**Week 2 Performance Measures and Performance Evaluation**

Introduction to Biometrics

**Projects**

- To be prepared in two stages:
  1. reading and literature survey; presented in class;
     - You should look up major papers in the area;
     - select 5-10, including some surveys, if there are any;
     - read and write a survey report
     - Prepare a 30-45 minute presentation
     - Also propose what you are going to do in the second part
     - Due: in the next month (end of March; April)
  2. implementation and presentation of results at the end of the semester
     - Find out if there are existing databases; obtain them
     - If necessary, collect data
     - Implement some techniques in the literature; modify them if possible
     - Design an experiment protocol and implement it
     - Write report
     - Present in class

**Introduction to Biometrics 2**

- Projects assigned last week
- Keystroke biometrics: A new project; assigned to newcomers
- Hand vein ageing: 1 person only; newcomers welcome

**Projects**

- Face Aging: Nese, Hande, Ozan, Baran, Senan
- Attacks: Efe, Oğuzen, Başak
- Voice: Farş, Mehmet, Ömer, Gürer, Levent
- Keystroke: Ömer, Ebru, Oğuz, Buğra
- Face profile: Nese, Onur, Burak, Melike, Murat, M. Yusufoglu
- Hand aging: Yusuf, Seniha

**Group presentations schedule**

<table>
<thead>
<tr>
<th>Project</th>
<th>March 11th</th>
<th>March 18th</th>
<th>March 25th</th>
<th>April 1st</th>
<th>April 6th</th>
<th>April 15th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face Aging</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Attacks</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Gait</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand Aging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keystroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Week of March 25th:**

All survey reports and presentations due

Actual presentation according to schedule above

**Introduction to Biometrics 3**

- Nese Hande Ozan Berna Serhan
- Efe Gökçen Başak
- Nese, Onur, Burak, Melike, Murat, M. Yusufoglu
- Yusuf, Seniha

**Introduction to Biometrics 4**

- Face Aging: Nese, Onur, Melike, Murat, M. Yusufoglu
- Attacking: Efe, Oğuzen, Başak
- Gait: Finnish, M. Yusufoglu
- Profile: Nese, Onur, Burak, Melike, Murat, M. Yusufoglu
- Hand Aging: Yusuf, Seniha

**Review: Biometric verification**

- Biometric data
- Pre-processing
- Feature extraction
- Feature vector
- Comparison score
- Classification
- Reference database

- Week of March 25th:
  - All survey reports and presentations due
  - Actual presentation according to schedule above
Review: Biometric identification

Comparison scores

- Similarity scores $s(x,r)$
  - also called matching scores
- Dissimilarity scores $d(x,r)$
  - e.g. distance measure, often $d(x,r) \geq 0$, $d(x,x) = 0$

Conversion

$$s = f(d),$$
$$s = -\log(d)$$
$$s \equiv \frac{1}{d}$$

Errors caused by:

- variations of biometric characteristics
  - Variation of permanence property
  - Aging
  - Spontaneous variations
- environmental factors
  - Illumination
  - Background noise
- Performance specified in error probabilities (error rates)
  - Assumes random nature. Correct?

What can go wrong?

- Acquisition errors:
  - Failure to acquire, capture – FTA, FTC (failure to acquire (capture) rate)
    - System cannot capture biometric
  - Failure to enroll – FTE (failure to enroll rate)
  - Failure to acquire during enrollment
  - Acquisition errors are not always random.
- Verification errors – algorithmic level
  - False match – FMR (false match rate)
    - Algorithm accepts impostor, "zero-effort attack"
  - False nonmatch – FNMR (false nonmatch rate)
    - Algorithm rejects true identity
- Verification error – system level
  - False acceptance – FAR, α, (false acceptance rate)
  - False rejection – FRR, β, (false rejection rate)

Verification errors and error rates

Relations between error rates

$$\text{FRR} = (1 - \text{FTA}) \text{FNMR} + \text{FTC}$$
$$\text{FAR} = (1 - \text{FTA}) \text{FMR}$$

- FRR, FAR and FNMR, FMR often confused
- In certain experiments FTC=0, distinction not relevant.

Inter-class vs intra-class variability

- Inter-class similarity (between classes)
  - Impostor vs genuine
- Intra-class variability (within class)
Relation of FMR and FNMR with threshold

\[ \text{FMR}(t) = \int_{-\infty}^{t} \phi_r(s) \, ds \]
\[ \text{FNMR}(t) = \int_{t}^{\infty} \phi_i(s) \, ds \]

Graphical representation: Error rates as function of threshold

More verification rates

- **Genuine match rate** – GMR
  - GMR = 1 - FNMR
- **Genuine accept rate** – GAR (detection rate)
  - GAR = 1 - FRR
  - GAR = (1 - FTC) \cdot GMR

Graphical representation: Operating characteristics

Verification performance quantifiers

- EER
- GMR@FMR=1%
- GMR@FMR=0.1%
- GMR@FMR=0.01%
- FTA, FTC
- FTE
- GAR@FAR=1%
- GAR@FAR=0.1%
- GAR@FAR=0.01%
- ...

