Multi-modal Biometric Authentication for mobile communication

Josef Bigun
http://www.hh.se/staff/josef

[1] J. Bigun, K. Choy, and H. Olsson. *Evidence on skill differences of women and men concerning face recognition*. *AVBPA*, LNCS-1206 Springer, pages 44-51, 2001.

[2] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. *Mulltimodal biometric authentication using quality signals in mobile communications*. In *Proc. of 12'th Int. conf. on image analysis and processing, Mantova, Italy*, pages 2-11. IEEE Computer Society Press, Piscataway, NJ, september 17-19 2003.

OUTLINE

- On Human face recognition skills
- Biometrics now: Face, fingerprint and speaker recognition
- Multi-modal biometrics: Smart wallets in your mobile phone!
- Conclusions

Multimodal Biometric Authentication using Quality Signals in Mobile Communications

J. Bigun[1], J. Fierrez Agullar[2], J. Ortega-Garcia[2], and J. Gonzalez-Rogidriguez[2]

www.hh.se

Halmstad University[1], Sweden

and

www.upm.es

Universidad Politecnica de Madrid[2], Spain

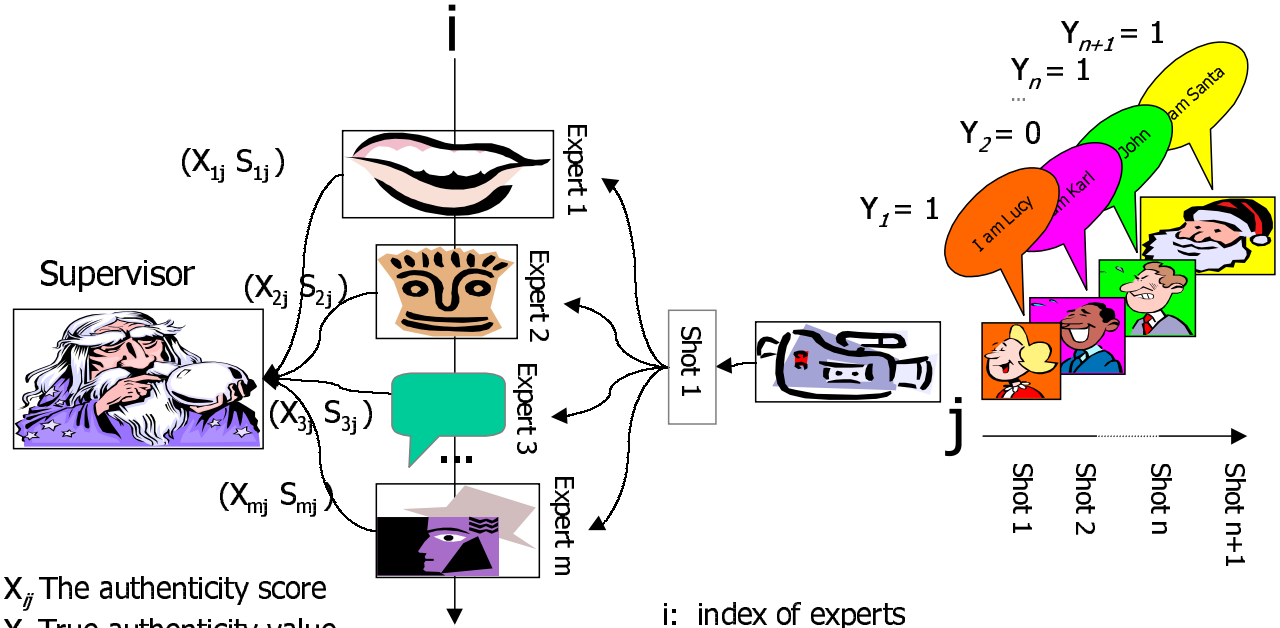
Overview

- How are human experts faring in biometrics tasks?
- Introduction and definitions in biometrics
- Supervisor
- Experiments: A) Face & Speech
- Experiments: B) Fingerprint & Mobile telephone speeech
- Experiments: C) Simulated experts
- Conclusions

Definitions

- Client
- Client identity claim (either true or false)
- Expert
- Expert opinion
- Supervisor
- Supervisor opinion
- Authentication (verification, 1:1) versus Identification (1:N)

System Model



X_{ij} The authenticity score

Y_j True authenticity value

$Z_{ij} = Y_j - X_{ij}$ The misidentification

S_{ij} = The variance of Z_{ij}

i: index of experts

j: index of shots

We can distinguish several supervisor categories.

