Chapter 6 1

Online Handwritten Signature Verification 2

Sonia Garcia-Salicetti, Nesma Houmani, Bao Ly-Van, Bernadette Dorizzi, 3
 Fernando Alonso-Fernandez, Julian Fierrez, Javier Ortega-Garcia, 4
 Claus Vielhauer, and Tobias Scheidat 5

Abstract In this chapter, we first provide an overview of the existing main ap- 6
 proaches, databases, evaluation campaigns and the remaining challenges in online 7
 handwritten signature verification. We then propose a new benchmarking frame- 8
 work for online signature verification by introducing the concepts of “Reference 9
 Systems”, “Reference Databases” and associated “Reference Protocols.” Finally, we 10
 present the results of several approaches within the proposed evaluation framework. 11
 Among them are also present the best approaches within the first international Sig- 12
 nature Verification Competition held in 2004 (SVC’2004), Dynamic Time Warping 13
 and Hidden Markov Models. 14

All these systems are evaluated first within the benchmarking framework and 15
 also with other relevant protocols. Experiments are also reported on two different 16
 databases (BIOMET and MCYT) showing the impact of time variability for online 17
 signature verification. The two reference systems presented in this chapter are also 18
 used and evaluated in the BMEC’2007 evaluation campaign, presented in Chap 11. 19

6.1 Introduction 20

Online signature verification is related to the emergence of automated verification 21
 of handwritten signatures that allows the introduction of the signature’s dynamic 22
 information. Such dynamic information is captured by a digitizer, and generates 23
 “online” signatures, namely a sequence of sampled points conveying dynamic in- 24
 formation during the signing process. Online signature verification thus differs from 25
 off-line signature verification by the nature of the raw signal that is captured: off- 26
 line signature verification processes the signature as an image, digitized by means 27
 of a scanner [25, 24, 8] while online signature is captured through an appropri- 28
 ate sensor sampling at regular time intervals the hand-drawn signal. Such sensors 29
 have evolved recently allowing to capture on each point not only pen position but 30
 also pen pressure and pen inclination in a three-dimensional space. Other pen-based 31

interfaces such as those on Personal Digital Assistants (PDAs) and Smartphones 32
operate via a touch screen to allow only a handwritten signature as a time sequence 33
of pen coordinates to be captured. 34

Actually, signature is the most socially and legally accepted means for person 35
authentication and is therefore a modality confronted with high level attacks. Indeed, 36
when a person wants to bypass a system, he/she will forge the signature of another 37
person by trying to reproduce as close as possible the target signature. The online 38
context is favorable to identity verification because in order to produce a forgery, 39
an impostor has to reproduce more than the static image of the signature, namely a 40
personal and well anchored “gesture” of signing—more difficult to imitate than the 41
image of the signature. 42

On the other hand, even if a signature relies on a specific gesture, or a specific 43
motor model [24], it results in a strongly variable signal from one instance to the 44
next. Indeed, identity verification by an online signature still remains an enormous 45
challenge for research and evaluation, but signature, because of its wide use, remains 46
a potential field of promising applications [37, 31, 23, 32, 33]. 47

Some of the main problems are related to signature intraclass (intrapersonal) 48
variability and signature’s time variability. It is well known that signing relies on 49
a very fast, practiced and repeatable motion that makes signature vary even over a 50
short term. Also, this motion may evolve over time, thus modifying the aspect of the 51
signature significantly. Finally, a person may change this motion/gesture over time, 52
thus generating another completely different signature. 53

On the other hand, there is also the problem related to the difficulty of assess- 54
ing the resistance of systems to imposture. Indeed, skilled forgery performance is 55
extremely difficult to compare across systems because the protocol of forgery ac- 56
quisition varies from one database to another. Going deeper into this problem, it is 57
hard to define what is a good forgery. Some works and related databases only exploit 58
forgeries of the image of the signature while in an online context [12, 22, 16, 4]; few 59
exploit forgeries of the personal gesture of signing, additionally to skilled forgeries 60
of the image of the signature [5, 6, 36, 2]. Finally, some works only exploit random 61
forgeries to evaluate the capacity of systems to discriminate forgeries from genuine 62
signatures [18]. 63

The first international Signature Verification Competition (SVC’2004) [36] was 64
held in 2004, and only 15 academic partners participated to this evaluation, a number 65
far behind the existing approaches in the extensive literature on the field. Although 66
this evaluation allowed for the first time to compare standard approaches in the lit- 67
erature, as Dynamic Time Warping (DTW) and Hidden Markov Models (HMMs), 68
it was carried out on a small database (60 people); with the particularity of con- 69
taining a mixture of Western and Asian signatures. Indeed, on one hand, this had 70
never been the case in published research works and therefore the participants to 71
SVC’2004 had never been confronted before to this type of signatures. On the other 72
hand, it is still unclear whether a given cultural type of signature may be better suited 73
for a given approach compared to another and thus one may wonder: were all the 74
systems in equal conditions in this evaluation? All these factors still make difficult 75
for the scientific community to assess algorithmic performance and to compare the 76
existing systems of the extensive literature about online signature verification. 77

6 Online Handwritten Signature Verification

This chapter has different aims. First, it aims to make a portrait of research and evaluation in the field nowadays (existing main approaches, databases, evaluation campaigns, and remaining challenges). Second, it aims to introduce the new benchmarking framework for online signature for the scientific community in order to allow future comparison of their systems with standard or reference approaches. Finally, it aims to perform a comparative evaluation (within the proposed benchmarking framework) of the best approaches according to SVC'2004 results, Dynamic Time Warping (DTW) and Hidden Markov Models (HMMs), relatively to other standard approaches in the literature, and that on several available databases with different protocols, some of which have never been considered in the framework of evaluation (with time variability).

The benchmarking experiments (defined on two publicly available databases) can be easily reproduced, following the How-to documents provided on the companion website [11]. In such a way they could serve as further comparison points for newly proposed research systems. As highlighted in Chap. 2, this comparison points are multiple, and are dependent of what we want to study and what we have at our disposal. The comparisons points that are illustrated in this book regarding the signature experiments are the following:

- One possible comparison when using such a benchmarking framework is to compare different systems on the same database and same protocols. In such a way, the advantages of the proposed systems could be pinpointed. If error analysis and/or fusion experiments are done furthermore, the complementarity of the proposed systems could be studied, allowing to a further design of new more powerful systems. In this chapter, five research signature verification systems are compared to the two reference (baseline) systems.
- One comparison point could be obtained by researchers if they run the same open-source software (with the same relevant parameters) on different databases. In such a way the performance of this software could be compared within the two databases. The results of such a comparison are reported in Chap. 11, where the two online signature reference systems are applied on a new database, that has the particularity of being recorded in degraded conditions.
- Comparing the same systems on different databases is an important point in order to test the scalability of the reported results (if the new database is of different size or nature), or the robustness of the tested systems to different experimental conditions (such as robustness to degraded data acquisition situations).

This chapter is organized as follows: first, in Sect. 6.2 we describe the state of the art in the field, including the existing main approaches and the remaining challenges. The existing databases and evaluation campaigns are described in Sects. 6.3 and 6.4, respectively. In Sect. 6.5, we describe the new benchmarking framework for online signature by introducing the concept of "Reference Systems", on "Reference Databases" and associated "Reference Protocols". In Sect. 6.6 several research algorithm are presented and evaluated within the benchmarking framework. The conclusions are given in Sect. 6.7.

