ISSN: 0067-2904
GIF: 0.851

Hand Written Signature Verification based on Geometric and Grid Features

Baraa Kareem Abd, Nada A.Z. Abdullah* and Qaswaa Khaled Abood
University of Baghdad, College of Science, Computer Science department, Baghdad, Iraq

Abstract

The fact that the signature is widely used as a means of personal verification emphasizes the need for an automatic verification system. Verification can be performed either Offline or Online based on the application. Offline systems work on the scanned image of a signature. In this paper an Offline Verification of handwritten signatures which use set of simple shape based geometric features. The features used are Mean, Occupancy Ratio, Normalized Area, Center of Gravity, Pixel density, Standard Deviation and the Density Ratio. Before extracting the features, preprocessing of a scanned image is necessary to isolate the signature part and to remove any spurious noise present. Features Extracted for whole signature first, then extracted for every part after dividing the signature into four sections. For verification, statistical verification techniques are used (Euclidean Distance, Hellinger Distance and Square Chord Distance). The system is trained on three datasets of signatures. The first and the second datasets have English signatures while the third one is collected from people; it contains Arabic and English signatures. The system has been tested on every dataset. The experimental results show that the Euclidean Distance has the average accuracy of 94.42, the Hellinger Distance has the average accuracy of 95.27 and the Square Chord Distance has the average accuracy of 93.14. That result for whole the image and the following average accuracy for image using grid the Euclidean Distance has the average accuracy of 93.54, the Hellinger Distance has the average accuracy of 95.87, and the Square Chord Distance has the average accuracy of 95.93.

Keywords: handwritten, signature, verification, feature extraction.

برنامج التحقق من التواقيع المكتوبة بخط اليد باستخدام الخصائص الهندسية

براء كريم عبد ،ونداء عبدالزهرة عبدالله* ، وقصواء خالد عبود
جامعة بغداد ، كلية العلوم ،قسم علوم الحاسبات، بغداد ، العراق

المستخلص

يستخدم التوقيع كوسيلة للتحقق من هوية الشخص عليه نحتاج لنظام للتحقق من صحة التواقيع . التحقق يتم اما بصورة مباشرة او غير مباشرة . فنظام التحقق غير المباشر يعمل على صور للتواقيع . في هذا البحث اقترح نظام للتحقق من صحة التواقيع يعتمد استخراج الخصائص الهندسية للتوقيع. الخصائص الهندسية المستخدمة هي : المتوسط الحسابي، النسبة التي يشغلها التوقيع ، نسبة الكثافة ، مركز الثقل ، كثافة البيكسل ، الانحراف المعياري واخيرا نسبة العرض الى الارتفاع . وقبل استخراج هذه الخصائص يجب تجهيز الصورة من خلال عمل بعض الامور منها ازالة الضوضاء من الصورة ، تحويلها الى الابيض والاسود ، وازالة اي حواف غير مرغوب بها وما الى ذلك من امور لتكون جاهزة لاستخراج الخصائص منها. الخصائص استخلصت للصورة كاملة ومن ثم جزئت الصورة الى اربع اجزاء واستخلصت خصائص كل جزء . واما بالنسبة للتحقق من صحة التوقيع فسوف

*Email: nadaazah@yahoo.com

يتم ذلك عن طريق استخدام ثلاثة طرق احصائية هي Euclidean Distance, Hellinger DistanceSquare Chord Distance. فيتم تدريب النظام باستخدام ثلاث مجاميع بيانات من التواقيع مجموعتي البيانات الاولى والثانية تواقيع انكليزية اما بالنسبة للثالثة فتم جمعها من اشخاص وهي عبارة عن تواقيع عربية وانكليزية . واخيرا استخلصنا نتائج الدقة لل Eucliden distance 94.42 ولل Hellinger Distance هي 95.27 ولل Square Chord Distance هي 93.14 هذا في حالة استخلاص الخصائص للتوقيع مرة واحدة. و نتائج الدقة لل Eucliden distance هي 93.54 ولل Hellinger Distance هي 95.87 ولل Square Chord Distance هي 95.93 هذا في حالة استخلاص الخصائص بعد تقسيم التوقيع الى اربعة اجزاء .استنتجنا من هذه ان النتائج تعتمد على البيانات المختبرة فنلاحظ ان لمجاميع البيانات الاولى Hellinger Distanceهي الافضل وهكذا كما وضح سابقا في نتائج الدقة .

