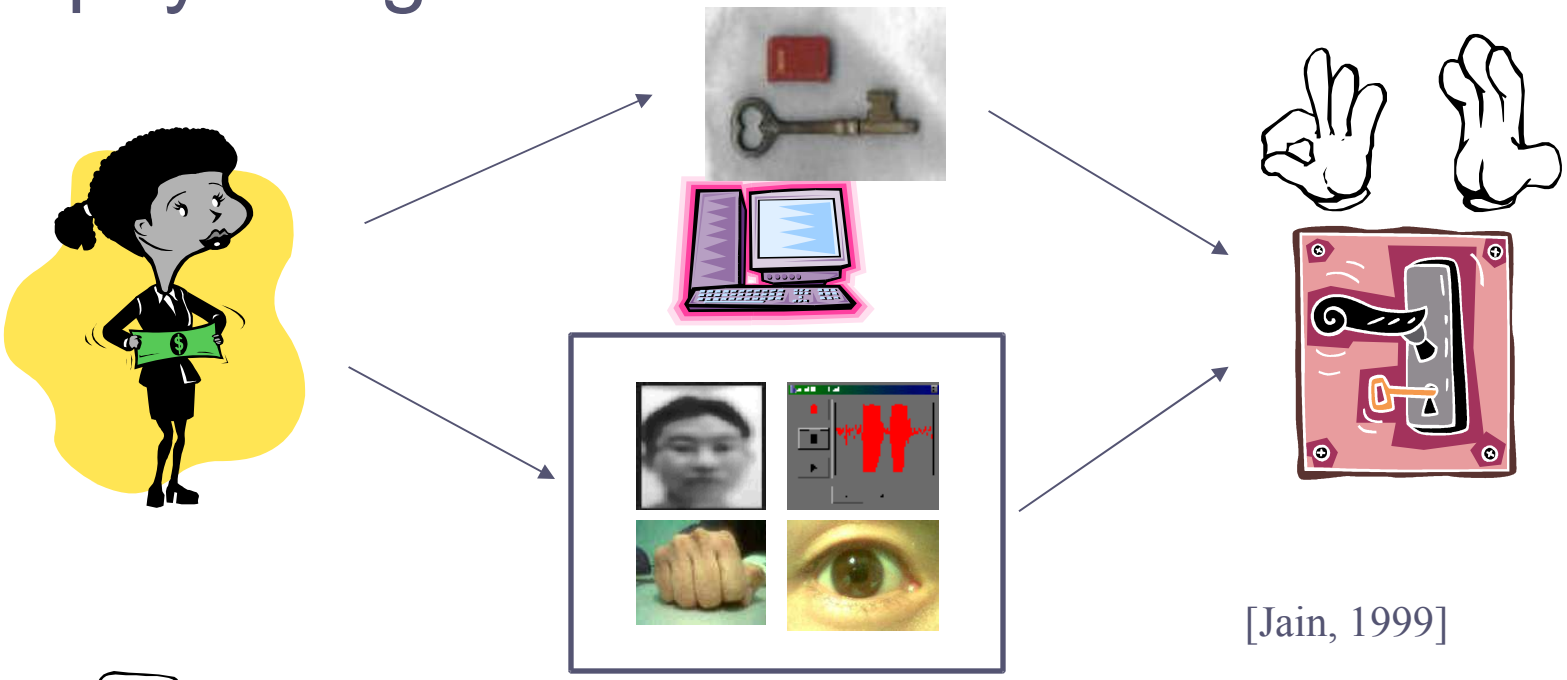
Can Chimeric Persons be Used in Multimodal Biometric Authentication Experiments?

Norman Poh
Samy Bengio

{norman,bengio}@idiap.ch

What is Biometric Authentication?

- A process of verifying an identity claim using a person's behavioral and physiological characteristics.



