and $d(R, R) = 0$

- Conversion
  - ▸ $s = f(d)$ where f is a monotonically decreasing function
  - ▸ Examples

  $$s = -d \qquad s = -log(d) \qquad s = \frac{1}{d}$$

Basic Metrics

Graphical representation of results

- Probability Density Function (PDF)
  for similarity scores *s(Q,R)*

  ▸ t - threshold = 45

  ▸ m - mated samples (genuine)

  ▸ nm - non-mated samples (impostor)
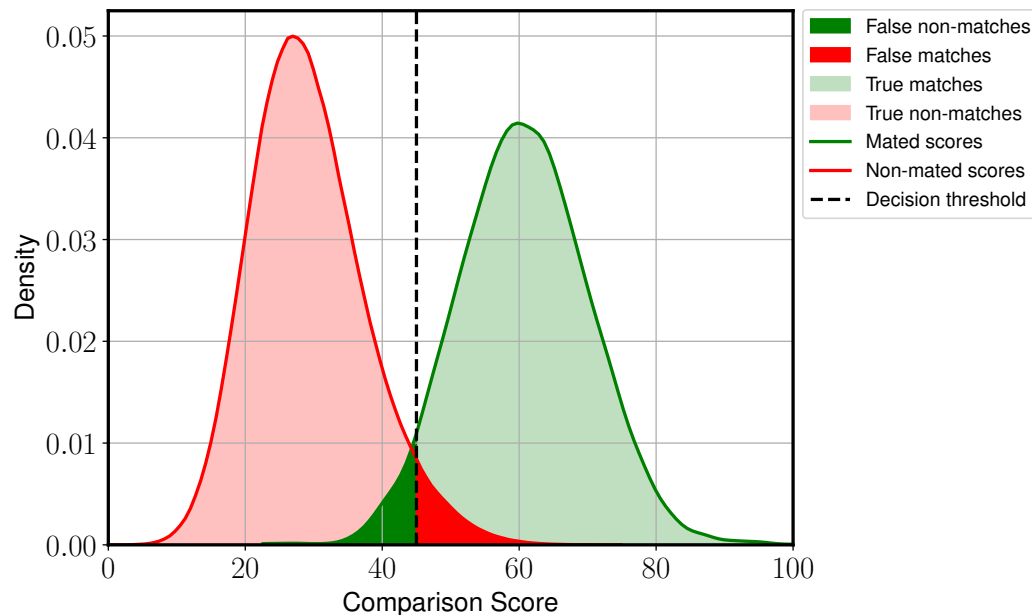
Reference Images R

	$s_m=70$ True positive	$s_{nm}=40$ True negative
Probe Image Q	$s_{nm}=55$ False positive	$s_m=42$ False negative

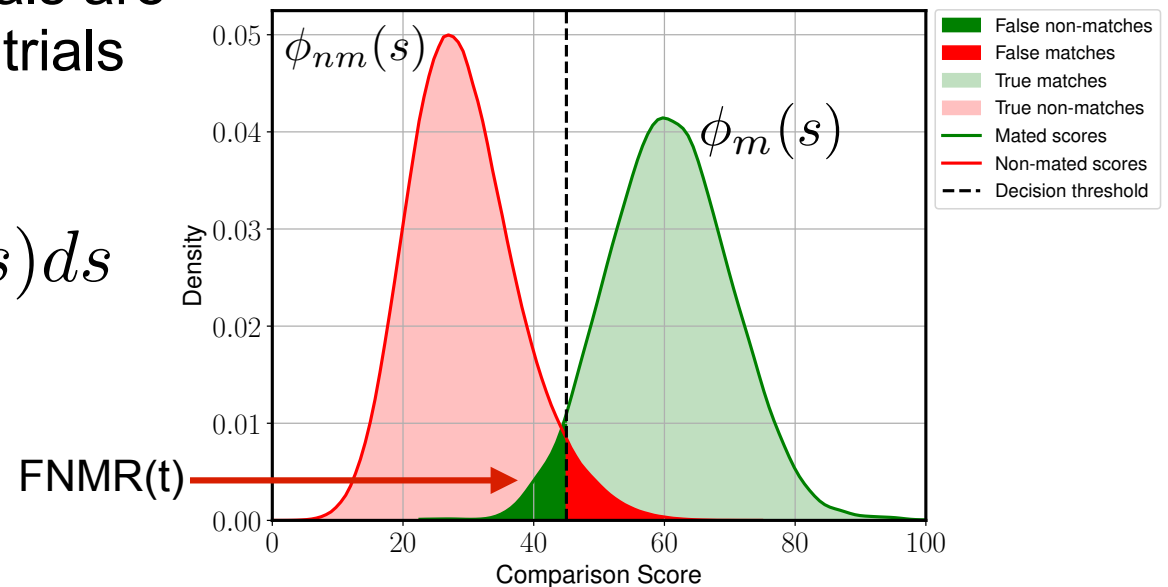## Probability density Distribution Function (PDF)

$\phi_m(s)$: PDF of mated similarity score $\quad s(Q, R)$

$\phi_{nm}(s)$: PDF of non-mated similarity score $s(Q, R)$

False-Match-Rate (FMR)

- **Def in ISO-HBV:** *proportion of the completed biometric non-mated comparison trials that result in a false match*

- Note: non-mated comparison trials are also referred to as impostor trials

$$FMR(t) = \int_t^1 \phi_{nm}(s) ds$$
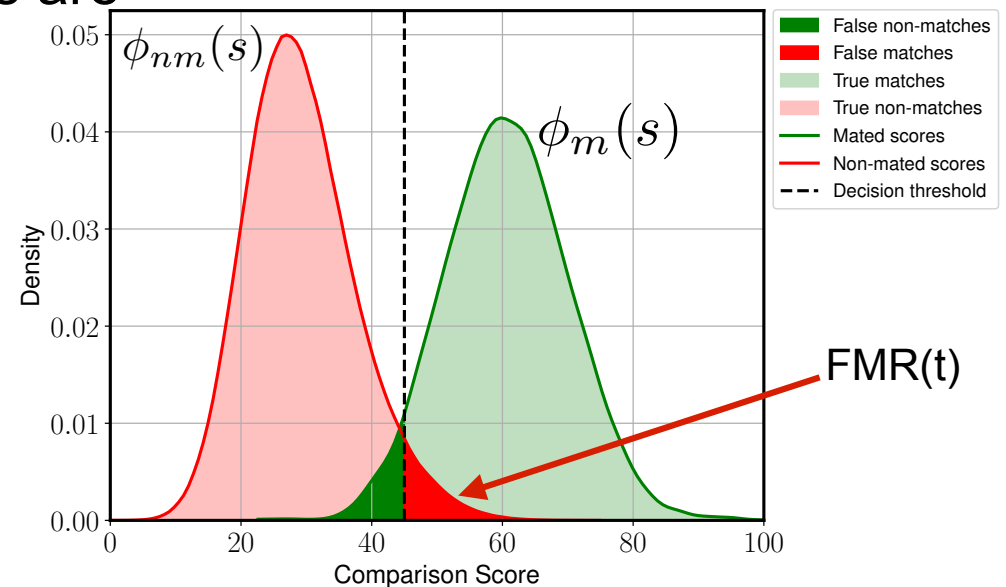
Performance Metrics - Convenience

Probability density Distribution Function (PDF)

$\phi_m(s)$: PDF of mated similarity score $\quad s(Q, R)$

$\phi_{nm}(s)$: PDF of non-mated similarity score $\quad s(Q, R)$

False-Non-Match-Rate (FNMR)

- **Def in ISO-HBV:** *proportion of the completed biometric mated comparison trials that result in a false non-match*

- Note: mated comparison trials are also referred to as genuine trials

$$FNMR(t) = \int_0^t \phi_m(s)\,ds$$

Probability density Distribution Function (PDF)