6.2 State of the Art in Signature Verification 121

In order to perform signature verification, there are two possibilities (related to the classification step): one is to store different signatures of a given person in a database and in the verification phase to compare the test signature to these signatures, called “Reference Signatures” by means of a distance measure; in this case a dissimilarity measure is the outcome of the verification system, after combining by a given function the resulting distances. The other is to build a statistical model of the person’s signature; in this case, the outcome of the verification system is a likelihood measure—how likely it is that a test signature belongs to the claimed client’s model. 122 123 124 125 126 127 128 129 130

6.2.1 Existing Main Approaches 131

In this section, we have chosen to emphasize the relationship between the verification approach used (the nature of the classifier) and the type of features that are extracted to represent a signature. For this reason, we have structured the description of the research field in two subsections, the first concerning distance-based approaches and the second concerning model-based approaches. Issues related to fusion of the scores resulting from different classifiers (each fed by different features) are presented in both categories. 132 133 134 135 136 137 138

6.2.1.1 Distance-based Approaches 139

There are several types of distance-based approaches. First, as online signatures have variable length, a popular way of computing the distance between two signatures is Dynamic Time Warping [26]. This approach relies on the minimization of a global cost function consisting of local differences (local distances) between the two signatures that are compared. As the minimized function is global, this approach is tolerant to local variations in a signature, resulting in a so-called “elastic matching”—or elastic distance—that performs time alignment between the compared signatures. 140 141 142 143 144 145 146 147

Among distance-based approaches, an alternative is to extract global features and to compare two signatures therefore described by two vectors of the same length (the number of features) by a classical distance measure (Euclidean, Mahalanobis, etc.). Such approaches have shown rather poor results. In [13], it is shown on a large data set, the MCYT complete database, that a system based on a classical distance measure on dynamic global features performs weakly compared to an elastic distance matching character strings that result from a coarse feature extraction. 148 149 150 151 152 153 154

Initially, Dynamic Time Warping was used exclusively on time functions captured by the digitizer (no feature extraction was performed) and separately on each 155 156

6 Online Handwritten Signature Verification

time function. Examples of such strategies are the works of Komiya et al. [18] and Hangai et al. [15].

In [18], the elastic distances between the test signature and the reference signatures are computed on three types of signals: coordinates, pressure and pen-inclination angles (azimuth and altitude), and the three resulting scores are fused by a weighted mean. On a private database of very small size (eight people) and using 10 reference signatures which is a large reference set compared to nowadays standards—as established by the first international Signature Verification Competition in 2004 (SVC'2004) [36]—of five reference signatures, an EER of 1.9% was claimed.

In [15], three elastic distances are computed on the raw time functions captured by the digitizer: one on the spherical coordinates associated to the two pen-inclination angles, one on the pressure time function and one on the coordinates time functions. Note that, in this work, the pen-inclination angles were claimed to be the most performing time functions when only one elastic distance measure is considered separately on a private database. The best results were obtained when a weighted sum was used to fuse the three elastic distances.

Other systems based on Dynamic Time Warping (DTW) performing time alignment at another level of description than the point level, were also proposed. On one hand, in [35], the fusion of three elastic distance measures each resulting from matching two signatures at a given level of description, is performed. On the other hand, systems performing the alignment at the stroke level have also appeared [4], reducing the computational load of the matching process. Such systems are described in the following paragraph.

Wirotius et al. [35] fuse by a weighted mean, three complementary elastic distances resulting from matching two signatures at three levels of description: the temporal coordinates of the reference and test signatures, the trajectory lengths of the reference and test signatures, and the coordinates of the reference and test signatures. The data preparation phase consists in the selection of some representative points in the signatures corresponding to the local minimum of speed.

On the other hand, Chang et al. [4] proposed a stroke-based signature verification method based on Dynamic Time Warping and tested the method on Japanese signatures. It is interesting to note that in Asian signatures, stroke information may be indeed more representative than in Western signatures in which intra-stroke variation is more pronounced. The method consists of a modified Dynamic Time Warping (DTW) allowing stroke merging. To control this process, two rules were proposed: an appropriate penalty-distance to reduce stroke merging, and new constraints between strokes to prevent wrong merging. The temporal functions (x and y coordinates, pressure, direction and altitude), and inter-stroke information that is the vector from the center point of a stroke to its consequent stroke, were used for DTW matching. The total writing time was also used as a global feature for verification. Tested on a private database of 17 Japanese writers on skilled forgeries, an Equal Error Rate (EER) of 3.85% was obtained by the proposed method while the EER of the conventional DTW was 7.8%.

Main works in this category of distance-based approaches are those of Kholmatov et al. [3] and Jain et al. [16], both using Dynamic Time Warping.

Jain et al. [16] performed alignment on feature vectors combining different types of features. They present different systems based on this technique, according to which type of feature or which combination of features is used (spatial with a context bitmap, dynamic, etc.) and which protocol is used (global threshold or person-dependent threshold). The main point is that the minimized cost function depends as usual on local differences on the trajectory but also on a more global characteristic relying on the difference in the number of strokes of the test signature and the reference signature. Also, a resampling of the signature is done but only between some particular points, called “critical” (start and end points of a stroke and points of trajectory change) that are kept. The local features of position derivatives, path tangent angle and the relative speed (speed normalized by the average speed) are the best features with a global threshold.

Kholmatov’s system [3] is the winning approach of the first online Signature Verification Competition (SVC) in 2004 [36]. Using position derivatives as two local features, it combines a Dynamic Time Warping approach with a score normalization based on client intra-class variability, computed on the eight signatures used for enrollment. On these eight enrollment signatures, three normalization factors are generated by computing pairwise DTW distances among the enrollment signatures: the maximum, the minimum and the average distances. A test signature’s authenticity is established by first aligning it with each reference signature for the claimed user. The distances of the test signature to the nearest reference signature, farthest reference signature and the average distance to the eight reference signatures are considered; then these three distances are normalized by the corresponding three factors obtained from the reference set to form a three-dimensional feature vector. Performance reached around 2.8% EER on the SVC test dataset, as described in detail in Sect. 6.4.

6.2.1.2 Model-based Approaches

Model-based approaches appeared naturally in signature verification because Hidden Markov Models (HMMs) have long been used for handwriting recognition [25, 24, 8].

One of the main pioneering and more complete work in the literature is Dolfing’s [5, 6]. It couples a continuous left-to-right HMM with a Gaussian mixture in each state with different kinds of features extracted at an intermediate level of description, namely portions of the signature defined by vertical velocity zeros. Also, the importance of each kind of feature (spatial, dynamic, and contextual) was studied in terms of discrimination by a Linear Discriminant Analysis (LDA) on the Philips database (described in Sect. 6.3). Dynamic and contextual features appeared to be much more discriminant compared to spatial features [5]. Using 15 training signatures (that is a large training set compared to nowadays standards of five training signatures) [36], with person-dependent thresholds, an EER of 1.9-2.9% was

6 Online Handwritten Signature Verification

reached for different types of forgeries (of the image by amateur forgers, of the dynamics, and of the image by professional forgers).