[Jain, 1999]

This is a binary classification problem

Motivations

- What is a chimeric user?
  - A virtual person created by combining a modality of a person with another modality from yet another person by random mix-and-match.
- Why the use of chimeric users?
  - Acquiring true multimodal data is expensive
  - It is justified by the modality « independence » assumption
- But is such practice acceptable (raised in a panel discussion in MMUA2003)?

Two Approaches to Investigation

- Theoretical:
  - So, what exactly do we mean by the independence assumption?
    - The analyses are given in the paper
  - Major conclusions: There can be various levels of independence (features, scores), that may decide if it is acceptable or not to use chimeric users, but experiments are needed to verify in practice.
- Empirical:
  - Is a performance measure calculated from datasets with chimeric user consistent with the one calculated from true users?

Empirical Approach

- H_0: The performance obtained from chimeric users is equivalent to the one obtained from true users
- H_1: The performance obtained from chimeric users is not equivalent to the one obtained from true users
- Data taken from 21 multimodal fusion datasets, documented in Pattern Recognition Journal, in press, and available for download at http://www.idiap.ch/~norman/fusion

Expected Performance Curve (EPC)

$$\mathrm{FAR}(\Delta_j) = 1 - P(Y(j)|I_j \leq \Delta_j)$$ ⟵ impostor scores

$$\mathrm{FRR}(\Delta_j) = P(Y(j)|C_j \leq \Delta_j)$$ ⟵ client scores

$$\mathrm{HTER}(\Delta_j) = \frac{1}{2}(\mathrm{FAR}(\Delta_j) + \mathrm{FRR}(\Delta_j))$$ (j is index to a particular client)

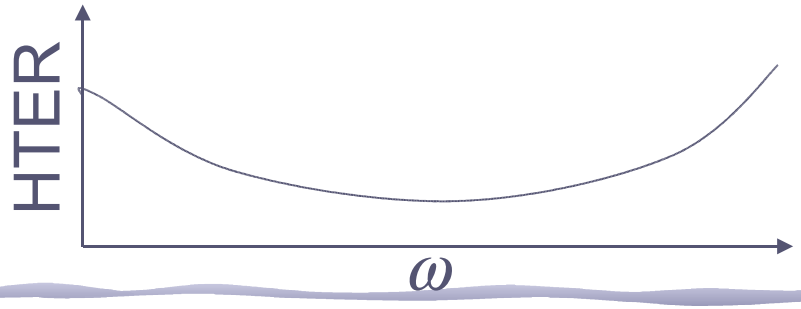
(User-independent threshold Δ is used)

For each ω in [0,1],

Calculate $\Delta_* = \arg\min_{\Delta} \omega\mathrm{FAR}(\Delta) + (1-\omega)\mathrm{FRR}(\Delta)$ from development set