1. Introduction

A handwritten signature can be defined as the scripted name or legal mark of an individual, executed by hand for the purpose of authenticating writing in a permanent form. The acts of signing with a writing or marking instrument such as a pen or stylus are sealed in the paper. Approaches to signature verification fall into two categories according to the acquisition of the data: On-line and Off-line. On-line data records the motion of the stylus while the signature is produced, and includes location, and possibly velocity, acceleration and pen pressure, as functions of time [1]. Online systems use this information captured during acquisition. These dynamic characteristics are specific to each individual and sufficiently stable as well as repetitive. Off-line data is a 2-D image of the signature. Processing Off-line is complex due to the absence of stable dynamic characteristics. The difficulty also lies in the fact that it is hard to segment signature strokes due to highly stylish and unconventional writing styles. The non-repetitive nature of variation of the signatures, because of age, illness, geographic location and perhaps to some extent the emotional state of the person, accentuates the problem. All these coupled together cause large intra-personal variation. A robust system has to be designed which should not only be able to consider these factors, but also detect various types of forgeries [2].

The system should neither be too sensitive nor too coarse. It should have an acceptable trade-off between a low False Acceptance Rate (FAR) and a low False Rejection Rate (FRR). [3]

2. The proposed system

As shown in Figure 1, the proposed system consists of three major parts: preprocessing, features extraction and verification. Preprocessing phase makes signature image ready for feature extraction. When the system is in learning mode, extracted features resulting from feature extraction step are used to learn system to the signature but, when the system is in testing mode, extracted features resulting from feature extraction step are used to classify the signature is original or forgery.

Separating data into training and testing sets is an important part . Typically, when separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing. Analysis Services randomly samples the data to help ensure that the testing and training sets are similar. By using similar data for training and testing, you can minimize the effects of data discrepancies and better understand the characteristics of the model.

Training model

1-input image

2- do preprocessing operation on image

*2.1 Convert to gray : convert image to gray using formula Gray color = 0.299*Red + 0.5876*Green +0.114*Blue*

2.2 Convert to binary :

if pixelvalue > 200 (threshold) then , change the value of pixel to 255 (white), else, change the value of pixel to 0 (black)

2.3 Crop and resize :- remove white area around image

2.4 Remove noise : using median filter

3-extract feature from signature

3.1 Compute mean and Standard Deviation

3.2 Compute pixel density : If c = 0 then (check the value of the pixel is black or not), C+=1

3.3 Compute slope : (gpx1- gpx2) / (gpy1- gpy2) where (gpx1, gpy1) gravity point of left part , (gpx2, gpy2) gravity point of right part

3.4 Compute baselineshift : (gpx1- gpx2)(gpy1- gpy2) where (gpx1, gpy1) gravity point of left part , (gpx2, gpy2) gravity point of right part
3.5 Compute occupancy ratio : pixel density/ Total pixel
3.6Compute density ratio :DR= xt / yt where xt pixel density in the top part of image , yt pixel density in the bottom part of image
3.7 Compute gravity center : gravity point = (gpx, gpy)
gpx= totx / count where totx sum the black pixel in the row
gpy = toty / count where toty sum the black pixel in the column
3.8 Compute aspect ratio
4- save features founded in Step 3 in access database

Test model

1- *Employ unknown Signature*
2- *Enhance image (binary, resize, crop and remove noise if image need).*
3- *Extract the features (same as features computed in training model)*
4- *Matching with the features stored in database.*
5- *Do the classification*
6- *Take decision as originals or forgeries*

2.1 Preprocessing

In order to improve the performance of the system, few preprocessing operations are carried out on offline signatures. The acquired signature images sometimes may contain extra pixels as noises which are due to some problems during scanning of signatures or due to non-availability of It is necessary to remove these extra pixels from the signatures; otherwise the signature may not be recognized correctly. The purpose of preprocessing phase is to make signatures ready for feature extraction. The preprocessing stage includes four steps: convert color image to gray, noise removal, cropping, Binarisation, image normalization, and grid.

Convert color image to gray

Gray scale images only contain brightness information. Compared with binary images, they contain richer information. Typically, gray scale images contain 8 bit data. The range of pixel values is from 0 to 255. These images can provide some sort of noise.
formula convert RGB to Gray level = 0.299*Red + 0.5876*Green +0.114*Blue …(1)

Remove noise

There are different kinds of noise in an image; to remove this kind of noise median filter can be applied to images, to apply mask is needed. This mask can be of 3*3 matrix.