At the same time, discrete HMMs were proposed by Kashi et al. [17] with a local feature extraction using only the path tangent angle and its derivative on a resampled version of the signature. It is a hybrid classifier that is finally used to take the decision in the verification process: another classifier using global features, based on a Mahalanobis distance with the hypothesis of uncorrelated features, is combined to HMM likelihood. The training set is of more limited size (six training signatures) in this work. The performance was evaluated on the Murray Hill database containing signatures of 59 subjects, resulting in an EER of 2.5%. A conclusion of this work is that fusing the scores of the two classifiers using different levels of description of the signature, gives better results than using only the HMM with the local feature extraction.

Discrete HMMs were also proposed for online signature by Rigoll et al. in [29], coupled to different types of features: the low-resolution image (“context bitmap”) around each point of the trajectory, the pen pressure, the path tangent angle, its derivative, the velocity, the acceleration, some Fourier features, and some combinations of the previously mentioned features. A performance of 99% was obtained with this model with a given combination of spatial and dynamic features, unfortunately on a private database which is not really described as often in the field.

More recently, other continuous HMMs have been proposed for signature by Fierrez-Aguilar et al. [7] by using a pure dynamic encoding of the signature, exploiting the time functions captured by the digitizer (x and y coordinates, and pressure) plus the path tangent angle, path velocity magnitude, log curvature radius, total acceleration magnitude and their first-order time derivatives to end with 14 features at each point of the signature. In the verification stage, likelihood scores are further processed by the use of different score-normalization techniques in [10]. The best results, using a subset of the MCYT signature database, described in Sect. 6.3, resulted in 0.78% of EER for skilled forgeries (3.36% without score normalization).

Another HMM-based approach, performing fusion of two complementary information levels issued from the same continuous HMM with a multivariate Gaussian mixture in each state, was proposed in [34] by Ly-Van et al. A feature extraction combining dynamic and local spatial features was coupled to this model. The “segmentation information” score derived by analyzing the Viterbi path, that is the segmentation given by the HMM on the test signature by the target model, is fused to the HMM likelihood score, as described in detail in Sect. 6.5.2. This work showed for the first time in signature verification that combining such two sorts of information generated by the same HMM considerably improves the quality of the verification system (an average relative improvement of 26% compared to using only the HMM likelihood score), after an extensive experimental evaluation on four different databases (Philips [5], SVC’2004 development set [36], the freely available subset of MCYT [22], and BIOMET [12]).

Besides, a personalized two-stage normalization, at the feature and score levels, resulted in client and impostor scores distributions that are very close from one database to another. This stability of such distributions resulted in the fact that testing the

system on the set composed of the mixture of the four test databases almost does not degrade the system's performance: a state-of-the-art performance of 4.50% is obtained compared to the weighted average EER of 3.38% (the weighted sum of the EER obtained on each of the four test databases, where the weights are respectively the number of test signatures in each of the four test databases).

Another hybrid approach using a continuous HMM is that of Fierrez-Aguilar et al. [9], that uses additionally to the HMM a nonparametric statistical classifier using global features. The density of each global feature is estimated by means of Parzen Gaussian windows. A feature selection procedure is used by ranking the original 100 global features according to a scalar measure of interuser class separability based on the Mahalanobis distance between the average vector of global features computed on the training signatures of a given writer, and all the training signatures from all other writers. Optimal results are obtained for 40 features selected from the 100 available. Performance is evaluated on the complete MCYT Database [22] of 330 persons. Fusion by simple rules as maximum and sum of the HMM score, based on local features, and the score of the non parametric classifier, based on global features, leads to a relative improvement of 44% for skilled forgeries compared to the HMM alone (EER between 5% and 7% for five training signatures). It is worth noticing that the classifier based on density estimation of the global features outperforms the HMM when the number of training signatures is low (five signatures), while this tendency is inverted when using more signatures in the training phase. This indicates that model-based approaches are certainly powerful but at the price of having enough data at disposal in the training phase.

Another successful model-based approach is that of Gaussian Mixture Models (GMMs) [27]. A GMM is a degenerated version of an HMM with only one state. In this framework, another element appears: a normalization process of the score given by the client GMM by computing a log-likelihood ratio considering also the score given on the test signature by another GMM, the "world-model" or "Universal Background Model" (UBM), representing an "average" user, trained on a given pool of users no longer used for the evaluation experiments.

This approach, widely used in speaker verification, was first proposed in signature verification by Richiardi et al. [28] by building a GMM "world-model" and GMM client models independently, in other words, with no adaptation of the world model to generate the client model. In this work, a local feature extraction of dynamic features was used (coordinates, pressure, path tangent angle, velocity). As experiments were carried out on a very small subset of MCYT [22] of 50 users, the world-model was obtained by pooling together all enrollment data (five signatures per user) and five forgeries per user done by the same forger; thus, the world model was not trained on a separate devoted pool of users. Similar performance was observed in this case with an HMM of two states with 32-Gaussian components per state, and a 64-Gaussian components GMM. More recently, the same authors have evaluated different GMM-based systems [2] (see also Chap. 11), some based only on local features, some based on the fusion of the outputs of GMMs using global features and GMMs using local features—obtaining very good results on the BioSecure signature subcorpus DS3 acquired on a Personal Digital Assistant (PDA).

6 Online Handwritten Signature Verification

Furthermore, Martinez-Diaz et al. [21] proposed in signature verification the use of Universal Background Model Bayesian adaptation to generate the client model. The parameters of the adaptation were studied on the complete MCYT database by using the 40 global features reported in [9]. Results reported show 2.06% of EER with five training signatures for random forgeries, and 10.49% of EER for skilled forgeries.

6.2.2 Current Issues and Challenges

Signatures are highly variable from one instance to another, particularly for some subjects, and highly variable in time. A remaining challenge in research is certainly the study of the influence of time variability on system performance, as well as the possibility of performing an update of the writer templates (references) in the case of distance-based approaches, or an adaptation of the writer model in the case of model-based approaches. Alternatively, the study of personalized feature selection would be of interest for the scientific community since it may help to cope with intra-class variability, usually important in signature (although the degree of such variability is writer-dependent); indeed, one may better characterize a writer by those features that show more stability for him/her.

From the angle of systems evaluation, the previous discussion shows that it is difficult to compare the existing approaches to the different systems in the literature and that few evaluations have been carried out in online signature verification. The first evaluation was SVC'2004 [36], on signatures captured on a digitizer, but on a database of very limited size (60 persons, only one session) mixing signatures of different cultural origins. More recently, the BioSecure Network of Excellence has carried out the first signature verification evaluation on signatures captured on a mobile platform [14], on a much larger database (713 persons, two sessions).

In this respect, the scientific community needs a clear and permanent evaluation framework, composed of publicly available databases, associated protocols and baseline "Reference" systems in order to be able to compare their systems to the state of the art. Section 6.5 introduces such a benchmarking framework.

6.3 Databases

There exist a lot of online handwritten signature databases, but not all of them are freely available. We will describe here some of the most well-known databases.

6.3.1 PHILIPS

365

The signatures in the Philips database [5, 6] were captured on a digitizer at a sampling rate of up to 200 Hz. At each sampled point, the digitizer captures the coordinates $(x(t), y(t))$, the axial pen pressure $p(t)$, and the “pen-tilt” of the pen in x and y directions, that is two angles resulting from the projection of the pen in each of the coordinate planes xOz and yOz .

