Persian Signature Verification Based on Fractal Dimension Using Testing Hypothesis

Atefeh Foroozandeh, Younes Akbari, Mohammad J. Jalili, Javad Sadri*
Faculty of Electrical and Computer Engineering, Department of Computer Engineering,
University of Birjand, Birjand, Iran.
Email: {at_foroozandeh@yahoo.com, akbari_younes@yahoo.com, mjavad_jalili@yahoo.com, javad.sadri@cs.mcgill.ca}
**Corresponding author*

***Abstract*—A new approach for verifying off-line Persian signatures is presented, in this paper. In our method, feature extraction step is conducted based on estimated Fractal Dimension (FD) of signatures images, and making decision about acceptance/rejection of test signature is formulated as testing hypothesis which is used for the first time in order to verify off-line Persian signatures. The proposed method has been tested on our new created database included 1000 genuine signatures and 200 skilled forgeries which have been collected from a population of 100 human subjects with different educational background. Obtained results confirm the effectiveness of the presented method.**

***Keywords*-Off-Line Signature Verification; Fractal Dimension; Testing Hypothesis; Database of Persian Signatures.**

I. Introduction

Automatic signature verification (SV) is one of the biometric approaches in person authentication. Compared to other biometrics such as: iris, voice, and fingerprint, the cost of personal verification using his/her signatures is very low. Analysis of signature to authenticate the identity of an individual is conducted through discrimination between genuine signature from a forgery [1]. Automatic SV has many applications including authenticating bank checks, contracts, and other security documents. SV can be conducted in two ways: on-line, and off-line. So many behavioural information such as: pen-point velocity, tremor information, and writing pressure which can be extracted in on-line SV, are not available in off-line SV and only some signatures signed on some sheets of papers are available in off-line SV. Mentioned restrictions lead to off-line SV be more difficult than on-line. Many attempts have been conducted in both off-line, and on-line SV in Latin language. Two comprehensive surveys of these conducted works have been done by Impedovo et al. [2], and Pal et al. [3].

During the last decade, some works have been conducted for verifying Persian signatures [4-6]. Compared to Latin there are fewer number of works on Persian signatures, so more attention is needed in this field. It should be noted that in spite of Latin signatures which are reshaped of their handwritten names, Persian signatures are usually made of cursive sketches [1]. Some samples of Persian and Latin signatures have been shown and compared in Figure 1.

Due to above mentioned issues, we decided to focus on off-line Persian SV and proposed a new approach for this target. Our presented approach uses two main concepts: Fractal Dimension (FD) as extracted feature from each signature image, and testing hypothesis used for discrimination between genuine signatures from skilled forgeries. Making decision about acceptance/rejection of signature in question using testing hypothesis yields to obtain more accurate results. Notably statistical tools have been used by some of researches in the process of SV specially in on-line SV systems such as [7]. Also, there are some similar works in off-line SV such as [8]. However, based on our searches, presented approach with us has not been proposed in Latin and Persian conducted works, till today. Our presented SV method uses FD values as global features and testing hypothesis as a statistical tool for final decision. The performance of the presented method has been evaluated using our new database. This database includes 1200 Persian signatures which have been signed by 105 writers.

(a) (b)

Figure 1. Some samples of Persian and Latin signatures are in (a), and (b), respectively. As shown, Persian and Latin signatures are essentially different.

The rest of this paper is organized as follows. Section 2 proposes some background information used in our method. Section 3 describes our presented approach. In Section 4, the experimental results are shown. Finally, our conclusions and future works are presented in Section 5.

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.254

II. Background

The presented SV method uses two main concepts: Katz's method, and Kolmogorov-Smirnov test. Before explanation about procedure of the presented method, two above mentioned concepts are explained briefly in the next subsections.