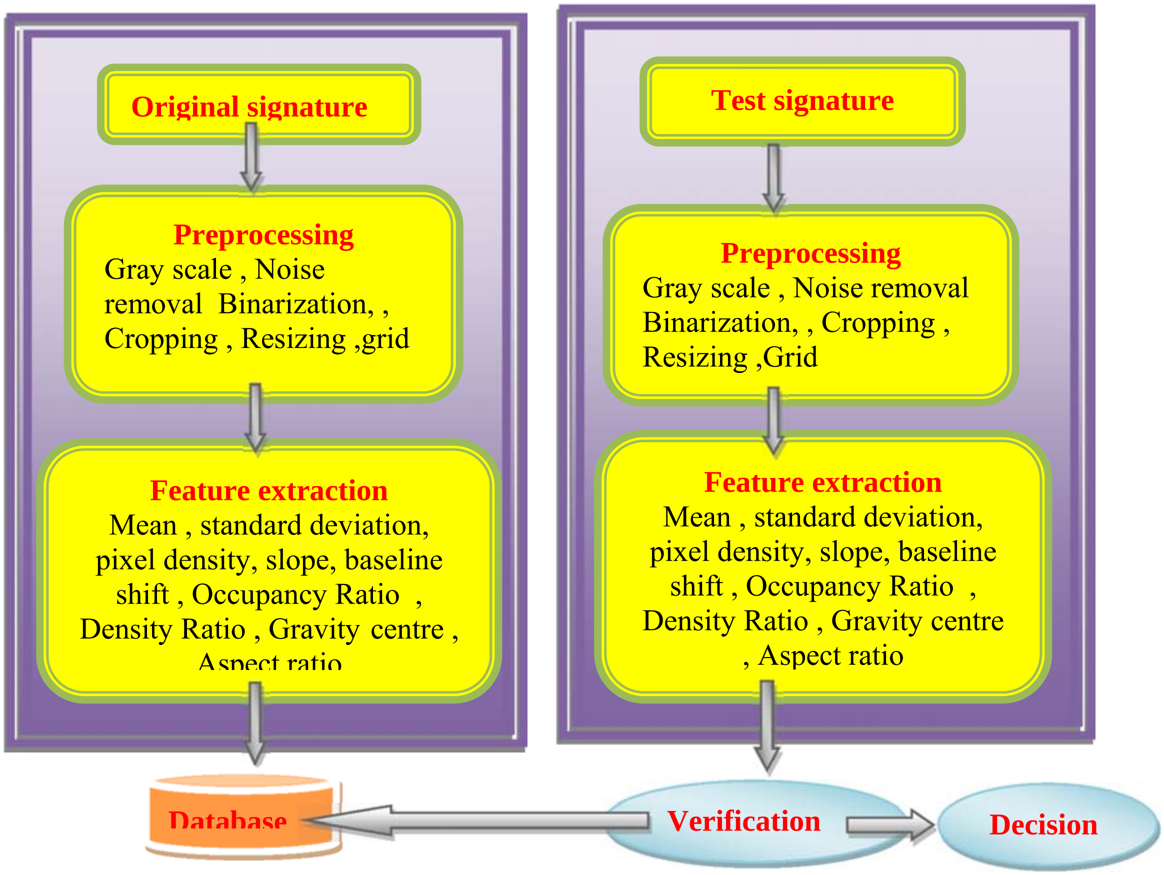

Figure 1-Block diagram of proposed system.

Binarisation

Conversion from gray scale image to a black and white (binary) image- For signature verification, the color of ink has no significance at all. Instead the form of two signatures must be compared. Hence all scanned images were converted to black and white images where white is represented by 255 and black by 0 [2]. Threshold in this algorithm is used to split the background from the signature the threshold value is (200) the value is chosen by trial and error.

Cropping and resizing

Cropping refers to the removal of the outer parts of an image to improve framing (presentation of visual elements in an image). The signature cuts from the image automatically by determining its beginning and ending. To compare two signatures they must be in the same size so that signature resize is required.[3]

2.2 Features Extraction

Features extraction process is an important step in developing any signature verification system since it is the key to identifying and differentiating a user's signature from another. In this system, the features extracted are: Mean, Standard Deviation, Pixel density, Occupancy Ratio, Density Ratio, Gravity, slope, baseline shift and aspect ratio.