This database contains data from 51 individuals (30 genuine signatures of each individual) and has the particularity of containing different kinds of forgeries. Three types of forgeries were acquired: “over the shoulder”, “home improved”, and “professional.” The first kind of forgeries was captured by the forger after seeing the genuine signature being written, that is after learning the dynamic properties of the signature by observation of the signing process. The “home improved” forgeries are made in other conditions: the forger only imitates the static image of the genuine signature, and has the possibility of practicing the signature at home. Finally, the “professional” forgeries are produced by individuals who have professional expertise in handwriting analysis, and that use their experience in discriminating genuine from forged signatures to produce high- quality spatial forgeries.

This database contains 1,530 genuine signatures, 1,470 “over the shoulder” forgeries (30 per individual except two), 1,530 “home improved” forgeries (30 per individual), and 200 “professional” forgeries (10 per individual for 20 individuals).

6.3.2 BIOMET Signature Subcorpus

385

The signatures in the online signature subset of the BIOMET multimodal database [12] were acquired on the WACOM Intuos2 A6 digitizer with an ink pen, at a sampling rate of 100 Hz. At each sampled point of the signature, the digitizer captures the (x, y) coordinates, the pressure p and two angles (azimuth and altitude), encoding the position of the pen in space. The signatures were captured in two sessions with five months spacing between them. In the first session, five genuine signatures and six forgeries were captured per person. In the second session, ten genuine signatures and six forgeries were captured per person. The 12 forgeries of each person’s signature were made by four different impostors (three per impostor in each session). Impostors try to imitate the image of the genuine signature. In Fig. 6.1, we see for one subject the genuine signatures acquired at Session 1 (Fig. 6.1 (a)) and Session 2 (Fig. 6.1 (b)), and the skilled forgeries acquired at each session (Fig. 6.1 (c)). As for certain persons in the database some genuine signatures or some forgeries are missing, there are 84 individuals with complete data. The online signature subset of BIOMET thus contains 2,201 signatures (84 writers $\times$ (15 genuine signatures + 12 imitations) – eight missing genuine signatures – 59 missing imitations).

6 Online Handwritten Signature Verification

Fig. 6.1 Signatures from the BIOMET database of one subject: (a) genuine signatures of Session 1, (b) genuine signatures of Session 2, and (c) skilled forgeries

AQ: Please provide better quality figure.

6.3.3 SVC'2004 Development Set

403

SVC2'004 development set is the database that was used by the participants to tune 404
their systems before their submission to the first international Signature Verification 405
Competition in 2004 [36]. The test database on which the participant systems were 406
ranked is not available. 407

This development set contains data from 40 people, both from Asian and Western 408
origins. In the first session, each person contributed 10 genuine signatures. In the 409
second session, which normally took place at least one week after the first one, each 410
person came again to contribute with another 10 genuine signatures. 411

For privacy reasons, signers were advised not to use their real signatures in daily 412
use. However, contributors were asked to try to keep the consistency of the image 413
and the dynamics of their signature, and were recommended to practice thoroughly 414
before the data collection started. For each person, 20 skilled forgeries were pro- 415
vided by at least four other people in the following way: using a software viewer, 416
the forger could visualize the writing sequence of the signature to forge on the com- 417
puter screen, therefore being able to forge the dynamics of the signature. 418

The signatures in this database were acquired on a digitizing tablet (WACOM 419
Intuos tablet) at a sampling rate of 100 Hz. Each point of a signature is characterized 420
by five features: x and y coordinates, pressure and pen orientation (azimuth and 421
altitude). However, all points of the signature that had zero pressure were removed. 422

Therefore, the temporal distance between points is not regular. To overcome this problem, the time corresponding to the sampling of a point was also recorded and included in signature data. Also, at each point of signature, there is a field that denotes the contact between the pen and the digitizer. This field is set to 1 if there is contact and to 0 otherwise.

6.3.4 MCYT Signature Subcorpus

The number of existing large public databases oriented to performance evaluation of recognition systems in online signature is quite limited. In this context, the MCYT Spanish project, oriented to the acquisition of a bimodal database including fingerprints and signatures was completed by late 2003 with 330 subjects captured [22]. In this section, we give a brief description of the signature corpus of MCYT, still the largest publicly available online western signature database.

In order to acquire the dynamic signature sequences, a WACOM pen tablet, model Intuos A6 USB was employed. The pen tablet resolution is 2,540 lines per inch (100 lines/mm), and the precision is 0.25 mm. The maximum detection height is 10 mm (pen-up movements are also considered), and the capture area is 127 mm (width) 97 mm (height). This tablet provides the following discrete-time dynamic sequences: position x_n in x -axis, position y_n in y -axis, pressure p_n applied by the pen, azimuth angle γ_n of the pen with respect to the tablet, and altitude angle ϕ_n of the pen with respect to the tablet. The sampling frequency was set to 100 Hz. The capture area was further divided into 37.5 mm (width) 17.5 mm (height) blocks which are used as frames for acquisition. In Fig. 6.2, we see for each subject the two left signatures that are genuine, and the one on the right that is a skilled forgery. Plots below each signature correspond to the available information—namely: position trajectories, pressure, pen azimuth, and altitude angles.

The signature corpus comprises genuine and shape-based highly skilled forgeries with natural dynamics. In order to obtain the forgeries, each contributor is requested to imitate other signers by writing naturally, without artifacts such as breaks or slowdowns. The acquisition procedure is as follows. User n writes a set of five genuine signatures, and then five skilled forgeries of client $n - 1$. This procedure is repeated four more times imitating previous users $n - 2, n - 3, n - 4$ and $n - 5$. Taking into account that the signer is concentrated in a different writing task between genuine signature sets, the variability between client signatures from different acquisition sets is expected to be higher than the variability of signatures within the same set. As a result, each signer contributes with 25 genuine signatures in five groups of five signatures each, and is forged 25 times by five different imitators. The total number of contributors in MCYT is 330. Therefore the total number of signatures present in the signature database is $330 \times 50 = 16,500$, half of them genuine signatures and the rest forgeries.

6 Online Handwritten Signature Verification

Fig. 6.2 Signatures from MCYT database corresponding to three different subjects. Reproduced with permission from Annales des Telecommunications, source [13]

6.3.5 BioSecure Signature Subcorpus DS2

462

In the framework of the BioSecure Network of Excellence [1], a very large signature subcorpus containing data from about 600 persons was acquired as part of the

multimodal Data Set 2 (DS2). The scenario considered for the acquisition of DS2 465
signature dataset is a PC-based off-line supervised scenario [2]. 466

The acquisition is carried out using a standard PC machine and the digitizing 467
tablet WACOM Intuos3 A6. The pen tablet resolution is 5,080 lines per inch and the 468
precision is 0.25 mm. The maximum detection height is 13 mm and the capture area 469
is 270 mm (width)×216 mm (height). Signatures are captured on paper using an 470
inking pen. At each sampled point of the signature, the digitizer captures at 100 Hz 471
sampling rate the pen coordinates, pen pressure (1,024 pressure levels) and pen 472
inclination angles (azimuth and altitude angles of the pen with respect to the tablet). 473

This database contains two sessions, acquired two weeks apart. Fifteen genuine 474
signatures were acquired at each session as follows: the donor was asked to per- 475
form, alternatively, three times five genuine signatures and two times five skilled 476
forgeries. For skilled forgeries, at each session, a donor is asked to imitate five times 477
the signature of two other persons (for example client $n - 1$ and $n - 2$ for Session 1, 478
and client $n - 3$ and $n - 4$ for Session 2). 479

The BioSecure Signature Subcorpus DS2 is not yet available but, acquired on 480
seven sites in Europe, it will be the largest online signature multisession database 481
acquired in a PC-based scenario. 482

