

Statistical Approach for Offline Handwritten Signature Verification

²Debnath Bhattacharyya, ¹Samir Kumar Bandyopadhyay, ²Poulami Das,
²Debashis Ganguly and ²Swarnendu Mukherjee

¹Department of Computer Science and Engineering, University of Calcutta,
Rajabazar Science College, India

²Department of Computer Science and Engineering, Heritage Institute of Technology,
Kolkata, India

Abstract: Signatures were considered an important tool for authenticating the identity of human beings. So, signature verification was one of the biggest uses for that. We proposed an algorithmic approach for the verification of handwritten signatures by applying some statistical methods. The research work was based on the collection of set of signatures from which an average signature was obtained based on our algorithm and then taking decision of acceptance after analyzing the correlation in between the sample signature and the average signature.

Key words: Signatures, authenticating, statistical methods, acceptance, analyzing, correlation, sample signature and average signature

INTRODUCTION

Signatures are composed of special characters and flourishes and therefore most of the time they can be unreadable. Also intrapersonal variations and the differences make it necessary to analyze them as complete images and not as letters and words put together^[1]. As signatures are the primary mechanism both for authentication and authorization in legal transactions, the need for research in efficient automated solutions for signature recognition and verification has increased in recent years. Various methods have already been introduced in this field and application of Statistical models is one of them in this regard.

Using statistical knowledge, we can easily perform the relation, deviation, etc between two or more data items. Strictly speaking, to find out the relation between some set of data items we generally follow the concept of Correlation Coefficients. In general statistical usage, correlation or correlation refers to the departure of two variables from independence, although correlation does not imply causation. Our approach is based on the above concept. To verify an entered signature with the help an average signature, which is obtained from the set of, previously collected signatures, we have followed the concept of correlation to find out the amount of divergence in between them. Based on some

predefined divergence, we have taken the decision of acceptance of the entered signature.

RELATED WORKS

Handwritten Signature Identification both offline and online is a classical work area in the line of Computer Science and Technology since last two decades. Numerous approaches have been proposed for Handwritten Signature Identification, Recognition and Authentication systems.

Apart from the all, the approach which makes our attraction is the application of Artificial Neural Network at the time of identification. An Artificial Neural Network is trained to identify patterns among different supplied handwriting samples. Handwritten signature samples are considered input for the artificial neural network model and typically weights also supplied for recognition.

In March, 2007, Debnath Bhattacharyya, Samir Kumar Bandyopadhyay and Poulami Das ^[12] have proposed a new recognition technique; an Artificial Neural Network is trained to identify patterns among different supplied handwriting samples. Handwritten signature samples are considered input for that artificial neural network model and typically weights also supplied for recognition^[2].

Another important method for offline signature verification is the application of Hidden Markov's rule. Justino, Bortolozzi and Sabourin proposed an off-line signature verification system using Hidden Markov Model^[3]. Hidden Markov Model (HMM) is one of the most widely used models for sequence analysis in signature verification. Handwritten signature is a sequence of vectors of values related to each point of signature in its trajectory. Therefore, a well chosen set of feature vectors for HMM could lead to the design of an efficient signature verification system.

Application of Support Vector machine (SVM) at the time of signature verification is also a new dimension in this field. Emre Ozgunduz, Tulin Şenturk and M. Elif Karşılığil has proposed an algorithmic approach according to which off-line signature verification and recognition can be done by Support Vector Machine^[4].

Statistical approach for offline handwritten signature verification also drives our attention towards it. In this method, generally using some statistical tool a deviation in between testing sample and the predefined samples is calculated and based on that value decision is taken up.

Debnath Bhattacharyya, Samir Kumar Bandyopadhyay and Deepsikha Chaudhury, 2007, proposed a scheme where the Standard Deviation for each byte of the Training Image Files (sample signatures) is computed and then each corresponding byte of Test Signature is compared to check whether it falls within the range of (Mean ± Standard Deviation). If 70% cases match, then the Test Signature is accepted^[5].