$\phi_m(d)$: PDF of mated dissimilarity score $\qquad d(Q, R)$

$\phi_{nm}(d)$: PDF of non-mated dissimilarity score $\quad d(Q, R)$

$$FNMR(t) = \int_0^t \phi_{nm}(d)dd$$

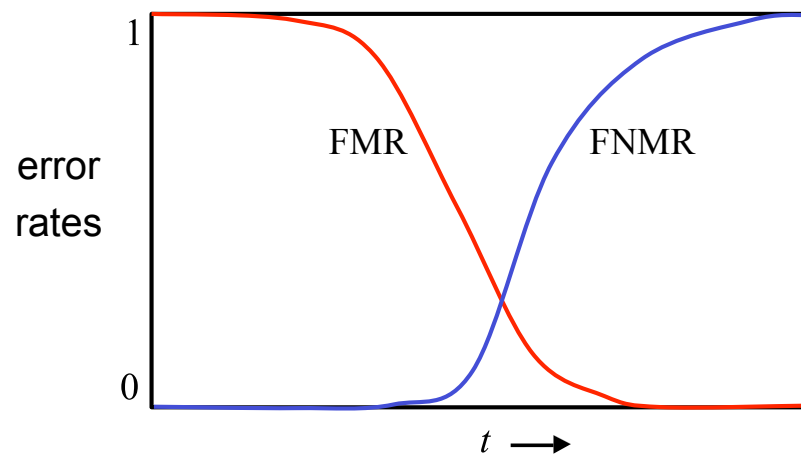$$FNMR(t) = \int_t^{max_d} \phi_m(d)dd$$
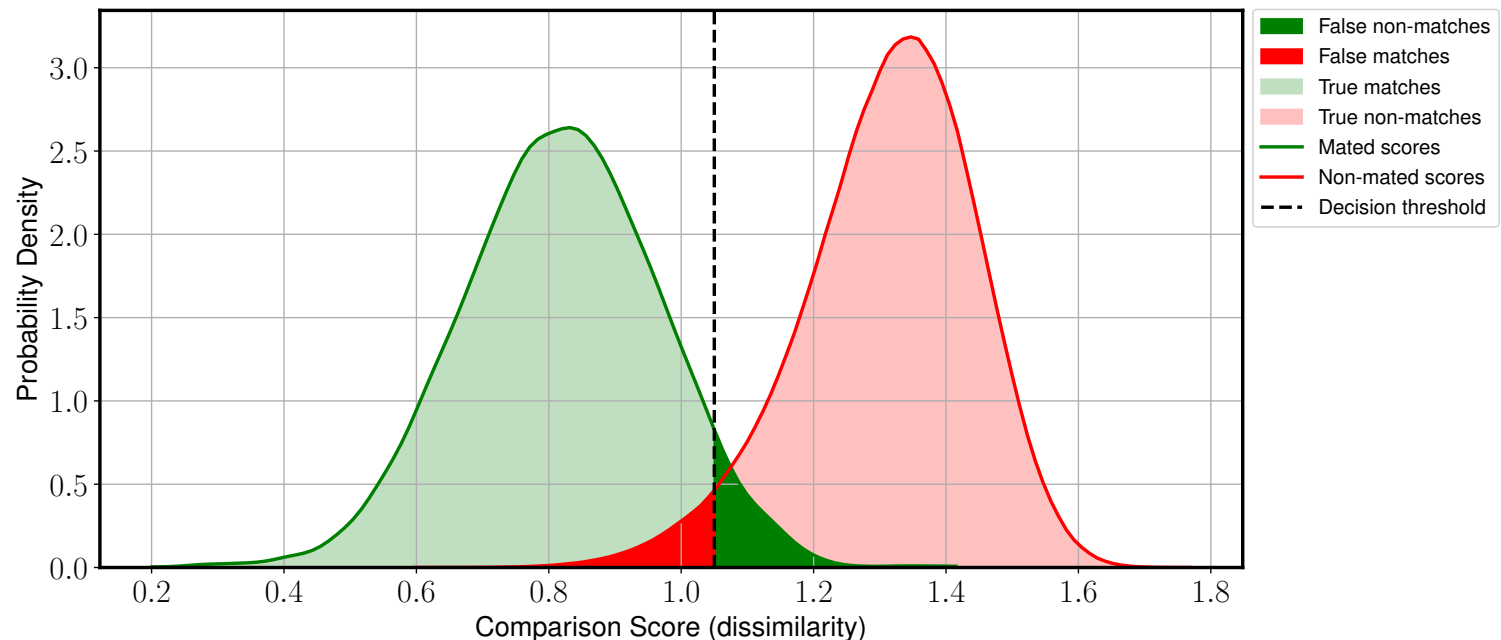
Performance Metrics

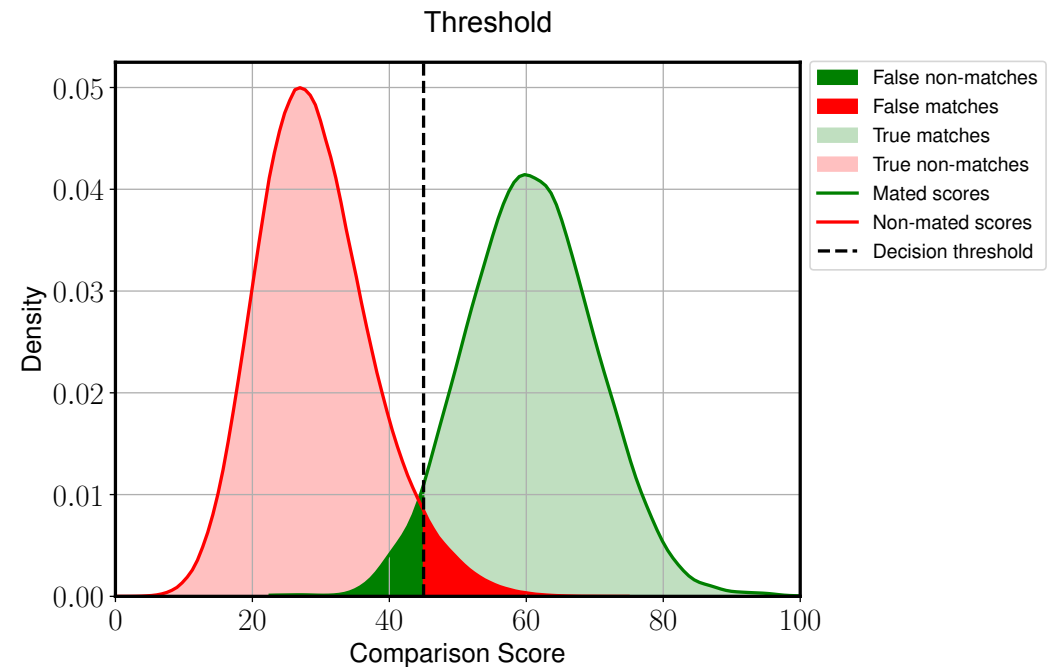## Algorithm error rates

- False-Match-Rate (FMR) - often confused with FAR

- False-Non-Match-Rate (FNMR) - often confused with FRR

- Equal-Error-Rate (EER)

Constraints:

$$FMR(0) = 1, FMR(1) = 0$$

$$FNMR(0) = 0, FNMR(1) = 1$$

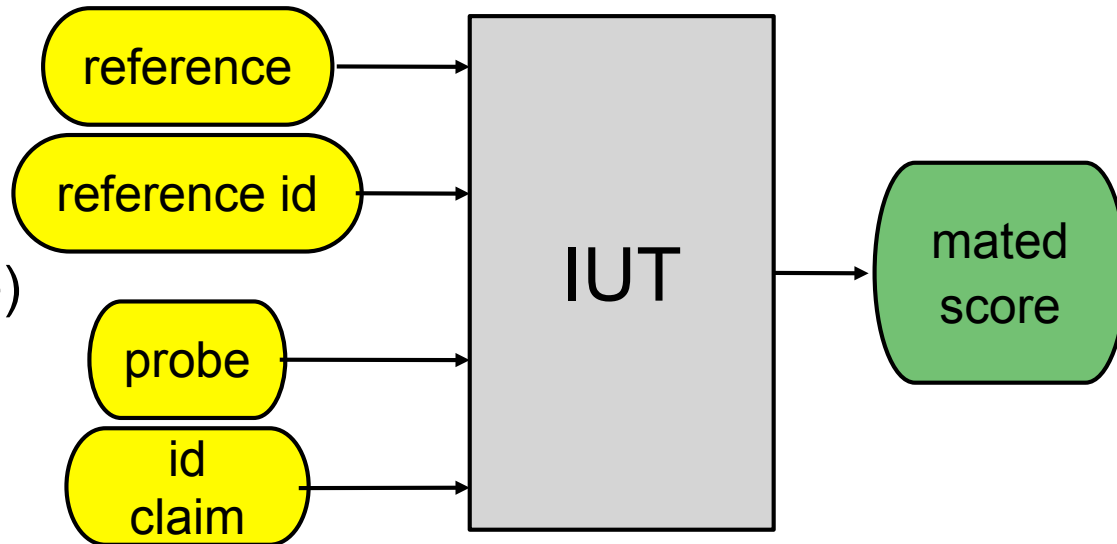# More Metrics

Verification rates
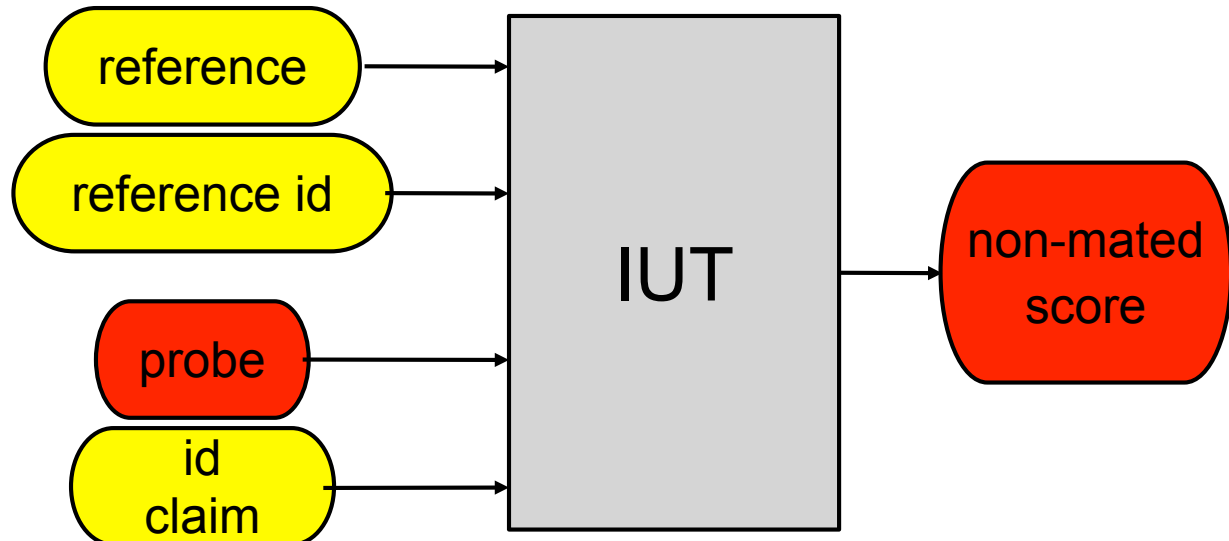
- Genuine-Match-Rate (GMR)
    - ▸ true positives

    $GMR = 1 - FNMR$

Performance Evaluation

**Mated
comparison trial**
(taken from the
 same biometric instance)

**Non-mated
comparison trial**
(taken NOT from the
 same instance)

References and probes are taken from a labeled test set

Similarity Matrix (closed set)

Similarity scores - comparator

- 3 subjects (3 instances) captured in 2 sessions
  - ▸ 3 mated comparison scores
  - ▸ (3-1) * 3 non-mated comparison scores

	face$_1$	face$_2$	face$_3$	enrolment samples
face$_1$	0.98	0.59	0.36	
face$_2$	0.71	0.65	0.43	
face$_3$	0.23	0.69	0.72	

probe samples

Similarity Matrix (closed set)

instance-ID

1   2 N

mated comparison trials (genuine scores)

non-mated comparison trials (impostor scores)

N — instances (IDs)
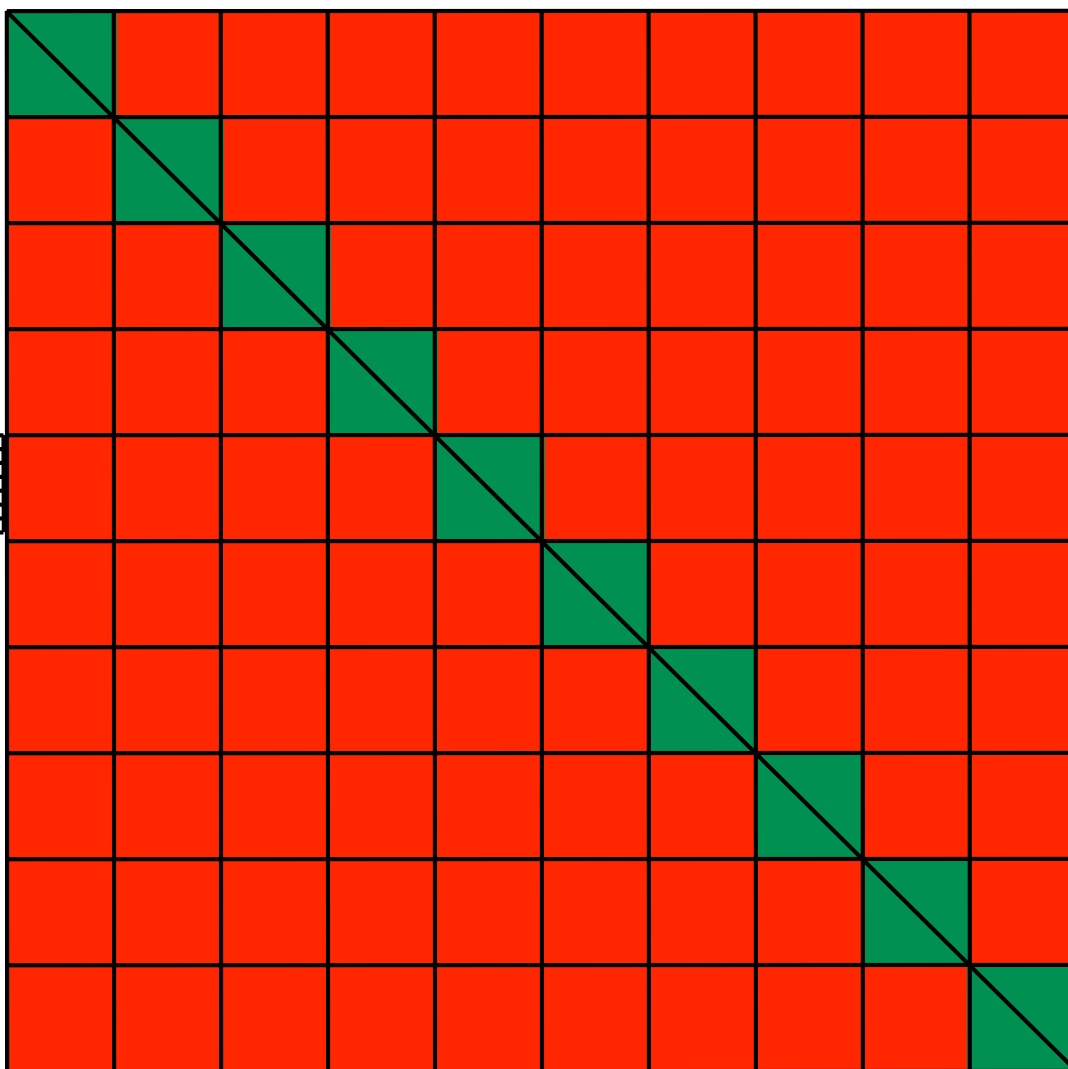
M — samples

$U = \dfrac{M}{N}$ — samples/instance

$M*M$ — scores in the similarity matrix

$(U^2 - U) * N$ — mated scores

$M * ((N - 1) * U)$ — non-mated scores

Computation of FMR, FNMR and GMR

Computing algorithm error on a given corpus

Ω_g set of all mated comparison scores (genuine)

Ω_i set of all non-mated comparison scores (impostor)

$\Omega_g(t)$ set of all mated scores $s > t$

$\Omega_i(t)$ set of all non-mated scores $s > t$

$\|\Omega\|$ number of elements in Ω

$$\mathrm{FMR}(t) = \frac{\|\Omega_i(t)\|}{\|\Omega_i\|}$$

$$\mathrm{GMR}(t) = \frac{\|\Omega_g(t)\|}{\|\Omega_g\|}$$

$$\mathrm{FNMR}(t) = 1 - \mathrm{GMR}(t)$$

Computing FNMR and FMR

From similarity scores to algorithm errors

- for a defined threshold t = **0.66**

- False-Match-Rate ?

- Genuine-Match-Rate ?

- False-Non-Match-Rate?

	face$_1$	face$_2$	face$_3$
face$_1$	0.98	0.59	0.36
face$_2$	0.71	0.65	0.43
face$_3$	0.23	0.69	0.72

Computing FNMR and FMR

From similarity scores to algorithm errors

- for a defined threshold t = **0.73**
- False-Match-Rate ?

- Genuine-Match-Rate ?

