Some remarks on a set of information theory features used for on-line signature verification

Margit Antal
Sapientia University
Faculty of Technical and Human Sciences
Department of Mathematics and Informatics
Tirgu Mures, Romania
Email: manyi@ms.sapientia.ro

László Zsolt Szabó
Sapientia University
Faculty of Technical and Human Sciences
Department of Electrical Engineering
Tirgu Mures, Romania
Email: lszabo@ms.sapientia.ro

***Abstract*—Recently a new set consisting of six information theory features was proposed for on-line signature verification by Rosso, Ospina and Frery. The proposed features were evaluated on the MCYT-100 on-line signature database resulting in the best performance ever measured on that dataset. In this paper we repeat their measurements and show that their result is erroneous. In addition, we evaluate the performance of the same on-line signature verification system using exactly the same number of state-of-the-art features. State-of-the-art features always outperform the information theory related features, regardless of the classification method used.**

***Keywords* —on-line signature verification, information theory features, feature consistency, performance evaluation**

I. Introduction

Handwritten signatures have been used in personal verification for centuries. Due to the proliferation of touchscreen mobile devices, the capture and the transmission of the signature is now available from anywhere. This in turn gives new dynamics to signature biometrics research.

Signature verification systems are divided into offline and on-line systems. While offline systems use the image of the signature, on-line systems use information related to the dynamics of the signature. This paper deals with on-line systems.

On-line signature verification systems can use three types of feature: local, global, and segmental[1]. Local features are extracted for each sample point of the signature (e.g. point-wise velocities). Global features are based on all sample points of the signature (e.g. duration or average velocity). In the case of segmental features the signature is divided into segments and one feature is extracted from each segment. This paper deals with global features.

More than 150 different global features have been proposed by several research papers. Fierrez-Aguilar, Nanni, Penalba, Ortega-Garcia and Maltoni [2] proposed 100 distinct global features. The set of features was sorted by individual discriminative power. Sae-Bae and Memon represented a signature as a set of histograms [3]. Histograms were derived from x-y trajectories, speed, angles, pressure and their derivatives.

Recently Rosso, Ospina and Frery [4] proposed time causal information theory features. They proposed features using Shannon entropy, statistical complexity, and the Fisher information evaluated over the Bandt and Pompe symbolization [5] of the horizontal and vertical coordinates of signatures. The authors claim that the use of these six information theory features and a one-class support vector classifier results in better performance than the use of state-of-the-art on-line systems that employ higher-dimensional feature spaces.

A great amount of research has been conducted in feature extraction and selection for on-line signature verification systems [6], [7], [8]. Good features are those having high consistency. Consistency as defined by Lei and Govindaraju [9] means that the values extracted from genuine signatures should be close to each other while the distances between genuine and forged features should be large. Lei and Govindaraju [9] presented a consistency model using a distance based measure. Both local and global features were examined. The authors conclude that the most consistent features are the x and y coordinates, the speed of writing and the angle with the x-axis.

In this paper we present our experiments using the six features proposed by Rosso, Ospina and Frery and show that their result is erroneous. The correct result is about 20% instead of the presented 0.19% equal error rate (EER) achieved for 5 training samples. We compare evaluation results using their six features with the same system using six state-of-the-art features. Finally we present evaluations related to the consistency of the used features. All data related to this paper are available on-line.[1]

Table I shows the best performance results reported on MCYT-100 dataset using 5 signatures for training.

The rest of this paper is organized as follows. Section II describes the feature extraction, feature selection and the anomaly detectors employed in this study. Experiments and results are presented in section III. The final section concludes the paper.

[1] http://mobio.ms.sapientia.ro/mcyt.html

TABLE I
PERFORMANCE EVALUATIONS ($ERR\,\%$) REPORTED ON THE MCYT-100 DATASET USING 5 SIGNATURES FOR TRAINING. RF - RANDOM FORGERIES. SF - SKILLED FORGERIES.

Author	Year	RF	SF	Description
Fierrez [2]	2005	0.24	2.12	Global (Parzen) and local (HMM) experts fusion
Pascual [10]	2008	0.29	1.23	Local (DTW-based) system
Rosso [4]	2016		0.19	Global system (one-class SVM) information theory features (6)

TABLE II
GLOBAL FEATURES. #FEAT: NUMBER OF FEATURES

Name	#feat
Duration	1
Average velocity	1
Average pressure	1
Average x velocity	1
Average y velocity	1
Sign changes of $X^1Y^1X^2Y^2P^1P^2$	6
Histogram of Θ sequence	8
Total	**19**

II. METHODS

A. Feature extraction

1) State-of-the-art global features: We computed position-, pressure- and time-based features [2] [11], such as duration, average, horizontal and vertical velocities, sign changes of different time series computed from the raw data, and histogram-based features. Several histograms were proposed by Sae-Bae and Memon [3]. We used some features extracted from the angles' histogram. The detailed description of the feature extraction process can be found in an earlier paper written by the authors [12].