A. Katz's Method

Fractal term was coined by Mandelbrot in 1975. Fractal theory is based on geometry and dimension theory. Quantity of information embedded in a pattern can be measured using Fractal Dimension (FD) values, which has so many applications in different fields of signal processing, and pattern recognition [9], [10]. In this work, feature extraction step has been conducted based on the estimated FD values of waveforms. Waveforms can be viewed as a sequence of points like: $P = \{p(1), p(2), ..., p(N)\}$, where $p(i) = (x(i), y(i))$, with $x(i) < x(i+1)$, $i = 1, 2, ..., N$, (N: number of points) which are special cases of planar curves that moving only forward in the x direction [10]. Several methods, such as Higuchi, Petrosian and Katz have been proposed for estimating the FD of waveforms [10], [11]. Between these, Katz's method is relatively insensitive to noise and also more convenient in practice [11]. Therefore, our study uses Katz's method. FD of a waveform by Katz's method is estimated using a sliding window which is moved over the waveform amplitude as overlapping or non-overlapping. The rate of neighboring sliding windows has a direct effect on precision of estimated FD. Non-overlapping windows will reduce the precision of estimating of FD. In the most of cases length of sliding window is determined empirically according to the length of waveform amplitude so that variations in each sliding window can be considered constant. Using Katz's method the FD of the samples within the current sliding window $\{P_i\}_{i=1}^{M}$, is estimated using the following equation:

$$FD = \frac{\log(n)}{\log(n)\log(d/l)}, \tag{1}$$

where l is total length of section of waveform which is in the current sliding window computed by:

$$l = \sum_{i=1}^{M} ||P_{i+1} - P_i||, \tag{2}$$

d is its diameter estimated the distance between the first point of the waveform which is in the current sliding window and the point in sliding window that provides the farthest distance from the first point:

$$d = max_i ||P_i - P_1||. \tag{3}$$

and n is the number of steps in the current sliding window which is 1 less than the number of points M. Also $||.||$ is the Euclidean distance [10]. Therefore using Katz's method, a vector of FD values (a FD profile) is obtained corresponding to each waveform. Since true waveforms can never become sufficiently convoluted to fill a plane, the waveforms will never have FDs approximating the dimensionality of a plane (D = 2.0) [12]. So the range of FD values of waveforms is between one and two.

B. Kolmogorov-Smirnov Test

One of the most useful nonparametric methods which is used for comparing two given distributions of a single independent variable is Kolmogorov-Smirnov (K-S) test. The K-S static for comparing distributions P_0, and P is computed as follows:

$$D_{mn} = (\frac{mn}{m+n})^{1/2} max_{-\infty<x<+\infty} |F_m(x) - G_n(x)|, \tag{4}$$

where $F_m(x)$, and $G_n(x)$ are empirical cumulative distribution functions (c.d.f.) of two given distance distributions P and P_0, respectively [13]. The p-value for this test of hypothesis is computed as follows:

$$p - value = P(D \geq D_{mn} | H_0) \approx 1 - H(D_{mn}), \tag{5}$$

where H(t) is the c.d.f. of K-S distribution with the following definition:

$$H(t) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 t}. \tag{6}$$

[13]. If the computed p-value greater than significance level (α), the null hypothesis is accepted, else is rejected. It should be noted that in testing hypothesis, two types of errors can be occur: type I, and type II [14]. If the null hypothesis is rejected when it is true, type I error occurs. Also, type II error occurs if the null hypothesis is accepted when it is false. The probability of committing a type I error in a decision rule is called the significance level (α) which is usually considered as 5%, or 1%.

III. Proposed Signature Verification Method

Block diagram of the presented method is shown in Figure 2. As shown in this figure our method includes three main steps: pre-processing, feature extraction, and final step for decision about genuine/forged of the test signature. These steps are explained in the next subsections.

A. Pre-Processing

In order to reduce inevitable differences between signatures samples related to each person existed due to special conditions of time of signing, and facilitating the feature extraction step, pre-processing step has been conducted on each of signatures images as follows.