Closed set identification errors and error rates

- Confusion: identity \( i \) is identified as \( j \)
- \( p(j|i) \): probability that identity \( i \) is identified as \( j \)
- Confusion matrix:

\[
C = \begin{pmatrix}
0.95 & 0.02 & 0.02 & 0.01 \\
0.05 & 0.93 & 0.00 & 0.02 \\
0.00 & 0.01 & 0.99 & 0.00 \\
0.05 & 0.04 & 0.01 & 0.90 \\
\end{pmatrix}
\]
Open set identification errors and error rates

- Confusion: identity \( i \) is identified as \( j \).
- False accepts – there is always a closest enrolled identity!
- Counter measure: include threshold.

- Increased FAR: \( \text{FAR}_N = 1 - (1 - \text{FAR})^N = N \times \text{FAR} \)

Identification performance quantifiers

- \( \text{FAR}_N \)
- Cumulative match graph - CMG
- Rank 1 identification index: CMG(1,N)

\[
\text{CMG}(M, N) = P(\text{rank true identity} \leq M \text{ in database of size } N)
\]

Performance evaluation

- Goal:
  - measuring performance quantifiers, DET, ROC, CMG
- Means:
  - controlled experiment
  - vary reference, reference id, probe, probe id
  - measure statistics of comparison scores
  - same setting for verification and identification

Similarity matrix – single enrolment case

\[
\begin{array}{c|ccc}
\text{Genuine scores} & 1 & 2 & M \\
M & 1 & 2 & \\
N & & & \\
R = \frac{N}{M} & & & \\
N \times N & & & \\
(R^2 - R)M & & & \\
\{N - \frac{1}{M}\}N & & & \\
M & & & \\
\end{array}
\]

Similarity matrix – multiple enrolment case

\[
\begin{array}{c|ccc}
\text{Genuine scores} & 1 & 2 & M \\
M & 1 & 2 & \\
N & & & \\
R = \frac{N}{M} & & & \\
N \times N & & & \\
(R^2 - R)M & & & \\
\{N - \frac{1}{M}\}N & & & \\
M & & & \\
\end{array}
\]

Experiment

- References and probes taken from labeled testing set
### Computation of FMR, FNMR and GMR

- \( \Omega_g \): set of all genuine scores
- \( \Omega_i \): set of all impostor scores
- \( \Omega_{gt} \): set of genuine scores \( x > t \)
- \( \Omega_{it} \): set of impostor scores \( x > t \)
- \( |\Omega| \): number of elements in \( \Omega \)

- \( \text{FMR}(t) = \frac{|\Omega_{gt}|}{|\Omega|} \)
- \( \text{GMR}(t) = \frac{|\Omega_{it}|}{|\Omega|} \)
- \( \text{FNMR}(t) = 1 - \text{GMR}(t) \)

### Considerations on data sets

- Data for training and testing must have no overlap!
- In case of multiple enrolment, enrolment set and probe set must have no overlap!
- More realistic results are obtained if users in training and testing set also have no overlap.
- More realistic results are obtained if data is collected in sessions with time intervals in between.
- More comparison scores lead to more accurate results.
- More comparison scores can be generated repeating the experiments after repartitioning the data in new sets for training and testing.

### Accuracy of estimated error rates

- Accuracy improves when number of scores increases
- Relative error increases for small values of FNMR and FMR
- Number impostor scores usually higher
  - Accuracy of FMR higher than that of FNMR

### Approximate confidence intervals

- A confidence interval is the interval around a measured parameter that the true parameter lies in with a specified probability, called confidence level.
- Usually the confidence level equals 90% or 95%.
- Confidence intervals of binomial distributions are outside the scope of this lecture.
- Approximate assumptions:
  - Number of scores large enough → estimate has normal distribution

\[
\frac{p}{\sqrt{|\Omega|}} - c \text{std}(\hat{p}) = \frac{1 - p}{\sqrt{|\Omega|}} \\
\frac{p}{\sqrt{|\Omega|}} + c \text{std}(\hat{p}) = \frac{1 - p}{\sqrt{|\Omega|}}
\]

- \( c \) determines probability of staying with bounds
Approximate confidence levels

\[ p_1 = \left(1 - \frac{1 - \rho}{\sqrt{2 \pi}} \right)^{-1} \]

\[ p_2 = \left(1 + \frac{1 - \rho}{\sqrt{2 \pi}} \right)^{-1} \]

