Towards Predicting Good Users for Biometric Recognition Based on Keystroke Dynamics

Aythami Morales(✉), Julian Fierrez, and Javier Ortega-Garcia

Biometric Recognition Group ATVS, EPS, Universidad Autonoma de Madrid,
C/Francisco Tomas y Valiente 11, 28049 Madrid, Spain
{aythami.morales,julian.fierrez,javier.ortega}@uam.es

Abstract. This paper studies ways to detect good users for biometric recognition based on keystroke dynamics. Keystroke dynamics is an active research field for the biometric scientific community. Despite the great efforts made during the last decades, the performance of keystroke dynamics recognition systems is far from the performance achieved by traditional hard biometrics. This is very pronounced for some users, who generate many recognition errors even with the most sophisticate recognition algorithms. On the other hand, previous works have demonstrated that some other users behave particularly well even with the simplest recognition algorithms. Our purpose here is to study ways to distinguish such classes of users using only the genuine enrollment data. The experiments comprise a public database and two popular recognition algorithms. The results show the effectiveness of the Kullback-Leibler divergence as a quality measure to categorize users in comparison with other four statistical measures.

Keywords: Keystroke · Typing patterns · Biometric · Authentication · Quality · Performance prediction

1 Introduction

Keystroke dynamics is a well-known biometric recognition technology which has attracted the interest of industry and researchers during the last decade [1][2]. The proliferation of web applications (e. g. e-banking or e-commerce) and the necessity of accurate and secure recognition methods has increased the interest in biometrics related with the user activity with the computer. Keystroke dynamics plays an important role in this context and its complementarity with other biometric modalities such as mouse dynamics has renovated the interest in these approaches [3]. The identification of people using their typing patterns can be applied to several scenarios including high security password authentication [1], text-independent authentication [4] and continuous authentication [5]. In summary, keystroke dynamics is an active research area with both scientific and industrial possibilities (e. g. DARPA, Active Authentication Program).

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 711–724, 2015.
DOI: 10.1007/978-3-319-16181-5_54

However, the accuracy of keystroke dynamics recognition systems is far from the performance achieved by the most competitive biometric traits and the error rates requested by international biometric standards (e. g. EN-50133-1). In terms of performance, keystroke dynamics can be considered halfway between hard and soft biometrics. As a behavioral biometric, it is highly user-dependent and it is difficult to generalize the performance among all population and scenarios. Previous works demonstrate the large user-variability of the error rates even with the most competitive recognition algorithms [6]. There are users with performances twenty times worst than others and therefore it is difficult to ascertain the overall accuracy. Predicting the performance of the users during the enrollment is a key to improve further recognition steps or the enrollment itself.

The performance of biometric recognition systems is strongly related with the quality of the samples [7]. Quality assessments have been studied for different biometrics traits such as fingerprint [8] or face [9] among others. Despite its well-known utility, the quality of keystroke dynamics has been scarcely studied [10]. The main reasons for this apparent disinterest could be found in the difficulties to establish a quality assessment of a behavioral biometric based only in timing between key events. It is not trivial defining the meaning of quality in keystroke dynamics technologies.

The term quality in the biometric literature have several meanings and applications. It is possible to distinguish between quality of biometric samples, quality of sensors and quality of the users among others. This paper focuses on quality of the users as a measure of their individual performance in terms of recognition accuracy. Low quality users imply users with low performances or high error rates while high quality users will be those users with high performances or low errors.

This paper studies different statistical measures for a reliable prediction of the quality of users for keystroke dynamics authentication. The purpose here is to analyze different ways to distinguish between users with well marked differences in terms of performance. The study assumes a scenario in which only the genuine enrollment data is used for both predicting the quality and enrolling the user. The experiments include a public benchmark dataset and two popular matchers for keystroke dynamics. The results suggest that it is possible to define a quality measure to categorize users correlated with their recognition performance.

The paper is organized as follows. Section 2 introduces the quality framework and proposes the Kullback-Leibler divergence as a feasible measure to establish the quality of the users of keystroke dynamics authentication systems. Section 3 presents the experimental protocol and results while Section 4 summarizes the main conclusions.