• Rotation normalization is conducted similar to [15].

•Size normalization is carried out using Aspect Ratio Adaptation Normalization (ARAN) method [16].

•Shift normalization is achieved by centering the Center Of Gravity (COG) of each image into a square frame of size $2 * w$, where w is the maximum of length and width

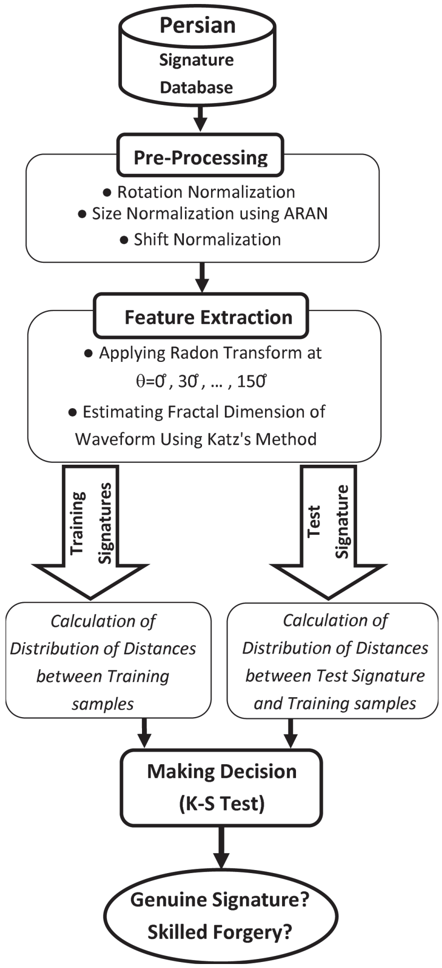

Figure 2. Block diagram of the presented signature verification method.

of all person's training signatures. The procedure of feature extraction step is explained in the next subsection.

B. Feature Extraction

In this paper, feature extraction step is conducted based on Radon Transform (RT), and Katz's method. It should be noted that RT is one of the most applicable methods in the field of image processing, and pattern recognition. Notably, focus of this paper isn't on RT. For more information about RT and it's applications refer to [17]. In order to extract feature vector from each signature image, following steps have been conducted on each pre-processed signature image.

I) In the first step, RT is applied at six angles = $0°$, $30°$,..., $150°$, on each signature. Therefore six projection profiles (waveforms) are obtained related to each signature image.

II) Obtained waveforms from Step I, are concatenated into a single vector so called RT Waveform (RTW) in this paper. Then FD of this vector is estimated using Katz's method (Subsection 2.1) [18]. In order to apply Kat's method, a sliding window of length 25 samples has been considered. Also, 10 samples have been considered as intersection of each two neighbouring sliding window. These values are determined based on the experiments. Here, according to the waveform amplitude, above considered size for sliding window enabled us to reasonably assume that the waveform is stationary during each sliding window. After applying Katz's method on RTW, a vector of FD values is obtained. This vector considered as feature vector related to each signature image in our method.

C. Decision Making Procedure

Here, making decision for acceptance/rejection of test signature is formulated as testing hypothesis i.e decision between two existed status: a given signature belongs to the person having claimed his/her identity (null hypothesis) or it is a forged signature (alternative hypothesis) [19]. In other words, mentioned test is decision between the following hypotheses:

$$H_0 : P = P_0 \quad vs. \quad H_1 : P \neq P_0, \tag{7}$$

where P_0 considered as the distribution of distances [8] between training samples of true person, and P considered as distribution of distances between the signature in question with all saved true samples of the person having claimed. Two distributions P_0, and P include pairwise distances between feature vectors of signatures images. These distances have been computed using cosine distance formula, in this paper. Cosine distance is one of the similarity measures which has been widely used in information retrieval applications including text analysis [20]. Cosine distance between two feature vectors X, and Y computed as follows, where x_i, and y_i show individual features of X, and Y, respectively.