MATERIALS AND METHODS

The algorithm proposed in this paper is basically deals with the scheme of handwritten signature verification in an offline system, but it can also be realized as the verification method of extended online verification of signatures for forgery control and security purpose.

The algorithmic approach mentioned in the following sections has the flexibility of choosing the number of signatures, i.e., no_of_Sign for testing purpose to generate a signature as Avg_Sign containing the specialized mean features set from the test signatures set. After collecting the signatures for testing, the algorithm converts them into a set of 2D arrays of binary data values-0 and 1. From these binary arrays using statistical methods of calculating expected mean an average data set is calculated using the formula given by:

$$\mu_x = E(X) = (\sum X_i)/N$$

For extracting the mean of data values a constant value, named as avgSignThresholdValue, is compared with the actual mean to take the decision for selection between the binary values to be placed and this constant value is to be calculated after conducting a survey on a test data set of known features and as per the strictness in the security concern of the institute in concern. From the binary data value calculated as above the Avg_Sign is generated in turn following the same algorithm.

The algorithm for verification of handwritten signatures is required an additional input of Sample_Sign, which is to be tested for verification of acceptance or rejection. Here, another constant value is maintained, named as decisionValue, which is also calculated and set by professional statistician and security administrator deciding the security concern and policies of the organization after conducting surveys and testing over some sample experimental data set with already known results. The algorithm compares between two binary data sets obtained after analysis of Avg_Sign and Sample_Sign and calculates out the correlation coefficient, r_{xy} between them using the following statistical formula:

$$r_{x,y} = \frac{(\sum X_i Y_i - \sum X_i \sum Y_i)}{((\sum X_i^2 - (\sum X_i)^2)^{1/2} ((\sum Y_i^2 - (\sum Y_i)^2)^{1/2})}$$

In turn, r_{xy} is compared with the decisionValue and accordingly return TRUE or FALSE as for acceptance or rejection respectively.

Saohsv_avgpiccalc (No of sign, sign1_pic,..): This is the main function in our algorithm. This function will be used for the verification of the handwritten signature by comparing with a standard signature, which is obtained by applying statistical analysis on a set of signatures.

This function will take Number of total signatures and corresponding signatures as argument and finally it will output the decision for the acceptance of the given signature:

- Declare one 2-D array namely statDataInputSign [no_of_Sign] [indexRowMajor] and a 1-D array as statDataAvgSign [indexRowMajor], first one stored binary data value for the number of sample signs given input and second one holds binary data values depicting and describing desired AVG_SIGN
- Consider the sample sign given input as SIGN_i_PIC and analyze its pixels

- If the pixel value corresponds to colour white then corresponding data value will be zero and if the colour is black that sign or scratch is present then it will be 1 in statDataInputSign []
- Step back to Step-2 if any other sample test signs are available else goes to next step
- Now calculate the sum of binary values for each rowMajorIndex of all the signs and then divide each one by no_of_Sign to calculate expected mean
- Compare the mean for each index with the constant avgSignThresholdValue calculated as conducting survey, if mean is greater and equals to the constant values, then corresponding value in AVG_SIGN will be 1 else 0 which is to be stored in 1-D array of statDataAvgSign[] in row major fashion
- From the statDataAvgSign[] form new picture as AVG_SIGN in such a fashion that pixel will be of white colour if value corresponding to an index in array is zero and black if value is 1

Saohsv_decision (avg_sign, sample_sign): This module will be used for taking the decision whether the input signature matches with the average signature, which is obtained in the main function.

