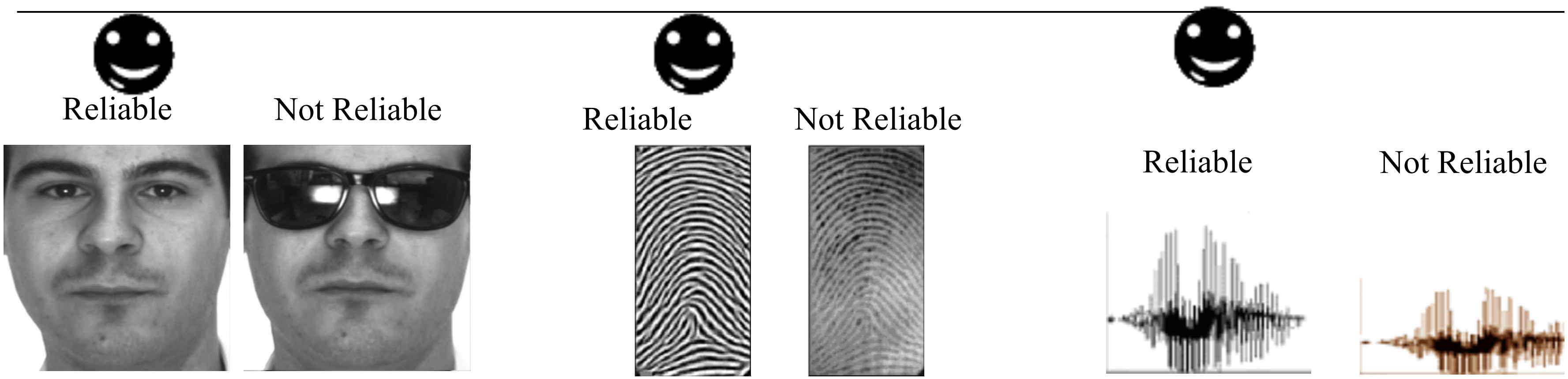
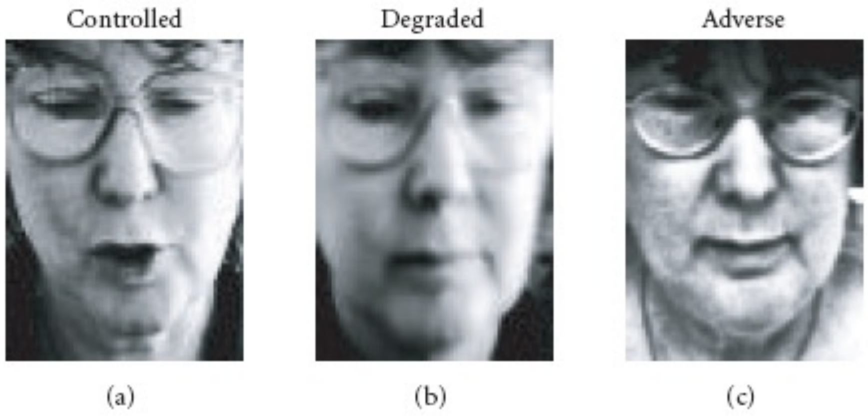
Reliability of an identification system

- Due to the possible different quality of input to different systems, and to possible accuracy in recognition procedures, it may happen that, notwithstanding global FOMs, not all responses are equally reliable.
- The definition of a reliability measure for each single response from a system provides further information to be used in setting up an operation policy (e.g., is f the identification is not reliable enough and if possible, repeat capture), but also to merge results from different systems (multibiometric architectures).



Some approaches: 1) image quality

- Margins based on quality
- (Kryszczuk, Richiardi, Prodanov and Drygajlo, 2006):