<table>
<thead>
<tr>
<th>Level</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>1.64</td>
</tr>
<tr>
<td>95%</td>
<td>1.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scores</th>
<th>FNMR</th>
<th>90% Interval</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.01</td>
<td>[0.007, 0.020]</td>
<td>[0.006, 0.025]</td>
</tr>
<tr>
<td>10000</td>
<td>0.01</td>
<td>[0.009, 0.012]</td>
<td>[0.008, 0.012]</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>[0.087, 0.117]</td>
<td>[0.085, 0.121]</td>
</tr>
<tr>
<td>10000</td>
<td>0.1</td>
<td>[0.085, 0.102]</td>
<td>[0.084, 0.106]</td>
</tr>
</tbody>
</table>

Boxplots: Nonparametric alternative

Biometric Evaluations

- **In-house—self-defined test:**
  - The database internally collected;
  - testing protocol is self-defined
  - Database not publicly released
- **In-house—existing benchmark:**
  - test is performed over a publicly available database,
  - according to an existing protocol.
  - Results are comparable with others
  - main drawback is the risk of overfitting the data
  - BANCA, XM2VTS

- **Independent—weakly supervised**
  - Database sequestered and made available just before the beginning of the test
  - Samples are unlabeled
  - test executed at the testee’s site and must be concluded within given time
  - Results are determined by the evaluator from the comparison scores
  - visual inspection of the samples, result editing etc., still possible
  - FERET, FRVT 2000, NIST Speaker Recognition Evaluation Program,
- **Independent—supervised:**
  - Same as above, but test is executed at the evaluator’s site on the testee’s hardware.
  - there is no way to compare computational efficiency; memory usage, etc.
  - there is no way to prevent score normalization and template consolidation

- **Independent—strongly supervised:**
  - Data are sequestered and not released before the conclusion of the test
  - Software components compliant to a given input/output protocol are tested at the evaluator’s site on the evaluator’s hardware.
  - main drawbacks: the large amount of time and resources necessary for the organization of such events
  - SVC2004: First International Signature Verification Competition,
  - Fingerprint Verification Competition (FVC2000, FVC2002, FVC-2004) is of this type.
Signals & Systems Group

Fingerprint Verification Competition (FVC 2004): organized by University of Bologna

Introduction to Biometrics 37

FVC 2004

Table: FVC 2004 Performance Indicators

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Left</th>
<th>Right</th>
<th>Left-Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template 1</td>
<td>1000</td>
<td>1500</td>
<td>2000</td>
</tr>
<tr>
<td>Template 2</td>
<td>1200</td>
<td>1700</td>
<td>2200</td>
</tr>
<tr>
<td>Template 3</td>
<td>1400</td>
<td>1900</td>
<td>2400</td>
</tr>
</tbody>
</table>

NIST Biometric Evaluation Campaigns

National Institute of Standards and Technology

1993 - 7 Evaluations
2000 - 5 Challenges Problems (Technology Development)
2004 - 3 Biometrics
2010 - 150,000+ Facial and Iris Images

Performance Improvement over time

FRVT images

FRVT images

FRVT images
Results from FRVT2006

- In 13 years, FR performance has improved by two orders of magnitude
- FR performance has improved due to
  - Algorithm improvement: by a factor of 4 to 6
  - Higher resolution images:
  - Greater consistency of lighting
- Human vs machine
  - Humans better at recognizing familiar faces
  - at false accept rates in the range of 0.05, machines can out-perform humans
- Iris, still face, and 3D face
  - performance of all three biometrics is comparable when all three biometrics are acquired under controlled illumination.
MBGC Goal

- Address face and iris recognition problems that are more relevant to those found in operational data
  - Low to medium resolution face
  - Still and video iris
  - Near Infrared (NIR) & High Definition (HD) video from portals
  - Unconstrained recognition from still & video
- Sequence of challenge problems – Sequence of challenge problems
  - Modeled after the FRGC and ICE 2005
  - Challenge problems and data distributed to researchers
  - Multiple Biometric Evaluation 2010

Assignment 1

Performance Evaluation of a Biometric System

In this assignment, you are required to analyze performance of a biometric system. Assume that a biometric matching algorithm produced a similarity matrix (SM). Since the SM is only available to compare:

a) Equal error rate (EER)
b) Provide FRR at the following FAR points: FRR@FAR=1%, FRR@FAR=0.1%, FRR@FAR=0.01% For each reported accuracy, provide confidence intervals for 95% confidence.
c) Plot ROC and DET curves.

What to submit:
- A one-page report containing the results of the assignment. Specifically
  1) EER obtained for the given similarity matrix
  2) A table containing FAR & FRR values with given confidence intervals (95%)
  3) Two figures for a) ROC and b) DET curves

In addition to your report, you’re required to give your matlab function (name it as analyze_similarity_matrix.m) that outputs the EER, FAR/FRR values, and plots the ROC/DET curves. Your function should have two variables as arguments:

- Input similarity matrix, SM. (A sample SM will be provided to you)
- Numerical class labels of every sample for the similarity matrix.

Submission Deadline: Next Week