We can make biomerics supervisors more “clever” by making them aware of the expertship level of the experts, the difficulty level of recognizing certain clients.

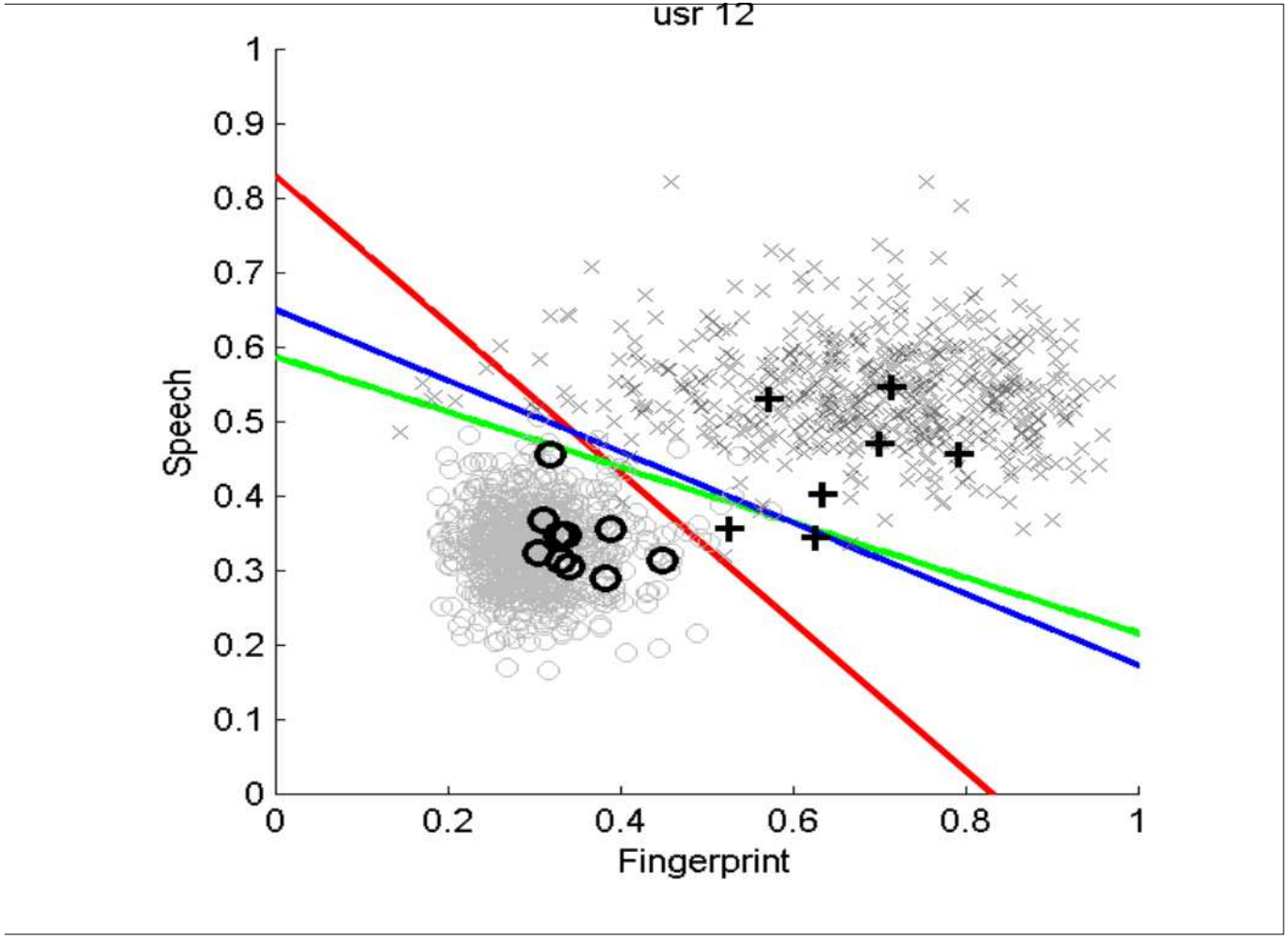
We could go one step further and make the supervisors be aware of the signal quality of the claim. In that case the supervisor behaviour would be appreciably different even if the expert and the client are unchanged e.g. when the image quality of the client would become poor due to lack of light.