Examples from BANCA database

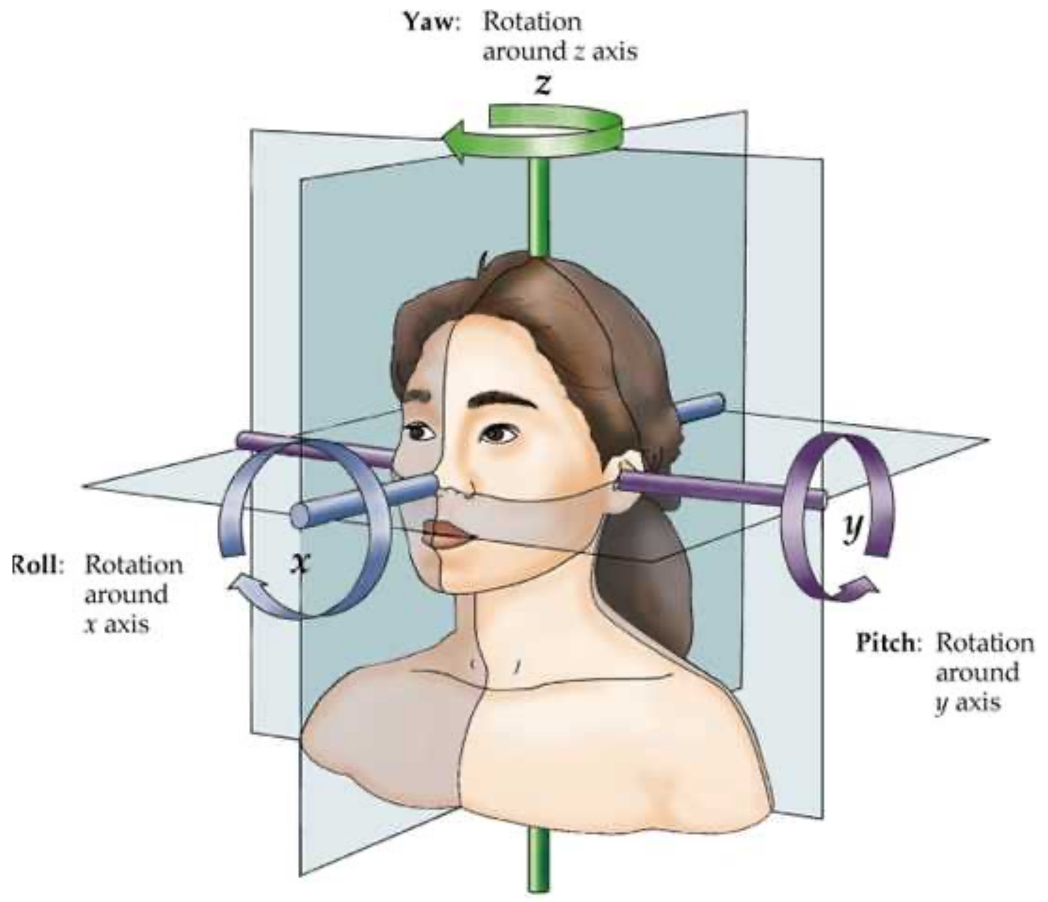
Correlation with an "average" image

The quality of training images can be modeled by creating an "average" template from all faces, the quality of which is taken as a reference.

Estimation of image sharpness

The lack of high frequency details in the image can be dscribed as a loss of sharpness (blurring)

Some approaches: 2) face "quality"
How to measure the "quality" of a face image



How to measure the "quality" of a face image

De Marsico, Nappi, Riccio (2011)

- SP: measure of distortion with respect to frontal pose, expressed in terms of misalignment of roll (α), yaw (β) and pitch (γ):

