Neurocomputing 248 (2017) 57–66

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Using the score ratio with distance-based classifiers: A theoretical and practical study in biometric signature recognition

Carlos Vivaracho-Pascual[1],*, Arancha Simon-Hurtado, Esperanza Manso-Martinez

Computer Science Department, Escuela de Ingeniería Informática, Universidad de Valladolid, Campus Miguel Delibes, Paseo de Belén, 15, Valladolid 47011, Spain

ARTICLE INFO

Article history:
Received 29 June 2016
Revised 25 November 2016
Accepted 26 November 2016
Available online 8 March 2017

Keywords:
Score ratio
Signature verification
Distance-based classifier

ABSTRACT

The binary classification problem where an input is classified as belonging or not to a certain class, the so-called Target Class (TC), is approached here. This problem can be stated as a basic hypothesis test: X is from the TC (H_0) vs. X is not from the TC (H_1), where X is the classifier input. When probabilistic models are used (e.g., Hidden Markov Models or Gaussian Mixture Models), the likelihood ratio, $p(X/H_0)/p(X/H_1)$, is an alternative widely used to improve the classification. However, as far as we know, this ratio is not usually applied with distance-based classifiers (e.g., Dynamic Time Warping). Following that idea, here we propose making the decision based not only on the score ("score" being the classifier output) assuming X to be from the TC (H_0), but also using the score assuming X is not from the TC (H_1), by means of the ratio between both scores: the score ratio. The proposal is tested in biometric person authentication using manuscript signature, with three different state-of-the-art systems based on distance classifiers. Different alternatives for applying the proposal are shown in order to reduce the computer load, should it prove necessary. Using the score ratio has led to improvements in most of the tests performed. The best verification results were achieved using our proposal, with the best ones without the score ratio being improved by an average of 22%.

1. Introduction

This paper focuses on the use of our "score ratio" proposal in binary classification problems where we have a Target Class, TC, and the goal is to classify an input sample as belonging or not to this TC. One representative, and very important example of this binary classification problem from the research and practical standpoint, is biometric person verification, with the score ratio being tested here in that field of work. Below, we single out our proposal in the biometric verification problem. Generalization to other fields is immediate.

The goal in biometric person verification is to authenticate the user or client C (the Target Class, here) by means of unique human characteristics (biometrics, e.g., iris, fingerprint, etc.). Fig. 1 shows the main modules (stages) in a biometric recognition system. In this figure, the signature is used as biometry, since it is the one approached in this paper, although the modules are the same for any other biometry. This work is focused on the Match stage. We have approached the Feature Extraction and Decision stages in previous works, [1] and [2].

Given a test sample (feature vector) X (Fig. 1), the problem of biometric verification can be stated as a basic hypothesis test between two hypotheses:

$$H_0: \ X \ is \ from \ client \ C \qquad H_1: \ X \ is \ not \ from \ client \ C$$

The decision between the two hypotheses can be made as shown in Eq. (1), using client information only, or can be made by means of the likelihood ratio test given by Eq. (2), also using impostor (the Non TC, NTC) information. $p(X/H_0)$ and $p(X/H_1)$ are, respectively, the probability density functions for hypotheses H_0 and H_1 evaluated for the observed biometric sample X, and θ is the decision threshold.

$$p(X/H_0) \begin{cases} \geq \theta & Accept \quad H_0 \\ < \theta & Reject \quad H_0 \end{cases} \tag{1}$$

$$\frac{p(X/H_0)}{p(X/H_1)} \begin{cases} \geq \theta & Accept \quad H_0 \\ < \theta & Reject \quad H_0 \end{cases} \tag{2}$$

Biometric verification in general, and signature verification in particular, are pattern recognition problems, where each client C is represented by means of a model λ_C, e.g., Hidden Markov Model

* Corresponding author.
E-mail addresses: cevp@infor.uva.es (C. Vivaracho-Pascual), arancha@infor.uva.es (A. Simon-Hurtado), manso@infor.uva.es (E. Manso-Martinez).
[1] www.infor.uva.es/~cevp.

http://dx.doi.org/10.1016/j.neucom.2016.11.080

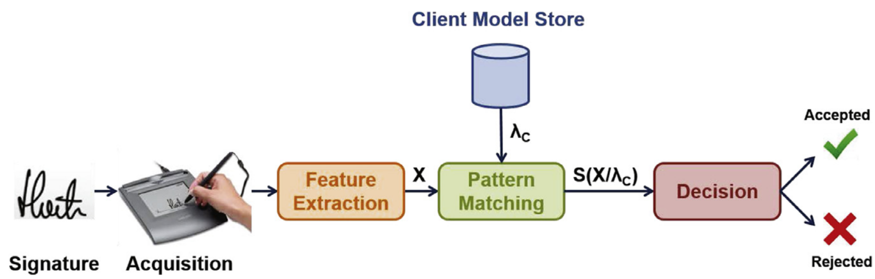


Fig. 1. Main modules (stages) in a usual biometric recognition system. Decision is performed with $s(X/\lambda_C)$, i.e., without score ratio.

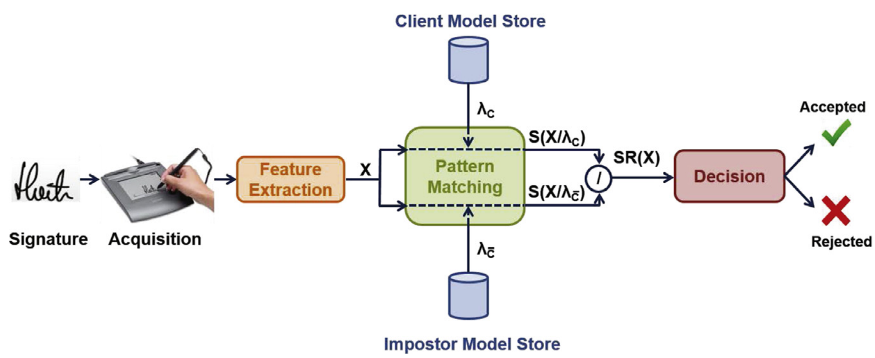


Fig. 2. Biometric recognition system with score ratio.

(HMM), Gaussian Mixture Model (GMM), Support Vector Machine, etc. The probability $p(X/H_0)$, whose calculation is not a straightforward task, is then estimated (approximated) by means of the classifier output (score) $s(X/\lambda_C)$ (Fig. 1). The likelihood denominator, $p(X/H_1)$, is estimated by means of the score $s(X/\lambda_{\overline{C}})$, where $\lambda_{\overline{C}}$ is the impostor model (Fig. 2), with the "impostor" being anybody other than the client. Since an accurate representation of the impostor class is impossible, different $\lambda_{\overline{C}}$ model estimation approaches can be found in the literature. Several of these will be shown in Section 3.

Following this pattern recognition approximation, Eq. (1) becomes Eq. (3), and Eq. (2) becomes the score ratio shown in Eq. (4).

$$s(X/\lambda_C)\begin{cases}\geq \theta & Accept \quad H_0\\ < \theta & Reject \quad H_0\end{cases} \tag{3}$$

$$\frac{s(X/\lambda_C)}{s(X/\lambda_{\overline{C}})}\begin{cases}\geq \theta & Accept \quad H_0\\ < \theta & Reject \quad H_0\end{cases} \tag{4}$$

When probability-based classifiers (e.g., HMM or GMM) are used, the classifier output can be interpreted as a probability, $p(X/\lambda_C)$, and the decision has typically been performed using the likelihood ratio $\frac{p(X/\lambda_C)}{p(X/\lambda_{\overline{C}})}$ test [3–5], since better performance is achieved. However, when distance-based classifiers are used, as far as we know, this score ratio (Eq. (4)) is not usually applied. Our proposal of score ratio is shown graphically in Fig. 2 in comparison with a usual biometric system where the final decision is only based on $s(X/\lambda_C)$ (Fig. 1).