6.3.6 BioSecure Signature Subcorpus DS3 483

The scenario considered in this case relies on a mobile device, under degraded con- 484
ditions [2]. Data Set 3 (DS3) signature subcorpus contains the signatures of about 485
700 persons, acquired on the PDA HP iPAQ hx2,790, at the frequency of 100 Hz and 486
a touch screen resolution of 1,280×960 pixels. Three time functions are captured 487
from the PDA: x and y coordinates and the time elapsed between the acquisition of 488
two successive points. The user signs while standing and has to keep the PDA in 489
her/his hand. 490

In order to have time variability in the database, two sessions between Novem- 491
ber 2006 and May 2007 were acquired, each containing 15 genuine signatures. The 492
donor was asked to perform, alternatively, three times five genuine signatures and 493
two times five forgeries. For skilled forgeries, at each session, a donor is asked to 494
imitate five times the signature of two other persons (for example client $n - 1$ and 495
 $n - 2$ for Session 1, and client $n - 3$ and $n - 4$ for Session 2). In order to imitate 496
the dynamics of the signature, the forger visualized the writing sequence of the sig- 497
nature they had to forge on the PDA screen and could sign on the image of such 498
signature in order to obtain a better quality forgery, both from the point of view of 499
the dynamics and of the shape of the signature. 500

The BioSecure Signature Subcorpus DS3 is not yet available but, acquired on 501
eight sites in Europe, it is the first online signature multisession database acquired 502
in a mobile scenario (on a PDA). 503

6.4 Evaluation Campaigns

504

The first international competition on online handwritten signature verification (Signature Verification Competition–SVC [36]) was held in 2004. The disjoint development data set related to this evaluation was described in Sect. 6.3.3. The objective of SVC'2004 was to compare the performance of different signature verification systems systematically, based on common benchmarking databases and under a specific protocol. SVC'2004 consisted of two separate signature verification tasks using two different signature databases: in the first task, only pen coordinates were available; in the second task, in addition to coordinates, pressure and pen orientation were available. Data for the first task was obtained by suppressing pen orientation and pressure in the signatures used in the second task.

The database in each task contained signatures of 100 persons and, for each person there were 20 genuine signatures and 20 forgeries. The development dataset contained only 40 persons and was released to participants for developing and evaluating their systems before submission. No information regarding the test protocol was communicated at this stage to participants, except the number of enrollment signatures for each person, which was set to five.

The test dataset contained the signatures of the remaining 60 persons. For test purposes, the 20 genuine signatures available for each person were divided into two groups of 10 signatures, respectively devoted to enrollment and test. For each user, 10 trials were run based on 10 different random samplings of five genuine enrollment signatures out of the 10 devoted to enrollment. Although samplings were random, all the participant systems were submitted to the same samplings for comparison. After each enrollment trial, all systems were evaluated on the same 10 genuine test signatures and the 20 skilled forgeries available for each person. Each participant system had to give a normalized similarity score between 0 and 1 as an output for any test signature.

Overall, 15 systems were submitted to the first task, and 12 systems were submitted to the second task. For both tasks, the Dynamic Time Warping DTW-based system submitted by Kholmatov and Yanikoglu (team from Sabanci University of Turkey) [3] obtained the lowest average EER values when tested on skilled forgeries ($EER = 2.84\%$ in Task 1 and $EER = 2.89\%$ in Task 2). In second position, we distinguished the HMM-based systems with Equal Error Rates around 6% in Task 1 and 5% in Task 2, when tested on skilled forgeries. This DTW system was followed by the HMM approach submitted by Fierrez-Aguilar and Ortega-Garcia (team from Universidad Politecnica de Madrid) [7], which outperformed the winner in the case of random forgeries (with $EER = 2.12\%$ in Task 1 and $EER = 1.70\%$ in Task 2).

6.5 The BioSecure Benchmarking Framework for Signature Verification

541

542

The BioSecure Reference Evaluation Framework for online handwritten signature is composed of two open-source reference systems, the signature parts of the publicly available BIOMET and MCYT-100 databases, and benchmarking (reference)

experimental protocols. The reference experiments, to be used for further comparisons, can be easily reproduced following the How-to documents provided on the companion website [11]. In such a way they could serve as further comparison points for newly proposed research systems.

6.5.1 Design of the Open Source Reference Systems

For the signature modality, the authors could identify no existing evaluation platform and no open-source implementation prior to the activities carried out in the framework of BioSecure Network of Excellence [13, 14]. One of its aims was to put at disposal of the community a platform in source code composed of different algorithms that could be used as a baseline for comparison. Consequently, it was decided within the BioSecure consortium to design and implement such a platform for the biometric modality of online signatures. The main modules of this platform are shown in Fig. 6.3.

Fig. 6.3 Main modules of the open-source signature reference systems Ref1 and Ref2

The pre-processing module allows for future integration of functions like noise filtering or signal smoothing, however at this stage this part has been implemented as a transparent all-pass filter. With respect to the classification components, the platform considers two types of algorithms: those relying on a distance-based approach and those relying on a model-based approach. Out of the two algorithms integrated in the platform, one falls in the category of model-based methods, whereas the second is a distance-based approach.

In this section these two algorithms are described in further detail. The first is based on the fusion of two complementary information levels derived from a writer HMM. This system is labeled as Reference System 1(Ref1) and was developed by TELECOM SupParis (ex. INT) [34]. The second system, called Reference System 2

6 Online Handwritten Signature Verification

(Ref2), is based on the comparison of two character strings—one for the test signature and one for the reference signature—by an adapted Levenshtein distance [19], developed by University of Magdeburg.

6.5.2 Reference System 1 (Ref1-v1.0)

¹ Signatures are modeled by a continuous left-to-right HMM [26], by using in each state a continuous multivariate Gaussian mixture density. Twenty-five dynamic features are extracted at each point of the signature; such features are given in Table 6.1 and described in more detail in [34]. They are divided into two subcategories: gesture-related features and local shape-related features.

The topology of the signature HMM only authorizes transitions from each state to itself and to its immediate right-hand neighbors. The covariance matrix of each multivariate Gaussian in each state is also considered diagonal. The number of states in the HMM modeling the signatures of a given person is determined individually according to the total number T_{total} of all the sampled points available when summing all the genuine signatures that are used to train the corresponding HMM. It was considered necessary to have an average of at least 30 sampled points per Gaussian for a good re-estimation process. Then, the number of states N is computed as:

$$N = \left\lfloor \frac{T_{total}}{4 \times 30} \right\rfloor \quad (6.1)$$

where brackets denote the integer part.

In order to improve the quality of the modeling, it is necessary to normalize for each person each of the 25 features separately, in order to give an equivalent standard deviation to each of them. This guarantees that each parameter contributes with the same importance to the emission probability computation performed by each state on a given feature vector. This also permits a better training of the HMM, since each Gaussian marginal density is neither too flat nor too sharp. If it is too sharp, for example, it will not tolerate variations of a given parameter in genuine signatures or, in other words, the probability value will be quite different on different genuine signatures. For further information the interested reader is referred to [34].