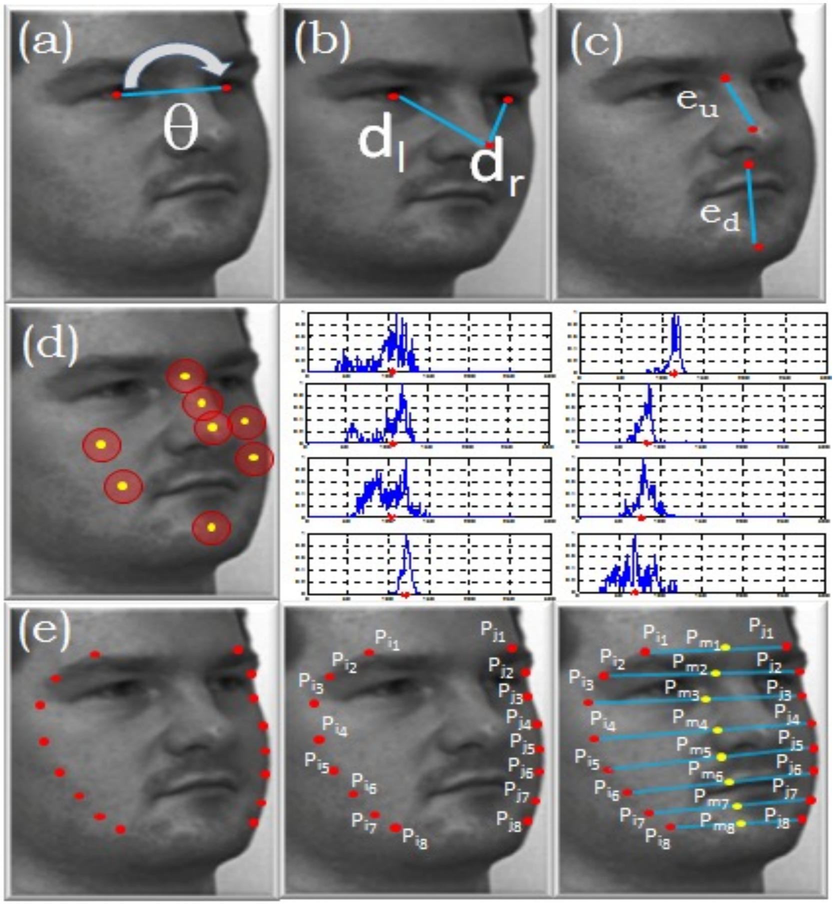
$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch)$$

- SI: is defined as a measure of homogeneity of grey levels in some pre-determined face regions:

$$SI = 1 - F(std(mc))$$

- SY: is defined as a measure of face symmetry.

$$SY = \sum_{(i,j)\in X} sym(P_i, P_j).$$



Other more general "quality" measures

<u>Universal Image Quality Index (UIQI)</u>: any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion

Let $x=\{x_i|i=1 \ldots N\}$ and $y=\{y_i|i=1 .. N\}$ be the original and test image respectively. The index is defined as

$$Q = \frac{4\,\sigma_{xy}\,\bar{x}\,\bar{y}}{(\sigma_x^2+\sigma_y^2)\,[(\bar{x})^2+(\bar{y})^2]}\,, \tag{1}$$

where

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i\,, \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i\,,$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\,, \quad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2\,,$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})\,.$$

Other “quality” measures

Sharpness Estimation Quality Index: In order to estimate the sharpness of an image I of x × y pixels, we compute the mean of intensity differences between adjacent pixels, taken in both the vertical and horizontal directions:

$$SE = \frac{1}{2}\left(\frac{1}{(x-1)y}\sum_{m=1}^{y}\sum_{n=1}^{x-1}\left|p_{n,m} - p_{n+1,m}\right| + \frac{1}{(y-1)x}\sum_{m=1}^{y-1}\sum_{n=1}^{x}\left|p_{n,m} - p_{n,m+1}\right|\right)$$

Example tests on face datasets

FERET: first 250 images of the fa (frontal) group, corresponding to 116 subjects

LFW (Labeled Face in the Wild) : 480 images, the first 6 of the first 80 subjects