Mean

The mean of a data set is simply the arithmetic average of the values in the set, obtained by adding the values and dividing by the number of values. The mean is a measure of the center of the distribution.

Standard Deviation

The standard deviation is a measure of the scatter of values within a set of data. SD is a standard for the distribution of a lot of information. The more spread apart the data, the higher the deviation.

Pixel density and Occupancy Ratio

Pixel density is the pixel belong to the signature it gives us the density of the signature. Occupancy ratio is the ratio of the number of pixels which belong to the signature to the total pixels in the signature image.

The pixel density feature xPD $\in$ [0, 1] . the pixel density of a signature segment is directly linked to stroke width, it is also commonly referred to as apparent pen pressure. For this reason, the pixel density feature is said to contain pseudo-dynamic signature information. [4]

PD= M / HW ……………(2)

Where M = number of black pixels

H = height of the image

W = width of image

Density ratio

Signature image is split into two halves horizontally. After that The pixel density of each part is calculated and divided it to each other. It provides information about the signature density ratio of two halves of the signature image which density part is greater than the other. It provides information about the signature density ratio of two halves of the signature image. [5]

DR = density of the left half /density of the right half ……..(3)

Gravity Center

Gravity center is the spot at which the total weight of the signature can be seen as focused. This feature is important because it is used in the other feature like slope and baseline shift feature extraction.

The Center of Gravity is the 2-tuple (X,Y) given by, [6]

$X = {}_{j=0}\sum^{N-1} P_V(j)*j/\Delta$

$Y = {}_{i=0}\sum^{M-1} P_H(i)*i/\Delta$ ……………………..(4)

where P_V and P_H are the vertical and horizontal projections respectively.

Baseline Shift

It is the difference between the vertical centers of gravity of the left, and the right part of the image. It was taken as a measure for the orientation of the signature.

Slope

In this feature the slope is found in the line joining the Centers of Gravity of the vertical splitting of the signature sample. The signature sample divided into left and right halves after that calculate the center of gravity of the two halves to distinguish signature samples.

Aspect Ratio Computation

The ratio is obtained by dividing the signature height to signature width. The height is the maximum length of the column in an image and similarly the width is the row of maximum length. This ratio may differ from person to person, but the ratio is constant for an individual.

2.3 Verification

The extracted features are fed to a classification system. This system compares the extracted features from the given signature with the features extracted from the corresponding signature in the database in order to verify the authenticity of the signature and makes a final decision for verification as a genuine or forged signature. A crucial parameter for verification is the choice of statistical distance metrics to measure the similarity or dissimilarity between two signature images. It is essential to explore the different similarity measures to find out the best distance metric for signature matching. In conventional signature matching technique, Euclidean distance is used to find the similarity between the test image and features database. Similarity score is used to find the best match of test image from the features database. Test image is more similar to image if the distance between the test image features and features in database is small. The proposed method is tested with three distance matrices: Euclidean distance, Hellinger distance, and Chord Distance. After trial and error the threshold value has been specified which is (0.6) for Euclidean distance, (0.3) for hellinger distance and (0.2) for square chord distance. Distance (D_{euc} , D_{heling}, D_{sqch}) can be calculated by using equations.

$D_{euc}=\sqrt{\sum} (p_j - Q_j)^2$ **……..(5)**

$D_{heling}=\sqrt{\sum} (\sqrt{p_j} - \sqrt{Q_j})^2$ **……..(6)**

$D_{sqch}=\sum (\sqrt{p_j} - \sqrt{Q_j})$ **……..(7)**

P and Q represents the feature for database image and test image respective

If this distance is below a certain threshold then the query signature is verified to be that of the claimed person otherwise it is detected as a forged one. [7] [8]

3 Signature Dataset

Data acquisition consisting of three datasets of signature images two of them collected from internet and the third one is taken from people . The first dataset consist of 55 signatures original and 55 signature forgery each writer have 22 sample of his sign , the second dataset consist of 30 signatures original and 30 signature forgery and the last dataset consist of 50 signatures original and 50 signature forgery . the first and the second dataset is English signature and the third one is collected from people it is contained Arabic and English signature .The figure 2 shows some sample of datasets.