In [6], an initial approach to this proposal was successfully made, showing that use of the score ratio in biometric systems based on distance classifiers can improve the system. In other words, it can be an interesting alternative. Here, a more in-deep twofold study is provided, that is:

1. Firstly, the same score ratio "basic approach" as in [6] is tested, but following a more realistic (Section 6.3) experimental setup.
2. Secondly, a practical study about the extra computing load introduced by the $s(X/\lambda_{\overline{C}})$ calculation is also undertaken. In addition to this analysis, different alternatives aimed at reducing the extra calculations are proposed and tested.

The study has been performed in biometric signature recognition. Of the several biometrics, signature is the second most important in behavioral biometrics. Here, *on-line* signature (the signature is written in a digitizing device) is used. Depending on the test conditions, two types of forgeries can be established:

- *Skilled forgery*, where the impostor imitates the client signature.
- *Random forgery*, where the impostor uses his/her own signature as a forgery.

Three different state-of-the-art signature recognition systems based on distance classifiers [7–9] have been tested.

The rest of the paper is organized as follows. In Section 2, a notation section is provided so as to make the paper easier to read. A brief theoretical background of the score ratio problem is shown in Section 3, then followed by our score ratio proposal (Section 4) together with proposals for reducing the computing load (Section 5). After describing the experimental setup in Section 6, the results achieved both with and without score ratio in all of the tested scenarios can be seen in Section 7. The score ratio computing load analysis and the performance of the proposal to reduce it are shown in Section 8. Conclusions and future lines of research are shown in Section 9.

2. Notation

For a better understanding of the proposal and of the various ways the score ratio may be implemented, it is important to establish the notation and terminology used. The aim of this section is to avoid becoming "lost in notation".

- ***C*** is used to refer to the user or client (Target Class) in general.
- ***C_i*** is a specific client (specific TC).
- *Cohort Set, ChS*. Set of representatives of the impostor class (Non TC), i.e., it is a set of signatories different from the client used to obtain $s(X/\lambda_{\overline{C}})$.

- $\boldsymbol{Ch_i}$. Each of the components of the cohort set: $ChS = \{Ch_1, Ch_2, \ldots, Ch_H\}$, $\boldsymbol{H}$ being the size of this set.
- $\boldsymbol{N}$: number of elements from the ChS ($N \leq H$) used to estimate $s(X/\lambda_{\overline{C}})$.
- $\boldsymbol{M}$: number of elements ($M < < H$) a priori selected from the cohort set to estimate $s(X/\lambda_{\overline{C}})$ in order to decrease the computing load of the score ratio application.
- *Normalization Set, NoS*. Set of signatories used to normalize the score (see Section 6.1).
- $\boldsymbol{Cn_i}$. Each of the components of the normalization set: $NoS = \{Cn_1, Cn_2, \ldots, Ch_R\}$, $\boldsymbol{R}$ being the size of this set.
- $\boldsymbol{S}$. The client, the cohort set elements and the normalization set elements are all of the same type, signatories in our case. All are extracted from the same corpus, and the difference is the role or task to be used. We then need a general identifier: S (signatory).
- $\boldsymbol{\lambda_S}$ is the S model, i.e., the model used to represent a signatory in general. $\boldsymbol{\lambda_C}$, $\boldsymbol{\lambda_{C_i}}$, $\boldsymbol{\lambda_{Ch_i}}$ and $\boldsymbol{\lambda_{Cn_i}}$ will be the client model in general, an i client specific model, an i cohort set component model and an i normalization set component model, respectively.
- $\boldsymbol{X}$ is used to refer to the generic input sample. This input sample is, in our case, the representation of a signature, i.e., the feature vector, $X = \{x_1, x_2, \ldots, x_Q\}$, extracted from the signature (Fig. 1), where, x_k is the k component of the feature vector (which will be a vector at the same time) and Q the number of them. $\boldsymbol{X_C^j}$ will be used to identify the j sample (signature feature vector) of the client. In the same way, $\boldsymbol{X_{Ch_i}^j}$ and $\boldsymbol{X_{Cn_i}^j}$ are the j samples of the cohort set and normalization set i element, respectively.
- Xa, Ya, Pr, Az and Al. A Wacom tablet is used to digitize the signature, obtained from each sampling instant (signature point): position in X-axis (Xa), position in Y-axis (Ya), pressure (Pr), azimuth (Az) angle and altitude (Al) angle.
- $\boldsymbol{d(A, B)}$. It is the distance between A and B.

3. Theoretical background

The concept of likelihood rate is strongly related with the statistic test. The problem to resolve could be defined as follows. Suppose that X can have one of two possible distributions ($p(X/H_0)$ and $p(X/H_1)$), that define two Hypotheses:

H_0: *X has probability density function* $p(X/H_0)$

H_1: *X has probability density function* $p(X/H_1)$

From a statistical point of view, the test can be constructed based on a simple idea: if we observe $X = X_{obs}$, with X_{obs} being an observation of X, $p(X_{obs}/H_1) > p(X_{obs}/H_0)$ is evidence in favor of alternative H_1. Otherwise, the evidence supports the alternative, H_0.

From the previous distributions, the likelihood ratio function $L(X)$ is defined as shown in Eq. (5), the statistical test now being constructed as: small values of $L(X_{obs})$ are evidence in favor of H_1 and, conversely, high values of $L(X_{obs})$ are evidence in favor of H_0. Thus, it seems reasonable to use $L(X)$, in short L, to decide which Hypothesis will be selected, determining a threshold value θ, such that H_0 is rejected if and only if $L < \theta$. The value θ is calculated with the significance level $\alpha = p(L < \theta/H_0)$, that is, setting the false negatives probability value.

$$L(X) = \frac{p(X/H_0)}{p(X/H_1)} \tag{5}$$

With the θ value determined, $p(L > \theta/H_1)$ can be calculated, that is the probability that H_0 will not be rejected if it is false, or the false positives probability. Neyman–Pearson [10,11] showed that, using the likelihood ratio, the above test is more powerful, meaning that it minimizes the probability of false positives.

On numerous occasions, we can have a set of distributions, $\{p_k(X/H_0)\}$ and $\{p_i(X/H_1)\}$, for Hypotheses H_0 and/or H_1, instead of a single one as until now. As will shortly be shown, this occurs in our problem. Neyman–Pearson generalized the likelihood ratio function as shown in Eq. (6).

$$L(X) = \frac{\max_k(p_k(X/H_0))}{\max_i(p_i(X/H_1))} \tag{6}$$

When the goal is to classify an input sample X as belonging or not to the Target Class or client C, relating specifically to biometry, the previous Hypotheses become the following:

H_0: *X is from client C* $\quad$ H_1: *X is not from client C*

In real systems, $p(X/H_0)$ is achieved by means of a statistical model, λ_C, (e.g., HMM or GMM) of the client, $p(X/\lambda_C)$, and $p(X/H_1)$ is obtained by modeling the impostor (Non Target) class, $p(X/\lambda_{\overline{C}})$. The problem is the calculation or estimation of this latter probability function.

The literature offers two ways of obtaining this likelihood: using a cohort set (representative set) of the NTC [3,5,12], or using a single model to explain NTC behavior [4,12].

