Identification Systems
Lecture 7

Piotr Syga

2021-04-27

Biometrics – idea

User identification based on the properties of his:

- physicality (e.g., face geometry, fingerprints, iris, palm print) – snapshot observation, often quick sample acquisition, static
- behavior (e.g., voice pattern, handwriting, gait) – observation of conscious action, time relations may be used

General assumption: we derive a set of features that are characteristic for a single subject and consistent for them

Challenges

- creating consistent measurement environment
- selecting a proper set of features
- establishing confidence intervals and thresholds to facilitate deviations
- updating the profiles
- establishing that the measurement was taken from a living subject
- ensuring repeatability, high processing speed and predictability

Selecting features

- There's no proof of uniqueness of biometric data – we extrapolate from known population.
- Features may be completely (e.g., DNA) or largely (e.g., hand geometry) determined by subjects genotype as well as moderately related (e.g., fingerprints) or mostly independent (e.g., handwriting).
- Feature set reliability is not dependent on genotype determination.
- Worldwide: 3-4 births in 1000 result in identical twins, in the US twin pregnancies occur in 1 in 80
- The percentage of Chimeras is difficult to estimate, yet including microchimerysm it may be as high as 10% [Wenk, 2018; Rinkevich, 2001]

Stages of system operation

1. Setup:
   - selecting mode
   - selecting and calibrating measurement device
   - ensuring uniform measurement environment
   - parameter selection
2. Enrollment:
   - acquiring biometric samples
   - generation templates from samples
   - storing the templates
3. User authentication:
   - deriving template of the subject being authenticated
   - matching the biometric template with a the one obtained during authentication
   - decision

Enrollment

Goal: obtaining good quality reference template, that is robust against subjects mood and health variance as well as devices' and measurements' issues.

- The process might be supervised.
- Multiple samples are collected enabling rejection and merging of the information.
- May require multiple passes and be time consuming.
- Feature extraction and template generation (cohesion vs. variance), storing auxiliary data.
- Storage of the templates (central vs distributed, authenticator-carried vs user-carried) and access information.

Authentication

Goal: reaching possibly fast and precise decision on the subject's identity.

- The process is (mostly) fully automated.
- Single measurement used (multiple measurement may be taken for quality control or auxiliary checks).
- Extracting features and generating a template in a similar manner to the ones created during enrollment.
- Liveliness check and additional verification.
- Matching: determining similarity scores and application of auxiliary data.
- Comparison with confidence intervals and thresholds, producing a decision.
- Optional reiteration.

Presentation attacks

- Aside of template matching, biometric system should present (to a various degree) robustness against presentation attacks (spoofing, forgeries, etc.).
- The goal is to determine whether there is a living subject being authenticated and that the read sample belongs to them.
- Common methods:
  - detection of recordings, irregularities
  - auxiliary biometry for liveliness detection
  - modification of the measurement's environment
  - surplus measurements (tasks)
  - manual verification

Sample types

- Genuine – the authentic samples of the subject
- Random forgeries – originating from a different subject or a different class/modality
- Skilled forgeries – prepared to imitate the authentic samples of a given class