Figure 2 - samples of different datasets

4. System Interfaces

The interface of the developed system shown in figure 3 .At first signatures have been taken from dataset, after that, the features are calculated from each signature in dataset and saved it in MS access database. Then the three distance methods are used to verify the tested signature. If the length of the test signature is within a certain limit, the system will classify it as authentic or else forged. Figure 4 shows system execution steps.

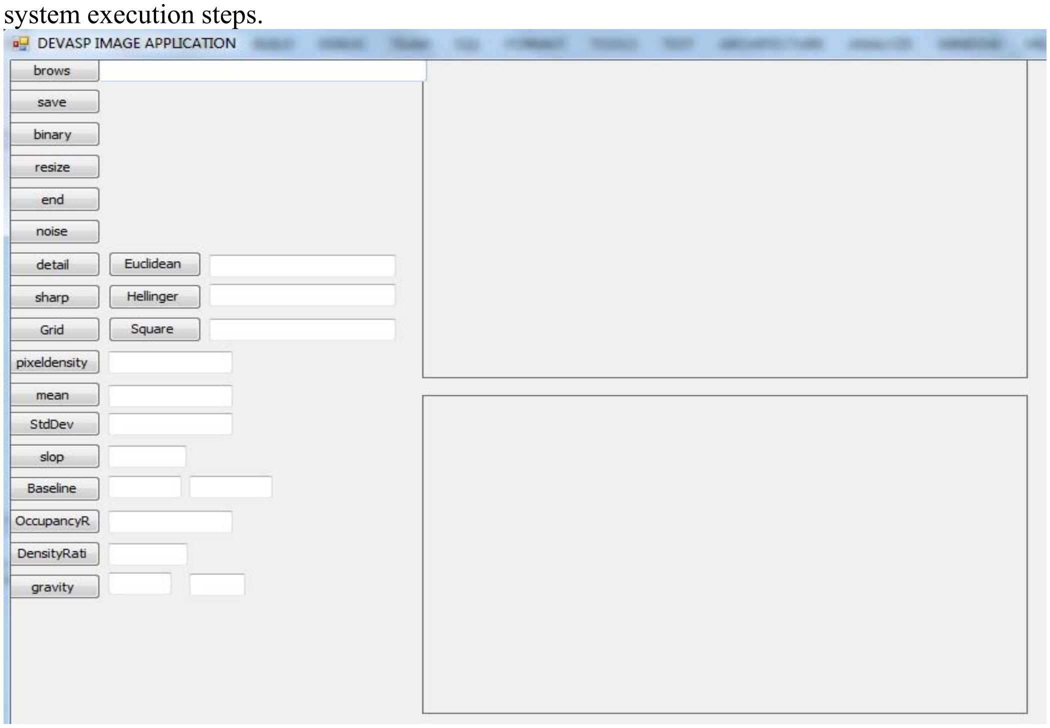

Figure 3- System interface

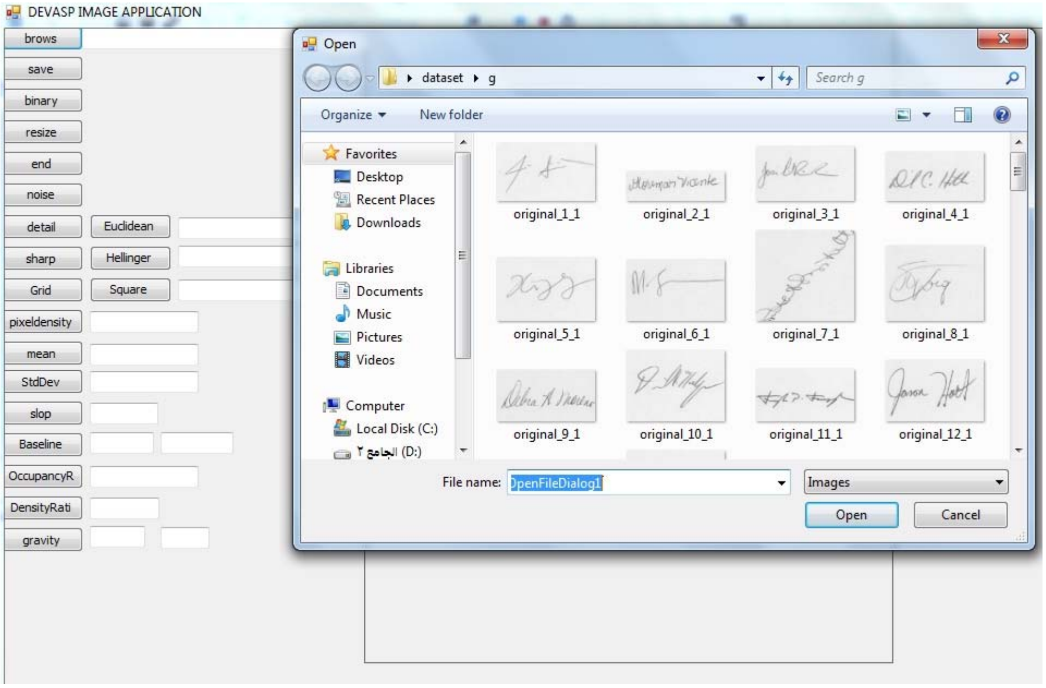

Figure 4-system execution steps

5. Experimental Results

The system has been tested for its accuracy and effectiveness on data from 55 users with 24 samples of each making up a total of 1320 signatures. The proposed verification algorithm is examined on both real and forged signature sample counterparts. We save the original signature in access database after finds each feature to the signature. after saving the signature we compare two signature one is new and the other in data base and check is the signature original or forged . False Acceptance Rate (FAR) and False Rejection Rate (FRR) are the two parameters used for measuring performance of any signature verification method. FAR and FRR are calculated by giving equations:

*FAR=Number of forgeries accepted/ Number of forgeries tested * 100*

*FRR= Number of originals rejected/ Number of originals tested****100***

The accuracy value also calculated:

accuracy = Number of originals accepted/ Number of originals ****100***

A crucial parameter for classification is the choice of an appropriate distance metric to measure the similarity or dissimilarity between two signature images. The results of forged and genuine signatures are indicated in the tables below. Table 4.1 shows the results of Euclidean distance measure for the first dataset. In this example the image features have been computed without using a grid (image is one segment)

Table 1- Euclidean distance value for the first data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0363 %	96.36
Forged	55	0.0545 %	--------	94. 54

In table- 2 the result of Hellinger distance for the first dataset are used and the images without grid.

Table 2- Hellinger distance value for the first data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0363 %	96.36
Forged	55	0.0545 %	--------	94. 54

Table-.3 shows the result of Square Chord Distance for first dataset without grid.