$$d(X,Y) = \sum_{i=1}^{d} \frac{x(i).y(i)}{\sqrt{\sum_{i=1}^{d} x^2(i)}\sqrt{\sum_{i=1}^{d} y^2(i)}} \tag{8}$$

Naturally, two distances distributions P_0, and P are slightly different. However, in signature verification process the goal is answering to the question that whether existed difference is significant for failing the null hypothesis (rejection the signature in question) or not. For this, K-S test with $\alpha = 5\%$ (Subsection 2.2) has been used. In our experiments. The signature in question has been claimed as a genuine sample (accept the null hypothesis) or skilled forgery (reject the null hypothesis) with confidence level 95%$(= 1 - \alpha)$ [14]. The experimental results are provided in the next section.

IV. Experimental Results

In this section, first the creation of our database is explained, then obtained results of our experiments are shown.

A. Database Description

In order to evaluate the performance of our presented method, a new database of Persian signatures has been created in this paper. In creation of our database 100 persons with different educational backgrounds have been participated. Each participant produced ten genuine signatures. Some samples of signatures in our database were shown in Figure 1. Also, the distribution of our participants has been shown in Table I. ince differences between Latin and Persian signatures which was mentioned in Section 1, in Persian SV, verifying skilled forgeries attracts more attention than random forgeries (which are usually seen in Latin SV). So, verifying skilled forgeries have been considered in this paper. For each participant two skilled forgeries have been copied without any restriction in time and number of iteration for exercise to copy from his/her genuine signatures. This imitation has been conducted by one of five forgers served in creation of the database, per participant.

Table I
THE OVERALL DISTRIBUTION OF SIGNATURES SAMPLES IN OUR DATABASE.

Range of participants	18-60(years)
No. of right handed signers	77
No. of left handed signers	23
Total no. of signers	100 (=77+23)
No. of forgers (men+women)	5 (=3+2)
No. of genuine signatures per participant	10
No. of skilled forgeries per participant	2
Total no. of signatures per participant	1000 (=100× 10)
Total no. of signatures in the database (genuine+skilled forgeries)	1200(=1000+200)

B. Results

In this paper, evaluation of the performance of the presented method has been conducted on our created database using Leave-One-Out Cross Validation (LOOCV) method. According to the structure of the database (Subsection 4.1) ten experiments have been carried out for each participant. The results of these experiments have been computed in terms of three error rates: False Rejection Rate (FRR) that is the ratio of the number of genuine test signatures rejected to the total number of genuine test signatures, False Acceptance Rate (FAR) that is the ratio of the number of accepted forgeries to the total number of forgeries, and also Average Error Rate (AER), which is the average of the FAR and FRR. Some of obtained results are presented in Table II, and also shown graphically in Figure 3. The overall performance of the presented method is reported as the true error rate which is equal to the average of error rates of all conducted experiments.

The influence of the number of genuine samples considered for training the similarities and variations which may be existed between signatures of each participant has been studied during some experiments as follows. For each participant, two (out of ten) samples of genuine signatures have been chosen randomly and let as genuine testing samples and considered constant during experiments related to the same participant. Now the influence of the number of training samples is examined using other eight genuine signatures by considering 4, 5, ..., 8 of them as training samples. Obtained results of these experiments are shown in Table III, and Figure 4. As shown, the values of FRR, and FAR approximately constant with slight decreasing till by considering eight signatures as training samples the values of FRR, and FAR obviously decrease. The last experiment has been conducted to compare performance of the presented method with method which extracted crossing counts, and curvature features as classical features, from each signature image. Obtained results which are shown in Table IV, confirm superiority of our method. Table V presents the results of some related works. Since each of these works used own database with different quality specially in forged signatures, the comparison between these results is a difficult task and not fair. However, according to the number of signatures in

Table II
SOME OF OBTAINED EXPERIMENTAL RESULTS.

