SID Signature Database:
A Tunisian Off-line Handwritten Signature Database

Imen Abroug Ben Abdelghani[1] and Najoua Essoukri Ben Amara[2]

[1] Higher Institute of Applied Sciences and Technology of Kairouan,
UR: SAGE-ENISo, University of Sousse, Tunisia
`abrougimen@yahooo.fr`
[2] National Engineering School of Sousse, UR: SAGE-ENISo, University of Sousse, Tunisia
`najoua.benamara@eniso.rnu.tn`

Abstract. Our research works are concerned with checking the off-line handwritten signature. We propose a base of Tunisian static handwritten signatures. In this paper, we go through the phase of designing the SID-Signature database and its various features. Afterwards, we present the results obtained by this base as part of our system of handwritten-signature verification.

Keywords: off-line handwritten signature, SID-Signature database, Tunisian signature, Planar modeling signature.

1   Introduction

The biometric verification of the handwritten signature remains one of the main approaches to authenticating the identity, which are present in various fields of the modern society such as financial and legal transactions. The static handwritten signature has always been used by different social classes to authenticate and certify all sorts of official documents and make individuals aware of their commitments (contracts, financial and legal transactions, etc.).

The signature is a draw resulting from a more or less voluntary movement that is specific to each individual and which allows certifying the documents they append. The signing action has existed for a very long time and has known a significant geometric variation since the Middle Ages. Indeed, to certify documents, people in the Middle Ages used the signatures, crosses or monograms, to which they could add initials or parts of their names [1]. Fig1 gives examples of signatures used at that era.
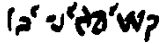

Fig. 1. .Sample signatures of the Middle Ages

The handwritten-signature morphology changes from one person to another as well as for the same one. In fact, the signatures of the same person may show shape variations depending on their age, physical and moral state, the material conditions, and the nature of the document (official or unofficial) [2-3].

However, according to the country's culture, signatures may be very different in their shape. The literature shows that we distinguish the classes of European, North American, Asian and Arabic signatures.

According to [4], the Arabic signatures have partly the characteristics of the cursive writing as well as those of the graphic initials, which makes them more complicated. In our case, we are interested in the Tunisian signatures. Table 1 shows some forms of signatures classified according to their origin.

Table 1. Examples of different-origin classified signatures

Origin/style	Samples		
North-Americans/ handwriting			
Europeans/ Graphical lines			
Asians/ Asian characters			
Arabs/ Graphics, writing			

In this context, several automatic systems of checking the handwritten signature have been proposed; the role of such a system is to determine the authenticity of a given signature by referring to a beforehand-conceived database. Various papers have run over the conducted research, [5-8].

In general, the systems checking the signature are learning-based, which requires to have a large database ideally representing the majority of the target population of the concerned application. The database samples must take into account all the possible variations in the real case of applications such as the intrapersonal variations. There are several databases of static signatures published for academic use and scientific research. We cite, for example the SIGMA [5], GPDS960 [6], GPDS960 Gray [7], MCYT-75 [8], SigComp09 [9], 4NSigComp2010 [10], SigComp11 [11] and 4NSigComp2012 [12] bases. We give in Table 2 an overview of the main features of some static signatures databases.

Table 2. The main characteristics of some static-handwritten-signature databases

Ref\ Base	Origin	Signatory Number		Type of signatures	Characteristics
[5] SIGMA	Malay-sian	213		Simulated	
[8] MCYT	Spanish	75			600dpi, 256 gray level
[6] GPDS960		960			300dpi, BMP
[7] GPDS960 gray		960 (881)			600dpi, PNG
[9] SigCom09		App: 12	Test: 100		300dpi, 600dpi, PNG
[10] 4NSigComp2010 (Subset of GPDS960)	Spanish	400			300dpi, BMP
[11] SigCom11	Chinese Dutch	12			400 dpi, PNG
[12] 4NSigComp2012		2	3	Simulated Disguised	300dpi, BMP

As showing in the literature and Table 2, most databases contain particularly the type of forgery by simulation; it is always the most difficult case to solve. Similarly, the bases are more and more composed of a large number of signatories to be closer to the real situation of the applications. We also note that the different research teams set up their own databases to validate their signature verification / identification systems. This can be explained by the lack of databases of international static handwritten signatures and the need to have a database tallying with the population of the country concerned with the corresponding real application.