Table 3- Square Chord Distance value for the first data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0545 %	94. 54
Forged	55	0.0545 %	--------	94. 54

Tables -4.4, 4.5, and 4.6 show the result of the second dataset, the verification techniques Euclidean distance, Hellinger distance, Square Chord Distance are used respectively.

Table 4- Euclidean distance value for the second data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0909 %	90.90
Forged	22	0.0909 %	--------	90.90

Table 5- Hellinger distance value for the second data set

Nature of Signature	No.of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0454 %	95.45
Forged	22	0.0454 %	--------	95.45

Table 6- Square Chord Distance value for the second data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0454 %	95.45
Forged	22	0.0909 %	--------	90.90

Tables - 7, 8, and 9 show the results for the third dataset which is collected from 50 people, each person have one signature original and one signature is forgery, when the verification techniques Euclidean distance, Hellinger distance, Square Chord Distance are used.

Table 7- Euclidean distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	--------	0.04%	96
Forged	50	0.04%	--------	96

Table 8- Hellinger distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	--------	0.06%	94
Forged	50	0.06%	--------	94

Table 9- Square Chord Distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	--------	0.06%	94
Forged	50	0.06%	--------	94

Euclidean distance gives good FAR and FRR values compare to all distances for these datasets.
In the proposed system the signature can be divided at preprocessing step into four equal segments for each segment feature. Tables 10,11 and 12 show the results for the first dataset when the verification techniques Euclidean distance, Hellinger distance, Square Chord Distance are used and image features are calculated using a grid.

Table 10- Euclidean distance value for the first data set

Nature of Signature	No.Of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0181%	98.18
Forged	55	0.0363%	--------	96.36

Table 11 -Hellinger distance value for the first data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0181%	98.18
Forged	55	0.0727%	--------	92.72

Table 12- Square Chord Distance value for the first data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	55	--------	0.0363%	96.36
Forged	55	0.2%0.0363%	--------	96.36

The tables- 13, 14, and 15 show the result of the second dataset.