No. of participant	FRR (%)	FAR (%)	AER (%)
1	50	0	25
2	20	50	35
3	20	0	10
4	20	40	30
5	10	0	5
⋮	⋮	⋮	⋮
100	0	0	0
True error rate	17.5	11	14.25

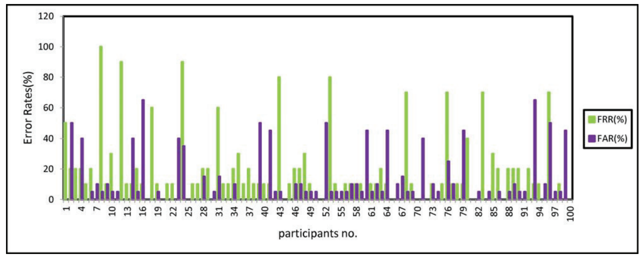

Figure 3. Obtained error rates of conducted experiments related to each participant in our database.

Table III
PERFORMANCE COMPARISON OF OUR METHOD USING DIFFERENT NUMBERS OF GENUINE SIGNATURES AS TRAINING SAMPLES.

No. of training samples	FRR (%)	FAR (%)	AER (%)
4	18	15	16.5
5	18.5	15	16.75
6	18	14	16
7	18	15	13.5
8	17	12.5	14.75

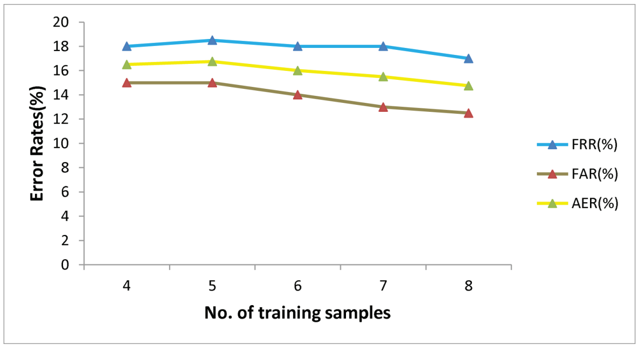

Figure 4. Graph of the performing the presented method using different number of training samples.

Table IV
PERFORMANCE COMPARISON OF OUR METHOD WITH CLASSICAL METHOD.

Method	FRR (%)	FAR (%)	AER (%)
Classical(local)	18	19.5	18.75
Katz(global)	17.5	11	14.25

our database and using only a vector of global features with low computational complexity compared with other methods shown in Table V, obtained results confirm the effectiveness of our presented method especially in reducing FAR which is very important in banking system, for example. Surely by reinforcing of each step of our method, more accurate results will be obtained.

V. CONCLUSION AND FUTURE WORKS

A new Persian signature verification method has been proposed in this paper. Feature extraction step is conducted using estimated FD of each signature image. We used testing hypothesis for making decision about state of signature in question (acceptance as genuine signature or rejection as skilled forgery) which doesn't have been used in process of verifying Persian signatures till today. Compared other similar works, our method leads to higher accuracy in FAR which confirms effectiveness of the method for its application in bank checks processing and other applications in security systems. In future, we are going to improve the performance of our method using modifying each of its steps and also by combining the method with other effective methods to give more and more accurate results. Also the independence of the presented method from the shape and style of signatures will be verified by testing on database of different nationalities in future with us.

Table V
COMPARISON OF THE PRESENTED METHOD WITH SOME RELATED WORKS.

Method	Database	FRR (%)	FAR (%)	AER (%)
Kiani (2009) [5]	600	4	17	10.5
Zoghi (2009) [4]	2000	7.5	17.3	12.4
Falahati (2011) [21]	200	8	13	10.5
Sigari (2011) [6]	600	15	15	15
Our Method	**1200**	**17.5**	**11**	**14.25**

REFERENCES

[1] A. Chalechale, and A. Mertins, *Line segment distribution of sketches for Persian signature recognition*, Conf. Convergent Technologies for Asia-Pacific Region, vol. 1, pp.11-15, 2003.