How?

In fact any two class classifier can function as supervisor provided that the opinions of the experts are merged to a feature vector

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and that the problem is to determine if this vector belongs to a client class or impostor class. Obviously, training data where experts assess both false and genuine identity claims are needed to construct classifiers. In supervisor training, the experts do not know the client identities, but the supervisors do, of course.



Examples

-Bayesian, SVM,

-Sum rule

Desiderata

-Simplicity of training

-Possibility to combine human opinions

Bayesian conciliation

-Developed to assess risks of catastrophes.

-Risks of certain events (e.g. an event which can initiate a catastrophe) are assessed.

-Objective data is rare, therefore the subjective human expert opinions as well as all available measured data must be fused,

-Classical statistical tools, e.g. density estimation via observations, require too much data to be usefull.

The attributes used to define **a catastrophe** are

- Lack of data,
- rareness,
- and undesirable consequences

-The authentication systems suffer from the *lack of data* on "good" impostors as well as clients.

-typically in a well designed system erroneous decisions are *rare*, but when they occur they usually result in very *undesirable consequences*.

-Sometimes, a human is in the decision loop

Consequently, an authentication system can be viewed as an automatic risk assesment device where **acceptance of an impostor or rejection of a client is, at least technically, a catastorophe.**

Lindley, V. D. (1985), "Reconciliation of Discrete Probability Distributions", Bayesian Statistics 2, pp. 375-390.

Winkler, R.L. (1981), "Combining Probability Distributions From Dependent Information Sources", Management Science, 27, pp. 479–488.

West, M. (1988), " Modelling Expert Opinion". Bayesian Statistics 3, pp. 493-508.

French, S. (1985), "Group Consensus Probability Distributions: A Critical Survey", Bayesian Statistics 2, pp. 183-202.

The supervisor (decision maker) estimates the quality of the authenticity scores delivered by an expert.

A different approach is to let the experts estimate the quality of the score besides that they give a score on the authenticity. This is interesting to system designers of biometrics because a significant component of the quality measure is directly dependent on the signal quality which may in turn be possible to measure. For example a rough quality measure of an image can be estimated by the image sensor itself or by the expert.

We use the ideas in

> E. S. Bigun (1995) "*Risk analysis of catastrophes using experts' judgments*" Eur. J. Operational Research, Vol. 87, pp. 599-612

that addresses the issue of conciliating opinions and confidence levels given by human experts through Bayes' theory.

In biometrics, we use machine modules as experts instead of human experts, but the methodology could also be used to combine machine authentication, which is what have been done. For various reasons a combination of human expertship and machine expertship could also be needed. For example in forensic person authentication.

The method results in two sum rules of calibrated expert scores.

Simultaneously

J. Kittler, Y. P. Li, J. Matas, and M.U. Ramos Sanchez. "Combining evidence in multimodal personal identity recognition systems". In *Audio and Video based Person Authentication - AVBPA97*, pages 327–334. Springer, 1997.