Table 13 - Euclidean distance value for the second data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0454%	95.45
Forged	22	0.0454%	--------	95.45

Table 14- Hellinger distance value for the second data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0454%	95.45
Forged	22	0.0454%	--------	95.45

Table 15 - Square Chord Distance value for the second data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	22	--------	0.0454%	95.45
Forged	22	0.0909%	--------	90.90

Tables-16, 17, and 18 show the result of the third dataset.

Table 16- Euclidean distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	-- ------	0.04 %	96
Forged	50	0.04 %	-- ------	96

Table 17 - Hellinger distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	--------	0.06 %	94
Forged	50	0.02%	--------	98

Table 18- Square Chord Distance value for the third data set

Nature of Signature	No. of Samples	False Acceptance Rate	False Rejection Rate	accuracy rate
Original	50	--------	0.04 %	96
Forged	50	0.06 %	--------	98

Conclusions and future work

In this paper an offline handwritten signature verification system has been developed which use set of simple shape based geometric features. Before extracting the features, preprocessing of a scanned image is necessary. Features Extracted for whole signature first, then extracted for every part after dividing the signature into four parts. For verification, statistical verification techniques are used. The system is trained on three datasets of signatures. The system has been tested on every dataset. After the development, testing, and evaluation of the offline handwritten verification system the following conclusions can be drawn:

1- The work described in this paper concerns the application of different statistical techniques to classify the signature as original or genuine.

2- After collecting the dataset and seeing that it needs enhancement, this is done in the preprocessing step. At this step, the signature needs crop, resize, etc. in order to be ready for feature extraction. Using crop and resize makes different sizes of the same signature (in different images) identical and the following processing deals with signature only.

3-The proposed signature verification system is based on some special features of geometric feature extraction. Geometric feature is the best feature because it keeps both their global as well as local feature properties.

4-To obtain better results (as shown in tables.10-18) the signature divided into four parts, for each part the features are extracted.

5- There are several approaches for offline signature verification, each technique has its different advantages and disadvantages, depending on feature set selected for different techniques that can be utilized to obtain optimum results.

6 -The developed system can verify English and Arabic signatures, because the system deals with signature as image. The third data set contains Arabic and English signatures.

7-Statistical techniques are used because they are simple to use and give better results.

8- The developed system can verify English and Arabic signatures, because the system deals with signature as image. The third data set contains Arabic and English signatures.

9. We can conclude the following comparison based on accuracy between the verification techniques used in this work.

	Euclidean distance	Hellinger Distance	Square Chord Distance
With grid	93.54	95.87	95.93
Without grid	94.42	95.27	93.14

Future work will include the automation of off-line handwritten signature trajectory recovery;

1- In preprocessing step, may use additional operations such as thinning.

2- In Feature extraction can use feature energy information from more different parts of the signature appropriately organizing the extracted information.

3- There are many Verification techniques, that depends on machine learning which can be used such as neural network or hidden Markova model.

References

1. Diana K. **2003.**" Personal Authentication Using Signature Recognition" Department of Information Technology, Laboratory of Information Processing Lappeenranta University of Technology
2. Qi.Y, Hunt B.R.**1994.**"Signature Verification using Global and Grid Features", *Pattern*
3. *Recognition*, Vol. 27, No. 12, pp. 1621-1629.
4. Nguyen V., Blumenstein M., and Leedham G. **2009**." Global features for the off-line signature verification problem". International Conference on Document Analysis and Recognition, ICDAR ’09, pages 1300– 1304, Washington, DC, USA,. IEEE Computer Society.
5. Jacques P. S., Bschons**.2009**." off-line signature verification using classifier ensembles and flexible grid features .", Department of Applied Mathematics, Stellenbosch University South Africa .
6. Ranjan J. , Rituparna S. , Debaleena D. **2013**. " offline signature verification using euclidian distance", *International Journal of Computer Applications* ,Volume 10– No.2, November
7. Hai R. L., Wen J. Y., Dong J . **2009** .“ offline signature verification based on deformable grid partition and hidden markov models” IBM China Research Lab
8. Pooja, Rajan G.**2014** .“Offline Signature Verification system” , Department of Computer Engineering Yadavindra College of Engineering , *International Journal of Computer Application and Engineering Technology* Volume 3-Issue2.
9. Hetal V. Davda, Sima K. Gonsai. **2014**. "Offline Signature Verification System Using Energy on Grid Level", *International Journal of Engineering Research* , Volume No.3, Issue No.2.