Calculate $\mathrm{HTER}(\Delta_*, \omega)$ from evaluation set

End

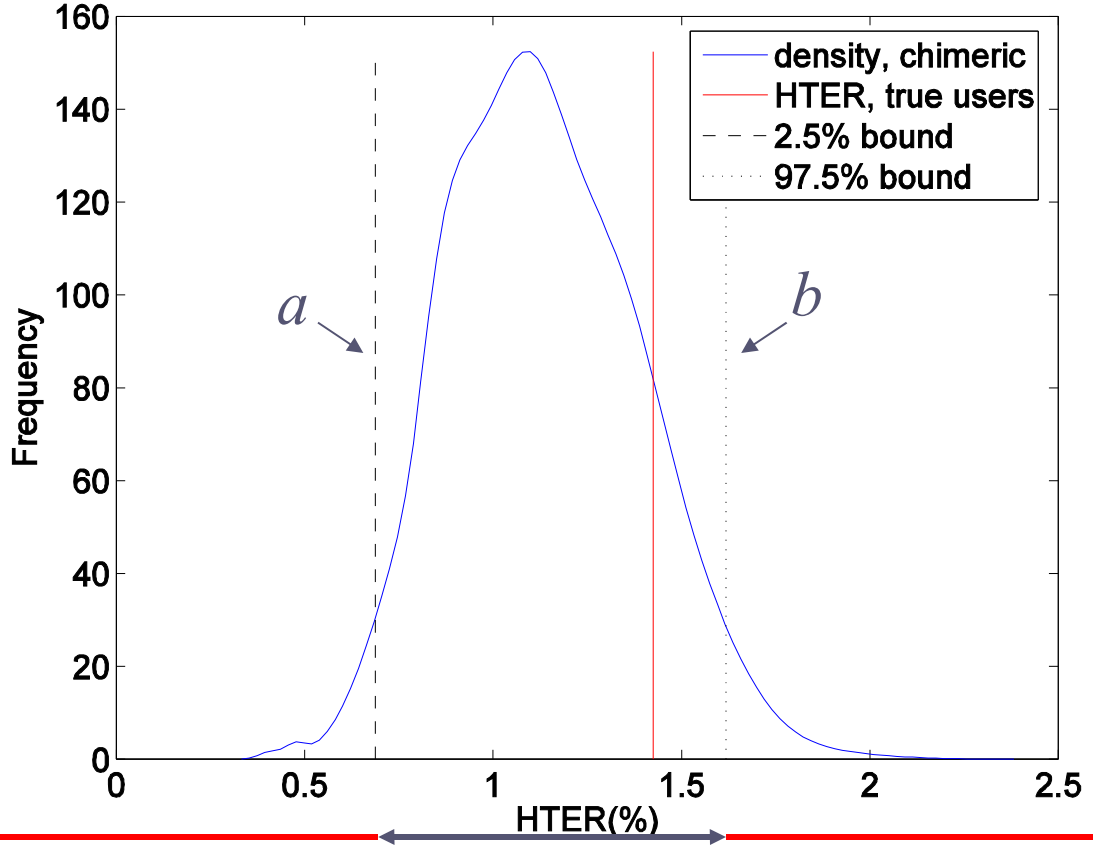


Statistical Confidence Interval

Statistics of interest; HTER in our case

$$p(v \in {}^{c}[a,b]|H_0) = \alpha$$

Significance level



α=95% in our case

a and b are chosen such that their tails have equal surface.

Supporting H_1 Supporting H_0 Supporting H_1

Experimental Procedure

- Build a fusion classifier on true user data set
- Calculate its *a priori* HTER*
- For each bootstrap i of chimeric users
  - Build a fusion classifier
  - Calculate its *a priori* HTER(i)
- Compare all HTER(i) with HTER* using statistical significance test outlined in slide 7

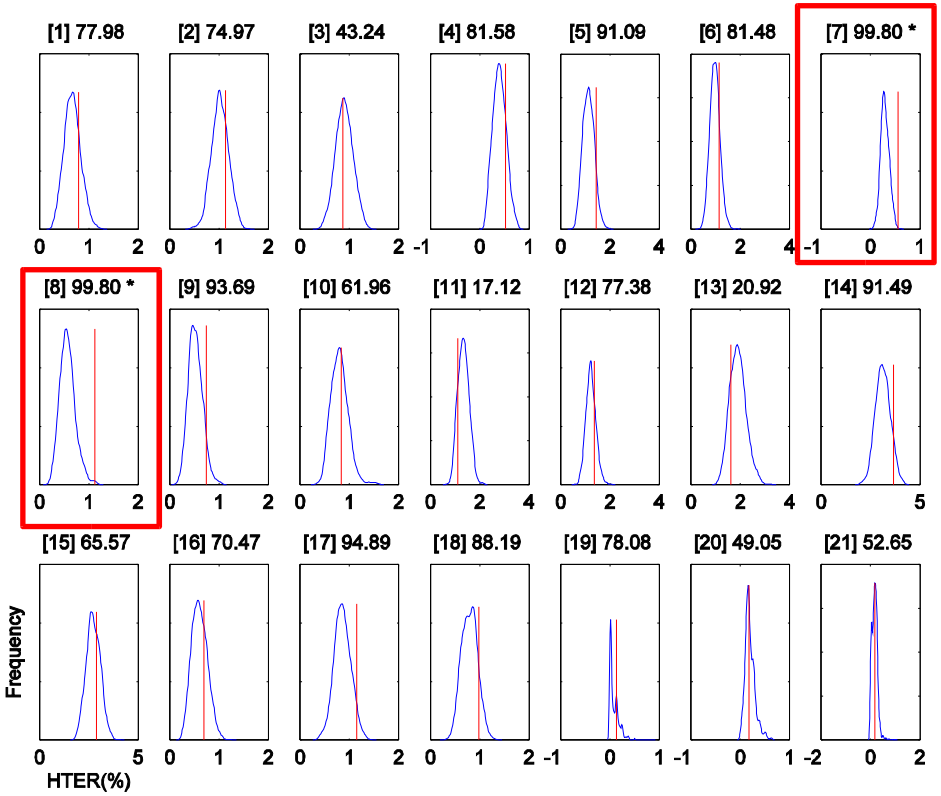
The « bootstrap » is modified as follows:

2. We sample chimeric users (made of combination of modalities)
3. This is done without replacement (number of combinations high)

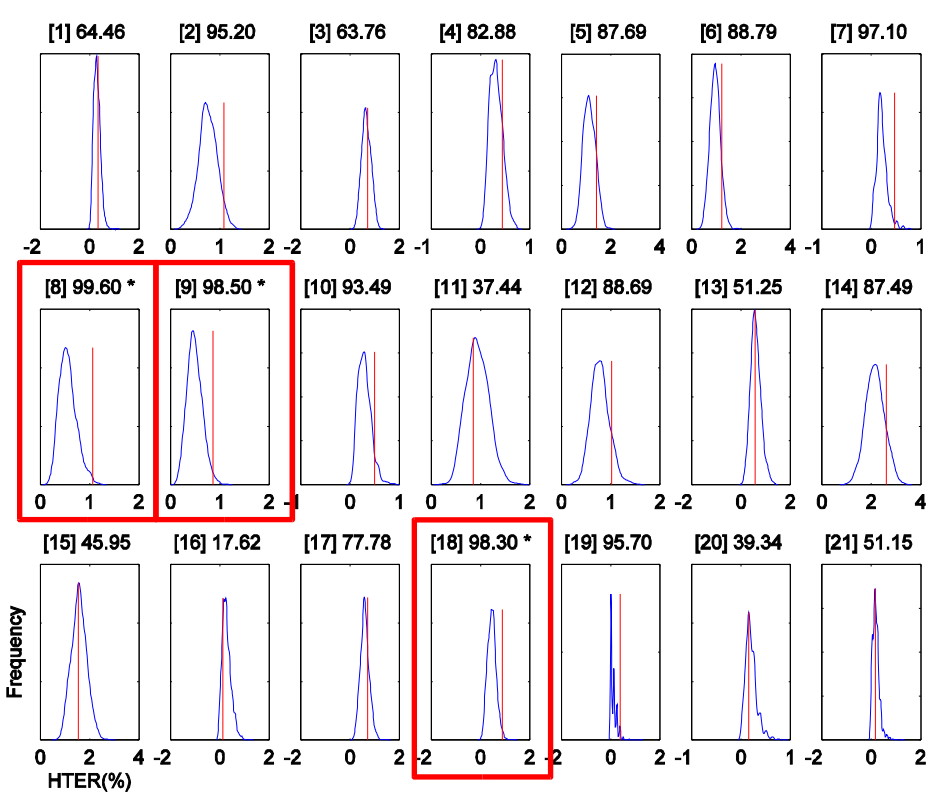
Distribution of HTER (ω=0.5)

(EER criterion)