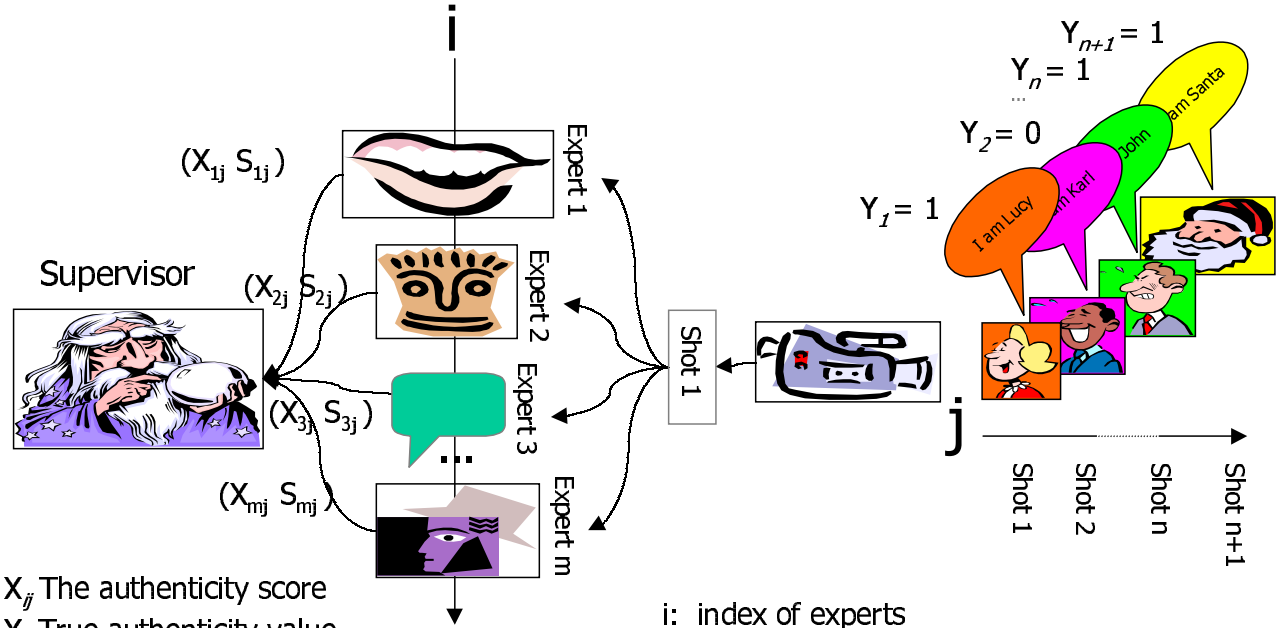
reported that fusion by sum rule is superior to a series other fusion rules in biometrics, including product rule.

For notational simplicity, let the index j represent the time instances or shots. When it is absent and replaced by "." at the subscripts, it takes all values in $\{1...n\}$, which will be considered to be the time instances up to the current time. The time instance $n+1$ represents the future.

Model: Given its mean and variance, the misidentification score is normally distributed

$$\begin{aligned}
(Z_{ij}|b_i, \sigma_{ij}) &\in N(b_i, \sigma_{ij}) \\
&\downarrow \\
(b_i|Z_{ij}, \sigma_{ij}) &\in N(M_i, V_i) \\
&\downarrow \\
(Y_{n+1}|Z_{i.}, \sigma_{i.}, x_{i,n+1}) &\in N(M_i', V_i') \\
&\downarrow \\
(Y_{n+1}|Z_{..}, \sigma_{..}, x_{..}, x_{i,n+1}) &\in N(M_i'', V_i'')
\end{aligned}$$

System Model



X_{ij} The authenticity score

Y_j True authenticity value

$Z_{ij} = Y_j - X_{ij}$ The misidentification

S_{ij} = The variance of Z_{ij}

i: index of experts

j: index of shots

Case 1: Only expert scores are available

Training

$$M_i = \frac{1}{n}\sum_{j=1}^{n} z_{ij} \qquad \text{and} \qquad V_i = \frac{\alpha_i}{n}$$

Testing/operation

Calibration $\quad M_i' = x_{i,n+1} + M_i \qquad \text{and} \qquad V_i' = V_i + \alpha_i$

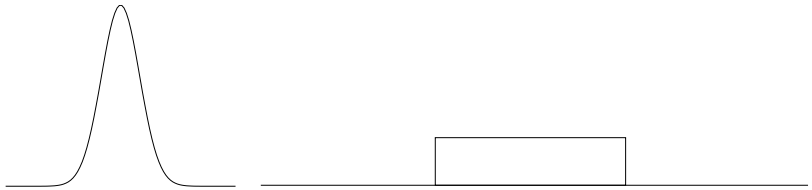
Fusion $\quad M'' = \dfrac{\sum_{i=1}^{m} \frac{M_i'}{V_i'}}{\sum_{i=1}^{m} \frac{1}{V_i'}} \qquad \text{and} \qquad V'' = \left(\sum_{i=1}^{m} \frac{1}{V_i'}\right)^{-1}$

where

$$\alpha_i = \frac{1}{n-3}\sum_{j=1}^{n} z_{ij}^2 - \frac{1}{n(n-3)}\left(\sum_{j=1}^{n} z_{ij}\right)^2$$

Case 2: expert scores and quality of scores are available

The quality of score, called s_{ij}, is provided by the expert and concerns a particular authentication assesment. It is not a general reliability measure for the expert himself, but a certainty measure based on entire unquantifiable knowledge of the expert and the data the expert assesses (e.g. it can be related to the quality of the image). Typically the quality of the score is chosen as the inverse of the width of the range in which one can place the score. Such intervals can be conveniently provided by a human expert too.



