BIOSECURE TOOL

PERFORMANCE EVALUATION OF A BIOMETRIC VERIFICATION SYSTEM

`version 1.0 - 01/12/2007`

BIOSECURE

Abstract. *This BioSecure tool plots traditional curves and gives some interesting values (with confidence intervals) in order to evaluate the performance of a biometric verification system. It uses the Matlab platform.*

Introduction

This 'how to' is divided into two main parts. In the first part, we present the criteria and methods used to evaluate the performance of a biometric verification system whereas in the second part, we explain how to use the tool in practical.

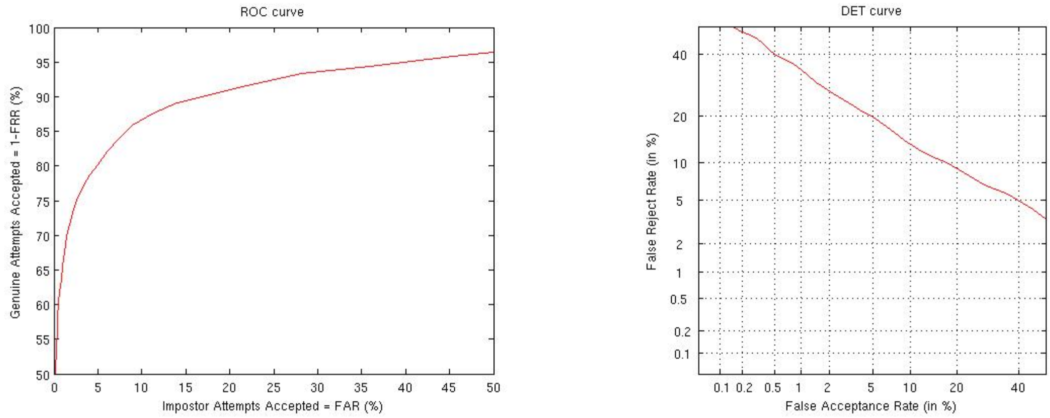
I.Performance evaluation

1.Criteria

Four criteria (two curves and two values) have been selected from the state-of-the-art [1] to evaluate the performance of a biometric verification system :

– **Receiver Operating Characteristic (ROC) curve** : This curve is used to summarize the performance of a biometric verification system. An ROC curve plots, parametrically as a function of the decision threshold, the percentage of impostor attempts accepted (i.e. false acceptance rate (FAR)) on the x-axis, against the percentage of genuine attempts accepted (i.e. 1 - false rejection rate (FRR)) on the y-axis (Fig. 1(a)). The ROC curve is threshold independent, allowing performance comparison of different systems under similar conditions.

– **Detection Error Trade-off (DET) curve** : In the case of biometric systems, the DET curve is often preferred to the ROC curve. Indeed, the DET curve plots error rates on both axes (FAR on the x-axis against FRR on the y-axis) using normal deviate scale (Fig. 1(b)), what spreads out the plot and distinguishes different well-performing systems more clearly [2].



(a) (b)

Fig. 1 – Traditional curves used to evaluate the performance of biometric verification systems : (a) ROC curve (b) DET curve