Fusion method: Mean



Fusion method: GMM

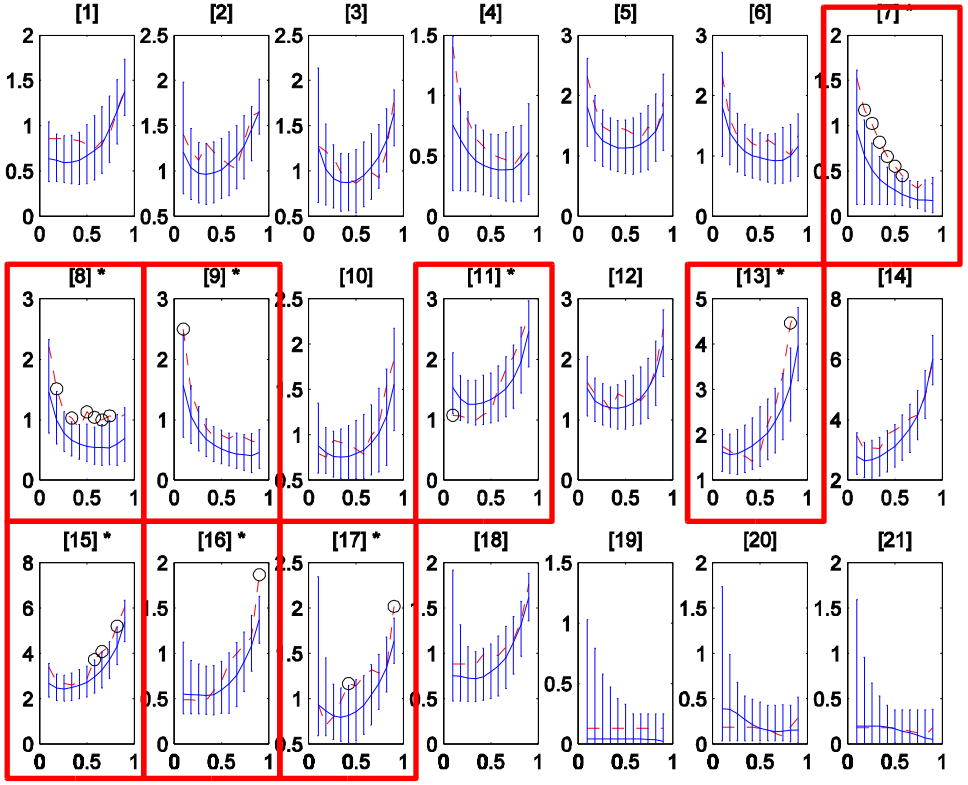


Some distributions are not Gaussian. This implies that 1000 bootstraps may not be enough. Each experiment involves 200 users. Hence, the total number of combinations is 200!=infinity!

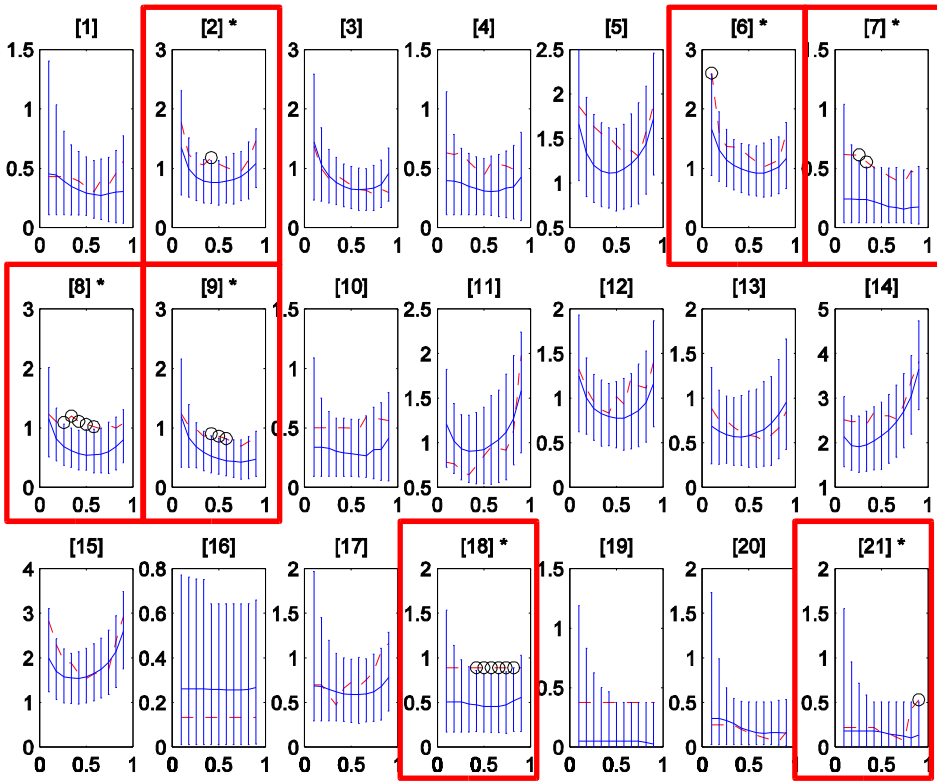
In practice, one can only make several runs of experiments.