- False-Non-Match-Rate?

	face$_1$	face$_2$	face$_3$
face$_1$	0.98	0.59	0.36
face$_2$	0.71	0.65	0.43
face$_3$	0.23	0.69	0.72

Open Set Identification Error Rates

Confusion: (non-enrolled) subject i is identified as subject j

- False positives – there is always a closest enrolled subject !
- Countermeasure: include threshold.
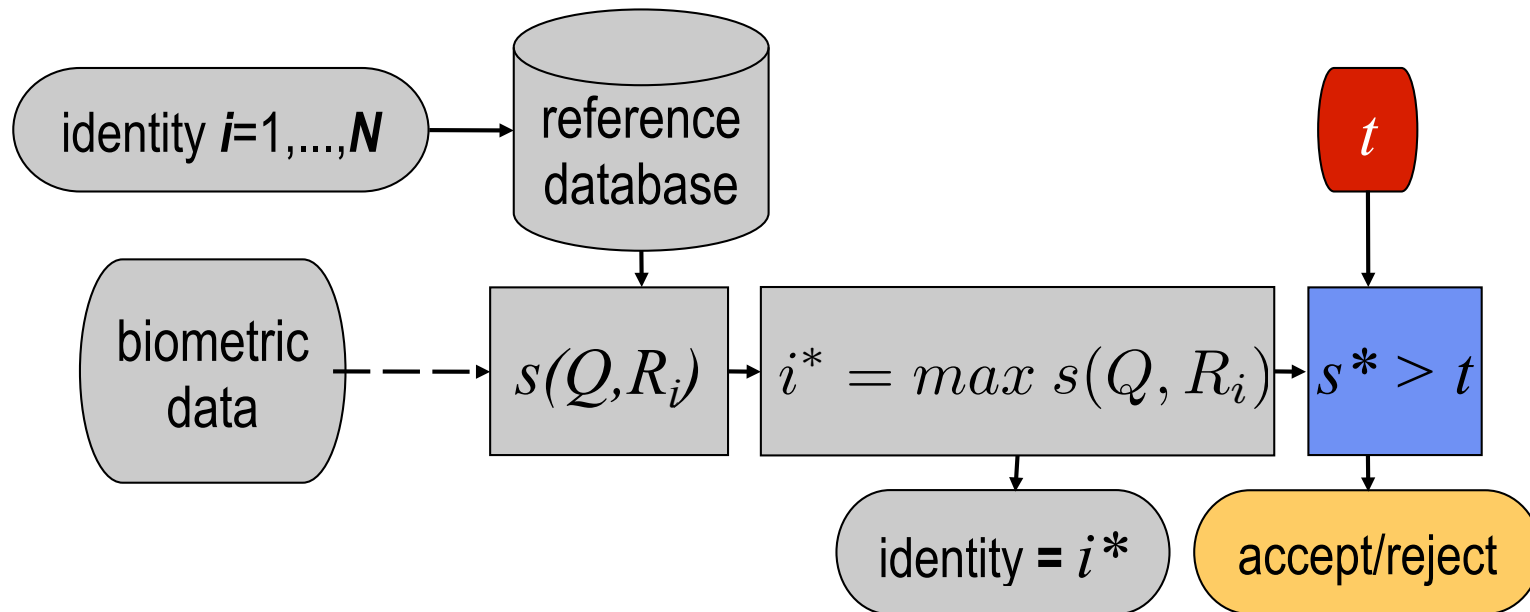
False Positive Identification Rate (FPIR):

proportion of identification transactions by capture subjects
not enrolled in the system for which a reference identifier is returned

Open Set Identification Error Rates

False-Positive-Identification-Rate (FPIR)

$$FPIR = (1 - FTAR) * (1 - (1 - FMR)^N)$$

- for small *FMR* we can substitute

$$(1 - FMR)^N \approx 1 - N * FMR$$

- and thus the *FPIR* can be approximated with

$$FPIR = (1 - 0) * (1 - (1 - N * FMR)$$

$$FPIR = N * FMR$$

Take care, when implementing identification systems, as the error will increase about linearly with the size *N* !

Overview Metrics

From algorithm testing to system level testing

- Consider different types of test
  - ▸ Technology testing
    - offline evaluation of one or more algorithms
  - ▸ Scenario testing
    - evaluation of simulated application
  - ▸ Operational testing
    - evaluation in which the performance of
      a complete biometric system is determined

Overview Metrics

From algorithm testing to system level testing

- Technology testing
  - ▸ Algorithmic level verification error
    - False-Match-Rate (FMR) - algorithm accepts „zero-effort" impostor
    - False-Non-Match-Rate (FNMR) - algorithm rejects true identity
- Scenario testing and operational testing
  - ▸ System level verification error
    - False-Accept-Rate (FAR)
    - False-Reject-Rate (FRR)
- System level error requires observation of:
  - ▸ Sample generation: Failure-to-Capture-Rate (FTCR)
  - ▸ Enrolment: Failure-to-Enrol-Rate (FTER)
    no reference for this subject
  - ▸ Verification: Failure-to-Acquire-Rate (FTAR)
    no probe feature vector

System Error Metrics

False-Accept-Rate (FAR)

$$FAR = FMR * (1 - FTAR)$$

False-Reject-Rate (FRR)

$$FRR = FTAR + FNMR * (1 - FTAR)$$

System Error Metrics
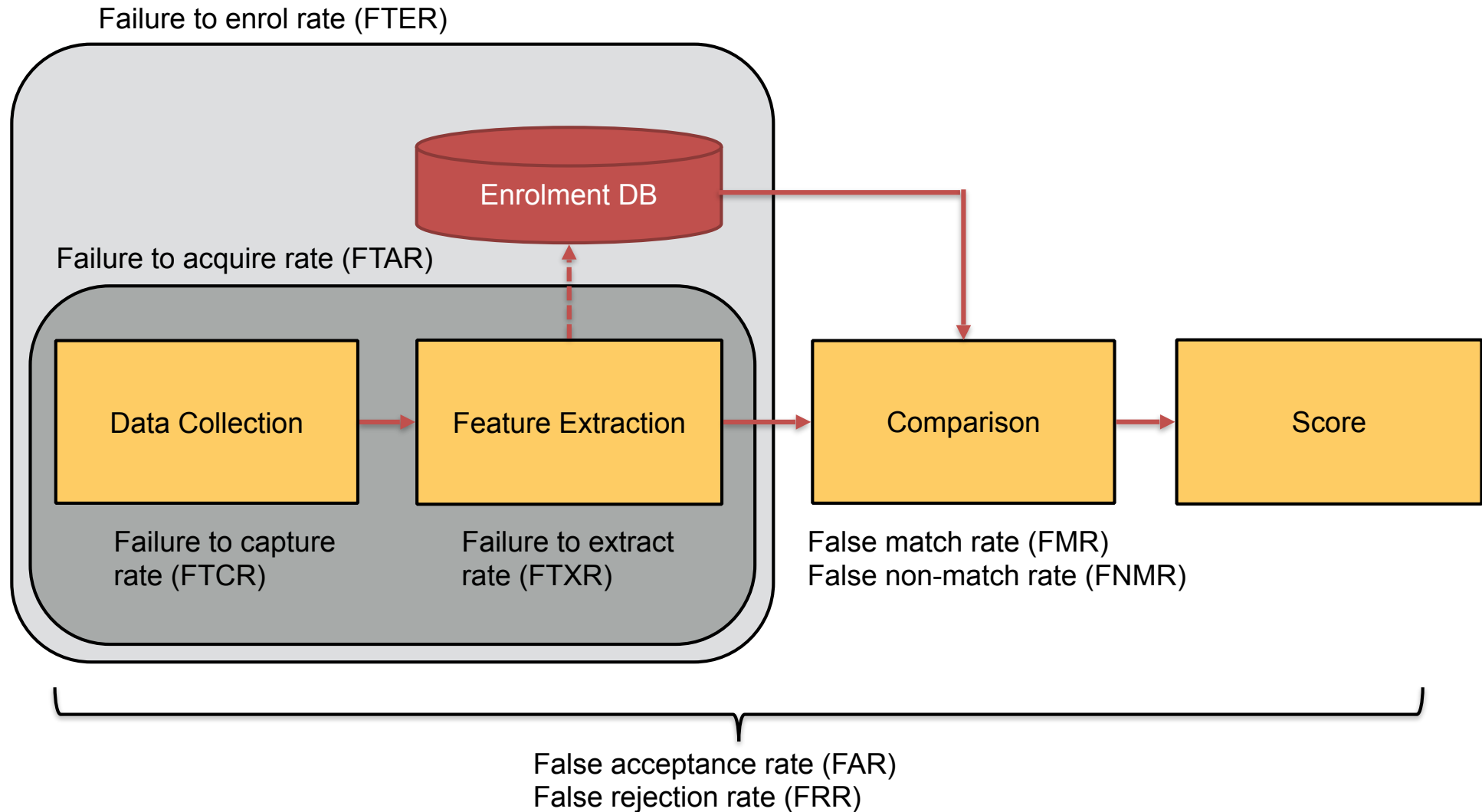
Generalized False-Accept-Rate:

$$GFAR = FMR * (1 - FTAR) * (1 - FTER)$$

Generalized False-Reject-Rate:

$$GFRR = FTER + (1 - FTER) * FTAR + (1 - FTER) * (1 - FTAR) * FNMR$$

## Potential failures in a biometric processing pipeline

Failure to enrol rate (FTER)
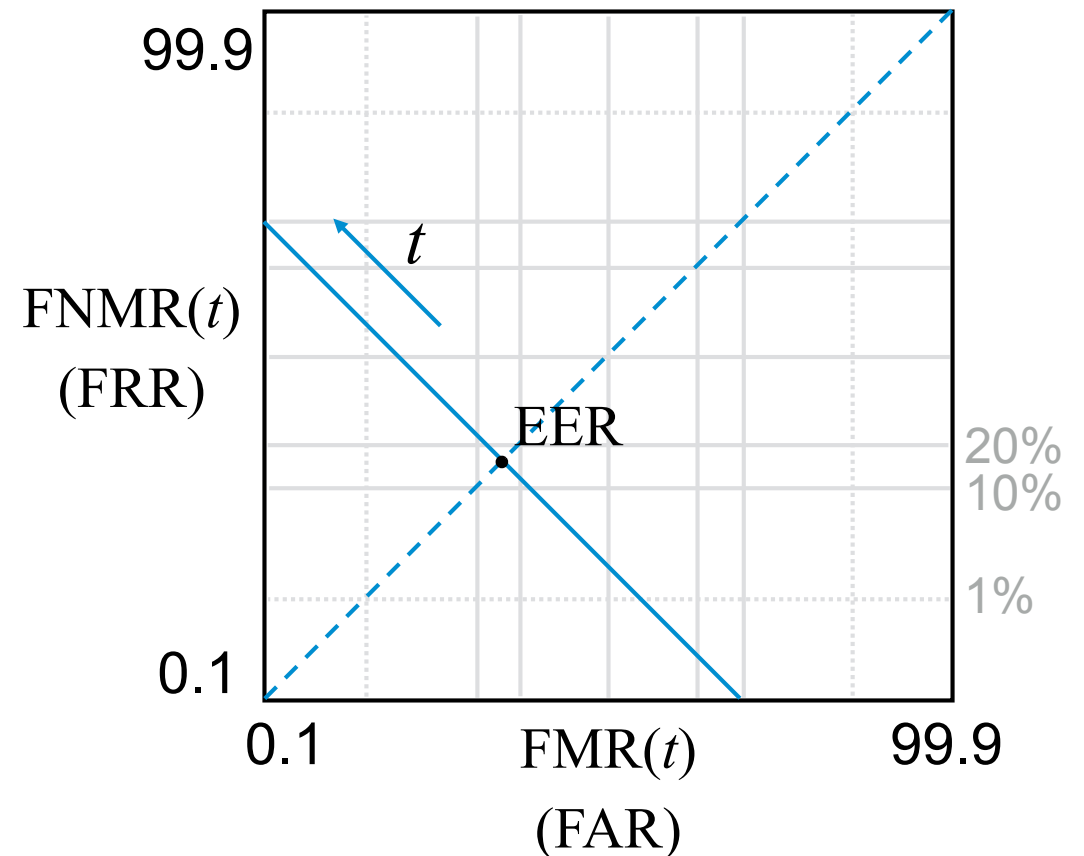
Failure to acquire rate (FTAR)

Enrolment DB

Data Collection

Feature Extraction

Comparison

Score

Failure to capture rate (FTCR)

Failure to extract rate (FTXR)

False match rate (FMR)
False non-match rate (FNMR)

False acceptance rate (FAR)
False rejection rate (FRR)

Reporting

Graphical Presentation

Receiver Operating Characteristic (ROC)

GMR(t) (GAR)

t

0

1

0   FMR(t)   1
(FAR)

20%
10%
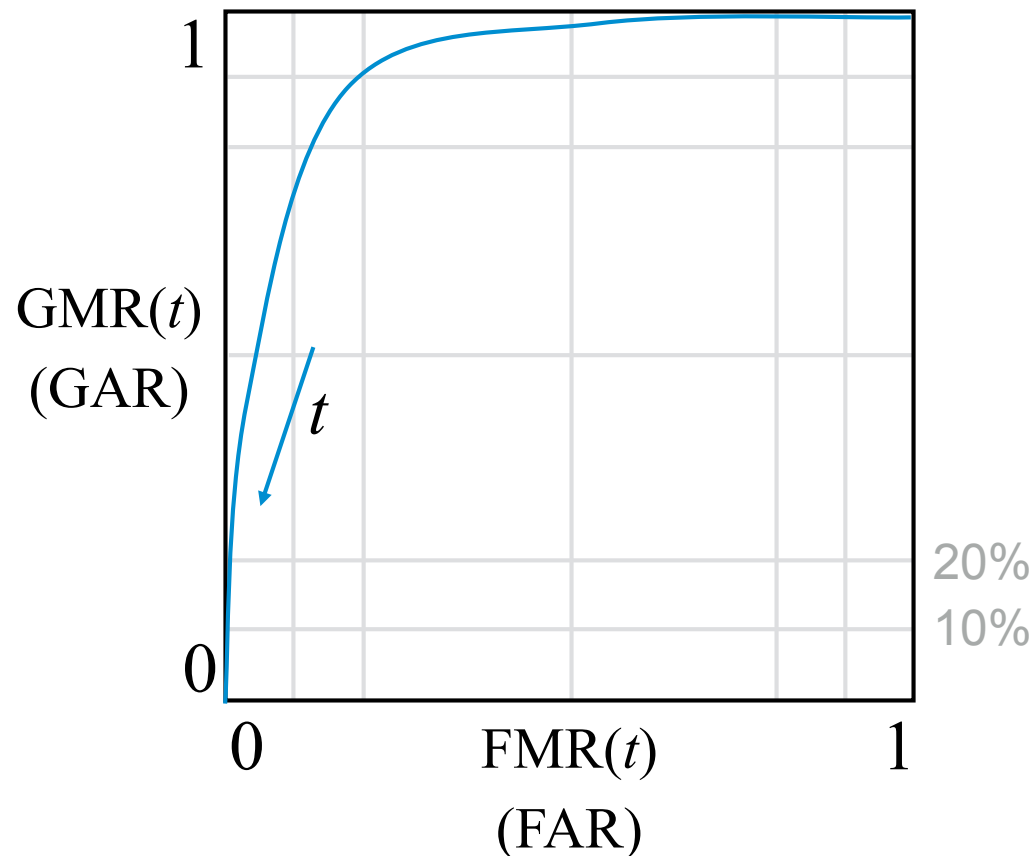
Detection Error Trade-off (DET) curve

FNMR(t) (FRR)

t

EER

99.9

0.1

0.1   FMR(t)   99.9
(FAR)

20%
10%

1%

Graphical Presentation

DET curve (detection error trade-off curve)

- plots error rates on both axes
  - false positives on the x-axis
 - false negatives on the y-axis

Source: ISO/IEC 19795-1:2006
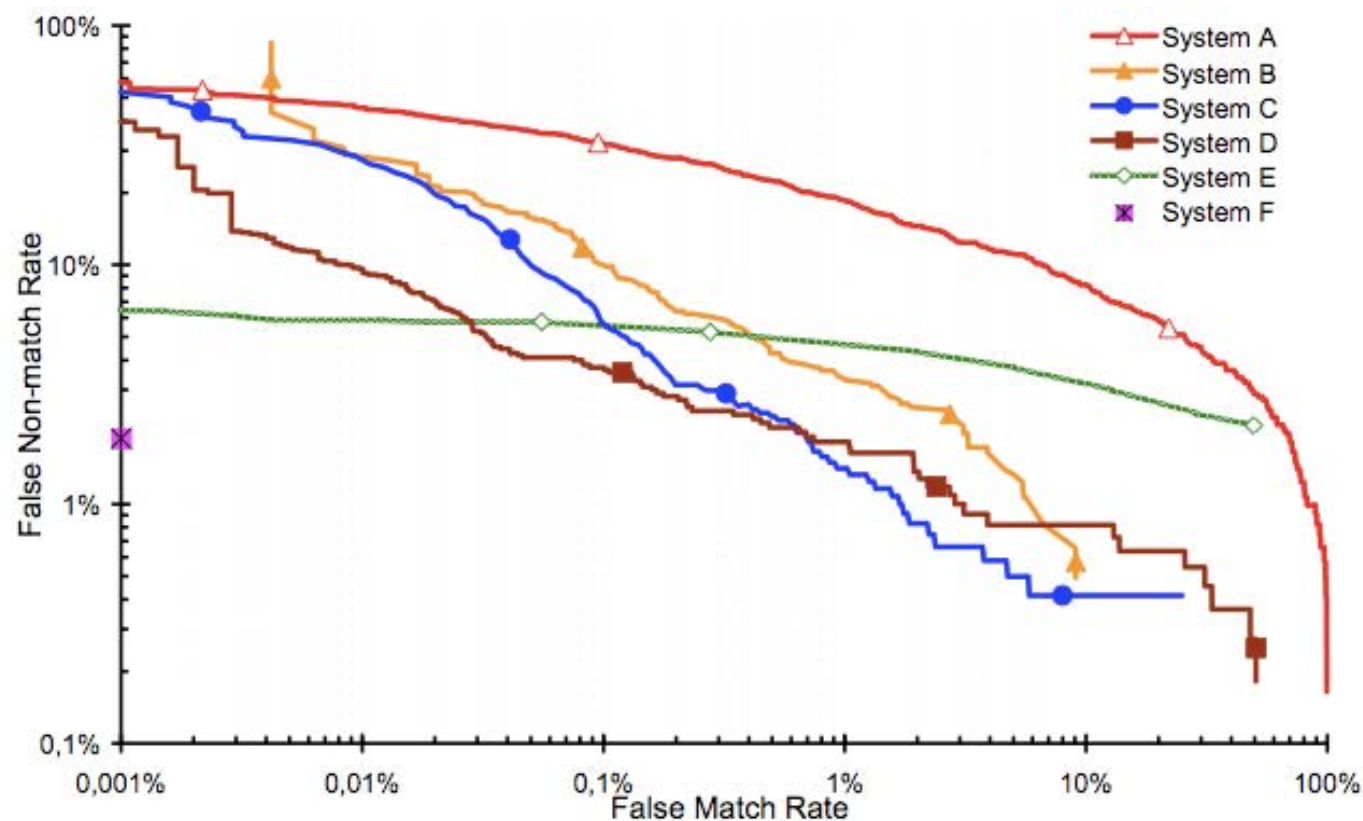
Graphical Presentation

