

The 4NSigComp2010 off-line signature verification competition: Scenario 2

M. Blumenstein¹, Miguel A. Ferrer², J.F.Vargas³

¹School of Information and Communication Technology
Griffith University, Gold Coast campus 4222 Q Australia

²Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones
Universidad de Las Palmas de Gran Canaria

Campus de Tafira s/n, E35017 Las Palmas de Gran Canaria, Spain

³Departamento de Ingeniería electrónica, GEPAR, Universidad de Antioquia, Medellín, Colombia.

Abstract—The objective of this competition (4NSigComp2010) is to ascertain the performance of automatic off-line signature verifiers to evaluate recent technology developments in the areas of document analysis and machine learning. The current paper focuses on the second scenario, which aims at performance evaluation of off-line signature verification systems on a newly-created large dataset that comprises genuine, simulated signatures produced by unskilled imitators or random signatures (genuine signatures from other writers). Ten systems were evaluated, and some interesting results are presented in terms of accuracy and execution time. The top ranking system attained an overall error of 8.94%. This result interestingly correlates with the top ranking accuracy achieved in a previous signature verification competition at ICDAR 2009.

I. INTRODUCTION

Automatic verification of a signature, a behavioural biometric, can be performed using a tablet with a stylus or using the signature of a scanned image. The former is called on-line verification and the latter is called off-line verification. Off-line verification has several advantages over its on-line counterpart. Firstly, it has widely been accepted in society. Secondly, it is more convenient as it does not require any special instruments. Thirdly, with the large amount of bank cheques, credit card authorisation forms, or legal documents still being signed every day, off-line verification can be considered commercially important.

Scenario 2 of the 4NSigComp2010 competition at ICFHR 2010 has been formulated given the intent interest in off-line signature verification. The second scenario has been primarily concerned with the detection of skilled versus non-skilled simulated signatures and aims at evaluating the performance of signature verification systems in a security-less critical environment. The questioned signatures can either be genuine (written by the reference writer), or forged (simulated by other writers than the reference writer), or a random forgery (genuine signature of other writers).

The current paper reports the outcomes for Scenario 2 of the 4NSigComp2010 competition. The remainder of this paper is divided into five sections. Section 2 outlines the participants in this competition, Section 3 discusses the signature database employed, Section 4 elaborates on Systems Evaluation and Section 5 presents the results obtained. Finally, Section 6 provides some concluding remarks.

II. PARTICIPANTS

After the competition announcement, 15 teams showed their interest in participating in the competition. Of those 15 teams, 7 submitted their programs: 3 from academia, 1 from a governmental institution and the remainder from industry. Two groups submitted several programs; therefore the competition evaluated 10 different systems. The teams are from 7 different countries: India, Canada, Turkey, U.S.A, Austria, Spain and France. Table 1 shows the participants' groups with the identification of the programs submitted.

TABLE I. 4NSIGCOMP 2010 SCENARIO 2 PARTICIPANTS

Institution	Country	Team coordinator	System id
Indian Statistical Institute	India	Rajesh Kumar	1
École de technologie supérieure, Montreal	Canada	Luana Bezerra	2
Sabancı University	Turkey	Berrin Yanikoglu	3,4,5
Parascript LTD	USA	Tim Strunkov	6
Anonymous	Austria	Anonymous	7
Universidad Autónoma de Madrid	Spain	Fernando Alonso-Fernandez	8
NIFISOFT	France	Ali Hassaine	9,10

A. Indian Statistical Institute

The team composed of Rajesh Kumar, Lopamudra Kundu and Bhabatosh Chanda, from the Electronics and Communication Science Unit of the Indian Statistical Institute, India, has sent an automatic signature verifier (system id 1) with three kinds of features namely morphological feature, invariant moment based feature and entropy based feature. After doing the necessary pre-processing, all the three features are extracted from off-line signature images. A pair of signatures is fed to the system and inference is made for their similarity or dissimilarity. Support vector machines (SVMs) are used as a classifier (verifier) for same. Results of three classifiers against the three features are fused to get the final result.