2 Quality Assessment for Keystroke Dynamics

Quality of biometric samples has become an important concern for the biometric community [7][11]. It is well known that the degradation of quality strongly affects the performance of biometric recognition systems and dealing with such

degradation is still an open challenge in many biometric traits. The quality of biometric samples is affected by many factors and it is difficult to generalize among all biometric technologies and sensors. The standard ISO/IEC 29794-1 has established normalizations and three main concepts related with the quality on biometric systems:

- **Character:** indicates the distinctiveness capability of the source.
- **Fidelity:** indicates the degree of similarity between the sample and its source.
- **Utility:** indicates the impact of a sample on the overall performance of the biometric system.

The quality measure of a biometric sample can be used for different purposes including: image enhancement [12], improving the matching algorithms [8] or optimized fusion strategies [13][11][14], see Fig. 1. Noteworthy, the quality is not exclusively related to a standalone sample and it is possible to measure the quality of a user or its enrollment set [14]. This quality evaluation of the users can be employed to improve the enrollment, the combination with other systems and the confidence on the authentication. The performance of the biometric recognition system is strongly influenced by the quality of the enrollment data and the evaluation of its utility is crucial in real applications [15].

Concerning keystroke dynamics, among the several factors that affect the quality of the biometric sample it is important to highlight:

- **Behavioral factors:** related with human emotional states, cooperativity or distractions. These factors also comprise the intrinsic characteristics of each user which include users particularly vulnerable to impersonation or users difficult to match, among others. The literature refer to this as biometric zoo [16] or menagerie [17].
- **Sensor factors:** related with the sensor, human-machine interactions, ease of use or maintenance. The proliferation of new portable devices and the necessity of interoperable schemes are important factors which affect the quality of the samples.

While the factors related with the sensor can be mitigated with hardware maintenance, the factors related with human behavior have more unpredictable consequences. How can we detect behavioral factors such as cooperativity or distraction from keystroke dynamic features? The features employed in keystroke dynamics are generally based on timing between key events [1][2] and the quality evaluation of these features arises several problems. In [10] the researchers analyzed six factors to explain different keystroke dynamics error rates (in order of relevance): algorithms, training amount, updating, typist-to-typist variation, feature set and impostor practice.

The quality evaluation of the keystroke dynamics has attracted scarce attention in the literature. The related works are focused mostly on outliers removal [18][19] and features improvement [20][21]. The outliers can be defined as samples with an unusual pattern in comparison with the available data from a