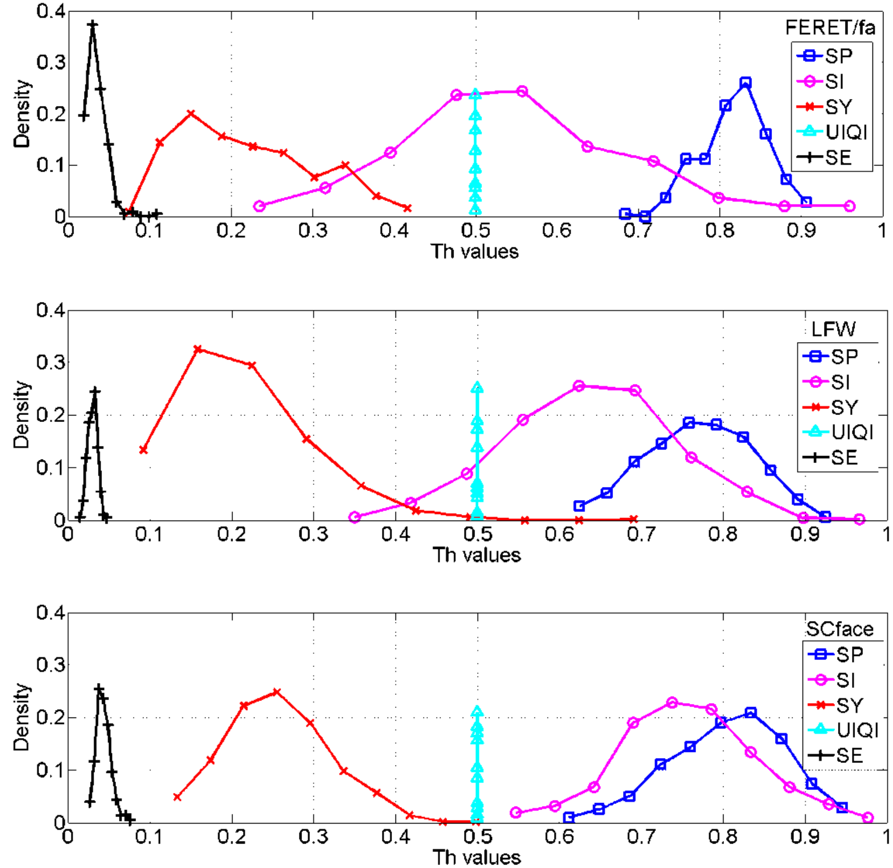
Scface: 650 images of 130 subjects, corresponding to groups cam1-5 (visible light) of the subgroup dist3 (greater distance)

How to measure the “quality” of a quality measure

- The first test for a quality measure is to check how the values of a dataset are distributed w.r.t. the returned values.

- This allows to understand which is the average level of quality of a face dataset w.r.t. to a specific measure.

- Two or more measures can be compared by computing the amount of correlation of the returned values w.r.t. to the images of given dataset.