When a single model is used to estimate $p(X/\lambda_{\overline{C}})$, the model is trained using samples provided by many users other than the client. An example can be found in [4] applied to biometric signature recognition. A User Adapted Universal Background Model (UA-UBM) based on a discrete HMM is proposed. First, a UBM trained using signatures from many signatories is obtained. The client model is then achieved by adapting the UBM with enrollment (training) client signatures. The score ratio is performed by means of the log likelihood: $\log \frac{p(X/\lambda_C)}{p(X/UBM)}$.

When the cohort set alternative is used, $p(X/\lambda_{\overline{C}})$ is modeled by means of a composite hypothesis, that is, a set of probability functions $\{p(X/\lambda_{Ch_i})\}$. Neyman–Pearson approach (Eq. (6)) shows the solution in this case, although, in practice, the N maximum probabilities are generally used [3,5,12], instead of the maximum one only, since better performance is achieved. This is because $\{p(X/\lambda_{Ch_i})\}$ is an estimation or approximation of $\{p_i(X/H_1)\}$. Using the $N > 1$ cohort set elements closest to the client can thus generally estimate $\lambda\bar{c}$, better than by using only the closest one, $N = 1$. The likelihood ratio function is then as shown in Eq. (7) [5,13], with $1 \leq N < H$ and $\{p(X/\lambda_{Ch_k})\} = N\max_i(p(X/\lambda_{Ch_i}))$ $1 \leq i \leq H$.

$$L(X) = \frac{p(X/\lambda_C)}{p(X/\lambda_{\overline{C}})} = \frac{p(X/\lambda_C)}{\frac{\sum_{k=1}^{N} p(X/\lambda_{Ch_k})}{N}} \tag{7}$$

4. Score ratio proposal: basic approach

With a distance based system, the problem when applying the score ratio (Fig. 3) is to estimate $s(X/\lambda_{\overline{C}})$ in Eq. (4). Based on the proposals for estimating $p(X/\lambda_{\overline{C}})$ seen in the previous section, with distance-based classifiers, it is not always possible to obtain a single impostor model, such that, the most general cohort set approximation is adopted here.

Thus, following Eq. (7), our proposal for score ratio is shown in Eq. (8). The same approach of using the N cohort set elements closest to the client is followed here, but, with distances, given an input sample X, the closest cohort set signatories to C will be those with the minimal scores ($\{s(X/\lambda_{Ch_k})\} = N\min_i(s(X/\lambda_{Ch_i}))$ $1 \leq i \leq H$), since low scores imply high similarity. These are selected to perform the score ratio (Eq. (8)).

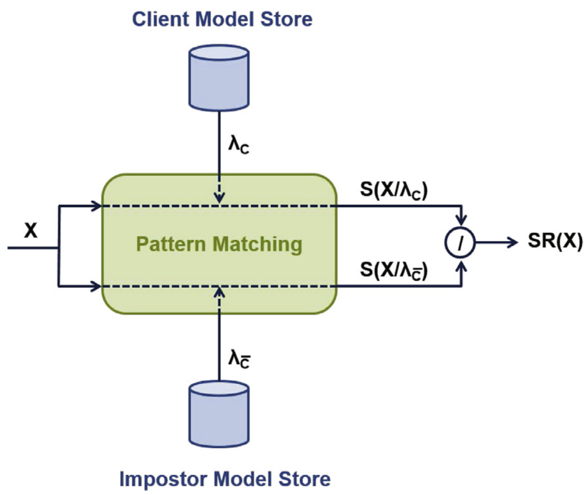


Fig. 3. Biometric system decision stage with score ratio.

This solution will be called *Score Ratio Basic Approach, SRBA*.

$$SRBA(X) = \frac{s(X/\lambda_C)}{s(X/\lambda_{\bar{C}})} = \frac{s(X/\lambda_C)}{\frac{\sum_{k=1}^{N} s(X/\lambda_{Ch_k})}{N}} \tag{8}$$

We performed the score ratio test by crossing two parameters: the cohort set size H (50 and 100) and different values of N (1, 3, 5 and 10). These values were the same as used to test the proposals in the following section.

5. Score ratio proposal: reduced calculations

Applying the score ratio seen in the previous section implies an increase in H (the size of the ChS) in the number of score calculations compared to using only $s(X/\lambda_C)$ to classify X. Moreover, ordering these scores to select the N lowest ones must be considered, although, the computing load of this latter operation is actually negligible due to the size of the ChS.

A study of the time cost of the extra calculations involved was performed (see Section 8.1) to determine whether this prevents real time system response. This is system dependent, although speeding up the system response is an interesting question in all cases. As a result, three proposals to reduce the computing load of $s(X/\lambda_{\bar{C}})$ estimation are put forward: RSR1, RSR2 and RSR3.

5.1. Basic approach modification (RSR1)

The first proposal involves replacing the model, λ_{Ch_i}, with a single sample (signature), $X^b_{Ch_i}$, in the $s(X/\lambda_{\bar{C}})$ estimation. Signature $X^b_{Ch_i}$ is randomly chosen from those used to build the model. Under this approach, Eq. (8) becomes Eq. (9).

$$RSR1(X) = \frac{s(X/\lambda_C)}{s(X/\lambda_{\bar{C}})} = \frac{s(X/\lambda_C)}{\frac{\sum_{k=1}^{N} s(X/X^b_{Ch_k})}{N}} \tag{9}$$

As in SRBA, given an X sample, the N lowest scores from $\{s(X/X^b_{Ch_i})\}$ $1 \le i \le H$ are selected to estimate $s(X/\lambda_{\bar{C}})$, i.e., $\{s(X/X_{Ch_k})\} = N\min_i(s(X/X^b_{Ch_i}))$ $1 \le i \le H$.

This approach will be called *Reduced Score Ratio 1, RSR1*.

5.2. A priori cohort selection (RSR2 and RSR3)

In the previous score ratio approaches, given a test sample X, all of the scores for each cohort set element must be calculated, using either model $(s(X/\lambda_{Ch_i})$ $1 \le i \le H)$ for SRBA or using a single sample $(s(X/X^b_{Ch_i})$ $1 \le i \le H)$ for RSR1. Once these scores have been calculated, the N lower ones are selected to estimate $s(X/\lambda_{\bar{C}})$.

Here, a different approach is proposed by means of a priori cohort set elements subset selection. The idea is to select a priori the $M \ll H$ cohort set elements closest to the client, and then, to perform the score ratio with this subset.

From an S element, its first five signatures (samples) in the corpus $\{X^1_S, X^2_S, \ldots, X^5_S\}$ are used to build the signatory model λ_S. If the cohort set element model is used to perform the score ratio (as in SRBA), then the distance between client C and the cohort set element Ch_i is defined as shown in Eq. (10). Using that distance, a subset of M ChS elements closest to the client is selected $(\{Ch_v\} = M\min_i(d(Ch_i, C))$ $1 \le i \le H)$. Score ratio is then performed by means of Eq. (8), but, now, using only the preselected cohort set elements subset $\{Ch_v\}$ $1 \le v \le M$. If $M = N$, it is not necessary to sort the $\{s(X/\lambda_{Ch_v})\}$ $1 \le v \le M$ scores in order to select the N lowest ones for each X. This proposal will be called *Reduced Score Ratio 2, RSR2*.

$$d(Ch_i, C) = \min_j(s(X^j_{Ch_i}/\lambda_C)) \quad with \quad j = 1, \ldots, 5 \tag{10}$$

If a single signature $X^b_{Ch_i}$ from each element of the cohort set is used to perform the score ratio (as in RSR1), the distance between client C and each cohort set element Ch_i is defined as shown in Eq. (11). Score ratio is performed as shown in Eq. (9), but, as in the previous approach, using only the preselected cohort set elements subset $(\{Ch_v\}$ $1 \le v \le M)$, i.e., those cohort set elements closest to the client, now using, the distance in Eq. (11). Likewise, if $M = N$, it is not necessary to sort the $\{s(X/X^b_{Ch_v})\}$ $1 \le v \le M$ scores in order to select the N lowest ones. This proposal will be called *Reduced Score Ratio 3, RSR3*.

$$d(Ch_i, C) = s(X^b_{Ch_i}/\lambda_C) \tag{11}$$

6. Experimental setup

6.1. Score normalization

Given a learning paradigm, its Match (client scores) and Non-Match (impostor scores) distributions vary from the classifier trained to distinguish one user from another (see Fig. 4a). For this reason, score normalization is essential to transform the scores of the client matchers into a common domain (Fig. 4b).