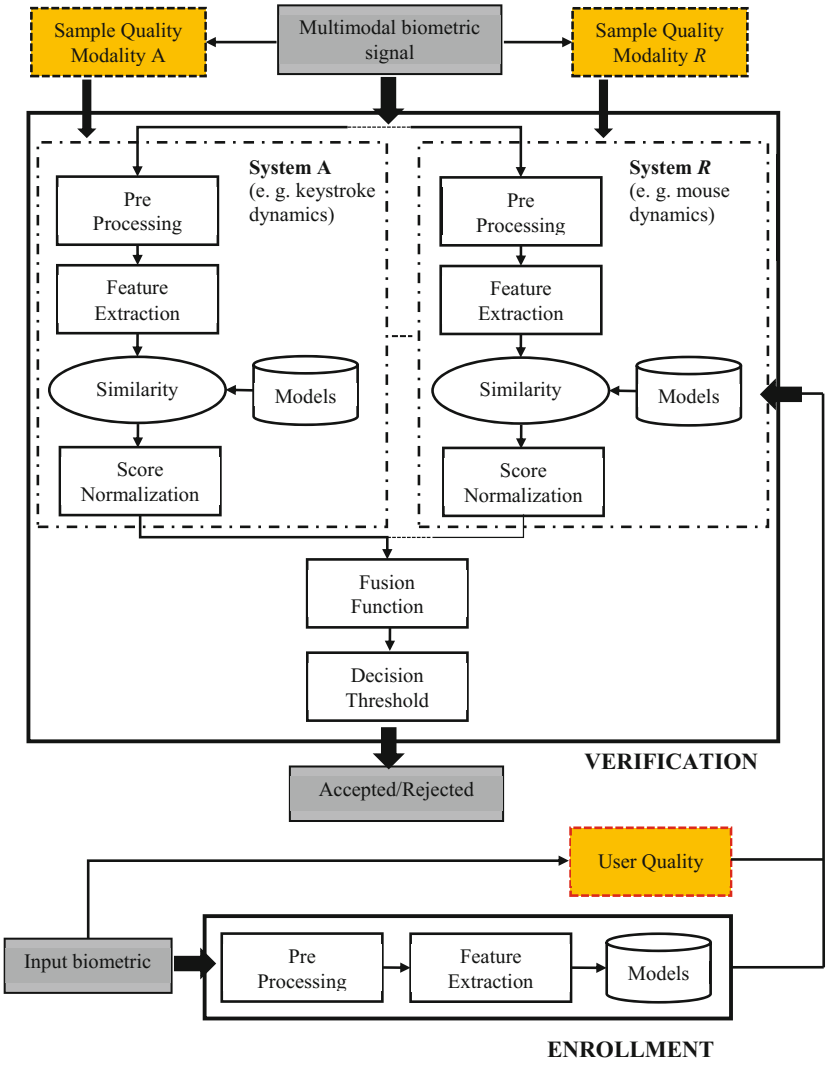

Fig. 1. Block diagram of multimodal biometric recognition/identification systems

specific user. The methodology used to discard these samples is traditionally based on statistical features (mean, variance, standard deviation) related with the distribution of the genuine data [18][19]. The inclusion of artificial rhythms and cues was proposed in [20] to improve the quality of data in terms of distinctive ability. In [21], the researchers established a quality classification in terms of uniqueness, inconsistency, and discriminability of the keystroke patterns. The main disadvantage of this classification is that all three measures need both gen-

uine and impostor data. Depending of the application, the impostor data may not be available (e. g. applications where the password is chosen by the user).

2.1 Measuring the Quality with the Kullback-Leibler Divergence

The entropy is a measure of the uncertainty in a random variable and it is related with the information present in any signal. Some researchers have studied the relationship between the performance of biometric recognition algorithms based on online signature and the entropy of the dynamic signals [22]. The researchers observed that high values of entropy implied higher error rates and low entropy values implied lower error rates. The reason of such behavior was explained with the stability of the genuine samples, which is greater for low entropy samples.

The Kullback-Leibler divergence (also called relative entropy or K-L divergence) is another information measure which have been proposed for biometric quality assessment [9]. The K-L divergence measures the difference between two probability distributions A and B in terms of the information needed to approximate A to B. In this paper we measure the K-L divergence from a feature vector $\mathbf{v}^Q = [v_1^Q, v_2^Q, ..., v_N^Q]$ (Query sample with N features) and the enrollment data mean $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_N]$ (generated with the enrollment data). The K-L divergence $D_{KL}(\mathbf{v}^Q||\boldsymbol{\mu})$ can then be defined as:

$$D_{KL}(\mathbf{v}^Q||\boldsymbol{\mu}) = \sum_{n=1}^{N} v_n^Q \log \frac{v_n^Q}{\mu_n} \tag{1}$$

where $\boldsymbol{\mu}$ is the enrollment data mean of the user obtained as:

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{v}_m^E \tag{2}$$

where each $\mathbf{v}_m^E$ is one out of the M enrollment samples (with N features each). Assuming $N = 31$ features and $M = 200$ enrollment samples, the Fig. 2 shows some examples of mean vectors (from the CMU benchmark dataset detailed in Section 3) as well as the $D_{KL}(\mathbf{v}^Q||\boldsymbol{\mu})$ obtained for each of the 50 query samples $\mathbf{v}^Q$ of the same users.

It can be seen that there are slight differences between the user enrollment data mean (note that the password was unique for all users). However, the K-L divergence between query samples and user background models shows different behavior and it is possible to find users with low stability (Fig. 2-Right black line) or users with very stable K-L divergence values (Fig. 2-Right grey lines). Next sections will analyze the correlation between K-L divergence and the performance of keystroke dynamics systems.