As part of our research works, we have been brought to build a database of North African offline handwritten signatures, taking into account the characteristics purely specific to the image of the North African signature. In this paper we are confined to the sub-base tallying with the Tunisian population; we present the SID-Signature database by exploiting it in the case of validating a planar modeling of the static handwritten signature. In the following section, we detail the phase of designing the database and its various features. Section 3 is devoted to the morphological study of the base. Section 4 presents the results of using this basis for validating a planar architecture of the handwritten signature. The conclusion and the perspectives are given in Section 5.

2   Design of the SID-Signature Database

To design this database we have proceeded by a collection of signature samples followed by a phase of digitizing and storing data. We describe in what follows the different adopted choices.

2.1   Collecting Signatures

The different signature samples have been affixed by means the same make of blue ballpoint pen on the same make of white A4 paper forms divided into 15 blocks of the same size. The choice of using 15 blocks to append the signatures has been the basis of studying a large number of signatures. The space reserved for appending is sufficient to draw different-sized signatures.

The database has currently 6,000 genuine signatures corresponding to 100 Tunisians with different ages and cultural and scientific levels. Each signatory has provided 60 samples, in different sessions, to avoid possible geometric variations related to the appending close to the time of signatures. Indeed, the signatory risks being tired from signing several times in succession or from being pressed to finish, which leads to significant variations in the form of their signature. Table 3 shows some statistics of   the SID-Signature database.
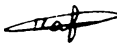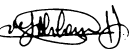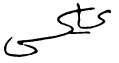
Table 3. Some statistics of the SID-Signature database

SID-Signature Database	
Gender distribution	64% male / 36% female
Age distribution	10% (<18); 34% (18-25); 28% (25- 35); 21% (35-50); 7% (> 50)
Scientific level	17% employee; 46% student; 37% teacher

10 forgers (persons different from writers) have been solicited to collect the forged signatures. Each forger has provided, in different sessions, 2 simple forgeries and 2 ones by a skilled imitation of each signature class.

To produce a skilled imitation, we have provided the forger with different static images of the genuine signature to imitate, and he produce the skilled forgery without practicing imitation.

Table 4. Samples of the SID-Signature database

Signatory	Genuine	Skilled Imitation	Simple Forgery	Random signature
Name: Sahnoun Surname: Mohamed				
Surname:Wafa				
Surname: كريم				

The simple forgery corresponds to writing the surname and / or the name of the genuine signatory (which have been provided to them); this gesture is often adopted by the majority of this type of falsifiers (Table 4).

It should be noted that the scrutiny of the two forgery types takes into account the highly variable shape of the Tunisian signatures (writing, graphics, and a blend between the two).

The SID-Signature database is subdivided into a learning sub-base and a test one. The learning base is composed of 4,000 genuine signatures (40 first samples corresponding to each signatory). The test sub-base consists of 2,000 authentic signatures, 2,000 simple forgeries, 2,000 skilled-imitated forgeries, and 2,000 random signatures: The random signatures correspond to the authentic signatures belonging to other writers (20 samples for each type) (Table 5).

Table 5. Distribution of samples of the SID-Signature database for modeling evaluation

| Sets of data	Number of samples	Type of signatures
Learning	4,000 (40x100)	Genuine
Test	2,000 (20x100)	Genuine
	2,000(20x100)	Random
	2,000(20x100)	Simple forgeries
	2,000(20x100)	Skilled-imitated forgeries

2.2   Signature Digitization

The different forms of the collected signatures have been digitized at a 300 dpi resolution with 256 gray levels. We have used for this three scanners: a HP 3200C, HP G2710 and EPSON DX4400.

Extracting the signatures of each form takes place automatically. To do this, we have developed a segmentation module based on exploring mainly the horizontal and vertical projections of the image of the considered form. The result images are stored in the SID-Signature database in a bitmap format in black and white (Fig 2). This database will be accessible to the public from the site http://www.sage-eniso.org/equipe/fr/1/SID-Signal-Image-et-Document.html

3   Morphologic Study of the Signature Database

The morphological study of the samples of the SID-Signature database shows the presence of the three signature styles previously mentioned in section 1: 14% of the
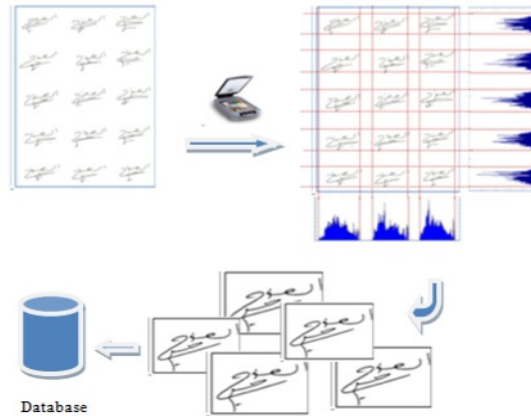
Fig. 2. Process of acquiring and extracting signatures database

signatories sign with writing (their surnames and / or forenames), 27% sign with pure-ly-random graphical forms, and the majority sign with a mixture parts of their sur-names and / or names linked to graphical lines.

