Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/3891
DOI URL: http://dx.doi.org/10.21474/IJAR01/3891

RESEARCH ARTICLE

ONLINE HANDWRITTEN SIGNATURE RECOGNITION BY PRINCIPAL COMPONENTS AND SUPPORT VECTOR MACHINE.

Fahad Layth Malallah[1], Zeyad T. Sharef [2], Kameran Hama Farj[1] and Zaid Ahmed Aljawaryy[3].
1. Department of computer Science, Cihan University, Sulaimaniya, Iraq.
2. College of Engineering, Ahlia University, Manama, Bahrain.
3. Faculty of Science and Technology University of Human Development, Sulaimaniya, Kurdistan Region, Iraq.

Manuscript Info

Manuscript History

Received: 08 February 2017
Final Accepted: 04 March 2017
Published: April 2017

Key words:-
Biometric, Authentication, Online Handwritten Signature, Principle Components Analysis, Support Vector Machine.

Abstract

With the rapid development of capture devices such as smart phone and tablets, there is a big trend towards online handwritten signature applications being used as behavioral biometrics. Online handwritten signature encounters difficulty in the verification process because an individual rarely signs exactly the same signature sample whenever he/she signs, which is referred to as intra-user variability. This paper presents a new technique for handwritten signature verification. The operation starts by normalizing the signatures samples to similar lengths of enrolled and authenticated samples without affecting to the signature shape. And then, Principal Component Analysis (PCA) is exploited for features' extraction and Support Vector Machine is utilized as classification operation. The experiment has been conducted on a SIGMA database on 200 users that comprises more than 6000 online handwritten signature samples, the result demonstrated 96%as successful recognition rate.

Introduction:-

Biometric system is deemed as pattern-recognition system that recognizes a user based on features extracted from behavioral or physiological descriptions that belongs to that user [1]. Two main modes of a biometric system are available nowadays [2]. First one is the identification Mode, which means matching the target biometric data with all the data available in the system, or meaning of this question “Who are you?”. The second mode of biometric system is the Verification Mode, which is meant by this question: “Are you who you claim to be?”. Here, the target biometric data is matched with the specific reference in the system to authenticate its identity [3].

Handwritten signature normally logically is comprises of the first and last name of someone. This type of signature is referred to as a paraph [4]. Signature can be defined as a behavioral type of biometrics that has a high legal value for document authentication. Moreover, it acts as a non-invasive and non-intrusive authentication process for the majority of the users, It is one of the most accepted biometrics, since most individuals have their own signatures that could be used as their own token [5]. The obstacle, which undermines the use of this type of biometric, is having property of the high intra-user variability. This property happens because individuals cannot originate a signature that is exactly the same as one of the previous versions. Another thing is that handwritten signature can be forged without using specialized hardware. Therefore, skilled forged signatures must be considered in the testing. Two types of Signature authentication named static or dynamic verification. The static is referred to as offline signature

Corresponding Author:- Fahad Layth Malallah.
Address:- Department of computer Science, Cihan University, Sulaimaniya, Iraq.

verification that performs user verification using scanned signature images. About dynamic verification is referred to as online signature verification system (such as this paper work) where signature samples are captured digitally usually by using digitized pen and graphical tablets. Here, a richer amount of information is captured, which often includes signals as a time series of $x[t]$ and $y[t]$ coordinates with pen pressure $p[t]$. Anyhow, achieving high correct matching accuracy in signature verification is not trivial task due to the high intra-user variability, which will increase the False Reject Rate (FRR).

This paper is organized as follows. Section II is dedicated for literature review related to signature verification. In section III, the signature verification methodology is proposed. In Section IV, the experiment and implementation are described and finally, in Section V, the conclusion is presented with future work.

Literature Review:-

As it has been noticed in the literature, online handwritten signature verification consists of four phases: data input, pre-processing (such as using signature length normalization in this paper), feature extraction and classification [6]. Usually, online signature samples are input by using tablets or Personal Digital Assistant (PDA) to capture the signature data. In preprocessing phase, some techniques that are adapted from signal processing algorithm are used. The advantage of pre-processing is to improve the input data in order to get a better recognition rate. types of preprocessing techniques are filtering, noise smoothing reduction, signature re-sampling. Such as online handwritten signature recognition by length normalization using Up-Sampling and Down-Sampling is proposed in [7], here normalization is based on interpolation operation for those samples which are less than set threshold length and down-sampling for those signature samples for more than threshold. It is essential to mention that proposed signature verification is depended on this normalization as in [7].