B. École de technologie supérieure, Montreal

This system received by the École de technologie supérieure (ÉTS), Montreal, Canada, was designed by Dominique Rivard, PhD candidate, with help of Luana Batista, Eric Thibodeau, Eric Granger and Robert Sabourin. Their automatic signature verifier (system id 2) is based on multiscale feature extraction, dichotomy transformation and boosted feature selection. Multiscale feature extraction increases the diversity of information extracted from the signature, thereby producing features that mitigate intra-personal variability, while dichotomy transformation ensures writer-independent classification, thus relieving the verification system from the burden of a potentially very large number of users. Finally, using boosted feature selection it allows for a low cost writer-independent classification system that selects features while learning. As such, the proposed system provides a practical framework to explore and learn from problems with numerous potential features.

C. Sabancı University

The Biometrics research group at Sabancı University has been active in online signature verification, as well as biometric privacy and template protection areas. It is a small group headed

by Prof. Yanikoglu, Dr. Kholmatov and graduate students. They have developed 3 base systems specifically for this competition. Two of them are closely related as they share the same normalization and feature extraction steps, but differ only in classifier training. The features consist of gradient orientation histograms obtained from the tessellated signature. For the Global-SVM (system id 3), they train a user-independent SVM to learn important elements of the high dimensional difference vector between the feature vectors of the query signature and the closest reference. For the User-SVM system (system id 4), we trained an SVM with the features of the reference signatures of a person, against random forgeries. The third system (system id 5) is based on a normalized correlation, so as to complement the two previous systems.

D. Parascript LTD

Parascript is a leading pattern recognition software company, providing high-performance solutions in many fields, including fraud prevention. They are a team of researchers and programmers with different backgrounds. Their submission (system id 6) used an ensemble of several verifiers based on different known techniques (neural networks, dynamic programming, the Radon transform, HMM, and others) and also on Parascript's proprietary techniques. The individual verifiers' results are then merged by a voter system that also uses the comparisons between the reference signatures to account for their stability.

E. Anonymous

The system id 7 uses three different approaches to measure the similarity of two signatures: The first similarity images compare the distance between the respective radon transforms found by dynamic time warping using the difference between the number of pixels in the histograms as the metric. The second measured image is based on blurring: The sample image is 'blurred' by marking all pixels within a horizontal or vertical distance of span from pixels belonging to the sample signature. The test image is then overlaid and the percentage of pixels in the test image that are on top of marked pixels is calculated. The third similarity measure maps the signature onto a grid of 100x80 and then onto a 'blurred' grid of the same dimensions. The distance of a test signature from the sample signature is given by the sum of the minimum distance between each cell in the sample 'blurred' grid and cells within a square of 9x9 cells fitted over the corresponding cell in the test 'blurred' grid. The result is calculated by weighting these 3 similarity measures.

F. Universidad Autónoma de Madrid

The Biometric Recognition Group - ATVS at Escuela Politecnica Superior of the Universidad Autonoma de Madrid (UAM) is devoted to research in the areas of biometrics, pattern recognition, image analysis, and speech and signal processing, with application to person authentication and forensics. The research activities of the ATVS group involve

several biometric traits: speaker recognition, fingerprint, signature verification, handwriting, hand biometrics, iris recognition and multimodal fusion.

The off-line system submitted (system id 8) is based on the fusion of three machine experts, one based on global information, a second one based on local analysis of the image and a third approach based on allographic analysis. To compute the similarity between the global and local features of two signature images, the χ^2 distance is used. The matcher based on allographic analysis is computed using a common codebook of shapes obtained by means of clustering techniques. Finally, fusion of the three machine experts is performed via linear combination of the individual scores. Linear regression is used to compute the optimal fusion weights.

G. Nifisoft

Nifisoft has submitted two systems. Nifisoft is a startup company headquartered in Saint-Etienne, France and provides solutions in handwriting recognition, graphology, signature verification and document image processing. Their automatic signature verifier computes several features based on the number of connected components, number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures and chain codes. Each feature F_i is computed for the questioned signature $F_i(q)$ and the N reference signatures $F_i(r)$ ($r=1..N$). The average absolute difference between the value of the feature F_i in the questioned signature and its values in the reference signatures is then computed:

$$D_i = \frac{\sum_{r=1}^N |F_i(q) - F_i(r)|}{N}$$

The differences obtained are combined via a logistic regression classifier trained either on the 4NSigComp2010 database (partial training method, system id 9) or on both 4NSigComp2010 and SigComp09 databases (full training method, system id 10).