3 Experiments

The experiments are conducted to analyze: i) the quality dependence of keystroke dynamics and; ii) the utility of the K-L divergence for predicting the

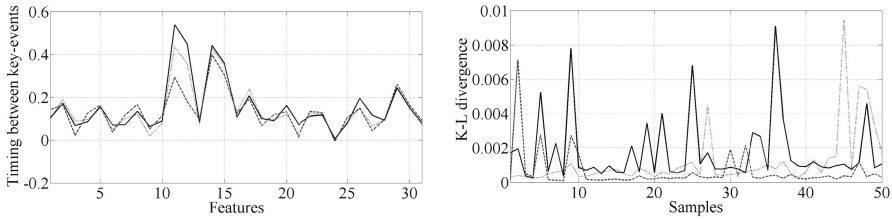

Fig. 2. The mean vectors of 3 different users (left) and the K-L divergence from 50 samples (feature vectors) of the same 3 users (right). $N = 31$ features and $M = 200$ enrollment samples.

performance of individual users. The experiments assume a scenario in which only genuine enrollment data is available (imposter data are not employed to model the users).

3.1 Database: the CMU Benchmark Dataset

The CMU benchmark dataset [19] comprises 51 subjects and 8 sessions with 50 repetitions per session. The time lapse between sessions is more than one day and the 400 typing samples were collected with an accuracy of 200 microseconds. The password was the same for all users and it consists of a ten characters typical strong password which includes uppercase, lowercase and symbols: **.tie5Roanl**. The feature data for each sample includes: hold time for each key (i.e. time between press and release); the keydown-keydown time between two keys (i.e. time between the press of the key 1 and the press of key 2); the keyup-keydown time between two keys (i.e. time between the release of key 1 and the press of key 2); the Enter key is included as a part of the password. The total number of features per samples is 31 (11 hold times, 10 keydown-keydown times and 10 keyup-keydown times).

The most attractive characteristics of the CMU benchmark dataset for this work can be summarized as: i) large number of samples per subject which allows an accurate modeling of the individual behavior; ii) publicly available benchmarks with several feature extraction and classification techniques [19].

3.2 Baseline Systems

The experimental protocol is the same as employed in popular benchmarks [18][19]. The 200 samples from the first 4 sessions are used as gallery/enrollment set. The genuine scores are obtained from the 200 samples corresponding to the last 4 sessions while the impostors are obtained from the first 50 samples of each subject in the database. The performance is evaluated in terms of Equal Error Rate (EER) for each of the 51 subjects in the database.

In this paper we evaluate two popular recognition algorithms for keystroke dynamics [18][19]. Both approaches have achieved the most competitive performances reported for the CMU benchmark dataset among more than 14 different systems. Both systems include training/modeling stages based exclusively on genuine data and other promising systems were discarded because they include impostor data during the training phase [23]. The approaches used in the experimental evaluation made in this paper are:

- **System A** - Modified Manhattan distance with Nearest Neighbor classifier [18]: this system is based on a combination of the Manhattan and the Mahalanobis distances. The method can be summarized as (see [18] for details): i) the feature vectors are normalized according principles inspired by the Mahalanobis distance (using the covariance matrix) and; ii) the normalized feature vectors are matched with the enrollment data using the Manhattan distance and a Nearest Neighbor classifier.
- **System B** - Scaled Manhattan distance [19]: this system is based on the simplicity of the Manhattan distance and its usefulness for decomposing into contributions made by each feature (see [19] for details). The distance is normalized by the average absolute deviation from the enrollment data.

Outlier removal is common in keystroke dynamics and it is a feasible method to improve the enrollment set. An outlier is a sample beyond the typical user variability and its inclusion in the enrollment set to model the user usually have a negative impact in the performance. The K-L divergence can be used to evaluate the stability of the enrollment set. Fig. 3 shows the mean K-L divergence for all the subjects in the CMU database for the different sessions available. The K-L divergence is estimated separately for each subject (there is one $\boldsymbol{\mu}$ for each subject which is calculated using all the samples available from the same user as it is described in section 2.1) and the results are averaged. Note that the users were not habituated to type the password (.tie5Roanl) and they needed a learning period in which they stabilized their typing patterns.

Fig. 3 clearly shows large differences between K-L divergences values from first and last sessions. The samples from the first session can be considered outliers. Table 1 shows how excluding such samples can improve the overall performance of the baseline systems. However, the improvement is slight (around 10% improvement of the average EER) and it is important to note that 150-200 enrollment samples could be considered excessive depending of the application.