The Baum-Welch algorithm described in [26] is used for parameter re-estimation. In the verification phase, the Viterbi algorithm permits the computation of an approximation of the log-likelihood of the input signature given the model, as well as the sequence of visited states (called “most likely path” or “Viterbi path”). On a particular test signature, a distance is computed between its log-likelihood and

¹ This section is reproduced with permission from Annales des Telecommunications, source [13].

t1.1 **Table 6.1** The 25 dynamic features of Ref1 system extracted from the online signature: (a) gesture-related features and (b) local shape-related features

t1.2	N^o	Feature name
t1.3	1-2	Normalized coordinates $(x(t) - x_g, y(t) - y_g)$ relatively to the gravity center (x_g, y_g) of the signature
t1.4	3	Speed in x
t1.5	4	Speed in y
t1.6	5	Absolute speed
t1.7	6	Ratio of the minimum over the maximum speed on a window of five points
t1.8	a)	7 Acceleration in x
t1.9		8 Acceleration in y
t1.10		9 Absolute acceleration
t1.11		10 Tangential acceleration
t1.12		11 Pen pressure (raw data)
t1.13		12 Variation of pen pressure
t1.14	13-14	Pen-inclination measured by two angles
t1.15	15-16	Variation of the two pen-inclination angles
t1.16	17	Angle α between the absolute speed vector and the x axis
t1.17	18	Sine(α)
t1.18	19	Cosine(α)
t1.19	20	Variation of the α angle: ϕ
t1.20	b)	21 Sine(ϕ)
t1.21		22 Cosine(ϕ)
t1.22		23 Curvature radius of the signature at the present point
t1.23		24 Length to width ratio on windows of size five
t1.24		25 Length to width ratio on windows of size seven

the average log-likelihood on the training database. This distance is then shifted to a similarity value—called “Likelihood score”—between 0 and 1, by the use of an exponential function [34].

Given a signature’s most likely path, we consider an N -components segmentation vector, N being the number of states in the claimed identity’s HMM. This vector has in the i^{th} position the number of feature vectors that were associated to state i by the Viterbi path, as shown in Fig. 6.4. We then characterize each of the training signatures by a reference segmentation vector. In the verification phase (as shown in Fig. 6.5) for each test signature, the City Block Distance between its associated segmentation vector and all the reference segmentation vectors are computed, and such distances are averaged. This average distance is then shifted to a similarity measure between 0 and 1 (Viterbi score) by an exponential function [34]. Finally, on a given test signature, these two similarity measures based on the classical likelihood and on the segmentation of the test signature by the target model are fused by a simple arithmetic mean.

6 Online Handwritten Signature Verification

Fig. 6.4 Computation of the segmentation vector of Ref1 system

Fig. 6.5 Exploitation of the Viterbi Path information (SV stands for Segmentation Vector) of Ref1 system

6.5.3 Reference System 2 (Ref2 v1.0)

617

²The basis for this algorithm is a transformation of dynamic handwriting signals (position, pressure and velocity of the pen) into a character string, and the comparison of two character strings based on test and reference handwriting samples, according to the Levenshtein distance method [19]. This distance measure determines a value for the similarity of two character strings. To get one of these character strings, the online signature sample data must be transferred into a sequence of characters as described by Schimke et al. [30]: from the handwriting raw data (pen position and pressure), the pen movement can be interpolated and other signals can be determined, such as the velocity. The local extrema (minimum, maximum) of the function curves of the pen movement are used to transfer a signature into a string. The occurrence of such an extreme value is a so-called *event*. Another event type is a gap after each segment of the signature, where a segment is the signal from one pen-down to the subsequently following pen-up. A further type of event is a short segment, where it is not possible to determine extreme points because insufficient

² This section is reproduced with permission from Annales des Telecommunications, source [13].

data are available. These events can be subdivided into single points and segments from which the stroke direction can be determined. The pen movement signals are analyzed, then the feature events ε are extracted and arranged in temporal order of their occurrences in order to achieve a string-like representation of the signature. An overview of the described events ε is represented in Table 6.2.

t2.1 **Table 6.2** The possible event types present in the Reference System 2 (Ref2)

t2.2	E-code	S-code	Description
t2.3	$\varepsilon_1 \dots \varepsilon_6$	$xXyYpP$	$x - \min, x - \max, y - \min, y - \max, p - \min, p - \max$
t2.4	$\varepsilon_7 \dots \varepsilon_{12}$	$v_x V_x v_y V_y vV$	$v_x - \min, v_x - \max, v_y - \min, v_y - \max, v - \min, v - \max$
t2.5	$\varepsilon_{13} \dots \varepsilon_{14}$	$g d$	gap, point
t2.6	$\varepsilon_{15} \dots \varepsilon_{22}$		short events; directions: $\uparrow, \nearrow, \rightarrow, \searrow, \downarrow, \swarrow, \leftarrow, \nwarrow$

At the transformation of the signature signals, the events are encoded with the characters of the column entitled 'S-Code' resulting in a string of events: positions are marked with x and y , pressure with p , velocities with v , v_x and v_y , gaps with g and points with d . Minimum values are encoded by lower case letters and maximum values by capital letters. One difficulty in the transformation is the simultaneous appearance of extreme values of the signals because then no temporal order can be determined. This problem of simultaneous events can be treated by the creation of a combined event, requiring the definition of scores for edit operations on those combination events. In this approach, an additional normalization of the distance is performed due to the possibility of different lengths of the two string sequences [30]. This is necessary because the lengths of the strings created using the pen signals can be different due to the fluctuations of the biometric input. Therefore, signals of the pen movement are represented by a sequence of characters. Starting out with the assumption that similar strokes have also similar string representations, biometric verification based on signatures can be carried out by using the Levenshtein distance.

The Levenshtein distance determines the similarity of two character strings through the transformation of one string into another one, using operations on the individual characters. For this transformation a sequence of operations (insert, delete, replace) is applied to every single character of the first string in order to convert it into the second string. The distance between the two strings is the smallest possible number of operations in the transformation. An advantage of this approach is the use of weights for each operation. The weights depend on the assessment of the individual operations. For example, it is possible to weight the deletion of a character higher than replacing it with another character. A weighting with respect to the individual characters is also possible. The formal description of the algorithm is given by the following recursion:

6 Online Handwritten Signature Verification

$$\left. \begin{aligned}
D(i, j) &:= \min \begin{cases} D(i-1, j) + w_d, \\ D(i, j-1) + w_i, \\ D(i-1, j-1) + w_r \end{cases} \\
D(i, 0) &:= D(i-1, 0) + w_d \\
D(0, j) &:= D(0, j-1) + w_i \\
D(0, 0) &:= 0
\end{aligned} \right\} \forall i, j > 0 \quad (6.2)$$

In this description, i and j are the lengths of strings S_1 and S_2 respectively. w_i , w_d 664
and w_r are the weights of the operations insert, delete and replace. If characters 665
 $S_1[i] = S_2[j]$ the weight w_r is 0. Smaller distance D between any two strings S_1 and 666
 S_2 denotes greater similarity. 667

6.5.4 Benchmarking Databases and Protocols

668

Our aim is to propose protocols on selected publicly available databases, for 669
comparison purposes relative to the two reference systems described in Sect. 6.5. 670
We thus present in this section the protocols associated with three publicly avail- 671
able databases: the BIOMET signature database [12], and the two MCYT signature 672
databases [22] (MCYT-100 and the complete MCYT-330) for test purposes. 673

6.5.4.1 Protocols on the BIOMET Database

674

On this database, we distinguish two protocols; the first one does not take into ac- 675
count the temporal variability of the signatures; the second exploits the variability 676
of the signatures over time (five months spacing between the two sessions). 677