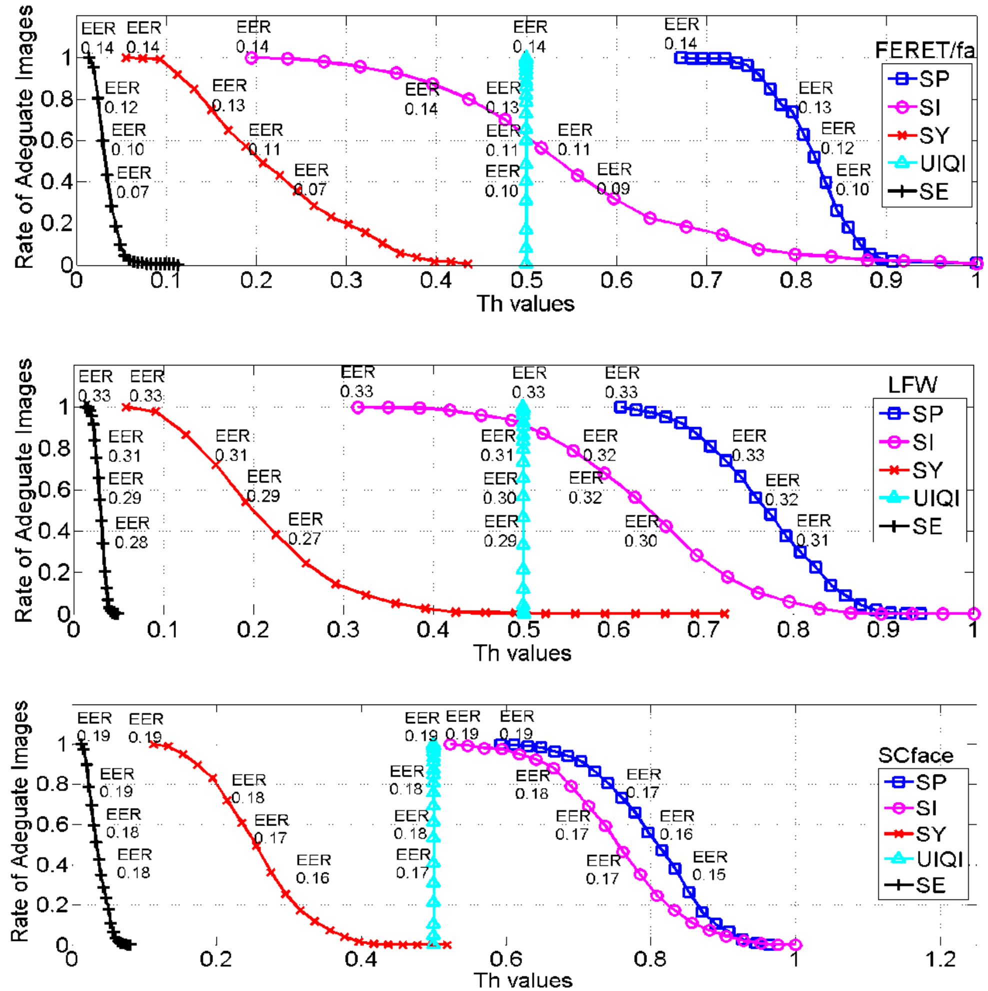


How to measure the “quality” of a quality measure

- A measure of quality of input samples allows to discard a-priori (before recognition) those affected by too high distortion which would lead to a wrong response or not reliable from the system.

- A further test for a quality measure is to evaluate how it affects the system performance (EER) by varying a tolerance threshold.

- A good quality measure must provide error good decrease by discarding as few samples as possible.



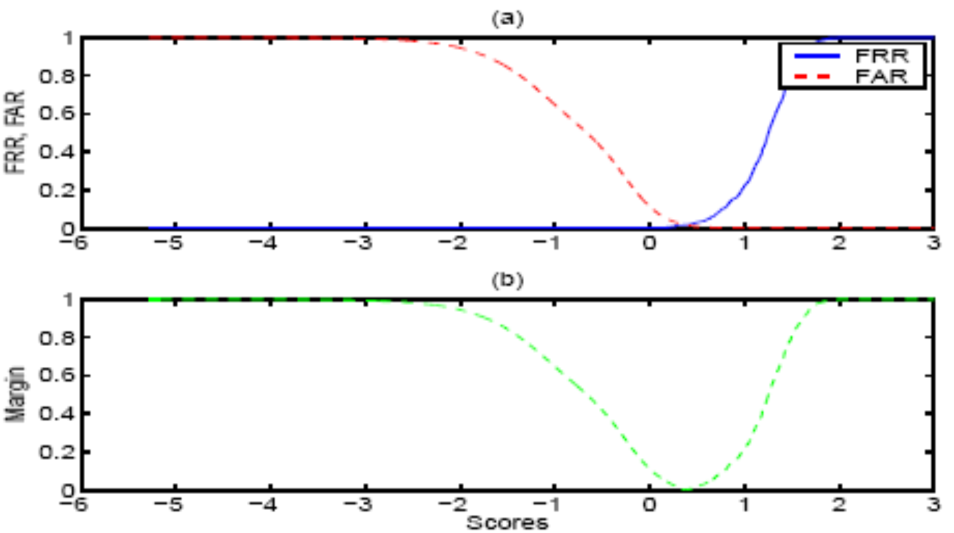
Some approaches: 3) margins based on error estimation

Poh and Bengio, 2004

System performance is measured in terms of:

$$\mathrm{FAR}(\Delta) = \frac{\text{number of FAs}(\Delta)}{\text{number of impostor accesses}},$$

$$\mathrm{FRR}(\Delta) = \frac{\text{number of FRs}(\Delta)}{\text{number of client accesses}}.$$



Margin $M(\Delta)$ is defined as:

$$\mathcal{M}(\Delta) = |\mathrm{FAR}(\Delta) - \mathrm{FRR}(\Delta)|.$$

Some approaches: 4) System Response Reliability (SRR)

There is a major difference between a quality measure for an input sample and a reliability measure for the response of a biometric system.

System Response Reliability ($srr \in [0,1]$) index measures the ability of an identification system to separate genuine subjects from impostors on a single probe basis.