Distribution of EPC of chimeric user datasets versus EPC of true users

Mean



GMM



Red box is triggered when there is at least one point on the EPC that supports H_1.

Conclusions

- The (theoretical) independence assumption does not necessarily imply that one can use chimeric users in experiments
- The empirical modified bootstrap does not support the use of chimeric users at 95% of confidence
- Hence, if chimeric users are to be used, one should be careful about the claim of performance w.r.t. the case where true users are used.
- Our proposal: make several runs of experiments with chimeric users to have an idea about the possible range of performance measures in question.
- Issues not investigated:
  - Relative performance (e.g., comparing two fusion classifiers)
  - how far the score distribution estimated with the independence assumption is from the one estimated with the dependence assumption?


I need your opinions!

Threshold-based Decision Function

Commonly used decision function:

$$\text{decision}(y(j)) = \begin{cases} \text{accept} & \text{if } y(j) > \Delta_j \\ \text{reject} & \text{otherwise} \end{cases}$$

OR

Log Posterior Ratio

$$\text{decision}(\text{LPR}_j) = \begin{cases} \text{accept} & \text{if } \text{LPR}_j > 0 \\ \text{reject} & \text{otherwise} \end{cases}$$

where

$$\text{LPR}_j \equiv \log\left(\frac{P(C_j|X)}{P(I_j|X)}\right) = \log\left(\frac{P(X|C_j)P(C_j)}{P(y|I_j)P(I_j)}\right)$$

$$= \underbrace{\log\frac{P(X|C_j)}{P(X|I_j)}}_{\text{Log Likelihood Ratio (LLR)}} + \underbrace{\log\frac{P(C_j)}{P(I_j)}}_{\text{threshold}} \equiv y(j) - \Delta_j$$

This is in contrast with scores assuming posterior probability, i.e.,

$$y(j) \equiv p(C_j|X) \qquad \Delta_j = 0.5$$

The LLR modeling will be used in the analysis that follows.

Four levels of Independence Assumptions

	Feature	Score (loose feature)
Dep.	[4] $y_{SD}(j) = \log \frac{p(X_1, X_2 \mid C_j)}{p(X_1, X_2 \mid I_j)}$	[3] $y_{LD}(j) = \log \frac{p(y_1(j), y_2(j) \mid C_j)}{p(y_1(j), y_2(j) \mid I_j)}$
Indep.	[1] $y_{SI}(j) = \log \frac{p(X_1 \mid C_j)p(X_2 \mid C_j)}{p(X_1 \mid I_j)p(X_2 \mid I_j)}$	[2] $y_{LI}(j) = \log \frac{p(y_1(j) \mid C_j)p(y_2(j) \mid C_j)}{p(y_1(j) \mid I_j)p(y_2(j) \mid I_j)}$

Assume feature independence here!

Levels of dependency

[1] [2] [3] [4]

weak → strong

Concerns chimeric users

$$\boldsymbol{y}_{chimeric} = [y_1(j), y_2(j')]^T \text{ where } j \neq j'$$

The assumption says nothing about chimeric users; they merely guide us how to combine the multimodal streams.