The main score normalization techniques for signature recognition [14] have been tested here:

Impostor-centric techniques:

IC1: $s_{norm} = s - \hat{\mu}^N_C$
IC2: $s_{norm} = s - (\hat{\mu}^N_C + \hat{\sigma}^N_C)$
IC3: $s_{norm} = (s - \hat{\mu}^N_C)/\hat{\sigma}^N_C$

where $\hat{\mu}^N_C$ and $\hat{\sigma}^N_C$ are the mean and the standard deviation of the Non-Match distribution for the client C classifier, estimated by means of the *Normalization Set*, $\{Cn_i\}$ (Section 6.3), as will be shown in Sections 6.1.1 and 6.1.2.

Target-centric techniques:

TC1: $s_{norm} = s - \hat{\mu}^M_C$
TC2: $s_{norm} = s - (\hat{\mu}^M_C + \hat{\sigma}^M_C)$
TC3: $s_{norm} = (s - \hat{\mu}^M_C)/\hat{\sigma}^M_C$

where $\hat{\mu}^M_C$ and $\hat{\sigma}^M_C$ are the mean and the standard deviation of the Match distribution for the client C classifier, estimated by means of the signatures used to build λ_C using the leave-one-out technique, as will be shown in Sections 6.1.1 and 6.1.2.

Target-impostor technique:

TI1: $s_{norm} = s - s_{EER_C}$

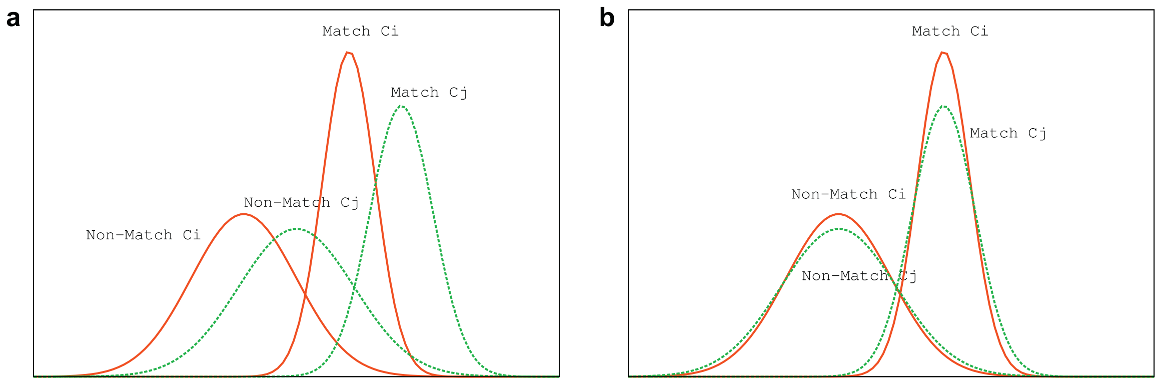


Fig. 4. Example of Match and Non-Match distributions of two users, (a) without score normalization and (b) with score normalization.

where s_{EER_C} is the a priori decision threshold of client C at the Equal Error Rate point (see Section 6.5), obtained from the Non-Match and Match distributions estimation.

6.1.1. Statistics estimation without score ratio

When score ratio is not used, $s = s(X/\lambda_C)$ in the previous equations.

As pointed out, the first five signatures of the signatory in the corpus (Section 6.2), $\{X_C^1, X_C^2, \ldots, X_C^5\}$, are used to build the client model, λ_C. As posited, Match distribution estimation is performed by means of this set. Using the leave-one-out technique, the following Match (client) Scores Set, MSS, is estimated: $MSS = \{s(X_C^g/\lambda_C^g)\}$, with $g = 1, \ldots, 5$ and the λ_C^g model built using the subset $\{X_C^j\}\ 1 \le j \le 5\ and\ j \ne g$. Then, $\hat{\mu}_C^M$ and $\hat{\sigma}_C^M$ are the mean and standard deviation of the MSS.

To estimate the Non-Match distribution, a specific set of signatories called the *Normalization Set* is used. From each element Cn_i of this set, a sample e is randomly selected, forming the so-called *Normalization Gallery*, $NoG = \{X_{Cn_i}^e\}\ 1 \le i \le R$, with R being the size of the *Normalization Set*. Given a client C, the scores for each element in the *NoG* set are calculated, obtaining the Non-Match (impostor) Score Set, $NMSS = \{s(X_{Cn_i}^e/\lambda_C)\}\ 1 \le i \le R$. This set is an a priori impostor score distribution estimation for client C. It is then used to estimate $\hat{\mu}_C^N$ and $\hat{\sigma}_C^N$, these being the mean and standard deviation of NMSS.

6.1.2. Statistics estimation with score ratio

Using the score ratio, $s = \frac{s(X/\lambda_C)}{s(X/\lambda_{\overline{C}})}$ in the score normalization equations of Section 6.1. The score ratio must then also be applied to the sets used to estimate the corresponding means and standard deviations.

More specifically, MSS here becomes the MSS_Score Ratio, $MSS\_SR = \{\frac{s(X_C^g/\lambda_C^g)}{s(X_C^g/\lambda_{\overline{C}})}\}\ 1 \le g \le 5$, with g and λ_C^g defined as in the previous section, and NMSS becomes the NMSS_Score Ratio, $NMSS\_SR = \{\frac{s(X_{Cn_i}^e/\lambda_C)}{s(X_{Cn_i}^e/\lambda_{\overline{C}})}\}\ 1 \le i \le R$. From MSS_SR, $\hat{\mu}_C^M$ and $\hat{\sigma}_C^M$ are estimated, while $\hat{\mu}_C^N$ and $\hat{\sigma}_C^N$ are estimated by means of NMSS_SR.

6.2. Corpus

The MCYT database [15] has been used. This database is one of the most popular and largest in signature verification, and can be considered a benchmark. Signatures were acquired with a WACOM graphic tablet. The sampling frequency of the acquired signals was set to 100 Hz, obtaining from each sampling instant: position in x-axis (Xa) and y-axis (Ya), pressure (Pr), azimuth (Az) and altitude (Al) angles. Samples from 333 different people were acquired. Each user produced 25 genuine signatures, and 25 skilled forgeries were also captured for each user. These skilled forgeries were produced by the five subsequent users by observing the static images of the signature to imitate, attempting to copy them (at least 10 times), and then producing the valid acquired forgeries fluidly (i.e., each individual acting as a forger is requested to sign naturally, without artifacts, such as breaks or slowdowns). In this way, shape-based natural dynamics of highly skilled forgeries are obtained. A study of the statistical significance of the results achieved with this database can be found in [16].