Errors

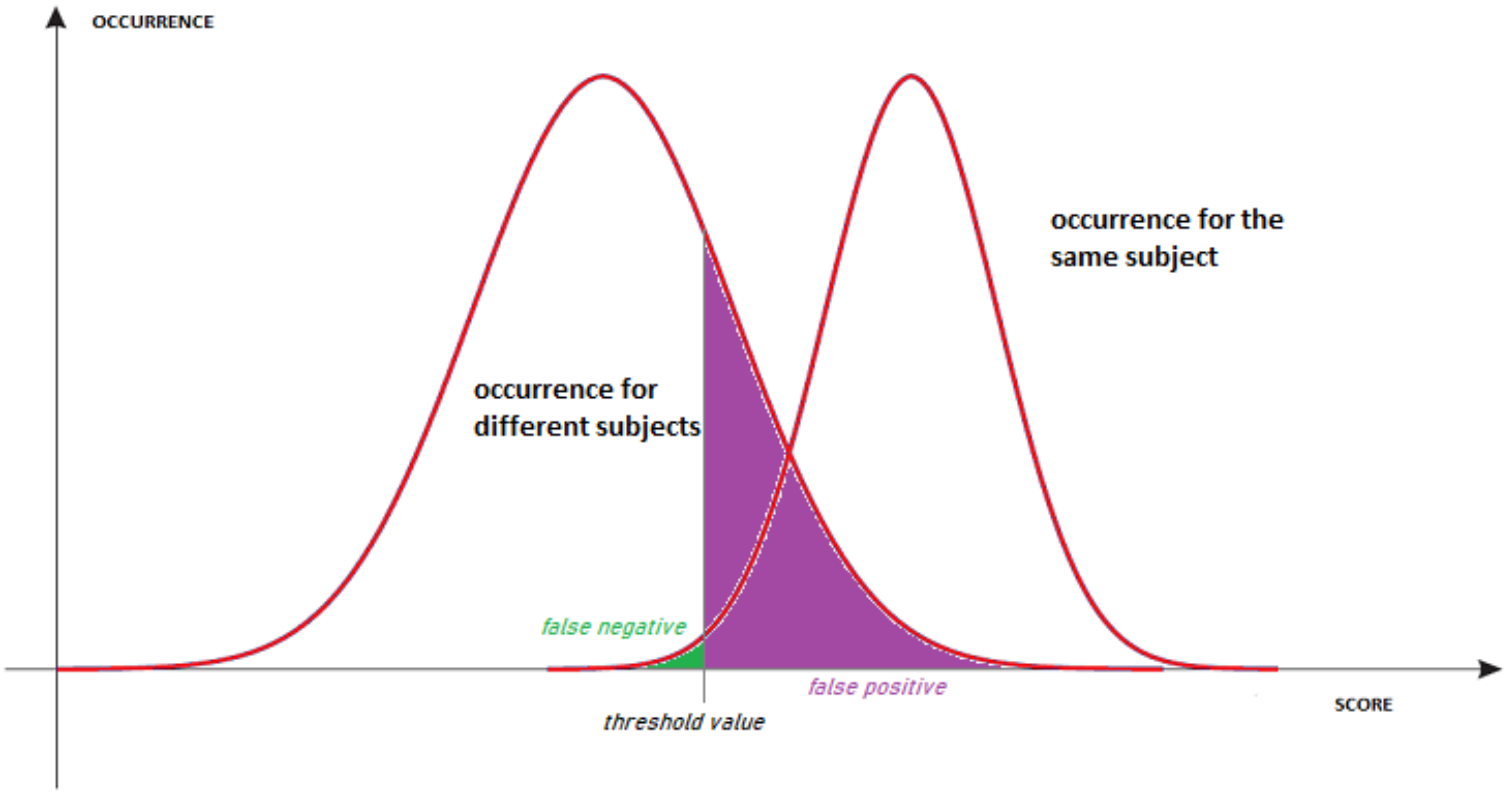
Due to the nature of biometric authorization multiple inaccuracies may occur – caused by the device inaccuracies, an environmental change, subject's mood, health or age, inaccurate model or features that are not unique.

Errors

Main errors:

- False-positive (False-match)– forgery matched a reference template
- False-negative (False-non-match) – genuine sample did not match a reference template

Threshold value



Issues

- We stated that there is no proof for uniqueness, moreover the distributions overlap.
- The devices in the system may differ, each with their own measurement inaccuracies.
- Changes in the environment and in the subject.

Main problems: How to describe the measurements uncertainties? How to reduce the errors? Can we set some guarantees? What should be the sample size? How to set acceptance thresholds based on the experiments? How to perform the experiments? How many are needed?

Typical metrics

- a pair False Positive Rate, False Negative Rate
- Equal Error Rate
- FPR@XFNR, FNR@XFPR (X depends on the problem)
- a pair Precision, Recall
- a pair Accuracy (Bias), Precision
- a pair Sensitivity, Specificity
- Failure to Acquire
- Failure to Enroll

Evaluation scenarios

- evaluation of the sample acquisition in an offline test
- evaluation of a single modality
- random subsample evaluation
- real-life scenarios
- deployment and operational evaluation
- standard: ISO/IEC 19795-1:2006

Samples

- We model each subject as a single (possibly multidimensional) random variable X_s.
- The population is a set of such variables $\mathcal{X}$.
- The contents of our database are a random sample $X = \{X_0, X_1, \ldots, X_k\}$ from $\mathcal{X}$.
- **Important assumption:** we assume that the sample is independent and identically-distributed.

Problem: we do not know the distribution, hence we cannot simply calculate the probabilities.