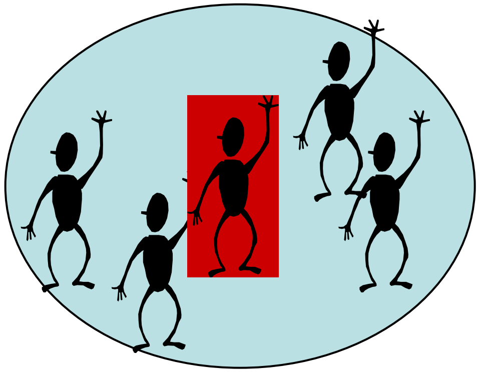
The SRR relies on different versions of function φ. We defined and tested two different φ functions:

- Relative distance;
- Density ratio;

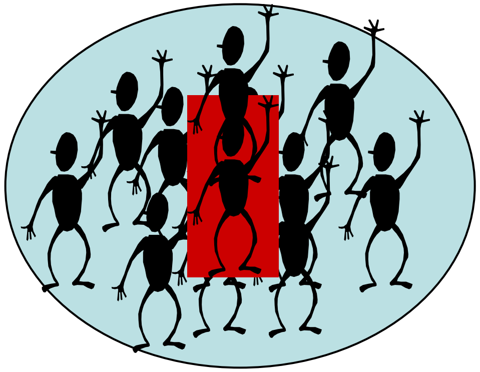
Both functions measure the amount of "confusion" among possible candidates.

We assume that the result of an identification operation is the whole gallery ordered by distance from the probe, or a short list at least.

SRR



Cloud around the returned subject
“less crowded” =
More reliable response



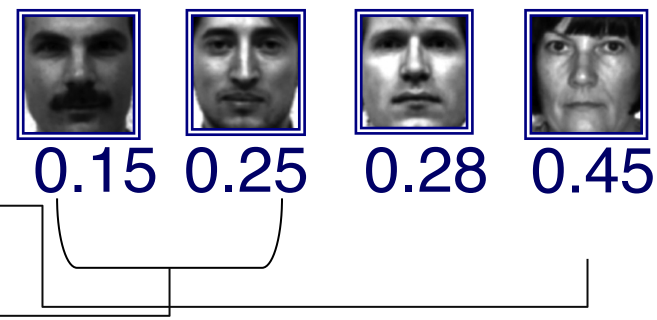
Cloud around the returned subject
“more crowded” =
Less reliable response

SRR

Given a probe *p* and a system *A* with gallery *G*, the ***relative distance*** is defined as:

$$\varphi(p) = \frac{F\left(d\left(p, g_{i_2}\right)\right) - F\left(d\left(p, g_{i_1}\right)\right)}{F\left(d\left(p, g_{i_{|G|}}\right)\right)}$$

0.25 – 0.15 = 0.10



The lower the difference in the numerator with respect to the denominator (the maximum computed difference with the probe), the higher the possible confusion related to the first two candidates, the lower the reliability.

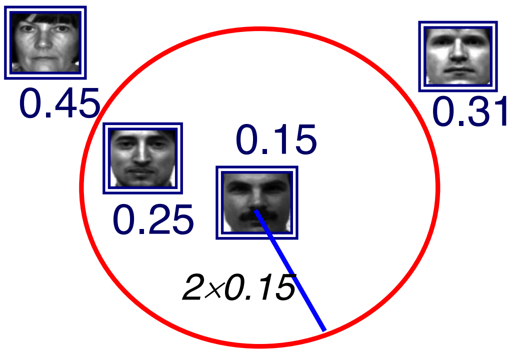
SRR

Given a probe *p* and a system *A* with gallery *G*, the ***density ratio*** is defined as:

$$\varphi(p) = 1 - |N_b| / |G|$$

With

$$N_b = \{ g_{i_k} \in G \mid F(d(p, g_{i_k})) < 2 \cdot F((d(p, g_1))) \}$$



$2 \times 0.15 = 0.30$

Density Ratio = 1 – 1/3
= 0.66 …

Both $|N_b|$ and $|G|$ are computed WITHOUT considering the element in the first list position in order to have a maximum value =1 (but this is not strictly necessary)

This function is less sensible to outliers, and in fact usually performs better than φ_1.

As a drawback, its definition takes to consider narrower clouds when the first retrieved identity is closer to the probe. On the contrary, a large distance takes to a larger cloud, which can be expected to be more crowded in any case. Any attempt to substitute 2 with an adaptable parameter did not achieve better results.

