Classification of Features into Strong and Weak Features for an Intelligent Online Signature Verification System

Saad Tariq, Saqib Sarwar & Waqar Hussain
Department of Electrical Engineering
Air University
Islamabad, Pakistan
tariq.saad@live.com

Abstract—**This paper presents an efficient algorithm for the classification of features into strong and weak features for every distinct subject to create an intelligent online signature verification system. Whereas Euclidean distance classifier is used for validation processes and low error rates obtained illustrate the feasibility of the algorithm for an online signature verification system.**

Keywords-Signature Biometrics; intellegient signature verification ; online signature verification; classification of features; strong features; weak features; dynamic signature verification; euclidean distance classifer

I. INTRODUCTION

Today, with the astonishing growth of the Internet and Intranet, E-commerce and E-finance become the hottest topics on this planet. Doing business through the public network makes personal identification data security more and more crucial as well. How to protect the private identification from being pirated is the key issue that the Internet and intranet clients would be concerned with before such E-business could be widely accepted since authentication has become an essential part of highly computerized services and/or security-sensitive installations in modern society.

Signature verification fulfills all the above described circumstances and can play a vital role in protection and personal identification as it is a popular means of endorsement historically. Although such signatures are never the same for the same person at diverse times, there appears to be no practical problem for human beings to discriminate visually the real signature from the forged one. It will be extremely useful when an electronic device can display at least the same virtuosity.

Signature verification systems are usually built following either on-line or off-line approaches, depending on the kind of data and application involved. On-line systems generally present a better performance than the off-line system but require the necessary presence of the author during both the acquisition of the reference data and the verification process limiting its use. In online signature verification systems, additional features such as pen pressure, pen speed and pen tilt angle have made the process of forging online signatures more difficult. Equal error rate of available online signature verification systems lies between 1 to 10%.

Still a lot of work is needed to be done to reduce Equal error rate (EER) to make online signature verification the most secure way of personal identification.

II. FEATURE EXTRACTION

Feature extraction phase is one of the crucial phases of an on-line signature verification system. The discriminative power of the features and their flexibility to the variation within the reference signatures of a writer, play one of the major roles in the whole verification process. While features related to the signature shape are not dependent on the data acquisition device, presence of dynamic features, such as pressure at the pen-tip or pen-tilt, depends on the hardware used.

Features may be classified as global or local, where global features identify signature's properties as a whole and local ones correspond to properties specific to a sampling point. For example, signature bounding box, average signing speed, trajectory length or are global features, and Local features include curvature change between consecutive points on the signature trajectory or distance are local features. Features may also be classified as temporal (related to the dynamics) and spatial (related to the shape).

These features can be referred as human traits, as they can vary from person to person and can be classified as strong or weak for every distinct individual. If we make a list of these features, more than 100 features are present and even new features can be derived depending on their discriminative power.

III. Database & Compilation

A. *System*

For the purpose of signature verification we made an experimental setup in which a person is enrolled in the database by taking some of his/her signatures and a template is created and stored against the name and ID of the specified person. A new signature from that person can then be checked against the enrolled template to validate the person. Furthermore we will discuss about the technique used in our system, database and how we optimized features as strong and weak features.

B. *Database Completion*

A comprehensive database was created by obtaining the signatures from the students. Signatures were gathered from a total of hundred subjects with ten signatures from each subject. So a total of thousand signatures were collected to create the original signature database. WACOM INtuous4 tablet with a sampling rate of 200 samples per seconds was used for this purpose.

To form the forgeries database we performed a total 10 forgeries per person, among which were five zero-effort forgeries and five skilled forgeries. The forgeries that are performed by first training the counterfeiter to copy the precise dynamics of the original signer are skilled forgeries. A forger is trained by showing him plots of the original signature being performed or by training the original signer himself.

IV. Signature Verification Technique

In the first phase, a signature verification technique was successfully put into operation for the classification of original and forged signatures using Euclidean Classifier. The technique is previously implemented by H. Dullink, B. Van Daalen, J. Nijhuis, L. Spaanenburg, and H. Zuuidhof [1].

A. *No Pre-Processing*

The technique we implemented did not use any preprocessing because the tablet used had a sampling rate of 200 samples per second. Therefore it was not essential to smooth or normalize the signature datasets, which were required if we had used the signatures collected from a tablets with low resolution. Re-sampling and resizing was also skipped considering the fact that valuable data is lost while pre-processing the data.