Statistical approach

- a statistic is any function of a random variable
- a statistic is a random variable by itself
- since the distribution of the random variables is not fully known we estimate the values
- authorization is reduced to testing a hypothesis

Estimators

Definition

Let us fix some unknown parameter θ. Any function $\hat{\theta}$ that matches sample space to the sample estimates is called an estimator. Based on a given sample X we may calculate the estimation error $\mathrm{Err}(X) = \hat{\theta}(X) - \theta$. In particular, we define estimator's bias as $B(\hat{\theta}) = E(\hat{\theta}) - \theta$. If $B(\hat{\theta}) = 0$ we say that the estimator is unbiased.

Estimators

Additional properties:

- Consistency – means that a sequence of estimators that converge in probability to the parameter as sample size grows, i.e., $\forall_{\varepsilon>0} \lim_{n\to\infty} \Pr[|\hat{\theta}_n - \theta| < \varepsilon] = 1$.
- Asymptotic normality – means that a consistent estimator has distribution around θ converges in distribution to a normal distribution.
- Efficiency – means that an ubiased estimator has a minimal variance, that is equal to $\frac{1}{I(\theta)}$.

We aim at creating *at least* an unbiased and consistent estimator.

FTA

Failure to Acquire estimator that models the difficulty of measuring given features $\mathrm{FTA} = \frac{\#\text{failed attempts}}{\#\text{all attempts}}$.
Main reasons include:

- in ability to present the feature – mutilation, illness, environmental constraints
- poor quality of the sensor
- sensor malfunction
- preprocessing malfunction (e.g., image segmentation failed)

FTE

Failure to Enroll estimator that models the difficulty of producing a template form the samples or storing the templates
$\mathrm{FTE} = \frac{\#\text{failed enrollments}}{\#\text{all enrollments}}$.
Main reasons include:

- poor quality or too large deviation
- lack of consistency in the sample
- error in template calculation
- failing the offline test
- database error

Note that by requiring the consistency we increase FTE, yet increase the authorization efficacy.

Positive and a negative verification

- When designing a biometric system we have to decide on the similarity score used.
- Both measuring similarity and dissimilarity are widely used. Dual verification is not common.
- The measures are typically not complementary, hence we test a hypothesis *the samples come from the same subject* or a hypothesis *the samples come from different subjects* – the area may overlap.
- Decisions are made after score comparison with test-derived thresholds. One approach does not translate directly to the other after the tests.