SRR

We need to identify a value fostering a correct separation between wrong rejections of enrolled subjects and wrong recognitions of not enrolled ones, both supported by the reliability value.

The critical φ_k is given by that value able to minimize the wrong estimates of function $\varphi(p)$, i.e. not enrolled subjects erroneously recognized (FA caused by a distance below the acceptance threshold or a similarity above) with $\varphi(p)$ higher than φ_k, or genuine subjects wrongly rejected (FR caused by a distance above the acceptance threshold or a similarity below) because recognized with $\varphi(p)$ lower than φ_k.

The distance between $\varphi(p)$ and φ_k is significant for reliability.

SRR

We also define as the width of the subinterval from to the proper extreme of the overall [0,1) interval of possible values, depending on the comparison between the current *φ(p)* and :

$$S\left(\varphi(p),\overline{\varphi}\right)=\begin{cases}1-\overline{\varphi} & \mathit{if} \quad \varphi(p)>\overline{\varphi}\\ \overline{\varphi} & \mathit{otherwise}\end{cases}$$

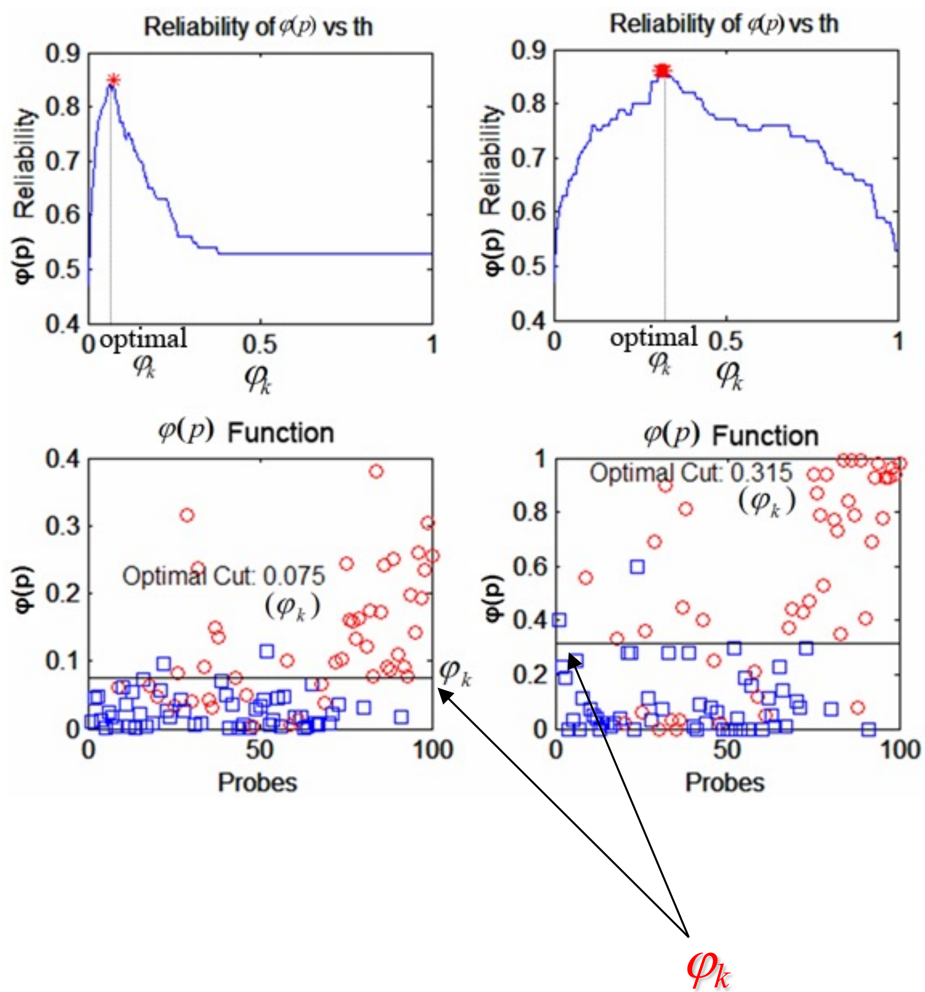
SRR index can finally be defined as:

$$\boxed{SRR=(\varphi(p)-\overline{\varphi})/S(\overline{\varphi})}$$

SRR

Density Ratio

Relative Distance



Red circles = GAs
Blue squares = FAs

Blue squares **above** the critical value = FAs **confirmed** by a high value of φ