DET curve

- plots are impacted by the scale

a) Normal deviate scale

b) Logarithmic scale (log$_{10}$)

c) Linear scale (not recommended)

d) Log-linear scale

Source: ISO/IEC 19795-1:2021
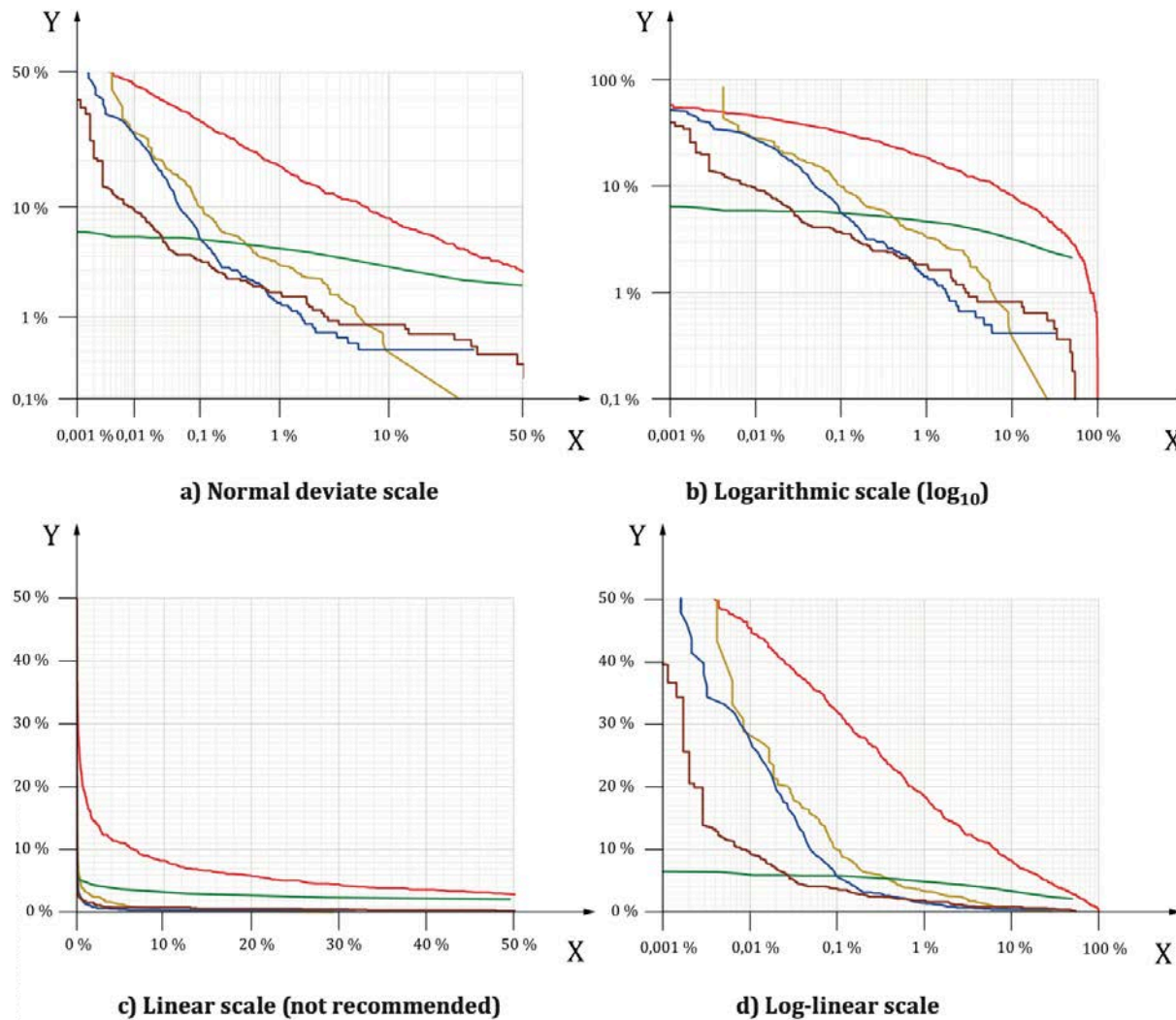
Graphical Presentation

