Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

Static Signature Authentication based on J48 and Random Forest

Ranjan Kumar Singh Sushila Maheshkar
Department of Computer Science
Indian Institute of Technology (ISM)
Dhanbad (Jharkhand - 826004), India

Vikas Maheshkar
Division of Information Technology
Netaji Subhas Institute of Technology
Delhi 110078

***Abstract*— With the exposure and development of technology, a better security system is needed to protect the data. Off-line signature verification is one of the prime area of research which is most widely recommended by the research community for security issues. For the Off-line signature verification from spontaneous handwritten signature image a precise and effective method is proposed which contributes significantly to that area and comprises of image prepossessing, feature extraction and decision tree classifier such as J48 and Random Forest. It exploit totally different features of writing like number of closed loop, standard deviation, skewness, number of edge points, kurtosis, centroid and number of cross points by computing a collection of options from writing samples at totally different levels of observations. These features are extracted from the standard signature database and extracted features are analyzed correspondingly. The method proposed here monitor an effective accuracy of the proposed algorithm.**

***Index Terms*— Decision Tree, J48, True Negative Rate, False Negative Rate, Accuracy, False Positive Rate, True Positive Rate.**

I. INTRODUCTION

The necessity for security is difficult often compounded by a tendency towards individual levels of general apathy. Biometrics has received recognition from a vivid range of applications available [1]. Signature has been the most important and useful behavioral biometrics that is widely accepted [1]. The surge in the application of authentication and identification for personal verification and security related application can be observed these years. The increasing studies and development on signature verification has drawn the attention of researchers from different fields of education.

Each person has their own traits on the basis of which they can be distinguished and hence authenticated [2].

One can verify and identify the claimed person by two mean of biometric system. Verification of signature means to verify whether the claimed person is the actual genuine person whereas signature identification means person's existing is there in the given database or not. Biometric system relies on specific data about unique biological traits [11]. It thus verifies or identifies the claimed users. Biometric identifier are usually classified as behavioral and physiological. Physiological characteristics refers to the form of body similar to fingerprints, face recognition, DNA, etc while behavioral characteristics are related to the pattern of behavior of a person including voice, gait etc. [6, 10].

Automatic Off-line Signature authentication and verification system occupy a very special position among the other automatic identification methods [13, 8]. Handwritten signatures are persistent and are being used for identity verification. They are used for getting permission in banks related works and transactions and thus have paved a way for criminal deception. Hence automatic signature verification is required for access grant in such areas. Depending upon ownership, signature can be genuine or insincere. Verifications are done on the basis of the traits of users that can neither be changed nor be remodeled [7].

Signature verification is advantageous since it is biologically linked to a specific individual [4]. Digitalized system may exacerbate the problem of technological obsolescence. A signature is more difficult to be forged then a fingerprint when a person is in unconscious state. Handwritten signature results in complex process and features. It is used as a legal mean of verifying individual's identity by financial sectors and administrative [4, 5].

II. RELATED WORK

A lots of research had already been examined and shown their pivot contribution in the field of biometric system mainly in off-line signature verification. Since off-line signature verification is the most important biometric verification in today's era, still it is very challenging area of research.

Bhargava et.al [11] proposed a discussion about univariate and multivariate approaches about decision tree and implement various algorithm in a data mining tool WEKA. Offline Handwritten Signature verification using Neural Network proposed by Hatkar et.al [9] proposed their work in which they used machine learning as classifier and have found accuracy 86.25%. Suryawanshi et.al [3] proposed as signature verification method that uses Artificial Neural Network as classifier. Jain et.al [1] proposed a method for Offline signature verification using Adaptive Resonance Theory 1. A standard database of 250 signatures is used for calculating the performance of system in this paper. The false rejection rate (FRR) and false acceptance rate (FAR) are found as 3.9% and 2.7%. Handwritten signature verification using neural network proposed by Ashwini and Shalini [2] having false rejection ratio (FRR) as 20%. The test is carried out with 150 as standard signature database. Ferrer et.al [7] proposed a method in which classification of the signature is done in very precise way and uses high intensity variation and cross over points. Prakash and Guru [14] proposed score level fusion method for offline signature verification and found AER for mean as 17.33. Grid Based feature extraction techniques has been optimize by using 2D Gaussian filter which is proposed by Nguyen and Blumenstein [15]. Vielhauer and Dittmann [12] proposed different behaviours related to various sets of reference signatures, a multi-level verification strategy is proposed which uses well-selected sets of reference signatures.