Hypothesis testing

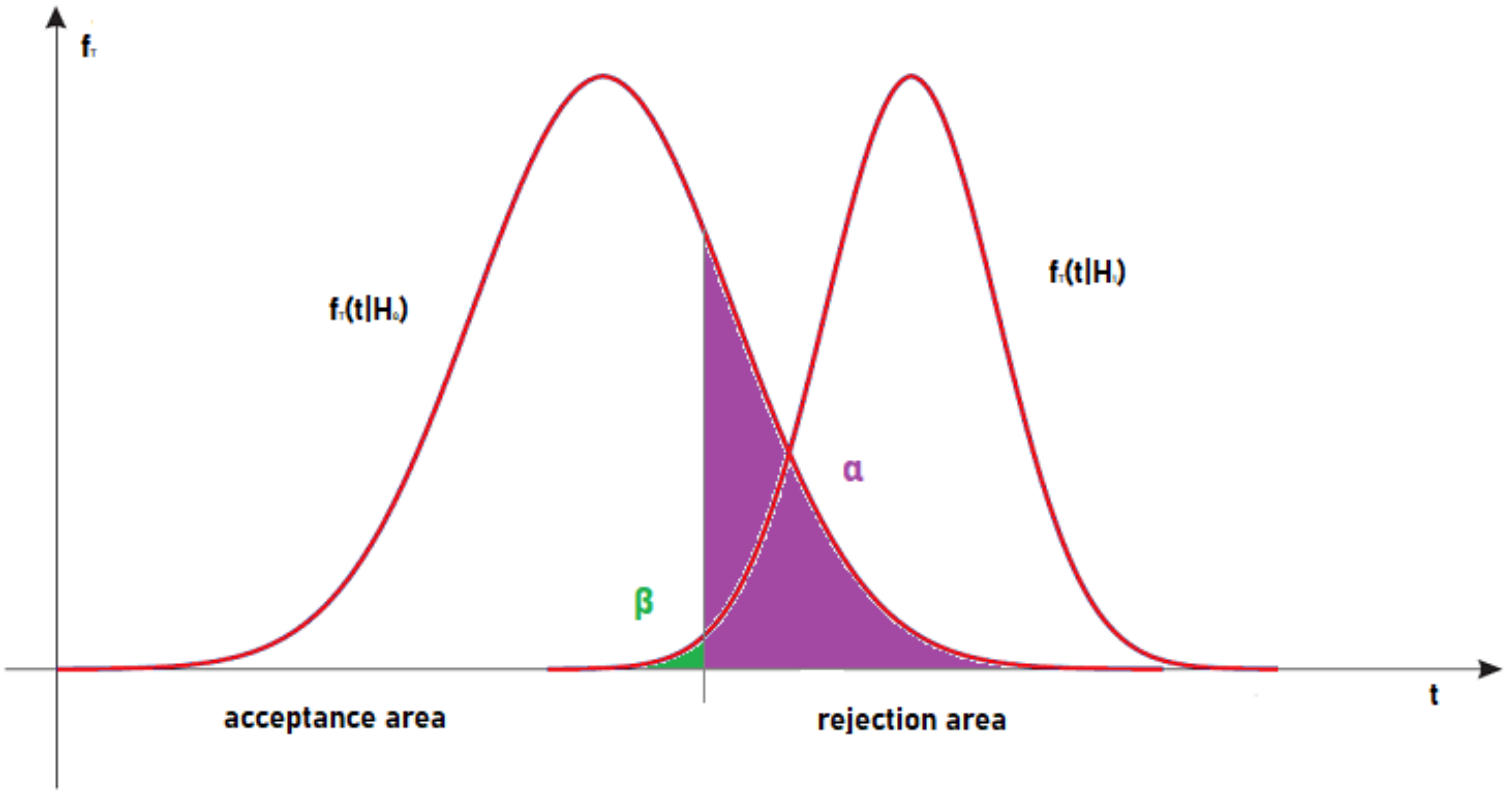
- We derive a null hypothesis (H_0) that is assumed to be true and try to find rationale to reject H_0.
  Note: Not finding the rationale (accepting H_0) does not mean that the null hypothesis is true.
- There is considered also an alternative hypothesis (H_1), that is accepted if we reject H_0.
- We assume that H_0 and H_1 completely specify the population distribution and *that we know this distribution*.

Note: In case of similarity score claiming that a sample belongs to the presented subject A is a simple hypothesis, claiming that a sample belongs to subject B despite presenting as subject A or claiming that it belongs to a one of the enrolled subjects is a simple hypothesis, yet claiming that a sample **does not** belong to subject A is a composite hypothesis.

Statistical tests

- A statistical test is any method of mapping samples into decisions.
- A test statistic is any function used to test the null hypothesis.
- The range of a test statistic is divided into rejection and acceptance region.
- Commonly used statistics: MSE, structural similarity, Hamming distance, absolute difference, bitwise XOR/AND