Training

$$M_i = \frac{\sum_{j=1}^{n} \frac{z_{ij}}{\sigma_{ij}^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_{ij}^2}} \qquad \text{and} \qquad V_i = \frac{1}{\sum_{j}^{n} \frac{1}{\sigma_{ij}^2}}$$

Testing/operation

$$M_i' = x_{i,n+1} + M_i \qquad \text{and} \qquad V_i' = V_i + \sigma_{i,n+1}^2$$

$$M'' = \frac{\sum_{i=1}^{m} \frac{M_i'}{V_i'}}{\sum_{i=1}^{m} \frac{1}{V_i'}} \qquad \text{and} \qquad V'' = \left(\sum_{i=1}^{m} \frac{1}{V_i'} \right)^{-1}$$

The variances are estimated by $\bar{\sigma}_{ij}^2 = s_{ij}^2 \cdot \alpha_i$ where

$$\alpha_i = \frac{1}{n-3}\left(\sum_{j=1}^{n}\left(\frac{z_{ij}^2}{s_{ij}}\right) - \left(\sum_{j=1}^{n}\left(\frac{z_{ij}}{s_{ij}}\right)\right)^2 \left(\sum_{j=1}^{n}\left(\frac{1}{s_{ij}}\right)\right)^{-1}\right)$$

Score Transformation

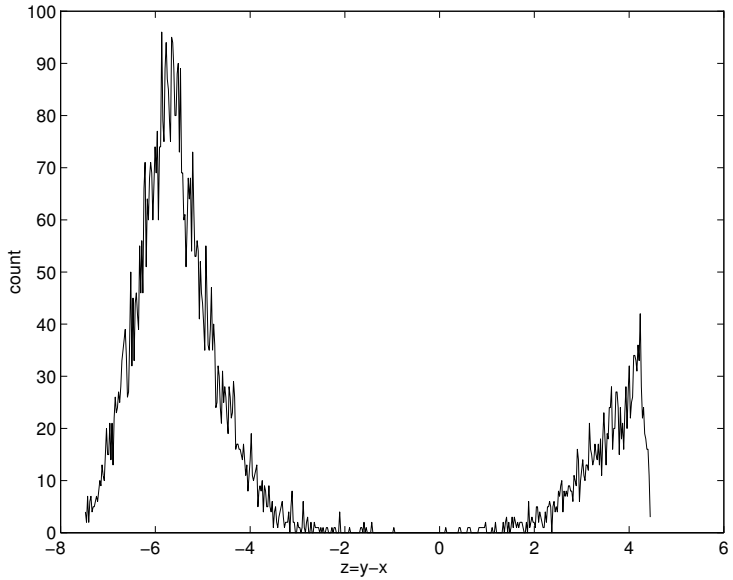
The scores of the experts X_{ij} may or may not be dimensionless or in the correct range i.e. $(-\infty, \infty)$. In this work the experts scores are in $[0+\epsilon, 1-\epsilon]$ and we use the **odds transformation**

$$X_{ij} = \log \frac{X'_{ij}}{1 - X'_{ij}}$$

in which the primed variables represent expert scores before the transformation. It can be shown that the previous formulas still hold with the difference that the conditional distribution of Y_{n+1} is log-normal. Finding the expectation value of Y'_{n+1} given the expert opinions in the past is not possible to obtain. Instead we used M'' as a decision parameter in our supervisor. We used $\epsilon = \exp^{-6}$.

Experiments

In experiments we observed that the misidentification error is not monomodal. Therefore, we assume that the positive and the negative parts of Z are normal and compute M'' twice (for positive and negative Z). We select the best M'' for decision making as follows.