3.3 Performance Evaluation

The standard ISO/IEC 29794 includes in the definition of the purpose of quality algorithms the next requirement: *"Quality algorithms shall produce quality scores that predict performance metrics such as either false match or false non-match".*

The quality is related with several factors including the acquisition, the features and the personal characteristics of the subjects among others. The performance of keystroke dynamics is highly user-dependent and it means that there

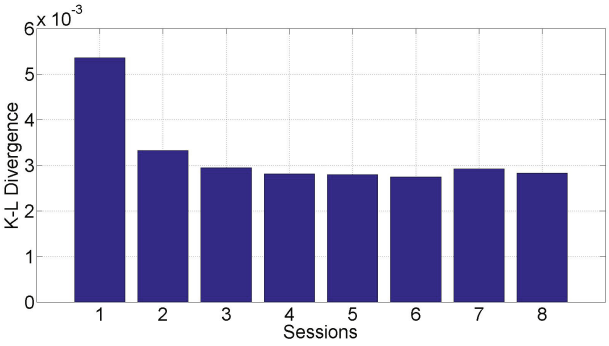

Fig. 3. Mean Kullback-Leibler divergence for each session on the CMU benchmark dataset

Table 1. Performance (EER in %) for different enrollment sets employed

	Enrollment Set	
	Sessions 1 to 4	**Sessions 2 to 4**
System A	8.89	8.20
System B	9.60	8.55

are users who exhibit an EER below 1% and others with EER greater than 20%, see Fig. 4.

Fig. 4 shows that different users present large variations in terms of performance. The reasons of such different performances vary with users who are easy to be recognized (Fig. 4c and Fig. 4d) and others are difficult to be recognized (Fig. 4a and Fig. 4b). The researchers analyzed and defined these classes of users as the biometric menagerie [17] or biometric zoo [16].

These performance evaluations are obtained a posteriori when the test data is compared with the enrollment data. Is it possible to predict the performance of each user based exclusively in her enrollment data? To answer this question it is necessary to determine if the enrollment data provided by the user contains enough information to ascertain the performance during the subsequent test phase.

Inspired by the methodology employed in [17], we divided the population in three groups according to their performance (obtained following the experimental protocol explained in Section 3.2) as Good (33% of users with lowest EER), Ugly (33% of users with highest EER) and Bad (the remaining 33% of users). This is not an ideal classification but allows us to group users with similar performances. See Table 2 for the resulting performance of the baseline systems in these three groups.

Fig. 5 shows the ROC curves for each of the performance groups of the two keystroke dynamics recognition algorithms employed in this work. The curves evidence the different performances for both recognition algorithms and the three groups considered.

Table 2. Performance (EER in %) according to the performance groups

System	All	Good		Bad		Ugly	
		Mean	Min/Max	Mean	Min/Max	Mean	Min/Max
A	8.20	4.31	0.45/6.65	8.58	6.90/9.75	15.31	9.80/24.50
B	8.55	4.27	0.90/6.90	8.65	7.10/11.10	16.66	11.10/27.55

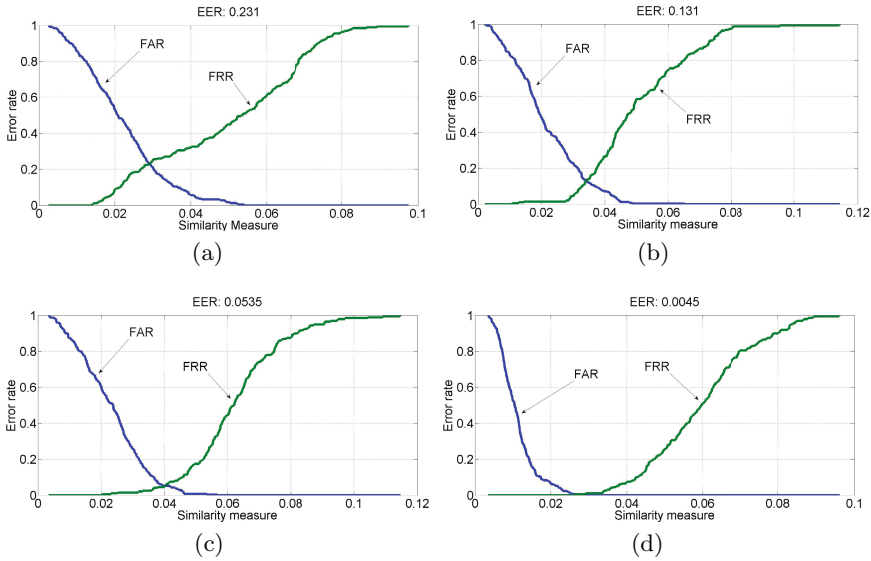

Fig. 4. False Acceptance and False Rejection curves for user 1 (a), user 2 (b), user 3 (c) and user 31 (d) from CMU benchmark dataset using the System A