III. SIGNATURE DATABASE

For this competition, a subset of the GPPDS960signature database has been used.

A. GPDS960Signature corpus

The off-line signature GPDS960signature database contains data from 960 individuals: 24 genuine signatures for each individual, plus 30 forgeries of his/her signature. The 24 genuine specimens of each signer were collected in a single day writing sessions. The forgeries were produced from the static image of the genuine signature. Each forger was allowed to practice the signature for as long as s/he wishes. Each forger imitated 3 signatures of 5 signers in a single day writing session. The genuine signatures shown to each forger are chosen randomly from the 24 genuine ones. Therefore for each genuine signature there are 30 skilled forgeries made by 10 forgers from 10

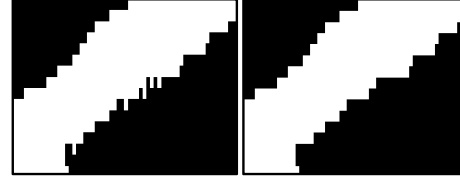


Figure 1. Eliminating the hair sticking out from signature strokes noise. Left: signature stroke detail of $I_{bw}(x, y)$ with noise, right: same signature stroke detail of $I_{NR}(x, y)$ without noise.

different genuine specimens. Each signer used their own pen.

The signatures has been scanned at 300 dpi in 256 gray scale levels, binarized and saved in "bmp" format. The files of the genuine signatures of xxx signer are named xxx\c-xxx-yy.bmp and the files of its forgeries are named xxx\cf-xxx-yy.bmp.

As the background of the scanned signatures is well contrasted with the darker signature strokes, the signature images where binarized by thresholding. Let $I(x, y)$ be a 256-level grey scale signature image of the database, a fixed threshold equal to 222 was selected to binarize the image obtaining:

$$I_{bw}(x, y) = \begin{cases} 0 & \text{if } I(x, y) > 222 \\ 255 & \text{otherwise} \end{cases}$$

The black and white image $I_{bw}(x, y)$, where the strokes are white and the background black, display a sort of hair sticking out from signature strokes as can be seen in Fig. 1. As this noise can fake the signature trace border we eliminate them as follows:

$$I_{NR}(x, y) = \begin{cases} 0 & \text{if } I_{bw}(x-1, y) = 0 \\ & \text{and } I_{bw}(x, y) = 255 \\ & \text{and } I_{bw}(x+1, y) = 0 \\ I_{bw}(x, y) & \text{otherwise} \end{cases}$$

$$I_{NR}(x, y) = \begin{cases} 0 & \text{if } I_{bw}(x, y-1) = 0 \\ & \text{and } I_{bw}(x, y) = 255 \\ & \text{and } I_{bw}(x, y+1) = 0 \\ I_{bw}(x, y) & \text{otherwise} \end{cases}$$

which corresponds to or-exclusive operations. In both cases $1 \leq x \leq N, 2 \leq y \leq M-1$ the above specified operation converts the white pixels to black if the left and right pixels are black or the upper and lower pixels are black. Fig. 1 shows an example of the above mentioned operation on a signature stroke

Finally, the signature traces were converted to black and the background to white, $I_s(x, y) = 255 - I_{NR}(x, y)$, and the image $I_s(x, y)$ was saved as a database image.

Some statistics of the GPDS960signature database: The age ranges from 16 to 73. The most of the user are belong 18 and 25. See Table II. The gender is more or less half male and half female. The left handwriting people are the 8% approximately.

The 300 first signer signatures of this database were already freely available with the name of the GPDS300signature corpus

(<http://www.gpds.ulpgc.es/download/index.htm>). The GPDS960signature corpus is freely available in the same conditions as the GPDS300signature database.