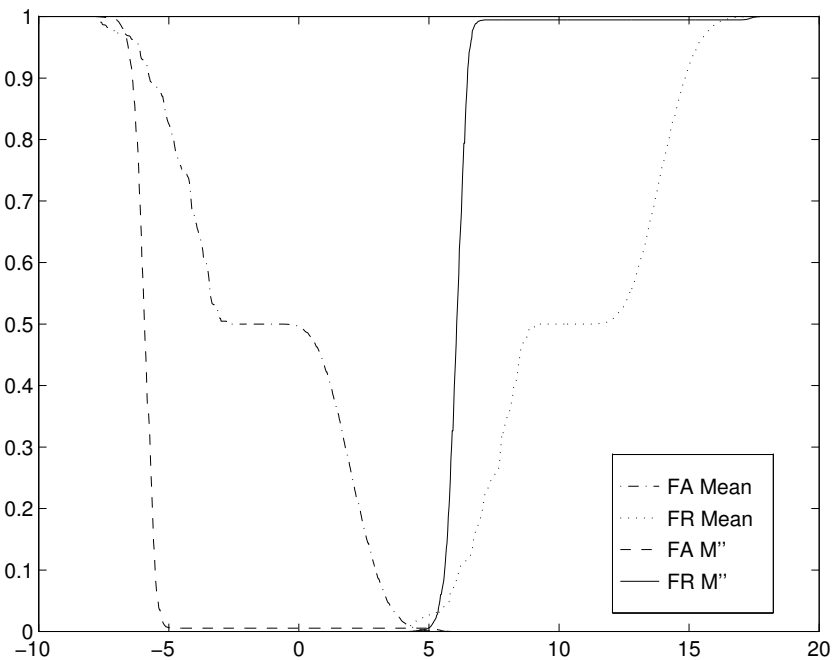
Algorithm implementation

1. (Supervisor Training) By using a training set i.e. x_{ij}, y_j, and p_{ij} with j up to n, (1), estimate the bias parameters of each expert, i.e. $\{M_i, V_i, \alpha_i\}$.
2. (Authentication Phase) The time is *always* $n+1$. The expert opinions are reconciled by computing M'' twice, once assuming that the user is impostor and once client. One chooses the best M'' by taking the one closest to its goal i.e. ϵ' and $(1-\epsilon)'$. M'' is ready to be thresholded to yield a definite decision.

#Training Shots:	2664
#Test Shots:	7992
bs_{n+1} **:**	Bayesian supervisor opinion ($= M''$)
$SR = 1 - (FA + FR)$**:**	Success rate

Expert 1 Speech (using HMM, text dependent)

Expert 2 Face (using Discriminative DLA)



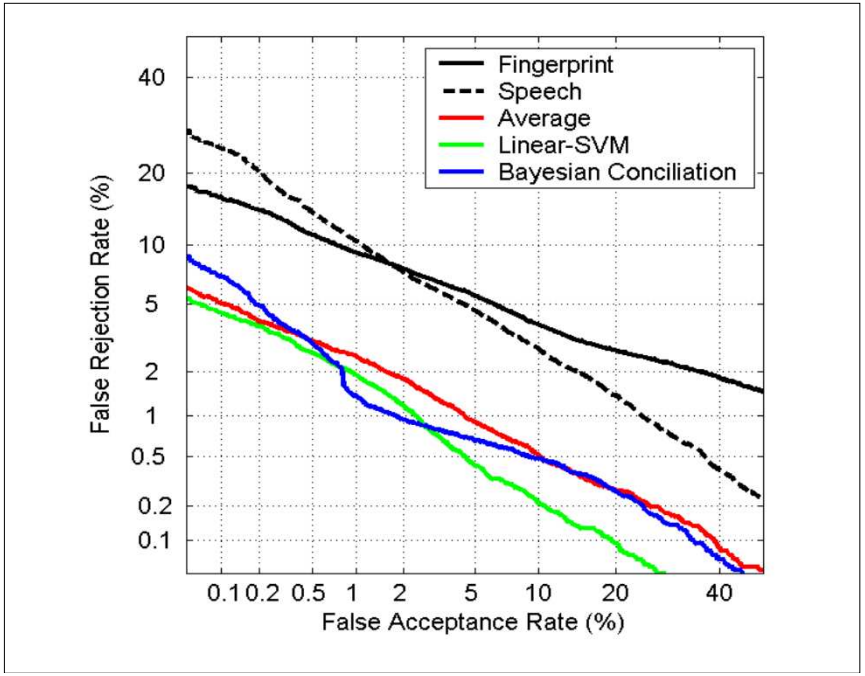
SR_1	SR_2	SR_{bs}	SR_{ms}
0.944	0.965	0.995	0.980

Left: FA and FR curves

Right: Success rates at $T = 0$.