The flow of the remaining part of paper: complete proposed work comprises in section III. Experiment results is shown in section IV and in section VI the conclusion of the work is given.

III. PROPOSED WORK

The proposed method is based on behavioral biometric authentication system and the core part of our discussion is static and offline signature verification. Now a days this signature is having broad application that includes banking application, student's credentials, government document, treaties between two countries etc. Due to demand of such mode of verification we need a system which is most reliable which is having proficient success rate as well as least time consumable. Such an effective method whose success rate is 75% will help to eradicate and eliminate the verification which is to be done manually. The signatures are stored in database and finally compared with specimen signature using feature like kurtosis, skewness, entropy, edge points, cross points and centroid so as to verify whether the specimen's signature is forged of genuine.

The three main phases of signature verifications used here include:-

1. Pre - Processing
2. Feature extraction
3. Classifier (Machine Learning Approach)

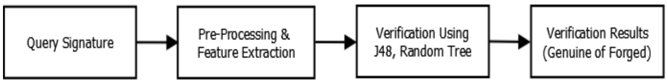
Query Signature
Pre-Processing & Feature Extraction
Verification Using J48, Random Tree
Verification Results (Genuine of Forged)

Fig. 1. Block Diagram of Authentication of Signature based on Decision Tree

Block diagram of the work is depicted in Fig.1. The proposed method is having different stages. The flow of the proposed method is given by Fig.2 which is having separate procedure for training the supervised machine.

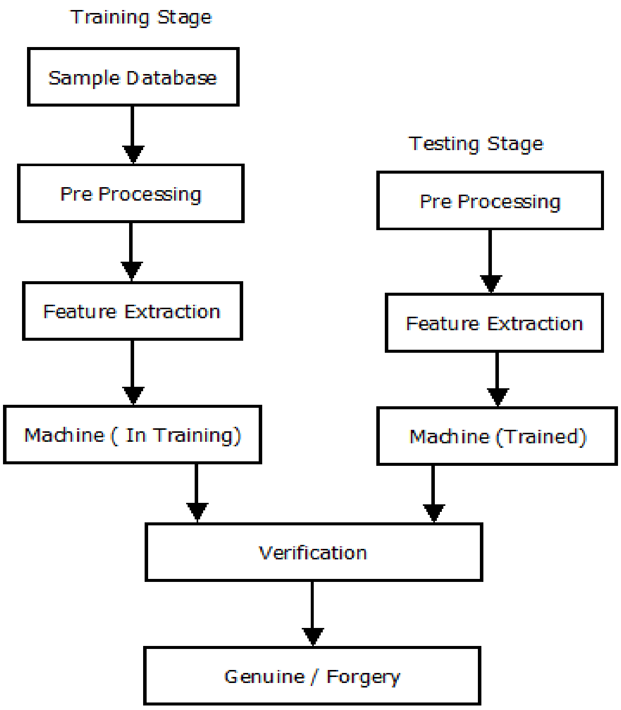

Fig.2. Flow Diagram of Authentication of Signature based on Machine Learning

A. Signature Pre-Processing

The preprocessing step is applied for already stored signature in database as well as the signature is to be test. The purpose of this step is to make the signature normalized in one form and improve the quality of image which is suitable for feature extraction. The preprocessing stages includes

1) Binarization

Binarization of gray scale signature is obtained by adaptive thresholding value. It begin with a threshold value for different signature which is being calculated by the method of thresholding. For each signature pixel element it computes associate intensity gradient by selecting a maximum of difference of left and right signature pixel intensity and upper and lower signature pixel intensity respectively to calculate the corresponding threshold. Finally, thresholding method is exploited for Binarization. During this technique intensity of every signature pixel intensity is compared with the threshold value. The signature pixel intensity is set to 1 for the pixel intensity value higher than the threshold while for lower value it is set to 0.