TABLE II. AGE DISTRIBUTION OF GPDS960SIGNATURE DATABASE

Age	Number	Percentage
<18	219	22.81
18-25	426	44.38
25-32	138	14.38
32-39	76	7.92
39-50	71	7.4
>50	30	3.13
Total	960	100

TABLE III. GENDER DISTRIBUTION OF GPDS960SIGNATURE DATABASE

Gender	Number	%
Male	470	48.96
Female	490	51.04

TABLE IV. HAND WRITING DISTRIBUTION OF GPDS960SIGNATUREDATABASE

Hand writing	Number	Percentage
Left handed	82	8.54
Right handed	878	91.46

B. The 4nSigComp 2010 Scenario 2 database

The database used for the 4nSigComp 2010 scenario 2 consists of signatures of the signers 301 to 960 of the GPDSSignature corpus, which are not public. For training, 4 genuine signatures of 301 to 700 from the GPDS960Signature corpus were provided to each participant. The files of the genuine signatures are named xxx\c-xxx-yy.bmp being xxx the *id* of the signer which goes from 301 to 700 and yy the repetition from 01 to 04

A Matlab script to read and display the images of the database can be seen in Program1.

The testing data contains 30000 questioned signature images obtained from the GPDS960signature database. The test files has been named c-xxxx-yyy.bmp being xxxx the number of file from 00001 to 30000 and yyy the *id* of the signer identity claimed from 301 to 700.

The test data includes original signatures of GPDS960signature signers 301 to 700, random signatures and simulated forgeries of each user. Hereby, random signatures are genuine signatures belonging to different writers out of the genuine users. A simulated forgery is a reasonable imitation of the genuine signature model.

Concisely, the test genuine signatures are the remainder 20 genuine signatures, no submitted for training, of the 400 signers, the simulated forgeries were the 30 forgeries of each signer, and the random forgeries are randomly extracted from the genuine and forgeries of users 701 to 960. So the shape of the random forgeries is not seen in the training. Therefore the test consists of $400 \times 20 = 8000$ genuine tests, $400 \times 30 = 12000$ simulated

PROGRAM 1. Matlab script to read and display the images of the training database

```

path='TrainingSet\';
for ift=301:700
    fprintf('signer: %g\n',ift)
    for irf=1:4
        f1=[path,num2str(ift,'%3d')];
        f2=['\c-',num2str(ift,'%3d')];
        f3=['-',num2str(irf,'%2d')];
        nfichd=[f1,f2,f3];
        I=imread(nfichd,'bmp');
        imshow(I)
        drawnow
        pause
    end
end

```

forgeries tests and we selected 10000 random forgeries test.

The training and test databases will be freely available after the ICFHR2010 conference in the same conditions that GPDS300signature database.

IV. SYSTEMS EVALUATION

As the competition aim is to measure the performance of the automatic signature verifier (ASV) tools developed by the participants in an operational environment, the participants were asked for an ASV program, which should calculate a score of the questioned signature belonging to the claimed identity and compared the score with an own threshold giving a decision. As a decision, the outputs expected were number 1 in the case of 'accept' the questioned signature belonging to the pretending signer, and number 0 in the case of reject that the questioned signature belongs to the pretended signer.

Each participant submitted a software tool called *asv.exe* with:

Input parameter: the id of the pretended signer between 301 and 700. The questioned signature will be in the file "*signature.bmp*"

Output: the file "*decision.txt*" containing the number 1 in the case of accept or the number 0 in the case of reject.

The systems submitted can be evaluated using the Matlab script program 2. The file *4nSigCompSignIdent.mat* contains the matrix *sign* of dimension 30000 by 4. For each test signature it contains the first and second row, which in turn contains the real signature number and repetition in the GPDS960Signature corpus, the third row contains the identify claimed and the fourth row contains the code of the experiment: 0 for FRR test, 1 for FAR of simulated forgery, and 2 for FAR of random forgery test. When the code is 2, if the repetition number is greater than 24, it refers to the simulated forgery number of repetitions minus 24.