Expert 1 Speech (Speaker: MFCC + GMM/UBM)

Expert 2 Fingerprint (Minutiae + String Matching)



Expert protocol

3 TR, 7 FR, 10 FA) x 75 ind.

Supervisors

Sum Rule, Linear-SVM, Bayesian Conciliation

Superv. Protocol

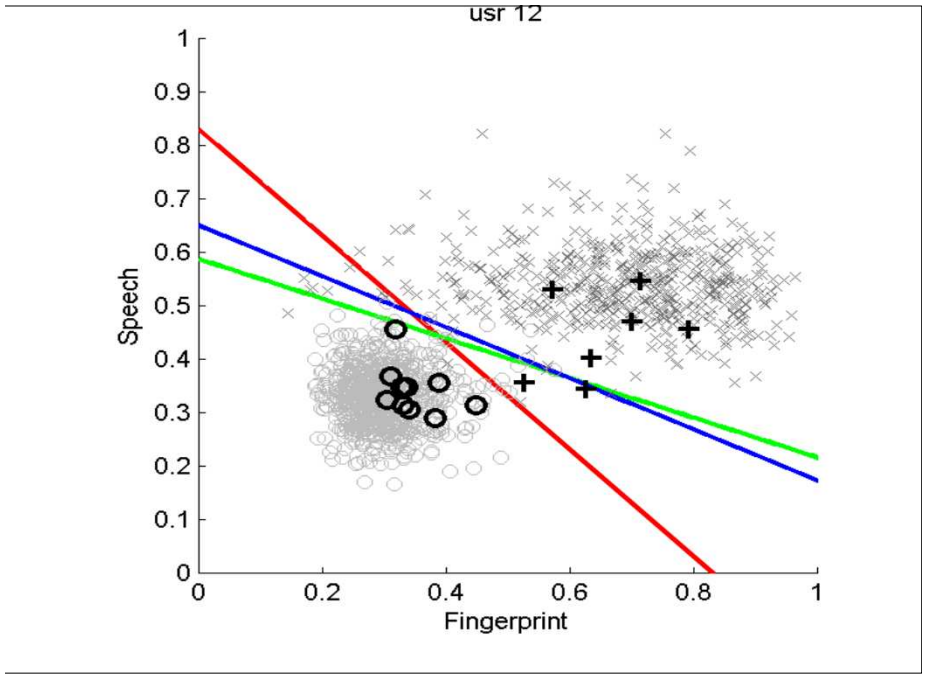
Leave-one-user-out

Database

75 ind. MCYT (Finger) and Mob-ATVS (Mobile)

Expert 1 Speech (Speaker: MFCC + GMM/UBM)

Expert 2 Fingerprint (Minutiae + String Matching)



Why simulating experts?