2) Complementation

Complement of a binarized signature means converting the zeros into ones and ones into zeros. Complementation of signature image results in better visibility of great difference of gray levels. It helps us to identify the fine detail in precise manner for signature image which helps to correctly calculate the features to classify about signature between genuine or forged. Complemented signature image is having more clarity for further operation since the lighter pixel in signature become dark and the dark area become lighter.

3) *Noise Reduction*

A noise is some specific portion of signature image which do not represent the part of signature. These type of portion need to be nullified. After getting complemented binary image filter is used which reduces the noise by excluding or ignoring the single black pixel of signature on the background of white. With the help of connectivity properties of pixels as 8-neighbors a chosen pixel is examined whether that is alone or not.

4) *Thinning*

Removal of some selected signature pixel from the binarized signature with morphological operation such as erosion or opening known as thinning. Skeletonization is the prime scheme for thinning with many other applications is remaining. In this work, it is used to precise the output by reducing the lines into the single pixel which results in better number of edge detection. Thinning means reducing binary objects or shapes to strokes that are single pixel wide. Thinning preserves the properties of Euler number 'En' where the number 'En' is the total number of objects in the image obtained by subtraction of number of Holes 'Hn' in the object from the connected component 'Cc' which is explained in (1).

$$En = Cc - Hn \qquad (1)$$

In the operation thinning, four condition (2), (3), (4) and (5) determines whether the pixel should be deleted or not which is listed below.

Condition 1:

$$y_H(p) = \sum_{i=1}^{4} a_i \qquad (2)$$

Where

$$a_i = \begin{cases} 1, if\, y_{2i-1} = 0 \wedge (y_{2i} = 1 \vee y_{2i+1} = 1) \\ \qquad 0, otherwise \end{cases}$$

$y_i(p) = y_1, y_1, \ldots, y_8$ represents eight neighbor of pixel (p) such that x_1 is the east pixel x_2 to x_8 can be obtained in traversing in counter clockwise direction.

Condition 2:

$$(y_2 \cup y_3 \cup y_8) \cap y_1 = 0 \qquad (3)$$

Condition 3:

$$(y_2 \cup y_3 \cup y_8) \cap y_1 = 0 \qquad (4)$$

Condition 4:

$$(y_6 \cup y_7 \cup y_4) \cap y_5 = 0 \qquad (5)$$

On the basis of these listed conditions pixel (p) is deleted in first sub iteration, if condition 1, 2 and 3 are satisfied whereas in the second sub iteration if condition 1, 2 and 4 are satisfied.

5) *Bounding Box of Signature*

Bounding box of signature basically draw a rectangle trace which form outline to the area of signature. With the help of this process one can reduce the effort and time to process the signature further. It is the possible region of interest of an image. Maximum and minimum coordinates value of x and y are calculated which can be fixed as the corner of rectangles and by this the bounding box is calculated. The coordinated represents as the top left and the bottom right of the bounding box.

B. *Feature Extraction*

It is quite difficult that which feature is to be choose for further verification of signature. Feature of signature image is thus been extracted and can be used as a reference points for further verification of signature. In our work, feature as a vector of seven entities which include entropy, number of closed loop and so on are used to precisely authenticate and verify the test signature. These features that has been depicted in Table 1.

TABLE 1
Extracted features from sample signature

Features	Sample Signature 1	Sample Signature 2	Sample Signature 3
Entropy	0.2586	0.2311	0.2586
No. of Closed Loop	1	1	2
No. of cross points	155	74	91
No. of edge points	8	13	23
Skewness	4.4691	4.8623	4.4691
Kurtosis	20.97	24.6424	20.9729
centroid	66.1576 77.5271	82.0974 85.2767	61.7636 58.7769

1) *Entropy*

Image entropy is a quantity which is used to describe the amount of information that must be coded for by an algorithm for better observation on image data. Image having low entropy must contain a lots of black pixel and having low amount of white pixel. Image having entropy as zero contain only flat pixel. Image having high entropy contain larger number of white pixel than that of the black pixel and cannot be compressed as much as low entropy images. Entropy that is calculated over here is the same formula used by Gallileo Imaging Team and which is calculated using equation (6).

$$Entropy = -\sum_i q_{(i)} log_2 q_{(i)} \qquad (6)$$

In above equation $q_{(i)}$ refers to the probability that the difference between two adjacent pixels is equal to i, and $\log_2$ is the base 2 logarithm.