In order to reduce the influence of the selected five enrollment signatures, we 678
have chosen to use a cross-validation technique to compute the generalization er- 679
ror of the system and its corresponding confidence level. We have considered 100 680
samplings (or trials) of the five enrollment signatures on the BIOMET database, as 681
follows: for each writer, five reference signatures are randomly selected from the 682
10 genuine signatures available of Session 2 and only the genuine test set changes 683
according to the protocol; in the first protocol (Protocol 1), test is performed on the 684
remaining five genuine signatures of Session 2—which means that no time variabil- 685
ity is present in data—as well as on the 12 skilled forgeries and the 83 random 686
forgeries. In the second protocol (Protocol 2), test is performed on the five genuine 687
signatures of Session 1—this way introducing time variability in data—as well as 688
on the 12 skilled forgeries and the 83 random forgeries. We repeat this procedure 689
100 times for each writer. 690

6.5.4.2 Protocol on the MCYT-100 Data Subset

691

On MCYT-100 database, we consider in the same way 100 samplings of the five 692
 enrollment signatures out of the 25 available genuine signatures. For each sampling, 693
 systems were tested on the 20 remaining genuine signatures, the 25 skilled forgeries 694
 and the 99 random forgeries available for each person. 695

6.5.4.3 Protocol on the Complete MCYT Database (MCYT-330)

696

As MCYT-330 is a much larger database than BIOMET and MCYT-100, for com- 697
 putational reasons, we could not envisage performing 100 samplings of the five 698
 enrollment signatures. Therefore, a study on the necessary number of samplings to 699
 reach an acceptable confidence on our results was carried out on this complete data- 700
 base. This study was carried out with the two reference systems (Ref1 and Ref2) 701
 described in Sect. 6.5. 702

To that end, we assume that the standard deviation (across samplings) of the 703
 Equal Error Rate (EER) (computed on skilled forgeries) is reduced when the num- 704
 ber of samplings increases. Thus, we search the number of random samplings that 705
 ensures that the standard deviation (across samplings) of the EER is significantly 706
 reduced. Figure 6.6 shows results on both reference systems. 707

Fig. 6.6 Standard deviation of the Equal Error Rate (EER) on MCYT-330 database, against the number of random samplings of the five enrollment signatures on skilled forgeries: (a) Reference System 1 and (b) Reference System 2

According to Fig. 6.6, we notice that at 15 random samplings, the standard devi- 708
 ation lowers for Reference System 1, but that 25 random samplings are required to 709
 reach the same result for Reference System 2. Besides, we notice in both cases that a 710
 substantial increase in the number of samplings (up to 100 samplings) does not lead 711

6 Online Handwritten Signature Verification

to a lower value of the standard deviation of the EER. Given these results, we assumed that 25 random samplings are sufficient to evaluate systems on the complete MCYT database, instead of considering 100 samplings as done on BIOMET and MCYT-100 databases which are smaller.

The resulting protocol is thus the following: five genuine signatures are randomly selected among the 25 available genuine signatures for each writer and this procedure is repeated 25 times. For each person, the 20 remaining genuine signatures, the 25 skilled forgeries and the 329 random forgeries are used for test purposes.

6.5.5 Results with the Benchmarking Framework

We compute the Detection Error Tradeoff (DET) curves [20] of Reference System 1 (Ref1 v1.0) and Reference System 2 (Ref2 v1.0). We report the average values of Equal Error Rate (EER) corresponding to Reference Systems, over 100 samplings on MCYT-100 database according to the protocol described in Sect. 6.5.4.2, and over 100 samplings on BIOMET database with two different protocols. In the first protocol, the test is performed only on the remaining five genuine signatures of Session 2 (Protocol 1, described in Sect. 6.5.4.1). In the second protocol, the test is performed on both the remaining five genuine signatures of Session 2 and the five genuine signatures of Session 1 (Protocol 3). Experimental results of the two Reference Systems on the BIOMET database according to Protocol 2, described in Sect. 6.5.4.1, are presented in Sect. 6.6.6. Two schemes are studied: one considering skilled forgeries, the other one considering random forgeries.

Fig. 6.7 DET curves of the two reference systems—Ref1 v1.0 and Ref2 v1.0—on the MCYT-100 database on skilled and random forgeries

t3.1 **Table 6.3** EERs of the two reference systems—Ref1 v1.0 and Ref2 v1.0—on the MCYT-100 database and their Confidence Interval (CI) of 95%, on skilled and random forgeries

t3.2	MCYT-100			
t3.3	Skilled forgeries		Random forgeries	
t3.4	System	EER% CI 95%	System	EER% CI 95%
t3.5	Ref1	3.41 ± 0.05	Ref1	0.95 ± 0.03
t3.6	Ref2	10.51 ± 0.13	DTWstd	4.95 ± 0.09

t4.1 **Table 6.4** EERs of the two reference systems—Ref1 v1.0 and Ref2 v1.0—on the BIOMET database and their Confidence Interval (CI) of 95%, on skilled and random forgeries

t4.2	BIOMET				
t4.3	Protocol 1			Protocol 3	
t4.4		Skilled forgeries	Random forgeries	Skilled forgeries	Random forgeries
t4.5	System	EER% CI 95%	EER% CI 95%	EER% CI 95%	EER% CI 95%
t4.6	Ref1	2.37 ± 0.06	1.60 ± 0.07	4.93 ± 0.07	4.06 ± 0.06
t4.7	Ref2	8.26 ± 0.15	6.83 ± 0.16	9.55 ± 0.15	7.80 ± 0.14

Fig. 6.8 DET curves of the two reference systems—Ref1 v1.0 and Ref2 v1.0—on the BIOMET database according to (a) Protocol 3 and (b) Protocol 1, on skilled and random forgeries

6.6 Research Algorithms Evaluated within the Benchmarking Framework

733

734

In this section, results within the proposal evaluation framework of several research systems are presented. Among them are also present the best approaches within the

735

736

6 Online Handwritten Signature Verification

SVC'2004 evaluation campaign (DTW and HMM). All these systems are evaluated first with the benchmarking framework and also with other relevant protocols.

6.6.1 HMM-based System from Universidad Autonoma de Madrid (UAM)

³This online signature verification system is based on functional feature extraction and Hidden Markov Models (HMMs) [7]. This system was submitted by Universidad Politecnica de Madrid (UPM) to the first international Signature Verification Competition (SVC'2004) with excellent results [36]: in Task 2 of the competition, where both trajectory and pressure signals were available, the system was ranked first when testing against random forgeries. In the case of testing skilled forgeries, the system was only outperformed by the winner of the competition, which was based on Dynamic Time Warping [3]. Below, we provide a brief sketch of the system, and for more details we refer the reader to [7].

Feature extraction is performed as follows. The coordinate trajectories (x_n, y_n) and the pressure signal p_n are the components of the unprocessed feature vectors, where $n = 1, \dots, N_s$ and N_s is the duration of the signature in time samples. In order to retrieve relative information from the coordinate trajectory (x_n, y_n) and not being dependent on the starting point of the signature, signature trajectories are preprocessed by subtracting the center of mass. Then, a rotation alignment based on the average path tangent angle is performed. An extended set of discrete-time functions is derived from the preprocessed trajectories. The resulting functional signature description consists of the following feature vectors $(x_n, y_n, p_n, \theta_n, v_n, \rho_n, a_n)$ together with their first order time derivatives, where θ, v, ρ and a stand, respectively, for path tangent angle, path velocity magnitude, log curvature radius and total acceleration magnitude. A whitening linear transformation is finally applied to each discrete-time function so as to obtain zero mean and unit standard deviation function values.