It is a convenient way of finding out what happens when

i) experts are unequally skilled

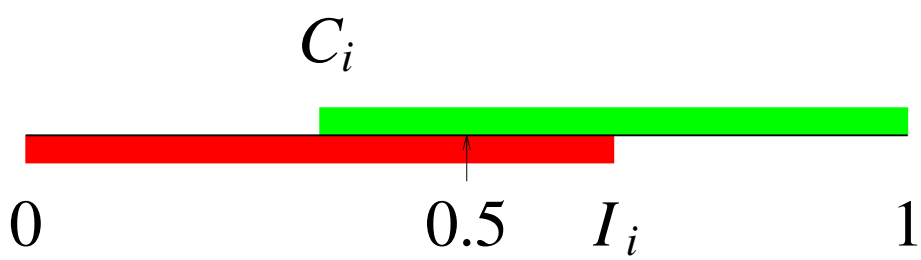
ii) the performance of the supervisor

- if the number of experts increase
- if the number of shots increase

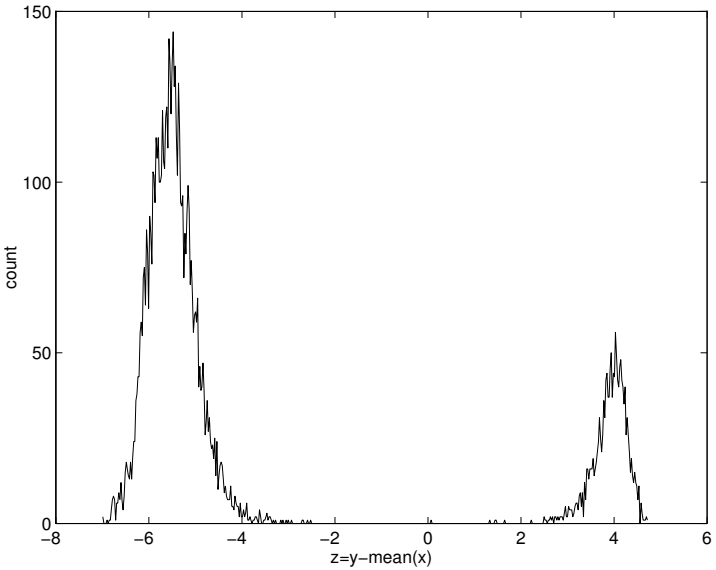
iii) the number of impostors versus clients vary.

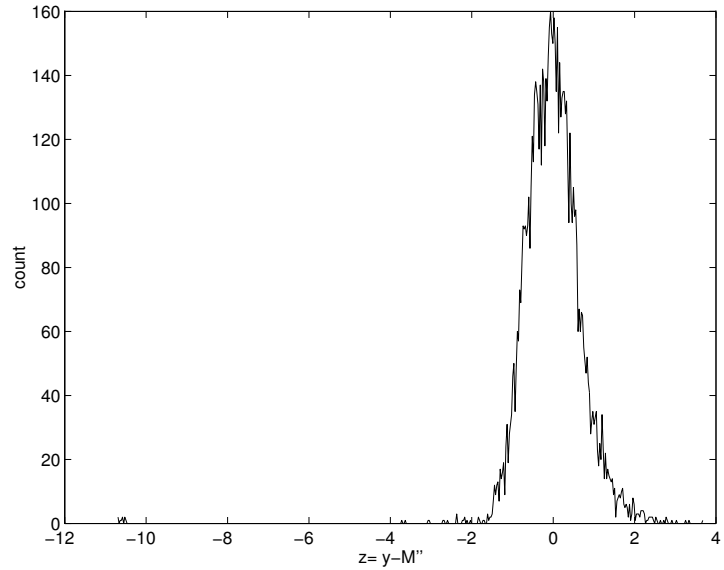
How?

- Generate C_i and I_i according to distributions which are uniform in the intervals (0.25, 0.5) and (0.5, 0.75).
- Generate imposter and client users randomly with $P(\text{impostor}) = 0.8$ $P(\text{client}) = 0.2$.
- Let $f(X_{ij}| \quad \text{user} = \text{impostor})$ and $f(X_{ij}| \quad \text{user} = \text{client})$ be uniformly distributed in $[0, I_j)$ and $[0, C_j)$. Generate X_{ij} accordingly.



Unequally skilled experts





(Left) Mean value supevisor (Right) Bayesian supevisor (4 experts)

SR_1	SR_2	SR_3	SR_4	SR_{bs}	SR_{ms}
0.664	0.821	-	-	0.985	0.815
0.763	0.732	0.701	0.681	0.999	0.872

Equal versus unequally skilled experts

SR_1	SR_2	SR_3	SR_4	SR_{bs}	SR_{ms}
0.814	0.816	0.819	0.825	1.00	0.974
0.649	0.952	0.640	0.608	0.991	0.801

These and other experiments indicate that the Bayesian supervisor is more successful in decision making due its capability of adapting to the miss identification densities well.

Conclusions

We concluded previously (Bigun et. al. 97) that supervisors aware of expertship levels of the experts (trained with weighted-sum rule) outperform unskilled supevisors (simple sum-rule) as soon as expert skills differ.

We conclude here that using the quality signals improves the performance of the decision making which can be crucial for the robustness a mobile use of biometrics would require.

In other words, in this context "weighted voting power" is a better guarantee for a wiser decision than "equal (democratic?) voting power" of the experts.

We can calibrate expert scores by means of the observed misidentifications in the past. The adopted approach can be viewed as a way to calibrate the views of overly optimistic and overly pessimistic experts before fusing them in a weighted sum rule.

The fusion of the calibrated expert opinions is carried out by parameter enstimation of fused probabilities which in turn can be estimated by Bayes theorem.

Simulated experts quickly uncover the weaknesses of supervisors. They can be used to cut the lead time in constructing efficient supervisors.