Red circles **below** the critical value = GAs **not confirmed** due to a low value of φ

A threshold for SRR

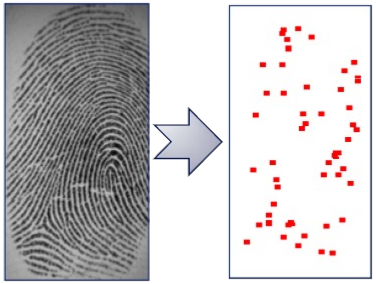
- The reliability threshold *th* can be automatically estimated by exploiting a certain number M of successive observations.

- We would desire to have a high average (the system is generally reliable) and a low variance (the system is stable)-

- We can summarize as:

$$th_i = \left| \frac{\mathrm{E}[\bar{S}_i]^2 - \sigma[\bar{S}_i]}{\mathrm{E}[\bar{S}_i]} \right|$$

Among possible solutions to increase quality/reliability: Template Updating



- Features extracted from a sample of a biometric trait, which are labeled with the individual's identity, represent its *template.*
  - Matching exploits the template, not the sample
  - A template "should not allow ro reconstruct" a valid sample
  - Size aids codings and storing on more devices
  - Different template are generated any time the individual provides a biometric sample
- During operation of the recognition system much more biometric data become available, which were acquired over time. The system can use such data to *update the templates* in the gallery on a regular basis in order to address
  - Template ageing
  - Template enhancing

Among possible solutions: Template Updating

- Label assignment
  - Supervised systems
    - They require a supervisor to assign identity labels to newly acquired data during recognition system operation.
    - They usually work offline
  - Semi-supervised
    - They use the union of labeled and unlabeled data.
    - They work both online and offline.
- Most representative template selection to perform.
  - Online
    - Selection is performed as soon as new input data is acquired by the recognition system.
  - Offline
    - Selection is performed after a certain amount of data has been acquired during a specific time elapse.

Some references

- R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, A. W. Senior, "The Relation between the ROC Curve and the CMC," AUTOID, pp.15 -20, Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05), 2005

- G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheeps, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST1998 speaker recognition evaluation. In Proc. ICSLD, 1998. http://www.dtic.mil/dtic/tr/fulltext/u2/a528610.pdf

- Mohammad Nayeem Teli, J. Ross Beveridge, P. Jonathon Phillips, Geof H. Givens, David S. Bolme, Bruce A. Draper. Biometric Zoos: Theory and Experimental Evidence. http://www.csis.pace.edu/~ctappert/dps/2011IJCB/papers/327.pdf

- Yager, N.; Dunstone, T., "Worms, Chameleons, Phantoms and Doves: New Additions to the Biometric Menagerie," *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on* , vol., no., pp.1,6, 7-8 June 2007. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4263204&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4263204

- Torralba, A., & Efros, A. A. (2011, June). Unbiased look at dataset bias. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 1521-1528). IEEE.

Some references

- Wang, Z., & Bovik, A. C. (2002). A universal image quality index. IEEE signal processing letters, 9(3), 81-84.

- K. Kryszczuk, J. Richiardi, P. Prodanov and A. Drygajlo, "Reliability-based decision fusion in multimodal biometric verification", EURASIP Journal on Advances in Signal Processing 2006, Volume 2007 (2007), Article ID 86572, 9 pages.

- De Marsico, M., Nappi, M., & Riccio, D. (2011, November). Measuring measures for face sample quality. In *Proceedings of the 3rd international ACM workshop on Multimedia in forensics and intelligence* (pp. 7-12). ACM.

- N. Poh, S. Bengio, Improving Fusion with Margin-Derived Confidence In Biometric Authentication Tasks, IDIAP-RR 04-63, November 2004.

- M. De Marsico, M. Nappi, D. Riccio, G. Tortora. NABS: Novel Approaches for Biometric Systems. IEEE Transactions on Systems, Man, and Cybernetics — Part C: Applications and Reviews. Volume: 41 Issue:4, July 2011, pp. 481-493