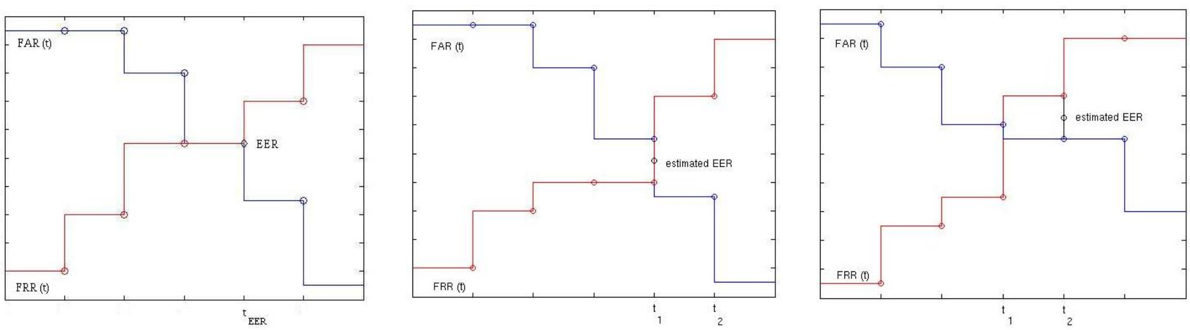
– **Equal Error Rate (EER)** : The equal error rate is computed as the point where $FAR(t) = FRR(t)$ (Fig. 2(a)). In practice, the score distributions are not continuous and a crossover point might not exist. In this case (Fig. 2(b),(c)), the EER value is computed as follows :

$$EER = \begin{cases} \frac{FAR(t_1)+FRR(t_1)}{2} & \text{if } FAR(t_1) - FRR(t_1) \leqslant FRR(t_2) - FAR(t_2) \\ \frac{FAR(t_2)+FRR(t_2)}{2} & \text{otherwise} \end{cases}$$

where

$$t1 = \max_{t \in S}\{t|FRR(t) \leqslant FAR(t)\},\ t2 = \min_{t \in S}\{t|FRR(t) \geqslant FAR(t)\}$$

and S is the set of thresholds used to calculate the score distributions.



(a) EER point exists (b) EER is estimated at t_1 (c) EER is estimated at t_2

Fig. 2 – FAR vs FRR curve : (a) example where EER point exists (b),(c) examples where EER point does not exist

- **Operating Point (OP)** : In practice, biometric systems operate at a low FAR instead of the equal error rate in order to provide high security. This operating point is defined in terms of FRR (%) achieved for a fixed FAR. The fixed value α of FAR depends on the security level required by the verification system. In practice, the OP is computed as follows :

$$OP_{\{FAR=\alpha\}} = FRR(t_{OP}) \mid t_{OP} = \max_{t \in S}\{t|\alpha \leqslant FAR(t)\}$$

  where S is the set of thresholds used to calculate the score distributions.

2. Confidence intervals

A 90% interval of confidence will be provided for the EER and the OP value. This allows for determining whether accuracy differences between systems are really statistically significant.
Using a parametric method [3], we could calculate error margins on $FAR(t)$ and $FRR(t)$ at some threshold t (see Annex) such as :

- $FAR(t) \in [\hat{\mathrm{FAR}}(t) - err_{\hat{\mathrm{FAR}}(t)}, \hat{\mathrm{FAR}}(t) + err_{\hat{\mathrm{FAR}}(t)}]$
  where $\hat{\mathrm{FAR}}$(t) is an estimation of FAR at threshold t and $err_{\hat{\mathrm{FAR}}(t)}$ is the error margin on this value.
- $FRR(t) \in [\hat{\mathrm{FRR}}(t) - err_{\hat{\mathrm{FRR}}(t)}, \hat{\mathrm{FRR}}(t) + err_{\hat{\mathrm{FRR}}(t)}]$
  where $\hat{\mathrm{FRR}}$(t) is an estimation of FRR at threshold t and $err_{\hat{\mathrm{FRR}}(t)}$ is the error margin on this value.

In this way, a confidence interval on the EER value is :

- $EER \in [\hat{\mathrm{EER}} - err_{\hat{\mathrm{EER}}}, \hat{\mathrm{EER}} + err_{\hat{\mathrm{EER}}}]$
  where $\hat{\mathrm{EER}} = \frac{\hat{\mathrm{FAR}}(t_{EER}) + \hat{\mathrm{FRR}}(t_{EER})}{2}$ and $err_{\hat{\mathrm{EER}}} = \frac{err_{\hat{\mathrm{FAR}}(t_{EER})} + err_{\hat{\mathrm{FRR}}(t_{EER})}}{2}$
  and t_{EER} is the threshold at which the EER has been evaluated.