Let $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_n\}$ be the x, y coordinates of a signature. We denote by $P = \{p_1, p_2, \ldots p_n\}$ the pressure measured in each touch point. First- and second-order derivatives of these sequences were computed as follows: $X^1 = \{x_i^1 | x_i^1 = x_{i+1} - x_i\}$, $Y^1 = \{y_i^1 | y_i^1 = y_{i+1} - y_i\}$, $P^1 = \{p_i^1 | p_i^1 = p_i\}$ where $i = 1, 2, \ldots, n-1$, and $X^2 = \{x_i^2 | x_i^2 = x_{i+1}^1 - x_i^1\}$, $Y^2 = \{y_i^2 | y_i^2 = y_{i+1}^1 - y_i^1\}$, $P^2 = \{p_i^2 | p_i^2 = p_{i+1}^1 - p_i^1\}$, $i = 1, 2, \ldots, n-2$. Angles $\Theta_i^1 = tan2^{-1}(y_i^1, x_i^1), i = \overline{1, n-1}$, were computed using the $tan2^{-1}$ trigonometric function which is a variation of the standard $arctan$ function.

The state-of-the-art features used in our experiments are presented in Table II. Typically these features have very different dynamic ranges. Therefore, these features were normalized in the subset used to train the models (separately for each subject) in order to meet real systems requirements. We applied standard min-max normalization:

$$f_i^{'} = \frac{f_i - min_i}{max_i - min_i}, i = 1, \ldots, D, \tag{1}$$

where f_i is the ith feature, min_i and max_i are the smallest, respectively the largest value of the ith feature across the training set, and D is the number of features.

2) Information theory features: Rosso, Ospina and Frery [4] proposed the use of the following time causal quantifiers based on information theory for handwritten signatures: normalized Shannon entropy (H), permutation statistical complexity (C) and permutation Fisher information measure (F). They extracted these three features both from x and y coordinates of signatures resulting in 6 features: $H_x, H_y, C_x, C_y, F_x, F_y$. Detailed description of these features is presented in their paper [4].

B. Feature selection

Feature selection was applied in the case of the 19 state-of-the-art features presented earlier in order to select 6 features.

Though signature recognition is a one-class problem, the existence of forgery samples in the MCYT-100 dataset allows us to formulate the feature selection problem in a similar way to a two-class classification scheme.

Consequently for each user of the dataset a two class subset was created, containing the genuine and the forgery samples of the user (25 genuine and 25 forgery samples). For all of these subsets, feature selection algorithms were applied separately, yielding as a result the most relevant features characterizing the user according to the selection algorithm. Finally, a majority vote was applied on all per user feature sets in order to select 6 features.

Feature selection methods were applied by using the Weka data mining framework. In the case of the feature set used in this study, a wrapper method was applied as a feature evaluation method (the *WrapperSubsetEval* method with the Random Forests classifier, 100 trees). As for search method the *BestFirst* algorithm was used with a forward search strategy starting from feature number 1: duration. Numerous previous tests confirmed that duration is one of the most relevant features that characterizes the user. The search termination parameter of consecutive non improvement nodes was set to 5.

Running the search for each user resulted in distinct sets of best performing features, next the first k features were selected from each per user feature set. Finally 6 features were selected from the resulting global set based on a majority vote. The feature set rf369 resulted, containing the same values irrespective of k=3, 6, 9.

It is worth mentioning that, by using other wrapper or information gain based feature selection methods, various 6 features set could be selected with classification results similar to rf369. In all of these sets, duration and average velocity were present as high ranked features, completed with features representing other velocities, pressure and sign changes.