We have noted an intrapersonal variation that is more or less important for all the signatories. This is a natural characteristic of the handwritten-signature drawing.

We have even noticed strong similarities between the shapes of some signatures classes, which can then bring about confusion during verification (Fig 3). The two dimensions of the various signatures of the base are highly diverse and range respec-tively from 86 to 670 and 137 to 812 (Table 6).

Table 6. Distribution of some characteristics of the SID-Signature database

Widths distribution	3% (131.5-190) ; 35% (190-365) ; 36% (365-482) ; 22% (482-598) ; 4% (598-715)
Heights Distribution	7% (74-118) ; 38% (118-205) ; 41% (205-336) ; 14% (336-510.5)
Slant distribution	4% (< -12.2°) ; 16% (-12.2°… -1.35°) ; 35% (-1.35°… 9.54°) ; 38% (9.54°- 42.2°) ; 7% (> 42.2°)

Signatories' n° 8 and 63   Signatories' n° 17 and 43

Fig. 3. Similarities between the shapes of signatures corresponding to different persons

4   Exploiting the Database by a Planar Architecture of the Handwritten Signature

Our research is about checking the handwritten signature by a planar modeling [13-14]. The choice of the planar modeling allows considering the two dimensions of the image in a simplified way. The signature image pre-processed beforehand (binarization, median-filtering, and size normalization) is segmented into three horizontal bands in order to bring the successive lines containing approximately homogeneous information together in the same band. The number of 'three' has been chosen following a study of the morphological changes in each class signature. Each band is modeled by a secondary modular MLP model. A vertical main modular MLP model ensures thereafter the correlation between the three secondary models. Thus, each signature class has been associated with a specific neuronal planar model.

We use two basic global feature types: geometric features and textural ones extracted from the application of the wavelet transform. Each secondary model has a six-characteristic input vector; these characteristics represent the number of black pixels and wavelet features of the corresponding band. As wavelet features, we retain the mean and the standard deviation of the approximation image and the standard deviation of the horizontal, vertical and diagonal detail images.
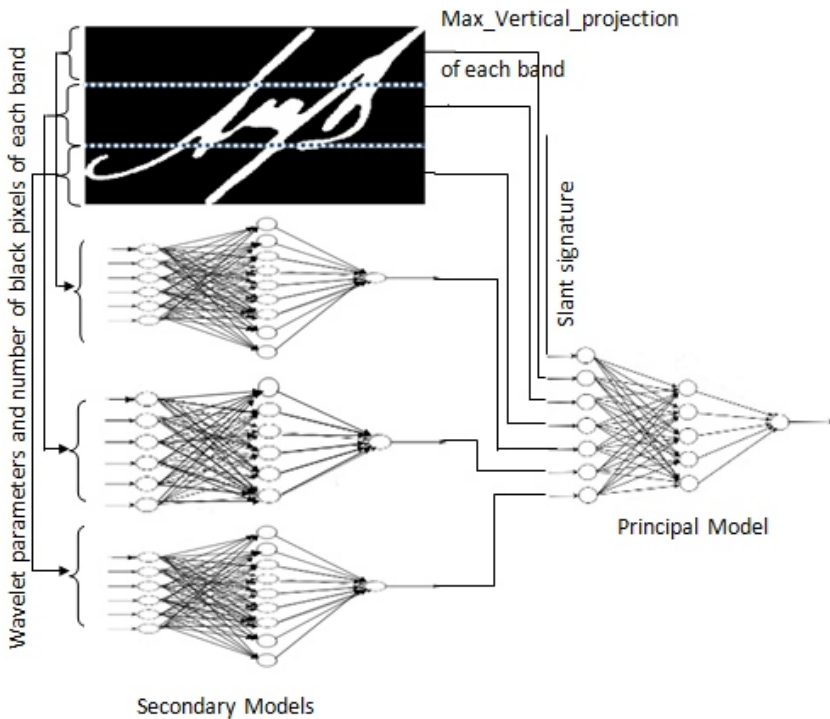

Fig. 4. The Planar model Architecture associated to the signatory n°100

The input vector of the principal model contains seven characteristics corresponding to the outputs of the three secondary models, the signature orientation, and the number of black pixels, in the highest line, of the vertical histogram for each band (Fig 4).