2) *Edge Points Number*

Edge detection is the identification of points (edge points) in digital images where brightness of the images changes sharply or does not continued. The pixel position where an abrupt changes occur are basically organized into a set of curved line segment which is termed as edges. The pixel which has only one neighbor, which belongs to the signature, in 8-neighbor is known as edge points.

3) *Skewness*

The symmetricity of distribution of pixel intensity is measured by skewness. The distribution is symmetric when there is equal distribution from the center point that is, it has identical distribution to the right and the left with respect to center. For normal data distribution the skewness will be zero and for symmetric data it closes to zero. If the data is partially distributed or distribution is tilted toward left than it value is negative. Skewness value is positive when the skewness of data is right. Hence it can be state that the surface having darkness tends to be have more skewness than that of the lighter surface.

The skewness of a random variable X is the third standardized moment γ_1, defined and explained in (7):

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{7}$$

4) *Kurtosis*

Kurtosis helps us to verify and compute whether the data are sharped or flat, when compare with normal distribution. If the data is having large kurtosis than it has a distinct number of peak near mean of the data. In normal distribution, it decline with high rate and larger tails which is the extreme case. Similarly data with flat top near mean is having low kurtosis with no sharp peak.

The Kurtosis is standardized and can be calculated as shown in (8).

$$\mathrm{Kt}[Y] = \frac{\mu^4}{\sigma^2} = \frac{E[(Y-\mu)^4]}{(E[(Y-\mu)^2])^2} \tag{8}$$

Where **μ** is the central moment and **σ** is the standard deviation.

5) *Centroid*

We find the geometric centroid of the image by traversing the whole signature image iteratively and consider the centroid of that moment. In this case we are having the centroid as the coordinate of both in x-coordinate as well as y-coordinate (center X, center Y).

C. *Classifier*

In this proposed method of off-line signature verification we use WEKA tools for classification of feature extracted. The most widely used data mining as an open source software is WEKA [11]. Full form of WEKA is Waikato Environment for knowledge analysis. The different operation like association, filtering, classification, clustering, visualization, regression etc. can be performed using this tools which is written using object oriented language java. J48 and Random Forest is used which is basically a type of decision tree.

1) *J48*

J48 is associated with open source Java implementation of C4.5 algorithm data mining tool WEKA. The improved version of ID3 (Iterative Dicho 3) is C4.5. This essentially produce a threshold so it split the list into two. First list consists of data having higher value than the threshold while second list comprises of those with equal or lesser value.

2) *Random Forest*

Random Forest is decision tree type machine learning algorithm. The features are randomly selected for replacement for each learner. In this algorithm, one modifies the training data. However, this modification is performed in the feature space.

D. *Proposed Algorithm*

Input: signature from database
Output: classified signature as forged or genuine
1. Extract signature image from a database.
2. Pre-processing the extracted signature.
3. Converting signature image into binary image.
4. Image complementation.
5. Noise Reduction of complimented image.
6. Finding bounding box of the thinned signature.
7. Thinning (I_{thin}).
8. Feature extraction of $I_{thin.}$
9. Finding Entropy of $I_{thin.}$
10. Calculate total number of edge points of $I_{thin.}$
11. Calculate total number of cross points of $I_{thin.}$
12. Count total number of closed loop for $I_{thin.}$
13. Find skewness and kurtosis of $I_{thin.}$
14. Find centroid of $I_{thin.}$
15. Feature vector is formed by combining extracted features.
16. Train a neural network with a normalized vector.
17. Steps 1 to 16 are repeated for testing signatures.
18. Feature vector is then apply for test signature to trained neural network using WEKA tools with two different algorithm.
19. Result generated by WEKA declaring a signature as a forged or genuine.