Visualising score distribution

- boxplots and whisker plot
  - ▸ the boxes represent the interquartile range of each set of scores
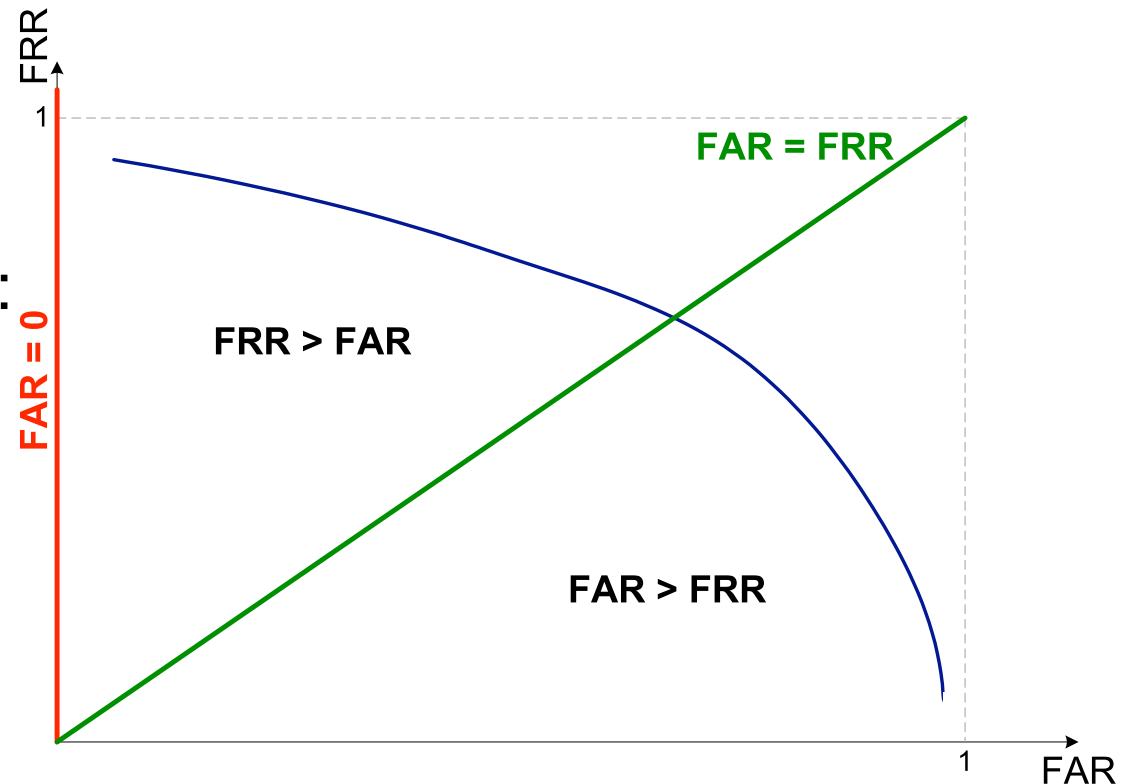 - ▸ the whisker extend from the box to the highest and lowest score
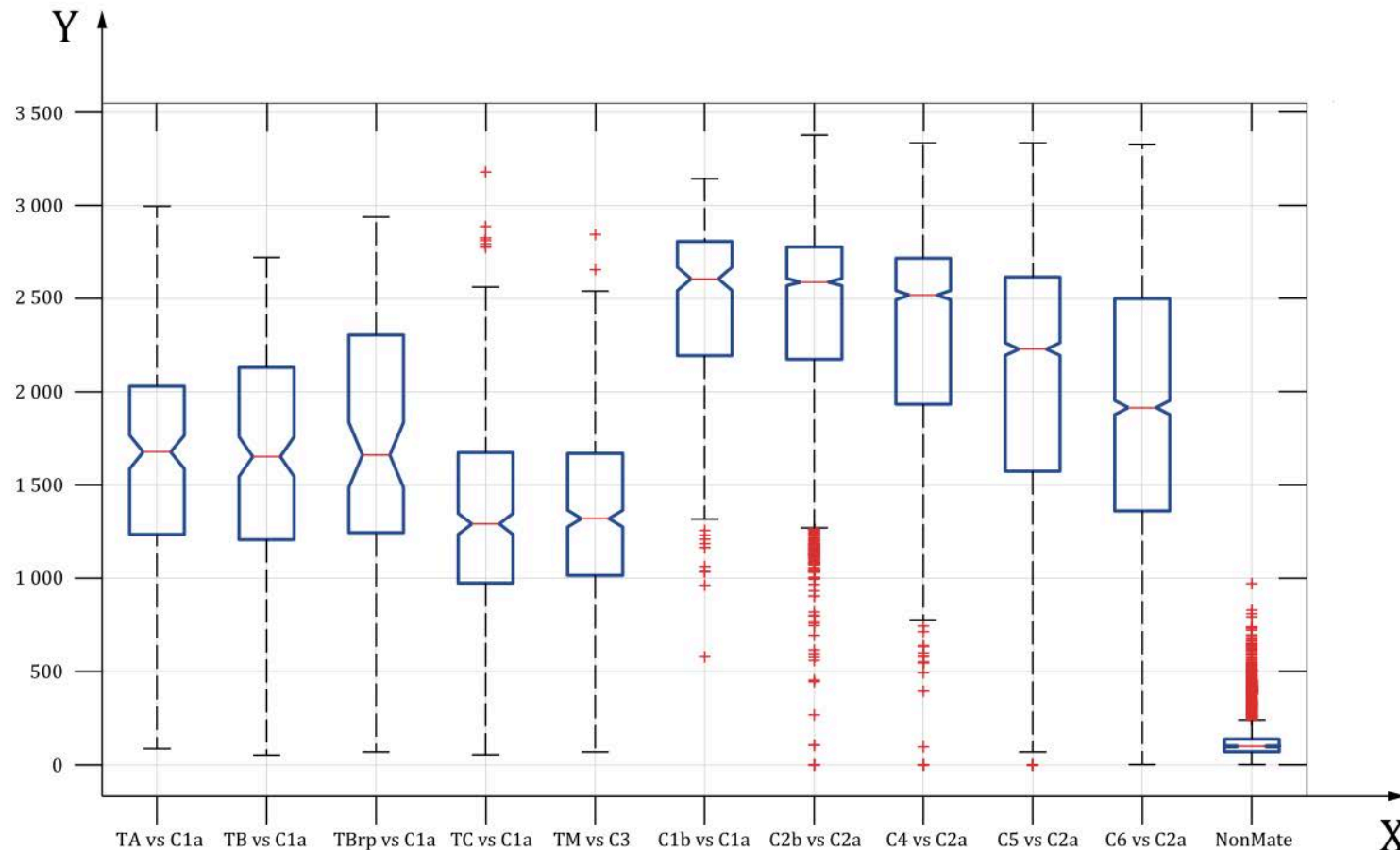
Applications and DET

Applications and appropriate operational points

- Nuclear power plant:
  - FRR > FAR (FAR $\cong$ 0)

- Studio membership card:
  - FAR > FRR

Test Standards

5 US Government Tests
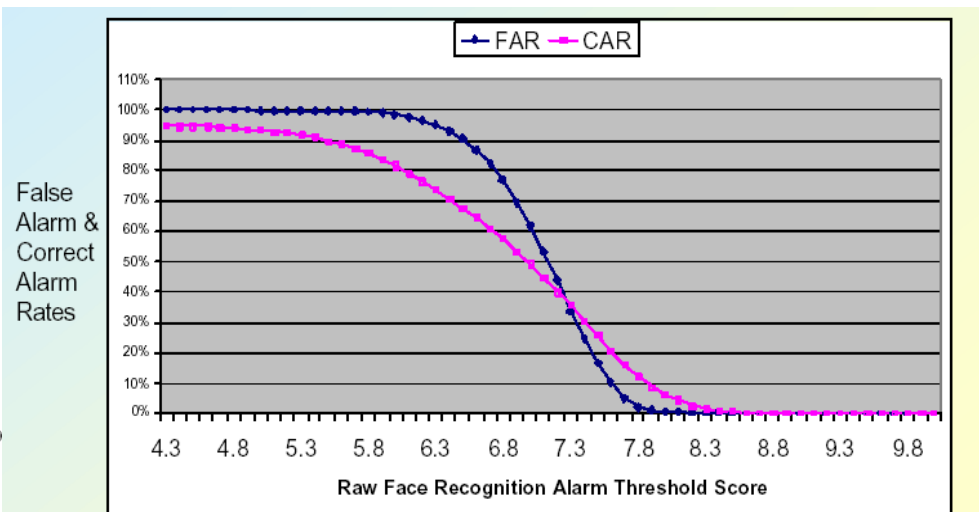
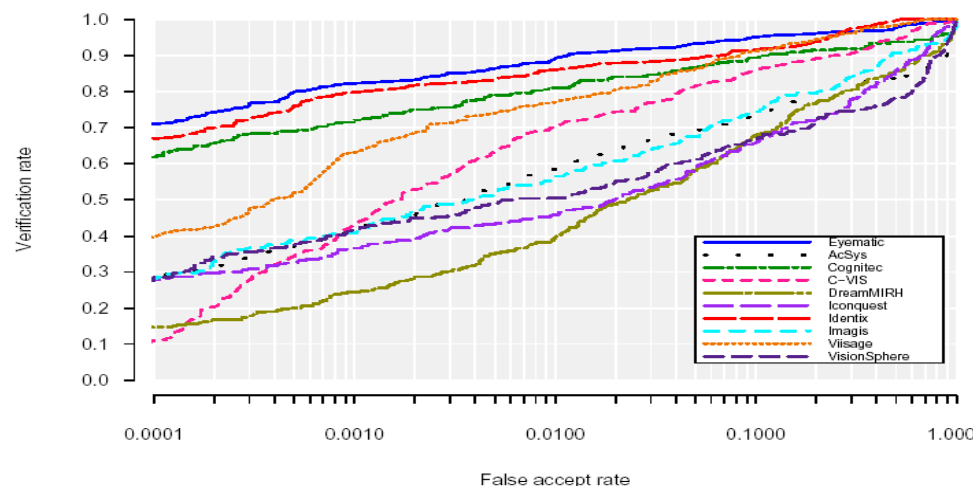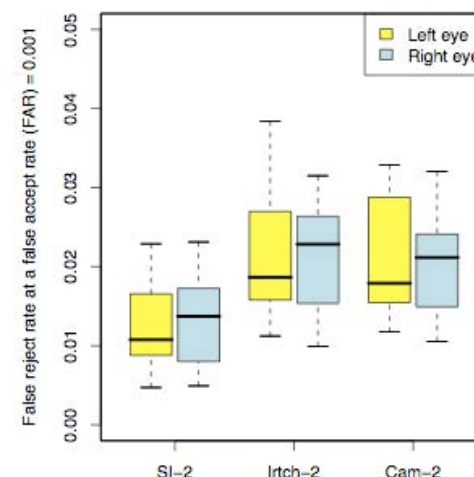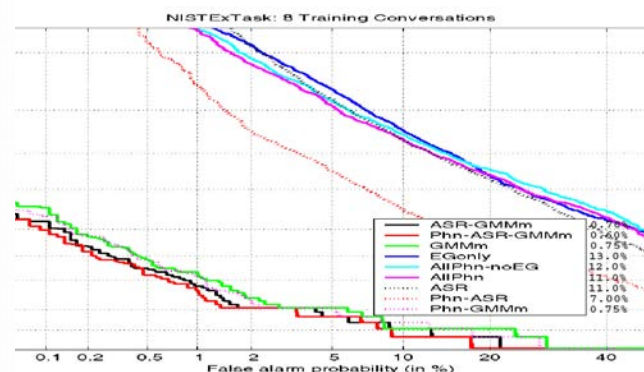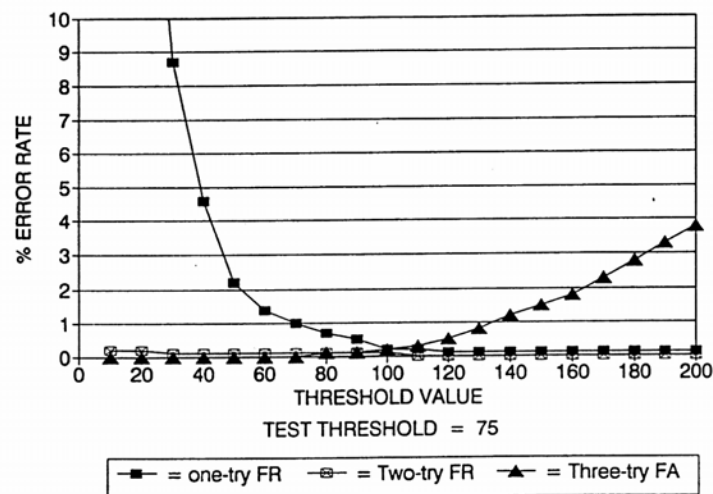Figure L.18 - Verification Performance for Indoor (different day) Probes. This figure corresponds to Figure 22 in the *FRVT 2002: Evaluation Report*.