3.4 Predicting the Quality

This paper analyzes five different measures for estimating the performance of keystroke dynamics users based exclusively on the enrollment data. Assuming that $\mathbf{v}^Q = [v_1^Q, v_2^Q, ..., v_N^Q]$ is the feature vector of the Query sample, N is the number of features ($N = 31$ for the CMU benchmark database) and $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_N]$ the enrollment data mean of the user (detailed in Section 2.1), the measures evaluated in this paper are defined as:

– **Variance:** the variance measures the stability of the data available for each user. A small variance indicates small differences between the query sample and the enrollment mean. A high variance indicates that the feature distribution is spread out around the mean. The variance is a valuable measure to characterize the stability of the data provided by the user. The variance is defined as:

$$\text{Variance} = \frac{1}{N} \sum_{n=1}^{N} \left(v_n^Q - \mu_n\right)^2 \tag{3}$$

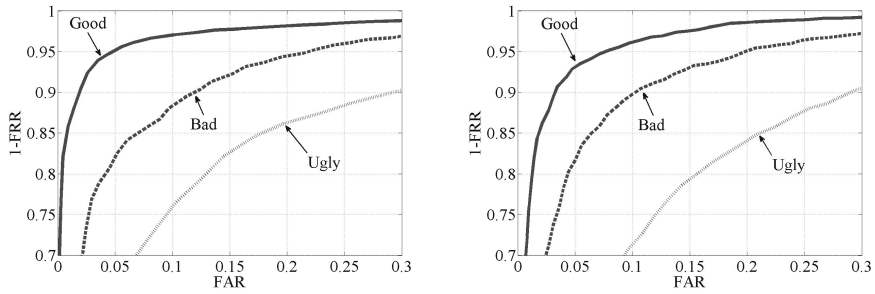

Fig. 5. ROC curves for different performance groups obtained using the System A (left) and System B (right)

– **Structural Content:** this is a popular quality measure for image analysis [24]. For a one-dimensional vector it is defined as the relative difference between the information of the query sample and the enrollment mean. The structural content is defined as:

$$\text{Structural Content} = \frac{\sum_{n=1}^{N} \left(v_n^Q\right)^2}{\sum_{n=1}^{N} \left(\mu_n\right)^2} \tag{4}$$

– **Entropy:** the entropy quantifies the expected value of the information contained in a sequence. The entropy value is defined as:

$$\text{Entropy} = -\sum_{n=1}^{N} v_n^Q \log v_n^Q \tag{5}$$

– **Kullback-Leibler Divergence:** as described in Eq. (1). The K-L divergence measures the amount of information needed to approximate two distributions.
– **Genuine scores:** it is possible to estimate the genuine score distribution of the enrollment data. From the 4 sessions available as enrollment, we used three sessions for training and the other one for validation. The protocol is repeated for all 4 sessions for a total number of genuine scores equal to 200.

The experiments conducted try to ascertain the capability of the proposed measures to predict the performance of each user in a keystroke dynamics system (using only its enrollment data). The protocol used to ascertain the prediction capability can be summarized in the following steps:

– Based on the performance obtained with each of the systems (performance reported in Table 2), we assign a quality value between 2 and 0, Q_i^A and Q_i^B, for each subject i (Good=2, Bad=1, Ugly=0, $i = 1, ..., 51$). Therefore there are two different quality values assigned for each user (one for each system).