6.3. Experimental sets

The corpus was split into three different subsets, as in [6], but in a different and more realistic way, since the three subsets are completely independent here. In [6], the elements from the cohort set not selected to perform the score ratio (i.e., those different from the N with the lower scores) were used for score normalization. This was done so as to obtain a large and significant test set (183 elements) and a large ($H = 150$ elements) cohort set. From the results in [6], testing a cohort set size of 150 elements is not necessary, since the results do not improve those achieved with $H = 50$ and $H = 100$. As a result, this size is not tested here. Consequently, the corpus was split into:

- **Normalization Set, NoS.** 50 signatories were randomly selected from the corpus to create this set.
- **Cohort Set, ChS.** From the signatories in the corpus not used for the NoS, 100 were randomly selected to create this set. Two different sizes were tested in order to gauge how this parameter impacted on score ratio performance: (i) $H = 50$ signatories (ChS-50), (ii) $H = 100$ signatories (ChS-100).
- **Test set (TS).** This was used to test and consists of 183 signatories not used in the previous sets. The same TS was used in all of the tests performed for objective comparison purposes.
  The following tests were performed for each user of this set:
  - **Genuine test.** While the first five signatures were used to build the signatory (client C) model, the twenty remaining ones were used for genuine tests. That is, 3660 (20x183) genuine tests.
  - **Skilled forgery tests.** The 25 skilled forgeries captured for each user were used. That is, 4575 (25x183) skilled forgery tests.
  - **Random forgery tests.** One signature out of 100 randomly selected users in the TS (different from the client) were used for random forgery tests, that is, 18,300 (100x183) tests in total.

6.4. Signature verification systems

Before describing the systems tested, we define the Δ operator, since it is used in the feature extraction stage of two systems. Given a sequence $Z = \{z_1, z_2, \ldots, z_n\}$, Δz_k is defined as $\Delta z_k = z_{k+1} - z_k$. In our case, z_k will be a feature extracted from a signature point. The signature is a time series, and, in this kind of signals, dynamic information at time k can be extracted by means of the derivative at this point: $\frac{z_{k+1}-z_k}{t_k}$, with t_k being the time interval between z_{k+1} and z_k. Here, t_k is fixed for all of the points (10 ms.) and can be eliminated, converting the derivative into the Δ. Then, Δ is used to introduce dynamic information into the feature vector.

The following three state-of-the-art systems based on distances were tested:

Vector Quantization-based system (VQSys). The system shown in [8] is used. This system achieves very good performance with a reduced computational requirement, which is lower than DTW. In addition, the system improves database storage requirements due to vector compression. The feature vector $X = \{x_1, x_2, \ldots, x_Q\}$ (Section 2) comprises the sequence of features extracted from each signature point: $x_k = \{Xa_k, Ya_k, \Delta Xa_k, \Delta Ya_k, \Delta Pr_k, T_k\}$ $with$ $1 \le k \le Q$, where T_k is the point timestamp. From each of the model signatures, $\{X_S^j\}$ $1 \le j \le 5$, and using the Linde–Buzo–Gray (LBG) algorithm, 256 centroids (a codebook size of 256) are extracted. Thus, $\lambda_S^j = \{256\ codebook\}$ $\forall j$ and $\lambda_S = \{\lambda_S^j\}$ $1 \le j \le 5$. Then, $s(X/\lambda_S) = d(X, \{\lambda_S^1, \lambda_S^2, \lambda_S^3, \lambda_S^4, \lambda_S^5\})$. The Euclidean distance is used to calculate the nearest centroid of λ_S to each x_i component of X and to calculate the distortion (distance) between this component and the corresponding nearest centroid.

Dynamic Time Warping-based system (DTWSys). Our state-of-the-art system based on DTW is used here [9]. This was among the best in the latest signature recognition evaluation performed (ESRA'2011). It is an evolution of the system used to participate in the BSEC'2009 signature recognition evaluation, where it was the second best system [1]. Simple but highly effective feature extraction is accomplished, extracting from each signature point the following two deltas: $x_k = \{\Delta Xa_k, \Delta Ya_k\}$, then $X = \{\{\Delta Xa_1, \Delta Ya_1\}, \{\Delta Xa_2, \Delta Ya_2\}, \ldots, \{\Delta Xa_{Q-1}, \Delta Ya_{Q-1}\}\}$. In this system, the feature vector of each of the five signatures used to build the model is directly the S signatory model: $\lambda_S = \{X_S^1, X_S^2, \ldots, X_S^5\}$. In order to obtain $s(X/\lambda_S)$, the distances, using DTW between X and each of the model signatures is calculated, the minimum one being the final score: $s(X/\lambda_S) = min_i(d(X, X_S^i))$ $1 \le i \le 5$.

Fractional distances-based system (FraDisSys). Our low-cost proposal shown in [7] is used. This system has fewer computational and storage requirements than the previous ones. The signature points number is normalized to a fixed value (15 points here). Signatures can thus be matched by means of a simple distance calculation. Due to their better performance, fractional distances are used (Eq. (12)). Compared to that used in [7], improved feature extraction was achieved here, since new features have been added. In the original proposal, from each point in the normalized signature the following features were used: $x_k = \{Xa_k, Ya_k, Pr_k, Az_k, Al_k\}$. The new added ones for each point are: point number (Pn), signature section length in the x coordinate (SlXa) and in the y coordinate (SlYa), as well as direction changes per section in both coordinates (DcXa and DcYa). The signature duration (Sd) global feature is also used. The 151-dimensional feature vector with this system is thus: $X = \{Xa_1, Ya_1, Pr_1, Az_1, Al_1, Pn_1, SlXa_1, SlYa_1, DcXa_1, DcYa_1, \ldots, Xa_{15}, Ya_{15}, Pr_{15}, Az_{15}, Al_{15}, Pn_{15}, SlXa_{15}, SlYa_{15}, DcXa_{15}, DcYa_{15}, Sd\}$. The S signatory model is: $\lambda_S = \{X_S^1, X_S^2, \ldots, X_S^5\}$. The distance between two signatures can be seen in Eq. (12), with $p = 0.2$. To calculate the final score, the minimum function is used, as with DTW:

Table 1
Choosing the best score normalization technique for VQSys. Best results are bold face emphasized. System error is measured by means of the EER (%).

	No score ratio		Score ratio	
	ChS-100		ChS-100, $N = 5$	
Technique	Random	Skilled	Random	Skilled
No norm.	**1.15**	**6.56**	**0.71**	**4.56**
IC1	0.87	9.81	0.85	10.90
IC2	0.96	11.77	1.28	12.62
IC3	1.03	15.60	0.64	8.70
TC1	3.37	12.16	2.27	9.26
TC2	3.36	12.24	2.43	9.68
TC3	18.36	26.56	19.86	26.60
TI1	1.04	7.08	0.82	5.60

$$s(X/\lambda_S) = min_i(d(X, X_S^i))\ 1 \le i \le 5.$$

$$dist(Y, Z) = \left(\sum_{i=1}^{151} |Y_i - Z_i|^p \right)^{\frac{1}{p}} \tag{12}$$

6.5. Performance measure

Verification systems can be evaluated using the False Match Rate (FMR, situations where an impostor is accepted) and the False Non-Match Rate (FNMR, situations where a user is incorrectly rejected), also known in detection theory as False Alarm and Miss, respectively. A trade-off between both errors usually has to be established by adjusting a decision threshold. The performance can be plotted on an ROC (Receiver Operator Characteristic) or on a DET (Detection error trade-off) plot [17].

However, if the number of comparisons is high, using a single number measure is more useful and easier to understand. The most widely used one in the literature is The Equal Error Rate (EER), that is the error of the system when the decision threshold is such that the FMR, P_{fa}, equals the FNMR, P_{miss} (in the DET curve, the point where the diagonal cuts the curve). This is the measure used here.