C. Feature consistency

We used a simple feature consistency measure proposed by Lee, Berger and Aviczer [13]. This measure is defined as:

$$d_i(s) = \frac{|m_i(genuine) - m_i(forgery)|}{\sqrt{\sigma_i^2(genuine) + \sigma_i^2(forgery)}}, \tag{2}$$

where $d_i(s)$ denotes the consistency of feature i for the subject s. $m_i(genuine)$ is the sample mean computed for feature i in the case of genuine signatures and $m_i(forgery)$ is the sample mean for the same feature in the case of forged signatures. $\sigma_i^2(genuine)$ and $\sigma_i^2(forgery)$ are the sample variations of feature i for genuine and forged signatures. Consistency was computed for each feature using the genuine and forged signatures of each subject. This resulted in a set of N consistencies for each feature, where N is the number of subjects. Therefore, we report the mean and its standard deviation over subjects.

D. Anomaly detectors

Several anomaly detectors were implemented for template creation and matching. The Euclidean, the Manhattan and the kNN anomaly detectors have been described in an earlier paper of the authors [12] and were used for the evaluation of the DooDB on-line signature dataset [14]. Each anomaly detector uses a specific template and score computation (see below). Dissimilarity scores (D_{score}) were transformed into similarity scores (S_{score}) by using formula 3.

$$S_{score} = \frac{1}{1 + D_{score}} \tag{3}$$

Training for the Euclidean anomaly-detection algorithm consists of the calculation of the mean feature vector of the training samples. In the test phase the detector calculates the Euclidean distance between the mean feature vector and the feature vector extracted from the test sample. In the case of the Manhattan detector the Manhattan distance is used in the test phase.

The k-nearest neighbour (kNN) detector is one of the simplest anomaly detectors which directly uses the feature vectors extracted from the training samples without any other computations. Testing means the calculation of the Euclidean distance between each of the training feature vectors and the feature vector extracted from test samples. Having more than one distance, the anomaly score is calculated as the average of the distances to the k nearest training samples.

One-Class Support Vector classification was performed with R (The R Project for Statistical Computing) by using the interface for LIBSVM [15] from R package e1071. Tuning LIBSVM parameters in a real signature recognition application is difficult due to the small number of positive and the lack of negative samples. Though one could find strategies for this task by dividing the user set, we chose to run LIBSVM with parameters similar to [4], $nu = 0.1$ and $gamma = 0.05$ respectively. The classifier decision values returned by the R interface were used as scores, and error rates were calculated with the ROCR package from R.

E. Performance metrics

Two types of EERs were computed: (i) using a global or universal threshold for all subjects (EER_g) and (ii) using user-dependent thresholds (EER_u). In the latter case we computed the EER for each user and report the mean and variance of these values. In addition, performance is reported in terms of Receiver Operating Characteristics (ROC) curve.

TABLE III
INFORMATION THEORY AND STATE-OF-THE-ART FEATURE SETS EACH HAVING EXACTLY 6 FEATURES.

Inf. theory feat. (rosso6)	State-of-the-art feat. (feat369)
H_x	Duration
H_y	Average velocity
C_x	Average pressure
C_y	Average x velocity
F_x	Average y velocity
F_y	Sign changes of X^1

III. EXPERIMENTAL RESULTS

In this section, experiments to evaluate the efficacy of the two feature sets are described and signature verification performances are reported.

A. Dataset and evaluation protocol

All the experiments presented in this study were carried out on the freely available MCYT on-line signature dataset (specifically on the MCYT-100 subset) published by Ortega-Garcia [16]. This subset contains - for each signatory - 25 genuine and 25 forged signatures. The signatures were acquired on a WACOM digitizer tablet with 100 Hz sampling frequency. We trained the system with the first 5, 10, and 15 genuine signatures per subject. The remaining genuine signatures and the 25 available forged signatures were used for evaluation in the skilled forgeries case. As for the random forgeries case we used the same genuine signatures for templates and the first genuine signatures from each other user as forged signatures (resulting in 99 signatures for each subject).