Table 3. Distances between a posteriori user quality estimations $\hat{Q}_i^M$ and a priori quality prediction Q_i^A based on individual statistical measures from the enrollment data (Baseline System A)

Quality Feature	$\frac{1}{51}\sum \lvert Q_i^A - \hat{Q}_i^M \rvert$	# Large errors (out of 51 subjects)
Genuine scores	1.05	16
Structural content	0.58	5
Entropy	0.63	7
Variance	0.62	4
K-L divergence	**0.52**	**3**

Table 4. Distances between a posteriori user quality estimations $\hat{Q}_i^M$ and a priori quality prediction Q_i^B based on individual statistical measures from the enrollment data (Baseline System B)

Quality Feature	$\frac{1}{51}\sum \lvert Q_i^B - \hat{Q}_i^M \rvert$	# Large errors (out of 51 subjects)
Genuine scores	0.78	9
Structural content	0.66	6
Entropy	0.67	6
Variance	0.58	6
K-L divergence	**0.54**	**3**

- The five statistical measures are computed using the enrollment data. The average values across the 200 enrollment samples are computed for each subject i.
- For each average measure M ={Variance, Structural Content, Entropy, K-L Divergence, Genuine Scores}, the estimated quality $\hat{Q}_i^M$ of each user i is assigned as Good (33% of best M values), Ugly (33% of worst M values) and Bad (the remaining 33% of M values). The term best depends of the measure employed being the lowest values in case of {Variance, Entropy, Structural Content and K-L Divergence} and highest values in case of {Genuine Scores}.
- The mean distance between the real quality groups obtained with the test data, Q_i^A and Q_i^B, and the different estimated qualities $\hat{Q}_i^M$ is evaluated.

Tables 3 and 4 show the distances between real quality groups obtained with test samples and the predictions obtained with the enrollment data. Note that the quality, as employed in this section, depends of the performance of a specific matcher. The tables also show the number of large prediction errors, *i. e.* the number of good users estimated as ugly or vice versa.

As can be seen in Tables 3 and 4, the genuine scores obtained from the enrollment data are less competitive than other measures. The reason is that genuine scores are very sensitive to those users especially vulnerable to impersonation. The K-L divergence shows the most competitive performance with only 3 large errors (out of 51 subjects) and a mean estimation error around 0.5. Fig. 6 shows

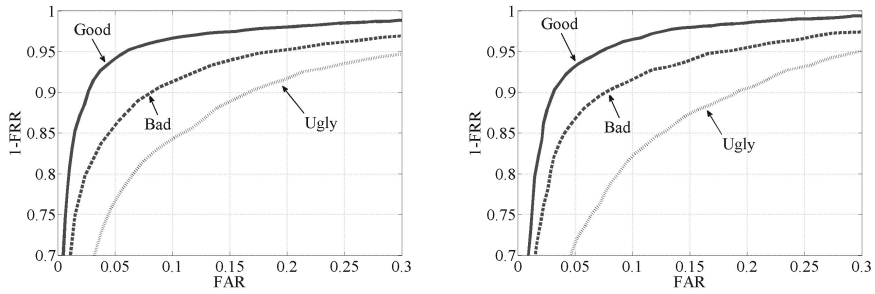

Fig. 6. ROC curves for different predicted qualities using the K-L divergence with Systems A (left) and B (right)

the ROC curves (averaged) of users classified by their predicted quality (using the K-L divergence).

The results show how the K-L divergence can be used to classify users apriori which will result in different performances groups in testing. The difference between classes is evident and the results suggest that the proposed measure is useful to predict the performance of keystroke dynamics using only the enrollment data.

4 Conclusions

This paper studied the feasibility of user quality prediction for biometric recognition based on keystroke dynamics. The usefulness of quality measures in biometrics is well-known and the scarce study on keystroke dynamics represents an open challenge for the scientific community. The performance of keystroke dynamics is highly user-dependent and it is usual to find large performance deviations among users even with the most competitive recognition algorithms. This paper analyzed five statistical measures for predicting the quality of users and the K-L divergence showed the most accurate results. The results showed that it is possible to ascertain the performance of users using exclusively the genuine enrollment data and encourage to further research in this area.

The work presented in this paper is focused on a limited dataset (i. e. same password and large amount of data per user) and future work includes other scenarios and databases. The prediction of performances when the password is different for each subject as well as text-independent keystroke dynamics are challenging scenarios to be studied.

Acknowledgments. Aythami Morales Moreno is supported by a Juan de la Cierva Fellowship from Spanish MINECO. This work has been partially supported by projects Bio-Shield (TEC2012-34881) from Spanish MINECO and BEAT (FP7-SEC-284989) from EU.