B. *Feature Extraction*

Among the list of features that can be extracted a total of 26 features were extracted. The features extracted were **standard deviation of x-acceleration, standard deviation of y-acceleration, average pressure, standard deviation of x-velocity, standard deviation of y-velocity, number of pen-up samples, pen down time/total time taken, standard deviation of y / change in y, pen down time, RMS velocity / maximum velocity, average jerk, jerk RMS, maximum sample point x-coordinate, maximum sample point of y-coordinate, zeros of x-velocity, standard deviation of x-coordinates, standard deviation of y-coordinates, total number of samples, time taken, length, zero crossings of x-velocity, zero crossings of y-velocity, zero crossings of x-acceleration, zero crossings of y-acceleration, zeros in x-acceleration, zeros in y-acceleration.**
A pressure sensitive tablet was used that records pressure at every sample taken, providing with a very strong local feature of pressure.

C. *Optimization & Experimental Setup*

Here is an important discussion that how we opted only 9 features out of those 26 features for our system. As we know that a large number of features have been proposed by researchers for online signature verification [2], [3], [4]. However, a little work has been done in measuring the consistency and discriminative power of these features [5], [6]. On the basis of consistency and discriminative power features can be divided into strong and weak features, where presence of the strong features decreases the FRR while on the other hand presence of some weak features also decreases FRR but increases FAR. Thus there is a need to select the best features set.

The approach we used for classification of strong and weak features is by using difference between mean to standard deviation ratio of each feature from the feature vector and from the forgeries features vector set. Thus the mean/standard-deviation difference of each feature from the template of 100 subjects was taken. The standard deviation of a feature shows how large a deviation from the enrolled template can be tolerated (i.e. large deviated signature could be classified as true for large standard deviation).

$$C = \sqrt{\left(\left(\frac{Mo}{STDo}\right) - \left(\frac{Mf}{STDf}\right)\right)^2} \qquad (1)$$

In (1), Mo/STDo is the mean/standard-deviation ratio of the feature of original signatures and Mf/STDf is the mean/standard-deviation ratio of the feature of forgery signature. The features with large value of mean/standard-deviation difference as compared to others were taken as strong features and others as weak features eliminating which results in considerable good results.

A number of original signature’s features have a large mean/standard-deviation ratio and of course it will decrease FRR but contrary to it forgery signature’s features having a large mean/standard-deviation will decrease FAR. So therefore to obtain best results we took the difference between the original signature and forgery signature.

D. *Optimization Results*

As computed using (1) nearly 14 features have greater C than other 16 features. As researchers have discussed earlier that too many features may decrease FRR but increase FAR

[7] therefore we have to choose between the best of them. The 14 features with greater C are **standard deviation in y-velocity, total samples, number of zeros in y-acceleration, number of zeros in x-acceleration, zero crossings in x-acceleration, zero crossings in y-acceleration, zero crossings in x-velocity, zero crossings in y-velocity, length, average pressure, total time, number of zeros in y-velocity, number of zeros in x-velocity and pen-down time.**

TABLE I. CALCULATIONS OF EQUATION (1)

Feature	Mo/STDo	Mf/STDf	C
Std Dev y/Δy	-4.8766	-1.9296	2.9
T(pen-down)/T(total)	23.3710	17.6752	5.4
N (pen-ups)	3.8719	0.8551	2.95
Standard Deviation vy	25.2692	13.9054	12.3
Standard Deviation vx	2.8116	1.9122	0.9
N(vy=0)	5.8355	1.2074	4.6
Average v/v(max.)	5.7595	3.3267	2.45
(x1-xmin)/average x	4.5197	2.8109	1.7
Total Samples	15.9329	2.1116	13.79
(x1-xmax)/average x	-7.4158	-8.2712	0.8
N(max. y)	15.9590	17.7610	1.81
Standard Deviation of ay	3.1448	4.0654	0.92
Standard Deviation of ax	1.6747	2.1500	0.48
Number of zeros in ay	7.7817	1.0288	6.78
Number of zeros in ax	8.5880	1.2653	7.30
Zero cross. X-acceleration	9.0654	1.3230	7.68
Zero cross. Y-acceleration	9.6669	1.2263	8.44
Zero cross. X-velocity	12.8354	1.5204	11.31
Zero cross. Y-velocity	13.5760	1.2228	12.35
Length	7.5981	1.7094	5.89
rms jerk	2.6554	1.9491	0.71
average jerk	2.7470	2.4410	0.26
N(max. x)	15.4440	13.6379	1.81
Average Pressure	12.1289	2.2516	9.87
Total Time	15.9329	2.1116	13.82
Number of zeros in vy	8.8355	1.2074	7.63
Number of zeros in vx	8.5746	1.2781	7.30
(y1-ymax)/average y	-3.9525	-3.1871	0.77
(x1-xmin)/average x	8.0218	5.4126	2.62
Pen-down Time	29.7766	2.9390	26.87