Several experiments have been carried out to determine the best parameters of each model tallying with each signatory (the number of hidden layers and hidden neurons of the principal and secondary models). We retain an MLP structure with three layers for both secondary and principal models. The number of hidden neurons varies according to the signatory class from 5 to 13. A particular learning procedure with the feed-forward-back propagation algorithm has been necessary [13-14]; we first carry out the learning of the three secondary models separately, and then we carry out the training of the principal model.

The training of the secondary models is carried out with an equal number of genuine and false samples (random genuine signatures of other persons) for each model. The training of the principal model is carried out in a particular way. In each algorithm iteration, the outputs of the three secondary models as well as the retained geometric features will then be injected as inputs for training the principal model. The experimentations give an FRR of 14.95%, a random FAR of 3.79%, a simple FAR of 1.25%, and a skilled FAR of 8.95%.

5   Conclusion

In this paper we have proposed a database of Tunisian static handwritten signatures for research purposes. The conceived database consists of 100 signatories and 10 forgers: 60 authentic samples and 20 forgeries for each signatory. The morphological study of the base has revealed the presence of the three signature styles (writing, graphics and hybrid), particularly as a characteristic of the Arabic signature. We have also noted the diverse intra-and-interpersonal variations between the different samples of the base. We have used this base to validate our system of checking handwritten signatures, which is based on a planar modeling of the signature image. The introduction of new signatures is underway to expand the database with North-African signatures.

References

1. Wikipedia, `http://wikipedia.fr`
2. Impedovo, D., Pirlo, G.: Automatic Signature Verification: The State of the Art. Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38(5), 609–635 (2008)
3. Plamondon, R., Srihari, S.N.: On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Transaction on Pattern Analysis and Machine Intelligence 22(1) (2000)
4. Sabourin, R., Genest, G.: Définition et évaluation d'une famille de représentations pour la vérification hors-ligne des signatures. Traitement du Signal 12(6) (1995)

5. Ahmad, S.M.S., Shakil, A., Ahmad, A.R., Agil, M., Balbed, M., Anwar, R.M.: SIGMA- A Malaysian Signatures Database. In: International Conference on Computer Systems and Applications, Doha, pp. 919–920 (2008)

6. Vargas, J.F., Ferrer-Aguilar, M., Travieso, C.M., Alonso, J.B.: Off-line Handwritten Signature GPDS-960 Corpus. In: The Ninth International Conference on Document Analysis and Recognition, Parana, Brazil, pp. 764–768 (2007)

7. Ferrer-Aguilar, M., Vargas, F., Morales, A., Ordoñez, A.: Robustness of Off-line Signature Verification based on Gray Level Features. IEEE Transactions on Information Forensics and Security 7(3), 966–977 (2012)

8. Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., Escudero, D., Moro, Q.-I.: MCYT baseline corpus: A bimodal biometric database. IEEE Proceedings Vision, Image and Signal Processing 150(6), 395–401 (2003)

9. Blankers, V.L., Van den Heuvel, C.E., Franke, K.Y., Vuurpijl, L.G.: The ICDAR 2009 signature verification competition. In: International Conference on Document Analysis and Recognition, Spain, pp. 1403–1407 (2009)

10. Blumenstein, M., Ferrer- Aguilar, M., Vargas, J.F.: The 4NSigComp2010 off-line signature verification competition: Scenario 2. In: 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, pp. 721–726 (2010)

11. Liwicki, M., Blumenstein, M., Heuvel, E.V.D., Berger, C.E.H., Stoel, R.D., Found, B., Chen, X., Malik, M.I.: SigComp11: Signature Verification Competition for On- and Offline Skilled Forgeries. In: 11th International Conference on Document Analysis and Recognition, Beijing, China (2011)

12. Liwicki, M., Malik, M.I., Alewijnse, L., van den Heuvel, E., Found, B.: ICFHR 2012 Competition on Automatic Forensic Signature Verification (4NsigComp 2012). In: 13th International Conference on Frontiers in Handwriting Recognition, Bari, Italy, pp. 819–824 (2012)

13. Abroug, I., Essoukri Ben Amara, N.: A Neuronal Planar Modeling for Handwriting Signature based on Automatic Segmentation. International Journal of Computer Applications 49(8), 29–34 (2012)

14. Abroug Ben Abdelghani, I., Ghardallou Lasmar, A., Essoukri Ben Amara, N.: Planar Hidden Markov/Neural Networks Modelling for Off-line Signature Vérification. In: The International Arab Conferénce on Information Technology, Jordan, pp. 455–460 (2005)