7. Results: Score Ratio Basic Approach

Here, the results with SRBA (Section 4) are shown. Score normalization performance is system dependent. So, firstly, we choose the best score normalization technique with and without the score ratio for each system. A cohort set size of 100 users was chosen for these tests, using $N = 5$ for the score ratio. A criterion based on the lowest average between random forgery EER and skilled forgery EER is used to choose the best score normalization technique in each case.

7.1. Vector Quantization-based system (VQSys)

When VQSys is used, the best system configuration is, exceptionally, without using score normalization for both with and without score ratio (Table 1). Once the best score normalization technique is fixed (here, none), the comparative study for this system with and without the score ratio can be seen in Table 2. The smallest error is achieved using the score ratio for both random and skilled forgeries, with a cohort set of 100 signatories and $N = 3$, and a cohort set of 50 signatories and $N = 10$, respectively.

7.2. Dynamic Time Warping-based system (DTWSys)

When DTWSys is used, the best score normalization techniques are TI1 without the score ratio and TC1 when the score ratio is

Table 2
With and without ($N = 0$ row) score ratio performance comparison for VQSys. The results for all of the cohort set sizes and N values are shown. System error is measured by means of the EER (%). Best results are bold face emphasized.

	ChS-50		ChS-100	
N	Random	Skilled	Random	Skilled
0	1.19	6.55	1.15	6.56
1	0.66	5.03	0.63	4.72
3	0.63	4.51	**0.57**	4.50
5	0.63	4.67	0.71	4.56
10	0.69	**4.45**	0.71	4.48

Table 3
Choosing the best score normalization technique for DTWSys. Best results are bold face emphasized. System error is measured by means of the EER (%).

	No score ratio		Score ratio	
	ChS-100		ChS-100, $N = 5$	
Technique	Random	Skilled	Random	Skilled
No norm.	14.45	20.55	0.79	9.01
IC1	1.75	7.65	21.83	39.74
IC2	2.49	9.66	33.00	44.23
IC3	2.34	10.18	1.09	20.54
TC1	2.54	5.87	**0.77**	**6.28**
TC2	2.67	6.04	1.48	8.26
TC3	7.65	9.70	8.52	15.54
TI1	**1.68**	**4.66**	1.26	10.73

Table 4
With and without ($N = 0$ row) score ratio performance comparison for DTWSys. The results for all of the cohort set sizes and N values are shown. System error is measured by means of the EER (%). Best results are bold face emphasized.

	ChS-50		ChS-100	
N	Random	Skilled	Random	Skilled
0	1.85	**4.64**	1.68	4.66
1	1.23	7.62	1.12	7.34
3	0.77	6.82	0.82	6.62
5	0.71	6.66	0.77	6.28
10	**0.66**	6.44	0.68	6.34

Table 5
Choosing the best score normalization technique for FraDisSys. Best results are bold face emphasized. System error is measured by means of the EER (%).

	No score ratio		Score ratio	
	ChS-100		ChS-100, $N = 5$	
Technique	Random	Skilled	Random	Skilled
No norm.	2.73	8.63	1.39	6.83
IC1	2.56	9.80	6.20	20.73
IC2	3.50	12.00	10.73	26.58
IC3	2.38	9.79	2.62	13.79
TC1	**2.00**	**6.08**	**1.33**	**6.33**
TC2	2.68	7.34	2.18	7.67
TC3	10.19	14.29	9.04	14.69
TI1	2.19	6.88	2.23	8.61

Table 6
With and without ($N = 0$ row) score ratio performance comparison for FraDisSys. The results for all of the cohort set sizes and N values are shown. System error is measured by means of the EER (%). Best results are bold face emphasized.

	ChS-50		ChS-100	
N	Random	Skilled	Random	Skilled
0	1.99	6.08	2.00	6.08
1	1.85	7.21	1.72	7.24
3	1.47	6.33	1.40	6.49
5	1.40	6.12	1.33	6.33
10	1.34	**5.90**	**1.31**	6.03

applied (Table 3). Once the best score normalization technique is fixed, the comparative study for this system with and without the score ratio is shown in Table 4. The smallest error for random forgery is achieved using the score ratio with a cohort set of 50 signatories and $N = 10$. However, for skilled forgery, the best error value has been achieved without the score ratio and with a cohort set of 50 signatories.

7.3. Fractional distance-based system (FraDisSys)

When FraDisSys is used, the best score normalization technique is TC1 for both with and without score ratio (Table 5). Once the best score normalization technique is fixed, the comparative study for this system with and without the score ratio can be seen in Table 6. The smallest error is achieved using the score ratio for both random and skilled forgeries, with a cohort set of 100 signatories and $N = 10$ and a cohort set of 50 signatories and $N = 10$, respectively.

7.4. Results analysis

From the previous results, the first important consideration is that the use of the score ratio has improved all of the cases studied except one, skilled forgeries with DTW. For random forgeries, great improvements have been achieved with all of the classifiers:[2] 50% for VQSys (ChS-100, $N = 3$), 61% for DTWSys (ChS-50, $N = 10$) and 34% for FraDisSys (ChS-100, $N = 10$), compared to the best results without the score ratio, respectively. For skilled forgeries, the following improvements have been achieved: 32% for VQSys (ChS-50, $N = 10$) and 3% for FraDisSys (ChS-50, $N = 10$) compared to the best results without the score ratio, respectively. It is interesting to note that the skilled forgery tests are used only in signature recognition, while, for the rest of biometrics, impostor tests are performed by means of random samples (i.e., samples of other users), which is where the biggest improvements have been achieved with the use of the score ratio.

From the systems without score ratio, the one based on DTW achieved the best results for skilled forgeries, 4.64% (this can also be seen in international competitions), while for random ones, the best results are achieved with VQ, 1.15%. This is typical in signature recognition: improvements in one forgery type usually worsen the other. However, here, the use of the score ratio has allowed a system to be achieved with the best results for both forgeries on average: 0.57% in random forgery and 4.5% in skilled forgery for VQSys, with ChS-100 and $N = 3$.

The size of the cohort set does not seem to have any great influence on the results with score ratio. In general, the influence of H is classifier and task (random-skilled forgery) dependent and differences in results are very small. With regard to the N value, the worst results were achieved, in general, with 1, although for the rest of the values tested, differences in results are also very small. These results show that the score ratio proposal can be applied with small cohort sets and with small N values, which is very important for real applications.

[2] The percentages have been calculated as follow: $\left(\frac{EER1-EER2}{EER1}\right) \cdot 100$ *with* $EER1 > EER2$. Here, $EER1 = best\ EER\ without\ SR$ and $EER2 = best\ EER\ with\ SR$. The same equation is used in the rest of the paper with the corresponding errors.

Table 7
System response (in seconds) to perform feature extraction and matching operations with SRBA for VQSys, DTWSys and FraDisSys.

	VQSys	DTWSys	FraDisSys
Score ratio	0.989	2.91	0.322

8. Results: reduced calculations

A new element in the study has been indirectly introduced at the end of the previous section: the computing load of the score ratio proposal. As posited, this is important for real systems, and thus merits analysis. A study of this question is shown in the following section, before showing the results with the proposals which aim to reduce the calculation number in the $s(X/\lambda_{\overline{C}})$ estimation (Section 5).

From the results analysis in the previous section, it can be concluded that cohort set size H does not have a significant impact on system performance. As a result, all the tests in this section were only performed with the smallest cohort set, i.e., ChS-50. All the results contained in this section are thus understood to be obtained from using said set, such that, for the purpose of clarity, they are not mentioned henceforth.