Highlighted features are with greater results

Features such as total time, pen-down time and total samples are all time dependent features so therefore for a versatile verification engine we opted total time to be the best among them. Moreover standard deviation of y-velocity is another feature having a greater result but on the standard deviation of x-velocity has a very small difference, therefore this ambiguous result made us step down with these features too.

V. INTELLIGENT ONLINE SIGNATUARE VERIFICATION

The experimental setup and optimization proposed above gave very good results but still as we have discussed earlier that signature and its features are personal traits and they may vary person to person. Thus to make this system efficient and intelligent we made it route person to person. As we had a list of 9 most efficient features, we decided to choose 5 out it but based on subject itself. These 5 features may vary person to person. While recording a template from a subject all these features were stored in the template but at the time of verification we proposed a system in which only 5 features were compared against its template based on the following results.

$$\mathbf{X = C/\ Vx\ \ -\ STDf} \qquad (2)$$

Where C is the difference between the **mean/standard-deviation** ratio of the feature of original signatures and the **mean/standard-deviation** ratio of the feature of forgery signature from (1) which is already calculated and Vx is current value of the sample and STDf is the standard deviation of the forgery signature already stored. So among the 9 features, only 5 features are opted which have a greater value of X from (2).

A. Comparison

For comparison we need a reference. So for the enrollment process we selected 5 original signatures from each signature extracted the 9 features described above to create a reference template. The template contains the mean, standard deviations and their difference stored in 3 vectors R, S and C respectively. If we want to compare a signature (original or forged) with the template we will first compute the feature vector of that signature and corresponding vector X using (2). Then the greater 5 features depending on the value of X will be stored in a vector T. To compare the signature we will simply opt out those 5 features from R and S and a distance vector D will be computed using Euclidean classifier.

$$\mathbf{D = R - T} \qquad (3)$$

Then the distance vector V will be normalized by dividing each value by the corresponding standard deviation in the vector S to obtain a vector Z whose mean is then computed and finally the computed norm is compared to a pre-defined threshold.

B. Results

Results for FRR, FAR of the template of 5 signatures of 100 subjects were computed with threshold from 4 to 9 for this intelligent online signature verification system and best results were obtained.

TABLE II. CALCULATIONS OF FFR AND FAR (1)

Threshold	FRR	FAR
4	**11.57%**	**0.72%**
5	**11.20%**	**3.92%**
6	**4.53%**	**8.02%**
7	**2.06%**	**13.62%**
8	**1.13%**	**19.89%**
9	**0.66%**	**27.02%**

Results obtained from our implementation are very better than a number of techniques implemented because we used very strong features and an intelligent system to classify them person to person. Anyways more work can be done on this system to make it more efficient by using other classifiers and updating signature over time with tablets with better sampling rates.

REFERENCES

[1] H. Dullink, B. van Daalen, J. Nijhuis, L. Spaanenburg and H. Zuidhof, *Implementing a DSP Kernel for Online Dynamic Handwritten Signature Verification Using the TMS320 DSP Family*, EFRIE, France December 1995, SPRA 304.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Charles E. Pippin, *Dynamic Signature Verification using Local and Global Features*, Georgia Institute of Technology, July 2004.

[3] T. S. Tolba, *A Virtual-Reality-Based System for Dynamic Signature Verification*, Digital Signal Processing Vol. 9, pp. 241-266, 1999. (article available online at http://www.idealibrary.com)

[4] V. S. Nalwa, *Automatic On-Line Signature Verification*, Proceedings of IEEE, vol. 85, pp. 215-239, 1997.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[5] Hao Feng and Chan Choong Wah, *Online signature verification using a new extreme points warping technique*, PRL(24), No. 16, pp. 2943-2951, December 2003.

[6] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.

[7] G. Lorette and R. Plamondon, *Dynamic approaches to handwritten signature verification*, Computer Processing of handwriting, World Scientific, 1990, 21-47.