References

1. Peacock, A., Ke, X., Wilkerson, M.: Typing patterns: A key to user identication. IEEE Security and Privacy **2**(5), 40–47 (2004)
2. Banerjee, S.P., Woodard, D.L.: Biometric authentication and identification using keystroke dynamics: a survey. Journal of Pattern Recognition Research **7**, 116–139 (2012)
3. Bailey, K.O., Okolica, J.S., Peterson, G.L.: User identification and authentication using multimodal behavioral biometrics. Computers and Security **43**, 77–89 (2014)
4. Gunetti, D., Picardi, C.: Keystroke analysis of free text. ACM Transactions on Information and System Security **8**(3), 312–347 (2005)
5. Sim, T., Zhang, S., Janakiraman, R., Kumar, S.: Continuous verification using multimodal biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligent **29**(4), 687–700 (2007)
6. Hocquet, S., Ramel, J.Y., Cardot, H.: User classication for keystroke dynamics authentication. In: International Conference on Biometrics, Seoul, Korea, pp. 531–539 (2007)
7. Alonso-Fernandez, F., Fierrez, J., Ortega-Garcia, J.: Quality measures in biometric systems. IEEE Security and Privacy **10**(9), 52–62 (2012)
8. Chen, Y., Dass, S.C., Jain, A.K.: Fingerprint quality indices for predicting authentication performance. In: International Conference on Audio and Video-Based Biometric Person Authentication, Hilton Rye Town, NY, USA, pp. 160–170 (2005)
9. Youmaran, R., Adler, A.: Measuring biometric sample quality in terms of biometric information. In: Biometrics Symposium, Baltimore, USA (2006)
10. Killourhy, K., Maxion, R.: Why did my detector do *That*?! predicting keystroke-dynamics error rates. In: Jha, S., Sommer, R., Kreibich, C. (eds.) RAID 2010. LNCS, vol. 6307, pp. 256–276. Springer, Heidelberg (2010)
11. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bigun, J.: Discriminative multimodal biometric authentication based on quality measures. Pattern Recognition **38**(5), 777–779 (2005)
12. Hong, L., Wan, Y., Jain, A.K.: Fingerprint image enhancement: Algorithm and performance evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**, 777–789 (1998)
13. Fiérrez-Aguilar, J., Chen, Y., Ortega-Garcia, J., K.Jain, A.: Incorporating image quality in multi-algorithm fingerprint verification. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 213–220. Springer, Heidelberg (2005)
14. Kumar, A., Zhang, D.: Improving biometric authentication performance from the user quality. IEEE Transactions on Instrumentation and Measurement **59**(3), 730–735 (2010)
15. Prabhakar, S., Pankanti, S., Jain, A.K.: Biometric recognition: Security and privacy concerns. IEEE Security Privacy Magacine **1**(2), 33–42 (2003)
16. Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: Sheep, goats, lambs and wolves: a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In: International Confenrence on Spoken Language Processing, Sydney, Austrailia (1998)
17. Yager, N., Dunstone, T.: The biometric menagerie. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(2), 220–230 (2010)
18. Zhong, Y., Deng, Y., Jain, A.K.: Keystroke dynamics for user authentication. In: IEEE Computer Society Workshop on Biometrics, Providence, USA (2012)

19. Killourhy, K.S., Maxion, R.A.: Comparing anomaly detectors for keystroke dynamics. In: International Conference on Dependable Systems and Networks, Estoril, Portugal, vol. 32, pp. 125–134 (2009)
20. Kang, P., Park, S., Hwang, S., Lee, H., Cho, S.: Improvement of keystroke data quality through artificial rhythms and cues. Computers and Security **27**, 3–11 (2008)
21. Cho, S., Hwang, S.: Artificial rhythms and cues for keystroke dynamics based authentication. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 626–632. Springer, Heidelberg (2005)
22. Houmani, N., Garcia-Salicetti, S., Dorizzi, B.: A novel personal entropy measure confronted with online signature verification systems performance. In: IEEE Conference on Biometrics: Theory, Applications and Systems, Washington, USA, pp. 1–6 (2008)
23. Deng, Y., Zhong, Y.: Keystroke dynamics user authentication based on gaussian mixture model and deep belief nets. ISRN Signal Processing **2013**, 1–7 (2013)
24. Killourhy, K.S., Maxion, R.A.: Image quality measures and their performance. IEEE Transactions on Communications **43**(12), 2959–2965 (1995)