In the feature extraction phase, two types of features can be used, which are function features and parameter features. In function features the signature is characterized as a time series signals [7], for instance, horizontal signal *x[t]* and vertical signal y[t] for positions, velocity signal, acceleration signal, pen pressure signal, and pen inclination signal. For the second type parameter features, the signature is a form of element vector that consists of a statistical and mathematical computation based on the acquired signature data, for instance signature time duration, number of pen up / pen down ,pen down ration, MAX/ MIN of positions, speed, and acceleration [8]. Generally, function features have better performance in compared to parameter features. In terms of the last phase, the classifier is used to differentiate based on extracted feature and then make a decision.

Normally, signature verification can be implemented using statistical or template matching approaches. In the case of template matching techniques, a queried sample is matched against templates of authentic / forgery signatures [8]. In this case, the most common approach used is the Dynamic Time Warping (DTW) technique [9]. In the case of statistical approaches, distance-based classifiers can be used to perform signature verification. For instance, Artificial Neural Networks (ANNs) are highly used for signature verification, because of their capabilities in learning and generalizing as in [10]. Hidden Markov Models (HMMs) is also used for online signature verification as in[11], Support Vector Machine (SVM) [8], Bayesian decision method [12].

Signature Verification Methodology:-

In this research, the dataset as online signature sample consists of time series signals of horizontal $x[t]$ and vertical $y[t]$ coordinates, with pen pressure $p[t]$ sampled at time *t*. the operation starts by doing normalization operation to the all signature samples according to the down-sampling and up-sampling operation to make all the signature samples as a similar length, the complete work is detailed in [7]. In this experiment, the length is chosen to be 256 trajectories as the average length of the SIGMA database [13] samples. the is applied to the length into our system for normalization and then verification. The second stage is feature extraction which is using Principal Component Analysis (PCA) features and thirdly, Support Vector Machine as a classifier. The operation of online signature verification is depicted as a diagram in Figure1. The verification operation starts by reading the $x[t]$, $y[t]$ and $p[t]$ signals, which are the horizontal trajectories, vertical trajectories and pen pressure respectively, where $t = 1,2....\overline{N}$, and $\overline{N}$ is the desired signature length, which is set to 256. Extracted the features, which is implemented using PCA, are stored in the database as a reference model to be used in the prospective matching with anyone who wants to verify the signature sample.

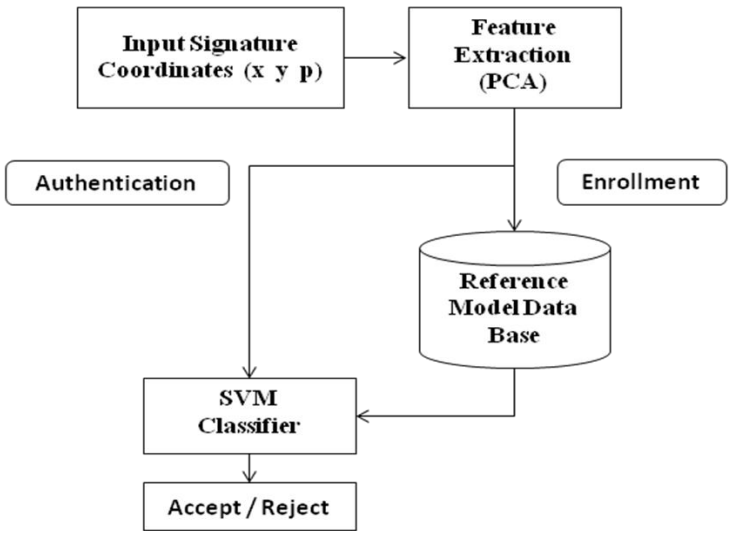

Figure 1:- Proposed Signature verification as biometric system.

In the authentication process, the queried identity signature will be read by the system. The same processes that have happened during the enrollment operation should also be applied to the queried signature sample. Figure 1 shows the main diagram of the proposed verification system. The signature verification system consists of two separate processes which are feature extraction by using Principal Components Analysis (PCA) and classification by using Support Vector Machine (SVM) as detailed in the following sub-sections.