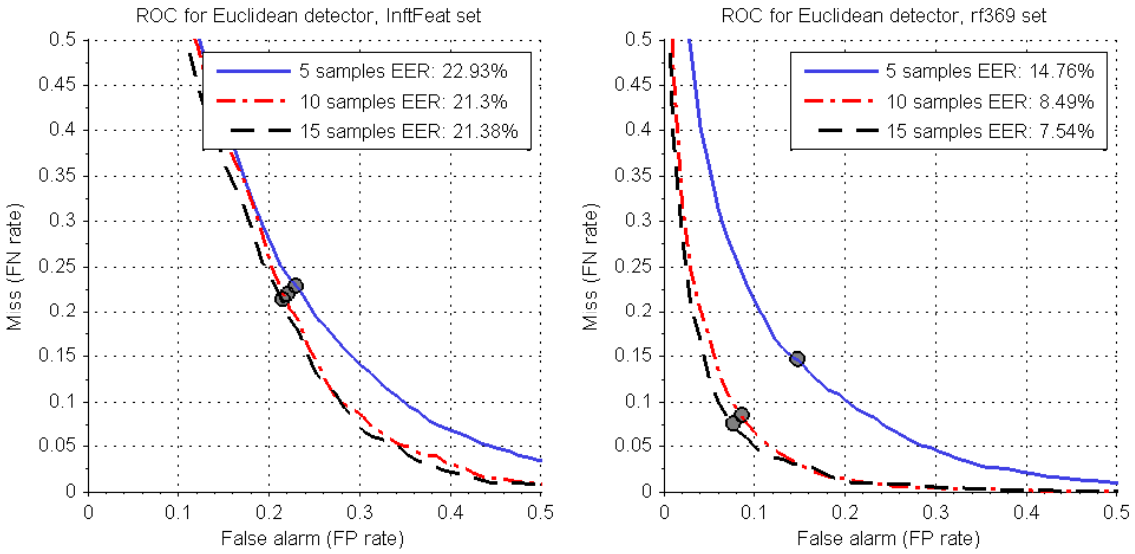
B. Verification results

Table III presents the two evaluated feature sets. Two types of evaluations were conducted: skilled- and random forgery cases. The corresponding verification performances are reported in Tables IV and V. In both cases four anomaly detectors were used and the evaluations were repeated three times using 5, 10 and 15 genuine signatures for template creation. In each case the state-of-the-art features proved to be significantly better than the information theory features.

Verification performance results for the two feature sets and skilled forgery case are depicted in the form of ROC plots in Figures 1a and 1b.

C. Feature consistency

Consistency (means and standard deviations – as presented in section II-C) of the employed features are presented in Table VI. The table presents the consistencies in decreasing order of mean consistency value. However, a good feature should have high mean and low standard deviation (in order to be stable across subjects). Although the information theory features have high mean consistency values, they also have high standard deviations. The low efficacy of these features was confirmed by the high EERs (see Table IV).



(a) 6 information theory features

(b) 6 state-of-the-art features

Fig. 1. ROC curves for Euclidean anomaly detector with 5, 10 and 15 training samples evaluated on the two feature sets (skilled forgeries case). The EERs are computed using global thresholds (EER_g).

TABLE V
VERIFICATION PERFORMANCE IN TERMS OF EER_g AND EER_u USING THE TWO FEATURE SETS. RANDOM FORGERIES CASE. [%]

	Inf. theory feat.		State-of-the-art feat.	
Detector	EER_g	EER_u (stdev)	EER_g	EER_u (stdev)
Enrollment: 5 samples				
EUCL.	19.11	16.24 (8.97)	10.11	4.54 (5.08)
MANH.	18.89	16.75 (9.31)	9.04	4.18 (4.80)
KNN	19.27	16.45 (9.77)	9.38	4.45 (4.74)
LIBSVM	23.20	16.40 (8.90)	17.70	4.40 (5.10)
Enrollment: 10 samples				
EUCL.	16.91	15.03 (8.66)	4.63	2.41 (3.29)
MANH.	17.32	15.69 (8.83)	4.76	2.43 (3.32)
KNN	20.58	18.16 (11.55)	5.11	3.03 (3.57)
LIBSVM	22.50	15.30 (8.30)	12.50	2.40 (3.40)
Enrollment: 15 samples				
EUCL.	15.90	14.47 (8.51)	4.17	2.03 (3.82)
MANH.	16.30	15.03 (9.03)	4.29	1.87 (3.29)
KNN	18.01	14.96 (9.44)	4.85	2.48 (4.57)
LIBSVM	30.20	14.50 (8.00)	28.40	2.10 (3.80)

TABLE VI
FEATURE CONSISTENCY

feature	mean	stdev
F_y	1.0745	0.6497
F_x	1.0272	0.6270
C_y	1.0243	0.5800
C_x	0.9962	0.5986
H_y	0.9786	0.5179
H_x	0.9527	0.5335
Duration	0.4219	0.4359
Sign changes of X^1	0.3095	0.2277
Average y velocity	0.3008	0.2403
Average velocity	0.3002	0.2186
Average x velocity	0.2870	0.2469
Average pressure	0.2740	0.1908