II.Presentation of the tool

This tool is made of several functions coded in Matlab. The main fuction `EER_DET_conf.m` calls all the other functions.

Description : the function `EER_DET_conf.m` plots traditional curves (ROC, DET and FAR vs FRR) and outputs some statistical values (EER, OP and error margins) in order to evaluate the performance of a biometric verification system.

Use : `[EER errEER OP errOP] = EER_DET_conf (clients, impostors,` α`, n)`

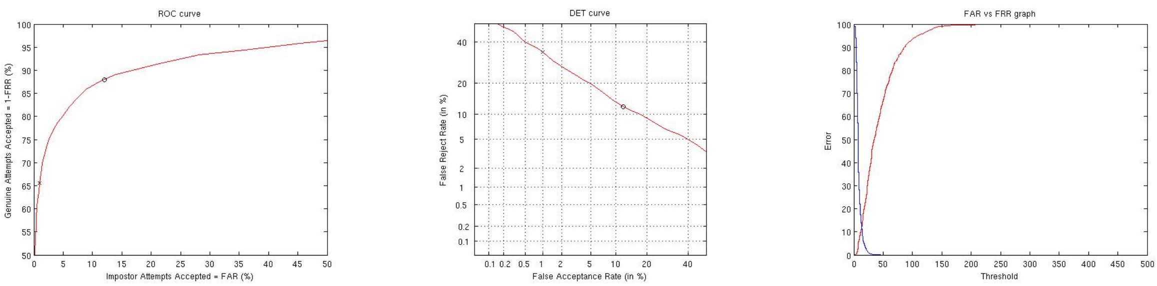
Inputs :
- `clients` : vector of genuine similarity scores
- `impostors` : vector of impostor similarity scores
- α : FAR value (in %) at which the OP value is estimated

– `n` : number of thresholds used to calculate the score distributions (10000 is the advised value for this parameter)

Outputs :
- `EER` : estimation of the EER value (in %)
- `errEER` : error margin on the estimation of the ERR value such as the 90% confidence interval around this value is [`EER-errEER` , `EER+errEER`]
- `OP` : estimation of the OP value (in %), i.e. estimation of the FRR value achieved for the fixed value α of FAR
- `errOP` : error margin on the estimation of the OP value such as the 90% confidence interval around this value is [`OP-errOP` , `OP+errOP`]

Furthermore, the function `EER_DET_conf.m` plots the next curves :

- ROC curve (Fig. 3(a))
- DET curve (Fig. 3(b))
- FAR vs FRR curve which plots FAR and FRR against the decision thresholds (Fig. 3(c))



(a) (b) (c)

Fig. 3 – Curves provided by the function `EER_DET_conf.m` : (a) ROC curve (b) DET curve (c) FAR vs FRR curve

Conclusion

This BioSecure tool has been described and fully tested. It can be used to easily evaluate the performance of every biometric verification systems.

ANNEX
Parametric confidence intervals

In this section, we present the parametric method used to estimate the confidence intervals on the FRR and FAR values. This method has already been explained by *R.M. Bolle et al.* in [4].

Suppose we have M client scores and N impostor scores. We denote these sets of scores by $\mathbf{X} = \{X_1, \ldots, X_M\}$ and $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ respectively. In the following, we suppose that available scores are similarity measures.
Let $\mathbf{S}$ be the set of thresholds used to calculate the score distributions.

For the set of client scores, $\mathbf{X}$, assume that this is a sample of M numbers drawn from a population with distribution F, that is, $F(x) = Prob(X \leqslant x), x \in \mathbf{S}$.
Let the impostor scores $\mathbf{Y}$ be a sample of N numbers drawn from a population with distribution $G(y) = Prob(Y \leqslant y), y \in \mathbf{S}$.
In this way, $FRR(x) = F(x)$ and $FAR(y) = 1 - G(y)$, x and $y \in \mathbf{S}$.
From now, we have to find an estimate of these distributions at some threshold $t_0 \in \mathbf{S}$ and then, we have to estimate the confidence interval for these estimations.