8.1. SRBA computing load analysis

From a practical point of view, the time response of the system is an important question, and more specifically, whether the system response is real time or not. In Table 7, the time required to perform a verification operation (to obtain the score s) using SRBA is shown from a technological point of view. This means that tests focus on the classification algorithm, i.e., these times include feature extraction and matching operations, not taking into account biometric sample acquisition and communication (for remote access) times.

Another important question with regard to the results in Table 7 is that the software used (developed with Java) is optimized to perform experiments. In other words, it is developed so as to easily make changes in the experimental setup and not for quick execution of authentication operations. The times in Table 7 are thus easy to improve in a real system.

As can be seen in Table 7, using the score ratio does not prevent the response from being considered real time for VQSys and FraDisSys, since the times in the table are, as posited, a pessimistic estimation. However, with the DTW based system, which has the largest computing load, the time response is not good enough for real systems and must be improved.

Having seen the system response with SRBA, the results of the proposals to reduce computing load are now addressed. Here, the goal is to reduce the times in Table 7, while seeking to maintain score ratio performance.

8.2. Results

For a better comparison, the results with all of the proposals for reducing computing load are shown in the same table (Table 8), together with the results with the *basic proposal* (SRBA). The results are achieved with the same score normalization techniques as the corresponding ones using SRBA. For greater clarity, results with $N = 1$, are not shown, since these are the worst.

Three different values of M (number of a priori selected elements from the Chs) have been tested: 5, 10 and 15, for the RSR2 and RSR3 approaches. It should be remembered that (see Section 5.2) if $M = N$ it is not necessary to sort the cohort set scores to estimate $s(X/\lambda_{\overline{C}})$.

In order to compare the computing load, the number of operations to obtain $s(X/\lambda_{\overline{C}})$ is added in each case (*N Ops* column), with "operation" being a distance computation. For example, for SRBA (Eq. (8)), estimating $s(X/\lambda_{\overline{C}})$ involves calculating 50 times the score $s(X/\lambda_{Ch_i})$, the size of ChS-50. Although the N lowest ones are those used to estimate $s(X/\lambda_{\overline{C}})$, all of the scores from the ChS must be calculated before performing said selection. Since $s(X/\lambda_S)$, generally, implies five distance calculations for all the systems:

- For VQSys: $s(X/\lambda_S) = d(X, \{\lambda_S^i\})\ 1 \le i \le 5$.
- For DTWSys: $s(X/\lambda_S) = min_i(d(X, X_S^i))\ 1 \le i \le 5$.
- For FraDisSys: $s(X/\lambda_S) = min_i(d(X, X_S^i))\ 1 \le i \le 5$.

The final number of operations required to obtain $s(X/\lambda_{\overline{C}})$ for SRBA will thus be 250 (50x5) for all of the systems.

What differs is the time needed to compute the distance in each case. As a result, a time response estimation, following the same assumptions as those in Section 8.1, is shown (*T* column).

From the results in Table 8, it can be said that the aim of decreasing the score ratio computing load has been achieved. In all cases, the number of operations required to obtain $s(X/\lambda_{\overline{C}})$ has been reduced without any loss in system performance. Real time response has been achieved in all cases. What does prove system dependent is the performance of RSR1, RSR2 and RSR3. We now provide a more detailed analysis.

RSR1 worked very well in all cases. The computing load has been reduced by a factor of approximately five, while the system EER has been maintained. In the worst case (VQSys, random forgery), system performance fell by only 8% (from 0.63% with SRBA to 0.68% with RSR1). System performance has even improved for skilled forgery with DTWSys and FraDisSys. Applying RSR1 has achieved a real time response for all the systems, and is thus a good solution in all cases.

VQSys is the only case where RSR2 and RSR3 performed better than RSR1. The best EERs for random forgeries were achieved with RSR2–$M = 15$, and were even better than those with SRBA. An interesting case is RSR3–$M = 15$, since the best EER for random forgeries, 0.62%, is even better than the best one with SRBA, 0.63%, and the EER for skilled forgeries was only 2% worse (from 4.45 to 4.56%), while the operation time fell by 92% (from 0.989 to 0.078 s).

For DTWSys and FraDisSys, RSR2 and RSR3 performed worse than RSR1. For FraDisSys, the quickest system, the best alternative is RSR1, since the best performance was achieved with only 0.017 s per operation. A more detailed analysis is advisable for DTWSys, the slowest system.

For DTWSys, a good rate between performance and computing load was achieved with RSR3 for $M = 10$ and $M = 15$, with the most interesting being the second. With $M = 15$, the best random forgery EER, 0.77%, is 15% worse compared to the best with SRBA, 0.66%, and RSR1, 0.67%, but is still much better than the EER without the score ratio (1.68%, in Table 4). Worsening in the skilled forgery EER is only 1% compared to the best with RSR1 (from 6.33 to 6.4%), and is better than the best EER with SRBA, 6.44%. However, operation time decreases by 93% compared to SRBA (from 2.91 to 0.195 s) and by 66% compared to RSR1 (from 0.557 to 0.195 s).

9. Conclusions and future works

To the best of our knowledge, the score ratio, generally applied in probabilistic-based systems, has not been used with distance-based ones. Here, we show that it can also be an interesting proposal for such systems. The present study was performed adopting both a theoretical and a practical approach.

Based on the likelihood ratio, a basic score ratio approach has been proposed and successfully tested with three different state-

Table 8
EER (in %) for all systems tested with the score ratio basic approach (SRBA) and the proposals RSR1, RSR2 and RSR3 to reduce the number of calculations. The extra operations number to obtain $s(X/\lambda_{\overline{C}})$ is added (*N Ops* column). An estimation to the time per authentication operation in seconds is also included (*T* column) for better comparison.

		Random			Skilled				
VQSys									
		N			*N*				
	M	3 (%)	5 (%)	10 (%)	3 (%)	5 (%)	10 (%)	*N Ops*	*T* (s)
SRBA		0.63	0.63	0.69	4.51	4.67	4.45	250	0.989
RSR1		0.68	0.71	0.71	4.63	4.62	4.63	50	0.205
RSR2	5	0.76	0.85		5.27	5.38		25	0.117
	10	0.66	0.60	0.70	4.70	4.67	4.86	50	0.213
	15	0.57	0.57	0.57	4.66	4.59	4.70	75	0.313
RSR3	5	0.75	0.68		5.00	5.14		5	0.041
	10	0.68	0.71	0.71	4.74	4.68	4.75	10	0.060
	15	0.68	0.62	0.68	4.56	4.59	4.59	15	0.078
DTWSys									
		Random			Skilled				
		N			*N*				
	M	3 (%)	5 (%)	10 (%)	3 (%)	5 (%)	10 (%)	*N Ops*	*T* (s)
SRBA		0.77	0.71	0.66	6.82	6.66	6.44	250	2.910
RSR1		0.77	0.67	0.68	6.91	6.72	6.33	50	0.577
RSR2	5	1.31	1.12		7.34	7.32		25	0.320
	10	1.03	0.90	0.83	7.08	7.01	6.58	50	0.557
	15	1.01	0.81	0.73	6.99	6.90	6.44	75	0.782
RSR3	5	1.22	1.06		7.18	7.21		5	0.108
	10	1.01	0.84	0.84	7.15	6.99	6.55	10	0.154
	15	0.90	0.77	0.77	6.99	6.85	6.40	15	0.195
FraDisSys									
		Random			Skilled				
		N			*N*				
	M	3 (%)	5 (%)	10 (%)	3 (%)	5 (%)	10 (%)	*N Ops*	*T* (s)
SRBA		1.47	1.40	1.34	6.33	6.12	5.90	250	0.309
RSR1		1.47	1.42	1.34	6.23	5.98	5.85	50	0.017
RSR2	5	2.24	2.08		7.38	6.95		25	0.036
	10	2.10	1.93	1.80	7.45	6.90	6.80	50	0.067
	15	1.77	1.69	1.65	6.83	6.49	6.35	75	0.097
RSR3	5	2.19	2.02		6.93	6.84		5	0.012
	10	2.21	2.12	2.02	7.23	6.93	6.58	10	0.012
	15	1.77	1.66	1.59	6.63	6.47	6.08	15	0.013

of-the-art biometric signature systems based on distance classifiers. Except for one case, improvements have been achieved in all of the tested scenarios. The best results are achieved using the score ratio for both random and skilled forgeries, improving the results obtained with the reference systems tested.