This function will take average signature and sample signature as arguments and finally it will output the decision for the acceptance of the given signature:

- Declare two 1-D arrays namely statDataAvgSign [indexRowMajor] and statDataSampleSign [indexRowMajor], where first one stores binary data corresponds to AVG_SIGN and second one for SAMPLE_SIGN
- Consider the sample sign given input as SAMPLE_SIGN and analyze its pixels
- If the pixel value corresponds to colour white then corresponding data value will be zero and if the colour is black that sign or scratch is present then it will be 1 in statDataSampleSign []
- Consider the average sign given input as AVG_SIGN and analyze its pixels
- If the pixel value corresponds to colour white then corresponding data value will be zero and if the colour is black that sign or scratch is present then it will be 1 in statDataAvgSign []
- Now calculate correlation coefficient, r_{xy} with bivariate data set x and y taken as statDataAvgSign [] and statDataSampleSign [] respectively
- Compare r_{xy} with a constant value decisionValue, calculated by statisticians and analyst keeping

security concern of organization. If former exceeds later then return TRUE else return FALSE

RESULTS AND DISCUSSIONS

The Algorithm stated in above section consists of two distinct divisions. a. Average Signature calculation, b. Decision: comparison between Average Signature and Sample Signature to find the correlation between them.

Complexity analysis of the stated algorithm: For conversion from bitmap pictures into 1D array: Considering the rowSize and columnSize as width and height of the pixel matrix of the bitmap pictures depicting the signatures, the total number of pixels, i.e., size of the pixel matrix is given by (row size×column size). So, for conversion from pixels to binary data as per suggested by algorithm, the time complexity required for each picture is given by O (row size×column size) \equiv O(sizeOfPixelMatrix) and thus total time complexity of conversion for no_of_Sign is given by O (no_of_Sign×row Size×column size).

For calculation of AVG_SIGN: For calculation of mean out of no_of_Sign number of pictures corresponding to the sample signatures total time complexity incurred is as follow:

$$O(\text{no_of_sign} \times \text{row size} \times \text{column size})$$

For SAOHSV_DECISION: In this module correlation coefficient is being calculated between two bivariate data sets of binary values corresponding to AVG_SIGN and SAMPLE_SIGN respectively. The time taken in this procedure is of order of O (row size×column×size).

Here in implementation of the above stated procedures through programs no extras variable spaces are required to be allocated in memory, so considering the space complexity of the algorithm, this is an inplace algorithm.

Test Results: In implementation of the algorithm proposed in above section database of 15 signatures for each of 100 different users is considered. Out of these 15 signatures, one Avg_Sign for each of the 100 users is calculated using the procedure described in Section III.A. Testing is done here with two trained forgery signatures and two true acceptable signatures, which are sent as Sample_Sign to the argument of the procedure of Section III.B.

The Array of 15 signatures for an arbitrarily chosen user from database is represented in the Fig. 1.

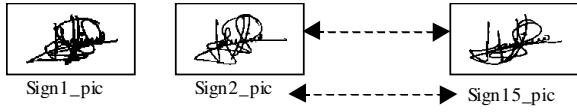


Fig. 1: Set of Sample Signatures

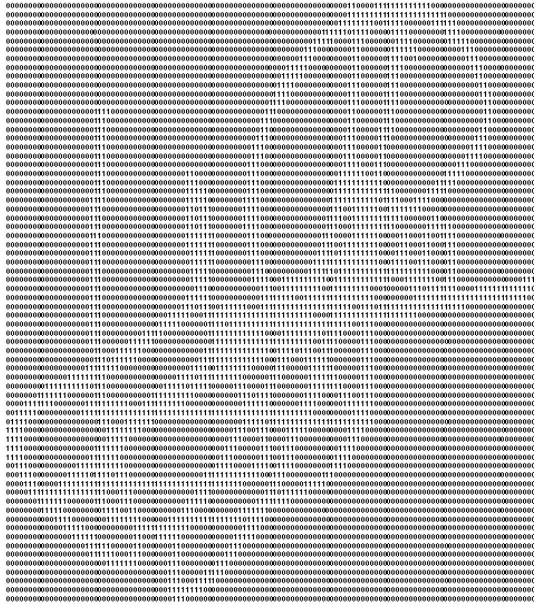


Fig. 2: Stat Data Avg Sign (After pixel by pixel transformation of input signatures into binary data values and calculating their mean)