Statistical tests



Decision matrix for Hypothesis testing

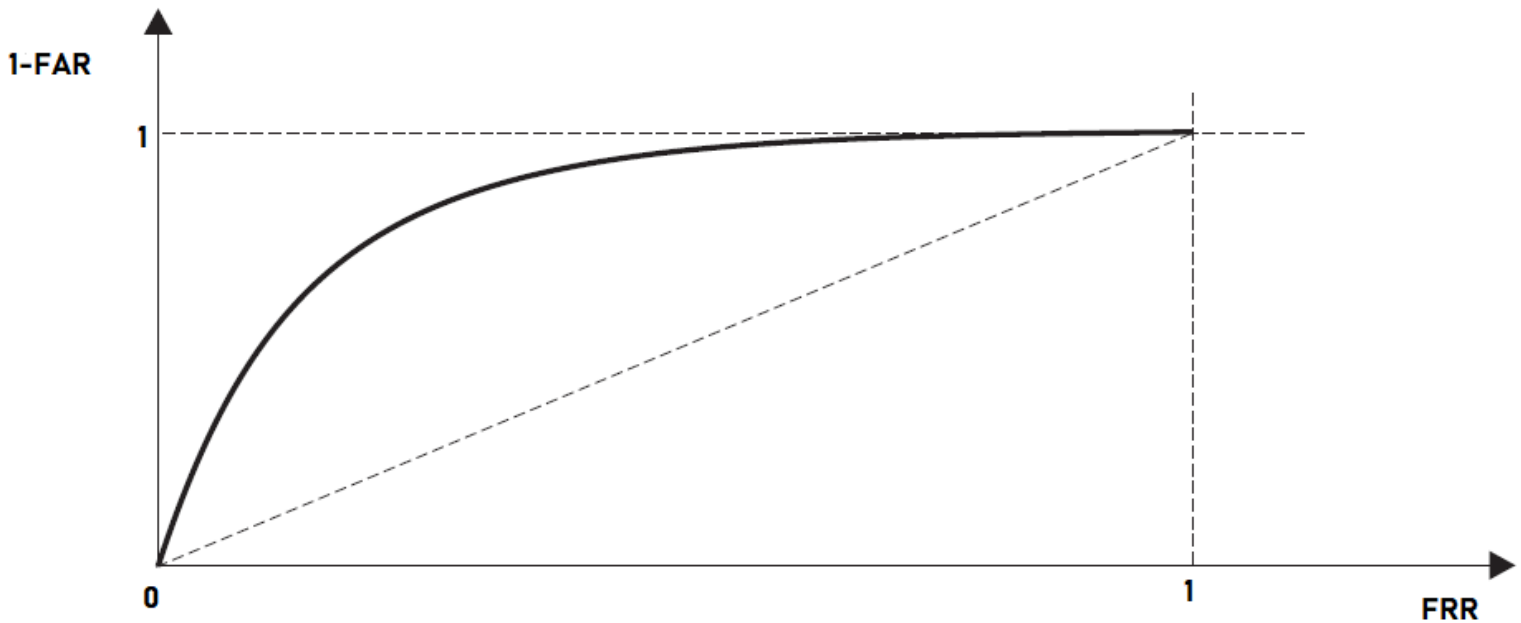
- The parameter α is called significance level.
- The parameter $1 - \beta$ – the power of the test.
- If H_0 is true and we find no grounds to reject it, the decision is correct and is made with ppb $1 - \alpha$.
- If H_0 is false and we find grounds to reject it, the decision is correct and is made with ppb $1 - \beta$.
- If H_0 is true and we find grounds to reject it, the decision is incorrect, we make type I error (with ppb α).
- If H_0 is false and we find no grounds to reject it, the decision is incorrect, we make type II error(with ppb β).
- The hypothesis testing directly translates into correct decisions (TPR, TRR) and errors (FPR, FRR).

FAR and FRR in an algorithm vs in a system

- We have to distinguish the evaluation process of the authentication algorithm form the evaluation of the system.
- For the algorithm Rejection is equivalent to the case of Non-match, and Acceptance to a Match.
- For the system we have to introduce the factor of failed acquisition form the presenting subject (no the subjects that have their templates stored in the database).
- Hence FAR=(1-FTA)FMR=$\frac{\#\text{accepted impostor tests}}{\#\text{all impostor tests}}$, and FRR=FTA+(1-FTA)FNMR=$\frac{\#\text{rejected genuine tests}}{\#\text{all genuine tests}}$

Receiver Operating Characteristic

A characteristic of the algorithm (or a system) may be described as a ROC, with acceptance threshold as a parameter. The area under the curve (AUC) estimates the efficacy of the classifier – $\mathrm{AUC} = 1$ means a perfect classification, $\mathrm{AUC} = \frac{1}{2}$ a random decision, $\mathrm{AUC} < \frac{1}{2}$ invalid classifier.



Interval estimation

- We estimate an interval by two random variables θ_1, θ_2.
- The variables are both a function of a random sample X, i.e., $\theta_1 = f_1(X)$, $\theta_2 = f_2(X)$.
- The functions are independent of the estimated parameter θ.
- After each experiment we get single estimates for the bounding variables θ_1, θ_2, hence a single confidence interval $[\hat{\theta_1}, \hat{\theta_2}]$, that may include θ but doesn't need to (since it is a single sample).
- For a significance level α the probability of inclusion in a sequence of (overlapping) intervals from repeated experiments is equal to $1 - \alpha$.

Interval estimation

- We interpret independent matches as Bernoulli tries with the failure probability p.
- Hence, probability of observing at most N failures in t tries equals $\Pr[e \leq N] = \sum_{0 \leq n \leq N} \binom{t}{n} p^n (1-p)^{t-n}$.
- Substituting p by its upper bound allows us to calculate the significance level.
- So if we achieve $N = 0$, we derive $(1-\tilde{p})^t = \alpha$, so (for small $\tilde{p}$) we obtain minimum error rate $\tilde{p} = \frac{3}{t}$.
- The analysis does not hold for dependent matches, i.e., including a subject into multiple sets.