- The estimate of $F(t_0)$ using data $\mathbf{X}$ is the unbiased statistic :

$$\hat{\mathrm{F}}(t_0) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}(X_i \leqslant t_0) \tag{1}$$

  $\hat{\mathrm{F}}(t_0)$ is so obtained by simply counting the $X_i \in \mathbf{X}$ that are smaller than t_0 and dividing by M.

  In the same way, the estimate $G(t_0)$ using data $\mathbf{Y}$ is the unbiased statistic :

$$\hat{\mathrm{G}}(t_0) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(Y_i \leqslant t_0) \tag{2}$$

- In the following, let us concentrate on the distribution F. For the moment, let us keep $x = t_0$ and let us determine the confidence interval for $\hat{\mathrm{F}}(t_0)$.
  First define Z as a binomial random variable, the number of successes, where success means $(X \leqslant t_0)$ is true, in M trials with probability of success $F(t_0) = Prob(X \leqslant t_0)$. This random variable Z has binomial probability mass distribution :

$$P(Z = z) = \binom{M}{z} F(t_0)^z (1 - F(t_0))^{M-z}, \quad z = 0, \ldots, M \tag{3}$$

  The expectation of Z is $E(Z) = MF(t_0)$ and the variance is $\sigma^2(Z) = MF(t_0)(1 - F(t_0))$.
  From this, it follows that the random variable Z/M has expectation $F(t_0)$ and variance $F(t_0)(1 - F(t_0))/M$. When M is large enough, using the law of large numbers, Z/M is distributed according to a normal distribution, i.e., $Z/M \sim N(F(t_0), F(t_0)(1 - F(t_0))/M)$.

Now, it can be seen that $\hat{\mathrm{Z}}/M$=$\hat{\mathrm{F}}(t_0)$. Hence, for large M, $\hat{\mathrm{F}}(t_0)$ is normally distributed, with an estimate of the variance given by :

$$\hat{\sigma}(t_0) = \sqrt{\frac{\hat{\mathrm{F}}(t_0)(1 - \hat{\mathrm{F}}(t_0))}{M}} \tag{4}$$

So, confidence intervals can be determined. For example, a 90% interval of confidence is :

$$F(t_0) \in [\hat{\mathrm{F}}(t_0) - 1.645\hat{\sigma}(t_0), \hat{\mathrm{F}}(t_0) + 1.645\hat{\sigma}(t_0)] \tag{5}$$

Estimates $\hat{\mathrm{G}}(t_0)$ for the probability distribution $G(t_0)$ using a set of impostor scores $\mathbf{Y}$ can be obtained in a similar fashion. Parametric confidence intervals are computed by replacing $\hat{\mathrm{F}}(t_0)$ with $\hat{\mathrm{G}}(t_0)$ in equations (4) and (5).

REFERENCES

[1] A.J. Mansfield and J.L. Wayman. *Best Practices in Testing and Reporting Performance of Biometric Devices*. Version 2.01, August 2002.

[2] A. Martin, G.R. Doddington, T. Kamm, M. Ordowski, and M.A. Przybocki. The DET curve in assessment of detection task performance. In *Proceedings of Eurospeech*, pages 1895–1898, Rhodes, Greece, 1997.

[3] R.M. Bolle, N.K. Ratha, and S. Pankanti. Evaluation techniques for biometrics-based authentication systems (FRR). In *Proc. 15th Internat. Conf. Pattern Recogn.*, pages 835–841, 2000.

[4] R.M. Bolle, N.K. Ratha, and S. Pankanti. Error analysis of pattern recognition systems - the subsets bootstrap. *Computer Vision and Image Understanding*, 93 :1–33, 2004.