PROGRAM 2. Matlab script to evaluate the systems submitted.

```

path='TestSet\';
time=zeros(30000,1);

numberHITS=0;
nFRR=0;
nFARCasual=0;
nFARskilled=0;

load 4nSigCompSignIdent sign

for isign=1:30000
    f1=['c-',num2str(isign,'%0.5d'),'*'];
    file=dir([path,f1]);
    IdClaimed=str2num(file.name(9:11));
    fprintf('sign number: %g',isign)
    fprintf('Id Claimed: %g',IdClaimed)
    I=imread([path,file.name]);
    imwrite(I,'signature.bmp','bmp')
    tic
    dos(['asv ',num2str(IdClaimed)]);
    time(isign)=toc;

    % Check that decision.txt file
    % has been written later than
    % signature.bmp file
    tc=dir('decision.txt');
    tg=dir('signature.bmp');
    td=datetime(tc.date)-datetime(tg.date);
    if td<0;
        fprintf('decision no written\n');
        keyboard;
    end
    load -ascii decision.txt

    % evaluate decision taken by asv.exe
    A=ne(decision,1);
    B= sign(isign,4)==1
    B=or(B,sign(isign,4)==2);
    if and(decision==1,sign(isign,4)==0)
        fprintf(' OK\n')
        numberHITS=numberHITS+1;
    elseif and(A,B)
        fprintf(' OK\n')
        numberHITS=numberHITS+1;
    else
        if sign(isign,4)==0
            fprintf(' FRR\n')
            nFRR=nFRR+1;
        elseif sign(isign,4)==2
            fprintf(' FAR casual\n')
            nFARCasual=nFARCasual+1;
        else
            fprintf(' FAR skilled\n')
            nFARskilled=nFARskilled+1;
        end
    end
end
fprintf('HITS number:%g\n',numberHITS)
fprintf('FRR number:%g\n',nFRR)
fprintf('FAR Casual:%g\n',nFARCasual)
fprintf('FAR Skilled:%g\n',nFARskilled)

```

TABLE V. NUMBER OF GENUINE SIGNATURES FALSELY REJECTED (*nGFR*) AND NUMBER OF SIMULATED AND RANDOM FORGERIES FALSELY ACCEPTED (*nSFA* AND *nRFA* RESPECTIVELY)

ID	Coordinator name's	nGFR	nRFA	nSFA
1	Rajesh Kumar	834	918	4702
2	Luana Bezerra	1765	7	5552
3	Berrin Yanikoglu	3029	15	3408
4		3266	9	3403
5		3347	9	3522
6	Tim Strunkov	117	1	937
7	Anonymous	3686	24	3567
8	Fernando Alonso-Fernandez	1440	431	3046
9	Ali Hassaine	1719	122	2741
10		895	96	5736

TABLE VI. FALSE REJECTION RATIO (FRR), FALSE ACCEPTANCE RATIO OF SIMULATED FORGERIES (FARS) AND FALSE ACCEPTANCE RATIO OF RANDOM FORGERIES (FARR) ALONG WITH THE OVERALL ERROR (OE) IN PERCENTAGE (%).

ID	Coordinator name's	FRR	FARR	FARS	OE	Rank
1	Rajesh Kumar	10.43	9.18	39.18	17.31	4 th
2	Luana Bezerra	22.06	0.07	46.27	22.62	6 th
3	Berrin Yanikoglu	37.86	0.15	28.4	26.07	7 th
4		40.83	0.09	28.36	27.53	8 th
5		41.84	0.09	29.35	28.28	9 th
6	Tim Strunkov	13.96	0.01	7.81	8.94	1 st
7	Anonymous	46.08	0.24	29.73	30.53	10 th
8	Fernando Alonso-Fernandez	18	4.31	25.38	16.42	2 nd
9	Ali Hassaine	21.49	1.22	22.84	16.76	3 rd
10		11.19	0.96	47.8	17.79	5 th

V. RESULTS

Performance was evaluated in terms of Overall error which is calculated from Type I error (False Rejection) and Type II error (False Acceptance) with simulated and random forgeries.

The overall error (OE) is calculated as:

$$OE = \frac{1}{2} \times \frac{nGFR}{nG} + \frac{1}{4} \times \left(\frac{nSFA}{nSF} + \frac{nRFA}{nRF} \right)$$

supposing that:

nG: number of genuine signatures in the test set, equal to 8000.

nSF: number of simulated forgeries in the test set, equal to 12000.

nRF: number of random forgeries in the test set, equal to 10000.

nS: number of Signatures in the test set. Obviously: $nS = nG + nSF + nRF$ and equal to 30000.

and being:

nGFR: number of genuine signatures falsely rejected.

nSFA: number of simulated forgeries falsely accepted.