Several cohort set sizes and values of *N* (number of users used to obtain the impostor score) have been tested, showing that even with small values of both, good results can be achieved with the score ratio application.

From a practical point of view, studying the extra computing load introduced for use of the score ratio is important. Said study was performed and three different proposals (RSR1, RSR2 and RSR3) for reducing the extra calculations were successfully proposed and tested.

The computing load study, based on the system response time, has shown that use of the score ratio does not prevent real time system response, although this is system dependent. Regardless of this, however, speeding up system response is important for any practical application. In this sense, the proposals put forward for reducing calculations performed well, since they have drastically cut the number of operations required whilst scarcely reducing system performance compared to SRBA, and have even improved it in some cases. Real time response has been achieved with all of the system.

These results encourage us to continue with the proposal put forward in this work, and to test with other biometrics, where, in addition, impostor tests are carried out using only random forgeries, where the greatest improvements have been achieved applying the score ratio.

Acknowledgments

Thanks to Philip Jaggs B.A. Hons for revising the English grammar.

References

[1] N. Houmani, et al., Biosecure signature evaluation campaign (BSEC'2009): evaluating online signature algorithms depending on the quality of signatures, Pattern Recognit. 45 (3) (2012) 993–1003, doi:10.1016/j.patcog.2011.08.008.

[2] C. Vivaracho-Pascual, A. Simon-Hurtado, E. Manso-Martinez, J.M. Pascual-Gaspar, Client threshold prediction in biometric signature recognition by means of multiple linear regression and its use for score normalization, Pattern Recognit. 55 (2016) 1–13. http://dx.doi.org/10.1016/j.patcog.2016.02.007.

[3] T. Matsui, S. Furui, Likelihood normalization for speaker verification using a phoneme and speaker independent model, Speech Commun. 17 (1) (1995) 109–116. http://dx.doi.org/10.1016/0167-6393(95)00011-C.

[4] E. Argones Rua, J. Alba Castro, Online signature verification based on generative models, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 42 (4) (2012) 1231–1242, doi:10.1109/TSMCB.2012.2188508.

[5] L. Nanni, A. Lumini, A supervised method to discriminate between impostors and genuine in biometry, Expert Syst. Appl. 36 (7) (2009) 10401–10407. http://dx.doi.org/10.1016/j.eswa.2009.01.037.

[6] C. Vivaracho-Pascual, A. Simon-Hurtado, E. Manso-Martinez, On the Use of Score Ratio with Distance-Based Classifiers in Biometric Signature Recognition, Neural Information Processing, Lecture Notes in Computer Science 9489 (2015) 318–327.

[7] C. Vivaracho-Pascual, M. Faundez-Zanuy, J.M. Pascual, An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances, Pattern Recognit. 42 (1) (2009) 183–193, doi:10.1016/j.patcog.2008.07.008.

[8] J.M. Pascual-Gaspar, M. Faundez-Zanuy, C. Vivaracho, Fast on-line signature recognition based on VQ with time modeling, Eng. Appl. Artif. Intell. 24 (2) (2011) 368–377. http://dx.doi.org/10.1016/j.engappai.2010.10.015.

[9] N. Houmani, S. Garcia-Salicetti, B. Dorizzi, J. Montalvo, J.C. Canuto, M.V. Andrade, Y. Qiao, X. Wang, T. Scheidat, A. Makrushin, D. Muramatsu, J. Putz-Leszczynska, M. Kudelski, M. Faundez-Zanuy, J.M. Pascual-Gaspar, V. Cardeoso-Payo, C. Vivaracho-Pascual, E.A. Ra, J.L. Alba-Castro, A. Kholmatov, B. Yanikoglu, Biosecure signature evaluation campaign (ESRA'2011): evaluating systems on quality-based categories of skilled forgeries, in: Proceedings of the International Joint Conference on Biometrics (IJCB), 2011, pp. 1–10, doi:10.1109/IJCB.2011.6117487.

[10] E.L. Lehmann, Testing Statistical Hypotheses, second ed., Chapman and Hall, 1986.

[11] T.S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York, 1967.

[12] D.A. Reynolds, Speaker identification and verification using gaussian mixture speaker models, Speech Commun. 17 (1) (1995) 91–108. http://dx.doi.org/10.1016/0167-6393(95)00009-D.

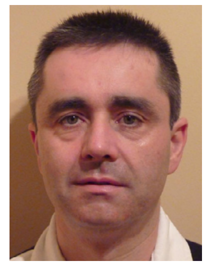
[13] T. Matsui, S. Furui, Similarity normalization method for speaker verification based on a posteriori probability, in: Proceedings of the Workshop on Automatic Speaker Recognition Identification and Verification, Martigny, Switzerland, 1994, pp. 59–62.

[14] C. Vivaracho-Pascual, A. Simon-Hurtado, E. Manso-Martinez, J. Pascual-Gaspar, A new proposal for score normalization in biometric signature recognition based on client threshold prediction, in: Proceedings of the IEEE Twelfth International Conference on Data Mining (ICDM), 2012, pp. 1128–1133, doi:10.1109/ICDM.2012.50.

[15] J. Ortega-Garcia, J. Fierrez, D. Simon, J. Gonzalez, M. Faundez-Zanuy, I. Hernaez, V. Espinosa, A. Satue, J.J.Igarza, C. Vivaracho, D. Escudero, Q. Moro, MCYT baseline corpus: a bimodal biometric database, IEE Proc. Visual Image Signal Process. 150 (6) (2003) 395–401, doi:10.1049/ip-vis:20031078.

[16] M. Faundez-Zanuy, On-line signature recognition based on VQ-DTW, Pattern Recognit. 40 (3) (2006) 981–992, doi:10.1016/j.patcog.2006.06.007.

[17] A. Martin, G. Doddington, T. Kamm, M. Ordowski, M. Przybocki, The DET curve assessment of detection task performance, in: Proceedings of EuroSpeech, 4, 1997, pp. 1895–1898.



Carlos Vivaracho-Pascual obtained a degree in Physics (1989) and the Ph.D.degree (2004) from the University of Valladolid (Spain), where he is currently working as a lecturer. He has published numerous papers in relevant congresses and journals and some book chapters. His research interests are mainly biometric and pattern recognition.

Arancha Simon-Hurtado obtained a degree in Physics (1989) and the Ph.D. degree (2002), both from the University of Valladolid, (Spain), where she is currently working as a lecturer. She has published several papers in international journals and conferences. Her current research interests include, mainly, pattern recognition, biometrics, and data mining.

Esperanza Manso-Martinez is a Lecturer at the Department of Computer Language and Systems at Valladolid University, Spain. She has a major in Mathematics and a Ph.D. from Valladolid University. She is interested in software maintenance, reuse and experimentation. Her works have been accepted in international journals and congresses.