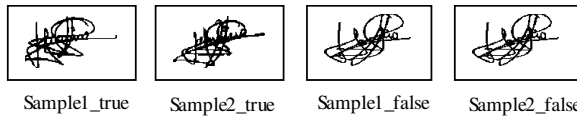


Fig. 3: Sample Signatures

Table 1 : Analyzed values for accepted sample signatures

Avg_sign	Sample_Sign	Correlation Coefficient, $r_{x,y}$
	Sample1_true	1.0000000000000000
	Sample2_true	0.870495723016037

(Acceptance: $r_{xy} > 0.50$)

The 1D array of binary data value corresponding to the Avg_Sign of the above taken user is same as in Fig. 2.

The Sample_Sign is taken as variable argument for the monochrome bitmap pictures, namely Sample1_true, Sample2_true, Sample1_false, Sample2_false and it is shown as in Fig. 3.

The test results of decision functions for the above given data set are shown in Table 1 and 2.

Table 2: Analyzed value and Rejected sample Signatures

Avg_sign	Sample_Sign	Correlation Coefficient, $r_{x,y}$
	Sample1_false	0.0812553839862523
	Sample2_false	0.0918058808861437

(Rejection: $r_{xy} < 0.50$)

CONCLUSION

Amongst the different biometric authentication schemes for security verification including voice detection, retina scan, fingerprint verification, handwritten signature verification, during monetary transactions and other security policies both on-line and off-line, is increasingly becoming popular. Our motivation behind this paper is to implement a simple statistical approach for such handwritten signature verification avoiding all such complexities of handling a huge database of monochrome pictures corresponding to signatures of each individual. Here to avoid complex image processing methods like thinning, scaling and other morphological schemes, the signatures taken in the form of monochrome bitmap images are firstly converted into 1D data arrays of binary values - 0 and 1. Then, Avg_Sign is calculated following Statistical formula for Expected Mean, though it gives the same result if the formula for R.M.S., i.e., Root Mean Square is being followed. The Recognition scheme is based on extensive Statistical Analysis of Correlation Coefficient between bivariate data set. In implementation of proposed algorithm to constant factors carry major impact on the validity of the method and the strength of the verification lies in the efficiency of selection of these constant parameters, namely avg Sign Threshold Value and decision value.

We hope that our Study and Research will definitely be focused to extend the above given approach from offline detection scheme to online one through realization of neural networks and artificial systems.

REFERENCES

1. Velez, J.F., A. Sanchez and A.B. Moreno, 2003. Robust off-line signature verification using compression networks and positional cuttings. Proceedings of 2003 IEEE Signal Processing Society Workshop on Neural Networks, 17-19 September, 2003, Toulouse, France, pp: 627-636.
2. Samir Kumar Bandyopadhyay, Debnath Bhattacharyya and Poulami Das, "A Flexible ANN System for Handwritten Signature Identification", Proceedings of the International MultiConference of Engineers and Computer Scientists 2007 Volume II, IMECS '07, March 21 - 23, 2007, Hong Kong, Lecture Notes in Engineering and Computer Science, pp. 1883-1887.

3. Justino, E.J.R., F. Bortolozzi and R. Sabourin, 2001. Off-line signature verification using hmm for random, simple and skilled forgeries. 6th International Conference on Document Analysis and Recognition, 10-13 September, 2001, Seattle, WA, USA, pp: 105-110.
4. Emre Ozgunduz, Tülin Şenturk and M. Elif Karslıgil, 2005. Off-line signature verification and recognition by support vector machine. Eusipco-2005, 4-8 September, 2005, Antalya, Turkey, pp. 113-116.
<http://www.eurasip.org/Proceedings/Eusipco/Eusipco2005/defevent/papers/cr2010.pdf>
5. Debnath Bhattacharyya, Samir Kumar Bandyopadhyay and Deepsikha Chaudhury, 2007. Handwritten signature authentication scheme using integrated statistical analysis of bi-color images, IEEE ICCSA 2007 Conference, Kuala Lumpur, Malaysia, August 26-29, pp: 72-77.