Feature Extraction (PCA):-

Feature extraction operation is used to transform the signature signals from its original domain to another domain to increase the variance and reduce the correlation among individuals. In this research Principal Component Analysis (PCA) is used to improve the recognition rate [14]. PCA has ability to transform a data set (signatures in this case) from correlated domain to another domain that is highly uncorrelated among the original data set [14]. In other words, PCA has ability to do variance maximizing between genuine and forged signature samples. In this paper, PCA is run on three columns of the input online signature and the results of the PCA operation are also three columns, the first column corresponds to the first Eigen vector component that belongs to the highest Eigen value. The second column is the second highest Eigen value and the third column is the lowest Eigen value. The length of each column is 256 features, which is the length of the signature after normalization. The proposed feature vector is comprised by combining the three component vectors of PCA into a single vector to represent a signature sample feature vector. After that, feature selection operation from PCA output is arranged as this procedure: the selection is implemented on each component vector by dividing the vector (256) into 8 segments (seg_xx); each segment size is 32 features. The selection is done by taking the 1st (seg_11), 4th (seg_14) and 8th (seg_18) segments among the 8 segments from the first component vector. The same goes for the second and third columns. Finally, the length of each signature represented vector is 300 features.

Classification (SVM):-

SVM is very beneficial for tow classes, which classifies data by finding the best hyper-plane that separates all data points of one class from those of the other class. The best hyper-plane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyper-plane that has no interior data points. In the Figure 2, the support vectors are the data points that are closest to the separating hyper-plane, these points that are on the boundary of the slab. Figure 2 illustrates these definition, with "+" label indicating data point of type 1 and "–" labels indicating data points of type -1 [15].

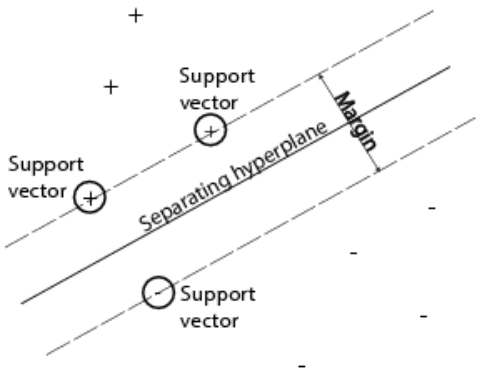

Figure 2:- SVM diagram explained hyper-plane.

Once the length of signature represented vector is ready, it will be stored in the data base for the future reference model to be used for SVM training dataset. Besides that, these signature vectors are matched later on against queried signature vectors, which have undergone the same operations as in the enrollment phase. The task of Support Vector Machine (SVM) is used to match between the stored and queried signature sample vector, in order to make a decision whether the signature is guanine or forge. The specification of SVM system, which has been implemented in this paper, are as follows: the kernel function of the used SVM, which is used to map the training dataset into kernel space, is linear type which is also named dot product, as well as the method of finding the separating hyper-plane is least squares method.

Experiment and Result:-

The implementation of the experiment is done on a SIGMA database which has more than 6000 online handwritten signature samples to test the verification accuracy of the proposed method. The steps of the experiment are as follows:

1. The training matrix is built by using signatures from the SIGMA database [13], the. The training matrix is comprised of signatures from 200 individuals. Each individual has 10 genuine and 10 forged samples (five of them are random forged and the other five are skilled forged samples). Each signature sample is represented by 300 features. Therefore, the training matrix size is $[300\times 20]$ (300 features for each sample with 20 samples for each individual). Training is run by SVM for the signatures of each individual.
2. The result produced by SVM is done by extracting the False Accept Rate (FAR) and the False Reject Rate (FRR) for each user separately. The testing matrix is built as the same as to the way of the training matrix that was built.
3. A label $+1$ in the training target (destination) of SVM is assigned to the first 10 signature samples of the trained matrix indicating to genuine for the first 10 samples, while -1 is assigned to the second 10 signature samples of the training matrix to mark and train the SVM that the second 10 are forged samples.
4. FRR is computed by evaluating the result scores of the first 10 samples. If any sign of the first 10 samples is less than the threshold (set to 0), False Rejection (FR) counter will be increased by one $(FR = FR + 1)$, because they are supposed to be as accepted (signs are larger than threshold) but they are wrongly rejected by the verifying system. Also, if the results of the second 10 samples have labels more than the threshold, they are deemed as False Accept (FA) and the counter will be incremented by one $(FA = FA + 1)$. The FAR and FRR are computed as in (1) and (2) respectively:

$$FAR = \frac{FA}{10} \times 100\% \tag{1}$$

$$FRR = \frac{FR}{10} \times 100\% \tag{2}$$

5. The accuracy of each individual has been computed using (3):

$$Accuracy\% = 100\% - \frac{FAR\% + FRR\%}{2} \tag{3}$$

6. An average of the 200 individuals' accuracy is computed by using (4) so as to take into consideration all individuals in the SIGMA database:

$$AVR\_Accuracy\% = \frac{1}{200}\sum_{n=1}^{200} user[n] \quad (4)$$

About the result of the experiment, Table 1 lists FAR, FRR and average error.

Table 1: Verification accuracies as an error rate (FAR and FRR)

Threshold	FRR%	FAR%	Average Error%
0	2.5%	1.5%	2%

The table lists the average errors in terms of zero threshold, which is the borderline between FAR and FRR. As shown in table 1, in the case of 0 threshold, the average error rate is 2% resulted from the FRR is 2.5% and FAR as 1.5%. In other words, the performance as a successful rate is 98% of 200 users of SIGMA database by using PCA and SVM system.

Conclusion:-

Intra-user variability is the major obstacle of the online handwritten signature verification as the same user cannot sign the same signature of the previous signature. In this paper, an acceptable recognition rate has been achieved as 98% as a result of the experiment that has been conducted on a SIGMA database for online handwritten signature which comprises more than 6000 signature samples. In this paper the verification operation is kicked off by doing normalization of the signature length as 256 trajectory points for the three time series signals *x[t], y[t] and p[t]*. Then, feature extraction operation is done by using principal components analysis (PCA) , as well as feature selection is done for preparing the feature vector to be input to the SVM classifier.

References:-

1. N. K. Ratha, *et al.*, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, pp. 614-634, 2001.
2. S. G. Kanade, *et al.*, "Cancelable biometrics for better security and privacy in biometric systems," in *International Conference on Advances in Computing and Communications*, 2011, pp. 20-34.
3. K. Radhika and S. Sheela, "Fundamentals of Biometrics—Hand Written Signature and Iris," in *Pattern Recognition, Machine Intelligence and Biometrics*, ed: Springer, 2011, pp. 733-783.
4. B. Miroslav, *et al.*, "Basic on-line handwritten signature features for personal biometric authentication," in *MIPRO, 2011 Proceedings of the 34th International Convention*, 2011, pp. 1458-1463.
5. E. Maiorana, *et al.*, "Cancelable templates for sequence-based biometrics with application to on-line signature recognition," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, pp. 525-538, 2010.
6. Z. Zhang, *et al.*, "A survey of on-line signature verification," in *Chinese Conference on Biometric Recognition*, 2011, pp. 141-149.
7. F. L. Malallah, *et al.*, "Online handwritten signature recognition by length normalization using up-sampling and down-sampling," *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, vol. 4, pp. 302-313, 2015.
8. D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, pp. 609-635, 2008.
9. A. G. Reza, *et al.*, "An efficient online signature verification scheme using dynamic programming of string matching," in *International Conference on Hybrid Information Technology*, 2011, pp. 590-597.
10. M. Alhaddad, *et al.*, "Online signature verification using probablistic modeling and neural network," in *Engineering and Technology (S-CET), 2012 Spring Congress on*, 2012, pp. 1-5.
11. M. R. Freire, "Biometric template protection in dynamic signature verification," MS thesis, Universidad Antonio de Nebrija, Madrid, Spain, 2008.
12. A. Kholmatov and B. Yanikoglu, "Identity authentication using improved online signature verification method," *Pattern recognition letters*, vol. 26, pp. 2400-2408, 2005.
13. S. M. S. Ahmad, *et al.*, "SIGMA-A Malaysian signatures' database," in *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, 2008, pp. 919-920.
14. C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, pp. 1-58, 2006.
15. J. Friedman, *et al.*, *The elements of statistical learning* vol. 1: Springer series in statistics Springer, Berlin, 2001.