IV. EXPERIMENTAL RESULTS

In our work we have implemented the algorithm with the help of data mining tool WEKA on the system having configuration i3-5005U CPU 2.00 GHz, 4.00 GB RAM of 64-bit operating system. We have performed the experiment with the help of standard J48 and Random Forest as an algorithm.

Signature Database

We have used MCYT corpus for our experiment and analysis. The dataset contain 28 users and for each user 5 genuine and 5 forged signatures are used. Thus in all 280 signature images are present in the data set.

A. *Experimental Configuration*

We have used WEKA (Waikto Environment for Knowledge Analysis) as a simulation tools for experimentation. WEKA provided a homogeneous interface platform to different number of machine learning algorithm for calculation of result of learning method on any given dataset. WEKA gives the performance of learning algorithm that can easily applicable to any dataset and also include diversity of tools for transforming the dataset. This tool allows huge support for the complete process of experimental data mining together with preparing the input data, calculate and evaluating learning method statistically, visualizing the learning data and the result of learning.

V.PERFORMANCE MEASURE

The list of symbols used for performance measure in experimentation is shown in Table 2. The performance measure of the signature verification is measured in term of different evaluation metric as shown in Table 2.

TABLE 2
Different Evaluation metric

Evaluation Metric	Equation
TPP(True Positive Rate)	TP/(TP+FN)
TNR(True Negative Rate)	TN/(FP+TN)
FPR(False Positive Rate)	FP/(FP+TN)
FNR(False Negative Rate)	FN/(TP+FN)
ACC(Accuracy)	(TP+TN) / (P+N)
F-Measure	2TP=(2TP+FN+FP)
MCC(Mathews Correlation Coefficient)	$\frac{(\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN})}{\sqrt{(\mathrm{TN}+\mathrm{FP})(\mathrm{TP}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})(\mathrm{TP}+\mathrm{FN})}}$

A. *Experimental Result*

In this proposed method, the experimentation is done excessively on two different classifier. The different feature obtained after pre-processing stage extracted from signatures are centroid, closed area, edge points, cross points, kurtosis, skewness and entropy.

1) *J48*

The time taken by J48 to build the model is 0.09 sec and to test the model on training data is 0.03 sec. The performance measure by J48 with 10 iteration and base learner is upto 71%.

2) *Random Forest*

The time taken by Random Forest decision tree to build the model is 0.27 sec. The performance measure by Random Forest with 10 iteration and base learner is up to 80 %.

Table 3 shows the final acceptance and rejection of signature of different classifiers in terms confusion matrix. Here, the number of genuine and number of forged signatures are 140 each.

TABLE 3
Measured value of confusion matrix for different classifier

Model	TP	FN	TN	FP
J48	104	36	94	46
Random Forest	110	27	113	30

The different accuracy measured by class of different classifier is shown in Table 4. It includes rate like TPR, FPR, FNR, TNR and ACC. Accuracy of J48 model is calculated as 70.71% and that for the Random Forest is 79.64%.

TABLE 4
Different accuracy by class of different classifier

Model	TPR	FPR	FNR	TNR	ACC
J48	0.74	0.32	0.25	0.67	70.71%
Random Forest	0.80	0.20	0.19	0.79	79.64%

The measured statistical value of different classifier are shown in Table 5. From the set of 280 instance J48 correctly classified 198 instance while 82 are misclassified. This results in 70.71% accuracy. In the case of Random Forest out of 280 instance 223 are correctly classified but still 57 are misclassified. This results in 79.64 as accuracy in Random forest classification. The different rate valued measured by two classifiers are shown in Table 5.

TABLE 5
Different rate valued measured by two classifier

Statistical Variables	J48	Random Forest
Correctly Classified Instances	198	223
Incorrectly Classified Instances	82	57
Kappa Statistics	0.4143	0.5929
Mean absolute error	0.3306	0.3331
Root mean square error	0.4951	0.3953
Relative absolute error	66.12%	66.62%
Root relative squared error	99.02%	79.08%

Table 6 depicts the comparison between the existing methods with the proposed method of the paper.