TABLE IV
VERIFICATION PERFORMANCE IN TERMS OF EER_g AND EER_u USING THE TWO FEATURE SETS. SKILLED FORGERIES CASE. [%]

	Inf. theory feat.		State-of-the-art feat.	
Detector	EER_g	EER_u (stdev)	EER_g	EER_u (stdev)
Enrollment: 5 samples				
EUCL.	22.93	20.32 (12.83)	14.76	7.90 (7.35)
MANH.	22.54	20.59 (13.48)	14.23	7.96 (6.05)
KNN	24.30	21.15 (14.09)	14.59	8.54 (7.96)
LIBSVM	32.80	20.30 (12.80)	31.80	8.80 (9.30)
Enrollment: 10 samples				
EUCL.	21.30	19.64 (12.32)	8.49	5.45 (6.71)
MANH.	22.31	19.71 (12.24)	8.64	5.51 (6.67)
KNN	25.10	21.83 (14.47)	9.07	6.23 (7.64)
LIBSVM	33.70	19.00 (11.90)	27.10	5.60 (6.50)
Enrollment: 15 samples				
EUCL.	21.38	18.47 (12.21)	7.54	4.76 (6.73)
MANH.	20.96	18.90 (12.42)	7.95	4.50 (6.70)
KNN	22.56	19.86 (13.93)	8.42	5.50 (8.40)
LIBSVM	38.70	18.00 (12.10)	33.80	4.70 (7.10)

IV. CONCLUSION

In this study we repeated the measurements presented by Rosso, Ospina and Frery [4], and showed that the performance obtained by their information theory features is erroneous. The same evaluation was performed for the same number of features, in this case using state-of-the-art features. State-of-the-art features always outperform the information theory related features, regardless of the classification method used. Consistency analysis of the used features was also presented. However, the information theory features are not the best choice for signature verification, although they might be proven useful in signature quality evaluation.

ACKNOWLEDGMENT

The Biometrics Research Lab (ATVS), Universidad Autonoma de Madrid, provided the MCYT-100 signature corpus

employed in this work. The information theory related features were provided by Raydonal Ospina. The work of Margit Antal was supported by a Domus Hungarica research grant, contract number 5634/1/2016/HTMT.

REFERENCES

[1] J. Richiardi, H. Ketabdar, and A. Drygajlo, "Local and global feature selection for on-line signature verification," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Aug 2005, pp. 625–629 Vol. 2.

[2] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni, "An on-line signature verification system based on fusion of local and global information," in *Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication*, ser. AVBPA'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 523–532.

[3] S.-B. Napa and N. Memon, "Online signature verification on mobile devices," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 6, pp. 933–947, June 2014.

[4] O. A. Rosso, R. Ospina, and A. C. Frery, "Classification and verification of handwritten signatures with time causal information theory quantifiers," *PLoS ONE*, vol. 11, no. 12, 2016.

[5] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, p. 174102, Apr 2002.

[6] R. Plamondon and F. Leclerc, "Automatic signature verication and writer identification. the state of the art." *Pattern Recognition*, vol. 22, no. 22, pp. 107–131, 1989.

[7] F. Leclerc and R. Plamondon, "Automatic signature verication: the state of the art: 1989-1993," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 8, pp. 643–660, 1994.

[8] R. Plamondon, "The design of on-line signature verfication system: From theory to practice." *International Journal of Pattern Recognition and Artficial Intelligence*, vol. 8, no. 3, pp. 795–811, 1994.

[9] H. Lei and V. Govindaraju, "A comparative study on the consistency of features in on-line signature verification," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2483 – 2489, 2005.

[10] J. M. Pascual-Gaspar, V. Cardeñoso-Payo, and C. E. Vivaracho-Pascual, *Practical On-Line Signature Verification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1180–1189.

[11] P. Bissig, "Signature verification on finger operated touchscreen devices," Master's thesis, ETH Zürich, Distributed Computer Group, 10 2011.

[12] M. Antal and L. Z. Szabo, "On-line verification of finger drawn signatures," in *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, May 2016, pp. 419–424.

[13] L. Lee and E. Berger, T.and Aviczer, "Reliable on-line human signature verfication systems," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 643–647, 1996.

[14] M. Martinez-Diaz, J. Fierrez, and J. Galbally, "The doodb graphical password database: Data analysis and benchmark results," *Access, IEEE*, vol. 1, pp. 596–605, 2013.

[15] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[16] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro, "Mcyt baseline corpus: a bimodal biometric database," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, Dec 2003.