TABLE VII. FALSE ACCEPTANCE AND FALSE REJECTION RESULTS WITH NO EXPERT PEOPLE AND FORENSIC EXPERT

	No. Expert	Forensic Expert	
			First 17 test
Number of Participants	14	1	
Number of Test	280	20	First 17 test
Number of Genuine Signatures Presented	145	11	10
Number of Simulated Forgeries Presented	135	9	7
Number of False Accepted Signatures	10	1	0
Number of False Rejected Signatures	15	2	0
Total Error (%)	8.88	14.65	0

$nRFA$: number of random forgeries falsely accepted.

The obtained results in terms of $nGFR$, $nSFA$ and $nFRA$ can be seen in Table V. The False rejection Ratio (FAR), False Acceptance Ratio of simulated forgeries ($FARS$) and False Acceptance Ratio of random forgeries ($FARR$) and the Overall Error (OA) are displayed in Table VI along with the rank in terms of OA .

As a curiosity, comparing the Table VI results with the on line ICDAR2009 Signature Verification Competition, Parascript LLC (USA, system id 1) has achieved first place in both Signature Competitions with a considerable difference in comparison to the second place attained in both competitions by the Biometric Recognition Group from Universidad Autónoma de Madrid (Spain, system id 8) [1].

Although the statistic relevance is limited, with the aim to know the human error with the database, the next two experiments were carried out.

The first experiment has been done with no expert people. We have developed a program that randomly selects a signer and presents his/her four training genuine signatures to the user. After giving as long as the user wishes for observing the signatures, a questioned signature is presented to the user. As the questioned signature is a genuine or simulated forgery, the user has to decide if the questioned signature is genuine or a forgery. Each user has to decide to accept or reject 20 signatures from 10 different signers. Each user has taken an average of 20 minutes to answer all the questions.

The second experiment was carried out performing the same test, but in this case the user is a trained forensic handwriting expert who works at the Spanish courts. The results with him are also in table VII. He takes about 90 minutes to answer all the questions, much longer than the non-expert people.

The result given in table VII with no expert people (8,88%) is similar to the OE of the Parascript LLC system (8,94%).

Obviously, it is a surprise that the error with the forensic expert is higher than the error with no

TABLE VIII. EXECUTION TIME FOR EACH SUBMITTED SYSTEM

ID	Coordinator name's	Averaged Execution Time (sec.)	Standard Deviation of Execution
1	Rajesh Kumar	4.43	1.45
2	Luana Bezerra	10.82	0.76
3	Berrin Yanikoglu	20.65	2.1
4		19.99	2.14
5		22.66	2.42
6	Tim Strunkov	2.27	0.45
7	Anonymous	2.6	0.66
8	Fernando Alonso-Fernandez	7.31	1.04
9	Ali Hassaine	1.64	0.62
10		1.66	0.62

expert people. It should be taken into account that the error with no expert people is distributed during the test but the mistakes of the forensic expert were done with the last 4 test signatures. As the forensic expert takes longer to decide about each questioned signature, he got tired and started to make mistakes at the end. With the first 17 signatures, the forensic expert error was 0%.

The execution time is another variable that we should take into account for realistic environments. The proposed evaluation script measures the execution time using *tic* and *toc* Matlab functions. Although this way of working includes some bias in the time measures, it may be useful for giving an order of each algorithm time requirements. Additionally, consider that while several systems could have been optimized (for instance, programmed in a language such as C directly) other ones have been compiled from Matlab, which is less efficient in terms of execution time. Table VIII shows some execution time statistics for the submitted systems.

In general, the execution time is not the same for each signature. The time of the *asv.exe* program with the same signature does not change when modifying the identity claimed input variable. So the execution time does not depend on whether the input signature is genuine or random. As a general rule, we have checked that the more complex a signature is the more time is used to undertake verification.

VI. CONCLUSION

In conclusion, scenario 2 of the 4NSigComp2010 competition had very positive participation with ten systems being submitted and evaluated on a newly created large database of off-line signatures. Overall, the systems performed well in terms of speed and accuracy, with the top results being attained by a system, which likewise produced the top result in the ICDAR 2009 Signature Verification competition.

REFERENCES

- [1] V.L.Blanker, C.E.van den Heuvel, K.Y. Franke, L.G.Vuurpjl, "The ICDAR 2009 Signature Verification Competition", in *10th International Conference on Document Analysis and Recognition*, pp. 1403-1407, Barcelona, Spain, 2009.