TABLE 6
Comparison of this work with existing methods

Author	Parameter	Matching Technique	ACC	Database
Prakash et al. [14]	Centroid Based	NN, HMM, SVM	NS	MCYT-75
Vu Nguyen et al. [15]	Gaussian Grid Based	SVM	NS	GPDS-960
Proposed Work	Local & Global Processing	J48, Random Forest	70.71% 79.64%	MCYT-75

VI. CONCLUSION

In this work, a new approach is proposed on the base of two classifier for offline signature verification. The proposed method classified good between the genuine and forged signature. The features like Skewness, Kurtosis, Centroid, Edge Number, Cross Points, Closed loop and Entropy are calculated of the standard signature corpus MCYT-75. The validation of these extracted feature is done with the help of machine learning approach. Now the performance of proposed method aimed at learning the ability to differentiate the given signature between genuine and forgery. We have use J48 and Random forest for classification of signature. The extracted features are now given to the both machine learning classifier one by one and come up with experimental results. J48 Classification gives 70% accuracy but Random Forest decision tree having the accuracy of 80%.

VII. REFERENCES

[1]. Charu Jain, Priti singh and Aarti chug, "An Offline Signature Verification using Adaptive Resonance Theory 1(ART1)" International Journal of Computer Applications (0975 – 8887) Volume 94 – No 2, May 2014.

[2]. Ashwini Pansare and Shalini Bhatia, "Handwritten Signature verification using Neural Network" International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 1– No.2, January 2012.

[3]. Rushikesh suryawanshi, Shantanu Kale, Rahul Kale, Rahul Pawar, Sidhartha Kadam and V.R. Ghule,"Offline Signature Cognition and Verification using Artificial Neural Network" International Journal of Advance Research in Computer and Communication Engineering Vol.5, Issue 3, March 2016.

[4]. Saeid Fazli, Shima Pouyan and HamedFathi, " High Performance Offline Signature Verification and Recognition Method using Neural Network" International Journal of Advanced Studies in Computers, Science and Engineering, Volume 4, Issue 6, 2015.

[5]. Rapanjot kaur, Gaganjeet Sing Aujla "Enhanced Offline Signature Recognition Using Neural Network and SVM" International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014.

[6]. Anil K Jain, Arun Ross and Salil Prabhakar, "An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no.1, pp. 1-29, 2004.

[7]. Migual A. Ferrer, Jesus B. Alonso and Carlos M. Travieso, "Off-line Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol.27, no.6, June 2005.

[8]. Lan-Rong Dung, Chang-Min Huang and Yin-Yi Wu, "Implementation of RANSAC Algorithm for Feature- Based Image Registration" Journal of Computer and Communications, 1, 46-50, 2013.

[9]. Pallavi V. Hatkar, Prof.B.T.Salokhe and Ashish A.Malgave, "Offline Handwritten Signature Verification using Neural Network", International Journal of Innovations in Engineering Research and Technology, Volume 2, Issue 1 Jan 2015.

[10]. Neeraj Bhargava, Girja Sharma, Ritu Bhargava and Manish Mathuria, "Decision tree analysis on J48 algorithm for data mining", International Journal of Advanced Research in Computer Science and Software Engineering, 2013.

[11]. Anil Jain, Patrick Flynn and Arun A Ross,"Handbook of Biometrics", Springer Science & Business Media, 2007.

[12]. C Vielhauer and J Dittmann."Biometrics for user authentication encyclopedia of multimedia", ed.b. furth, 2006.

[13]. R. Plamondon, S. Srihari, On-line and off-line handwriting recognition: a comprehensive survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) 63–84 (2000).

[14]. HN Prakash and DS Guru. Offline signature verification: An approach based on score level fusion. International journal of computer applications, pages 0975-8887, 2010.

[15]. Vu Nguyen and Michael Blumenstein. An application of the 2d gaussian filter for enhancing feature extraction in off-line signature verification. In 2011 International Conference on Document Analysis and Recognition, pages 339-343. IEEE, 2011.