Biometric Performance Testing Standard

ISO/IEC 19795-x, Information technology -
Biometric performance testing and reporting

▸ Part 1: Principles & Framework (revised in 2021)

- Guidance applicable to the broad range of tests

▸ Part 2: Testing Methodologies for
        Technology and Scenario Evaluation

▸ Part 3: Modality-Specific Testing

▸ Part 4: Interoperability Performance Testing

▸ Part 5: Framework for biometric device performance evaluation
        for access control

▸ Part 6: Testing methodologies for operational evaluation

▸ Part 7: Testing of on-card biometric comparison algorithms

▸ Part 9: Testing on mobile devices

▸ Part 10: Quantifying biometric system performance variation
        across demographic groups

Part 1: Principles & Framework

Content

- Definitions
  - ▶ Describe biometric system
  - ▶ Performance metrics
- Planning an evaluation
- Data collection
  - ▶ Enrolments
  - ▶ One-to-one comparison trails
  - ▶ Identification Trias
- Analyses
- Graphical presentation of results
- Record keeping
- Reporting performance results

Required Metrics

Fundamental (always reported)

- Failure-to-Enrol Rate
- Failure-to-Acquire Rate
- False-Match-Rate vs. False-Non-Match-Rate

Verification
As above plus …

- False-Accept-Rate vs. False-Reject-Rate

And perhaps …

- Generalised error rates incl. enrolment failures
- Distribution of errors in test population …

Identification (open-set)
As above plus …

- False-Positive-Identification-Rate vs. False-Negative-Identification-Rate
  - ▸ (dependence on database size)

If binning used …

- Binning-Error-Rate
- Penetration-Rate
  - ▸ FMR & FNMR are within-bin error rates

Binning

Partitioning based on soft-biometrics

- skin and hair color

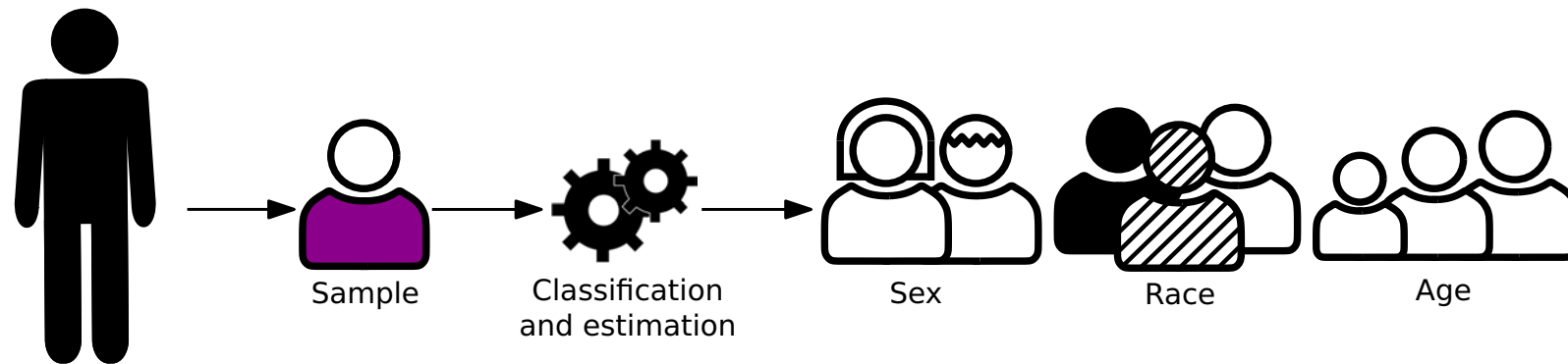

- eye color

- height

- weight

Image Source: A. Jain 2013

Binning

Partitioning based on soft-biometrics

- is done by classification algorithms that estimate membership in pre-defined soft-biometric classes

Sample → Classification and estimation → Sex   Race   Age

- Binning error / pre-selection error:
  error that occurs when the corresponding subject identifier is not in the preselected subs of candidates

- Penetration rate:
  average proportion of the total number of references that are pre-selected

Biometric Performance and Interoperability

Minutiae Interoperability Exchange Test - MINEX

- National Institute of Standards and Technology (NIST)
- Biometric Performance based on Images vs. Minutia-Templates
- Large test
  - ▸ Fingerprint images from 250.000 individuals
  - ▸ Live-Scan data
- DET-curves
  - ▸ False-Match-Rate
  - ▸ False-Non-Match-Rate
- Performance @ FMR 0,01
  - ▸ Image-1F: 0,0047
  - ▸ Minutia-1F: 0,0129
  - ▸ Image-2F: 0,0002

Source:: Minex-Report 03/06
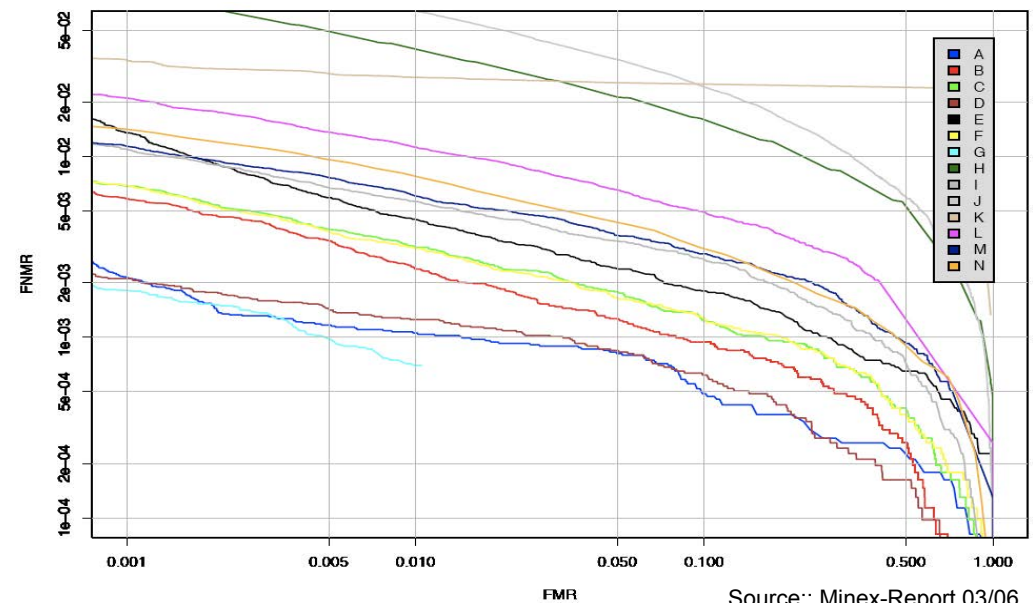
References

A script computing DET curves will be provided

- This script shall be used for term paper reports
  on biometric performance or
  on presentation attack detection performance

- Python-code

- Matlab-code

References

Web

- National Institute for Standards and Technology
  http://fingerprint.nist.gov/

- BEAT testing platform
  https://www.beat-eu.org/news/the-beat-platform-goes-open-source

Complementary reading

- ISO/IEC 19795-1, "Biometric performance testing and reporting - Part 1: Principles and framework", 2021

- B.D. Jovanovic, P. Levy, "A look at the rule of three", The American Statistican, pp. 137-139, 1997

- G. R. Doddington, M. A. Przybocki, A.F. Martin and D.A. Reynolds, "The NIST speaker recognition evaluation: Overview methodology, systems, results, perspective", Speech Communication, 2000