[2] D. Impedovo, and G. Pirlo, *Automatic signature verification: The State of the Art*, IEEE Trans. vol. 38(5), pp. 609-635, 2008.

[3] S. Pal, M. Blumenstein, and U. Pal, *Automatic off-Line signature verification systems: A Review*, ICWET. vol.14, pp.20-27, 2011.

[4] M. Zoghi, and V. Abolghasemi, *Persian signature verification using improved Dynamic Time Warping-based segmentation and Multivariate Autoregressive modeling*, 15th. Workshop on SSP. pp.329-332, 2009.

[5] V. Kiani, R. Pourreza, and H. Pourreza, *Off-line signature verification using local radon transform and support vector machines*, IJIP. vol. 3(5), pp. 438-443, 2009.

[6] M. H. Sigari, M. R. Pourshahabi, and H. R. Pourreza, *Off-line handwritten signature identification and verification using multi-resolution gabor wavelet*, IJBB. vol. 5, pp.1-15, 2011.

[7] S. J. Elliott, *Development of a Biometric Testing Protocol for Dynamic Signature Verification*, 7th ICARCV. vol. 2, pp. 782-787, 2002.

[8] H. Srinivasan, S. N. Srihari, and M. Beal, *Signature verification using kolmogorov-smirnov statistic*, Proc. Int. Graphonomics Society Conf. (IGS), pp.152-156, 2005.

[9] S. Spacic, A. Kalauzi, G. Grbic, L. Martac, and M. Culic, *Fractal analysis of rat brain activity after injury*, Med Biol. Eng. Comput. vol. 43, pp.345-348, 2005.

[10] G. E. Polychronaki, P. Ktonas, S. Gatzonis, A. Siatouni, P. A. Asvestas, H. Tsekou, D. Sakas, and K. S. Nikita, *Comparison of fractal dimension estimation Algorithms for epileptic seizure onset detection*, J. Neoral Eng. vol. 7, pp.1-6, 2010.

[11] M. J. Katz, *Fractals and the analysis of waveforms*, Computers in Biology and Medicine, vol. 18(3), pp.145-156, 1988.

[12] B. S. Raghavendra, and D. Narayana Dutt, *Computing fractal dimension of signals using Multiresolution Box-counting method*, Int. J. Info. Mathematical Sci. vol. 6, pp.50-65, 2010.

[13] http://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture14.pdf, Accessed 15, Feb. 2012.

[14] http://stattrek.com/help/glossary.aspx?target= significance _ level, Accessed 25, Feb. 2012.

[15] M. Kalera, S. Srihari, and A. Xu, *Offline Signature Verification and Identification Using Distance Statistics*, IJPRAI. vol. 18(7), pp.1339-1360, 2004.

[16] M. Cheriet, N. Kharma, C. Y. Suen, et al. *Character recognition systems a guide for students and practitioners*, *John Wiley, and Sons Inc.* 2007.

[17] T. Peter, *The Radon Transform - Theory and Implementation*, PhD thesis, Dept. of Mathematical Modelling Section for DSP of Technical University of Denmark.

[18] M. Radmehr, S. M. Anisheh, M. Nikpour, and A. Yaseri, *Designing an off-line method for signature recognition*, World Applied Sci. J. vol. 13(3), pp. 438-443, 2011.

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C: *The art of scientific computation*, Cambridge University Press, 1988-1992.

[20] C. Aggarwal, *Towards systematic design of distance functions for data mining applications*, In Proc. 9th Intl. Conf. on Knowledge Discovery and Data Mining (KDD), pp. 9-19, 2003.

[21] D. Falahati, M. S. Helforoush, H. Danyali, and M. Rashidpour, *Static signature verification for Farsi and Arabic signatures using dynamic time warping*, 19th. ICEE. pp. 17-19, 2011.