Given the parametrized enrollment set of signatures of a user, a continuous left-to-right HMM was chosen to model each signer's characteristics. This means that each person's signature is modeled through a double stochastic process, characterized by a given number of states with an associated set of transition probabilities, and, in each of such states, a continuous density multivariate Gaussian mixture. No transition skips between states are permitted. The Hidden Markov Model (HMM) λ is estimated by using the Baum-Welch iterative algorithm. Given a test signature parametrized as O (with a duration of N_s time samples) and the claimed identity previously enrolled as λ , the similarity matching score s :

$$s = \frac{1}{N_s} \log p(O/\lambda) \quad (6.3)$$

is computed by using the Viterbi algorithm [26].

³ This section is reproduced with permission from Annales des Telecommunications, source [13].

6.6.2 GMM-based System

773

This system is based on a coupling of a Gaussian Mixture Model [27] with a local feature extraction, namely the 25 local features used for Reference System 1, described in Sect. 6.5.2. Indeed, it is an interesting question whether the GMM viewed as a degenerate version of an HMM can be compared to an HMM-based approach in terms of performance on the same feature extraction. As for the HMM-based approach of Reference System 1, the number of Gaussians of the writer GMM is chosen in a personalized way, which is set as $4N$ where N is the number of states as computed in (6.1).

6.6.3 Standard DTW-based System

782

This system is based on Dynamic Time Warping, which compensates for local handwriting variations, and allows to determine the dissimilarity between two time sequences with different lengths [26]. This method, with a polynomial complexity, computes a matching distance by recovering optimal alignments between sample points in the two time series. The alignment is optimal in the sense that it minimizes a cumulative distance measure consisting of “local” distances between aligned samples. In this system, the DTW-distance between two time series $x_1, \dots, x_M$ and $y_1, \dots, y_N$, $D(M, N)$ is computed as:

$$D(i, j) = \min \left\{ \begin{array}{l} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{array} \right\} + w_p \times d(i, j) \quad (6.4)$$

where the “local” distance function $d(i, j)$ is the Euclidian distance between i^{th} reference point and j^{th} testing point, with $D(0, 0) = d(0, 0) = 0$, and equal weights w_p are given to insertions, deletions and substitutions. Of course, in general, the nature of the recurrence equation (which are the local predecessors of a given point), and the “local” distance function $d(i, j)$ may vary [26]. Standard DTW-based system aligns by Dynamic Time Warping (DTW) a test signature with each reference signature and the average value of the resulting distances is used to classify the test signature as being genuine or a forgery. If the final distance is lower than the value of the decision threshold, the claimed identity is accepted. Otherwise it is rejected.

6.6.4 DTW-based System with Score Normalization

800

This system is also based on Dynamic Time Warping (DTW), calculated according to (6.4). However, a score normalization following the principle of Kholmatov’s system [3], the winning system of the first international Signature Verification

6 Online Handwritten Signature Verification

Competition [36] (SVC'2004), is introduced. This normalization, previously de- 804
scribed in Sect. 6.2.1.2, relies on intraclass variation; more precisely, the system 805
only normalizes the output distance (defined as the average DTW distances between 806
the test signature and the five reference signatures) by dividing the latter by the aver- 807
age of pairwise DTW distances in the enrollment set. This results in a dissimilarity 808
score which is in a range from 0 to 1. If this final score is lower than the value of the 809
threshold, the test signature is authentic, otherwise it is a forgery. 810

We chose to consider this second version of a DTW-based approach because 811
Kholmatov's system obtained indeed excellent results in SVC'2004 with such intra- 812
class normalization, particularly in comparison to statistical approaches based on 813
HMMs [36]. 814

6.6.5 System Based on a Global Approach

815

The principle of this system is to compute, from the temporal dynamic functions 816
acquired by the digitizer (coordinates, pressure, pen-inclination angles), 41 global 817
features (described in Table 6.5), and to compare a test signature to a reference sig- 818
nature by the City Block distance. 819

During enrollment, each user supplies five reference signatures. Global feature 820
vectors are extracted from such five reference signatures, and the average value 821
of all pairwise distances is computed by the City Block distance, to be used as 822
a normalization factor. In the verification phase, a test signature is compared to 823
each reference signature by the City Block distance, providing an average distance. 824
The final dissimilarity measure results from the ratio of this average distance and the 825
normalization factor previously mentioned. If this final value is lower than the value 826
of the decision threshold, the claimed identity is accepted, otherwise it is rejected. 827

6.6.6 Experimental Results

828

We compute the Detection Error Tradeoff (DET) curves [20] of the following seven 829
systems: Reference System 1 (*Ref1*), Reference System 2 (*Ref2*), UAM's HMM- 830
based system (*UAM*), a GMM-based system with local features (*GMM*), a standard 831
DTW system (*DTWstd*), the DTW system with the score normalization based on 832
intraclass variance (*DTWnorm*) and, finally, a system based on a global approach 833
and a normalized City Block distance measure (*Globalappr*). 834

We report the average values of Equal Error Rate (EER) corresponding to the 835
seven systems, over 100 samplings on BIOMET and MCYT-100, and over 25 sam- 836
plings on the complete MCYT database (according to our previous study on the 837
required number of samplings in Sect. 6.5.4.3). Two schemes are studied: one con- 838
sidering skilled forgeries, the other one considering random forgeries. 839

t5.1 **Table 6.5** The 41 global features extracted from the online signature

t5.2	N°	Feature name
t5.3	1	Signature Duration
t5.4	2	Number of sign changes in X
t5.5	3	Number of sign changes in Y
t5.6	4	Standard Deviation of acceleration in x by the maximum of acceleration in x
t5.7	5	Standard Deviation of acceleration in y by the maximum of acceleration in y
t5.8	6	Standard Deviation of velocity in x by the maximum of velocity in x
t5.9	7	Average velocity in x by the maximum of velocity in x
t5.10	8	Standard Deviation of velocity in y by the maximum of velocity in y
t5.11	9	Average velocity in y by the maximum of velocity in y
t5.12	10	Root mean square (RMS) of y position by the difference between maximum and minimum of position in y
t5.13	11	Ratio velocity change
t5.14	12	Average velocity by the maximum of velocity
t5.15	13	Root mean square (RMS) of velocity in x by the maximum of the velocity
t5.16	14-15-16	Coordinates ratio
t5.17	17	Correlation velocity by square of velocity
t5.18	18	Root mean square (RMS) of acceleration by maximum of acceleration
t5.19	19	Average acceleration by the maximum of acceleration
t5.20	20	Difference between the Root mean square (RMS) and the minimum of x position by the RMS of x position
t5.21	21	Root mean square (RMS) of x position by the difference between maximum and minimum of position in x
t5.22	22	Root mean square of velocity in x by the maximum of velocity in x
t5.23	23	Root mean square of velocity in y by the maximum of velocity in y
t5.24	24	Velocity ratio
t5.25	25	Acceleration ratio
t5.26	26	Correlation coordinates
t5.27	27	Correlation coordinates by the square of the maximum of acceleration
t5.28	28	Difference between the Root mean square and the minimum of y position by the Root mean square of y position
t5.29	29	Difference between the Root mean square and the minimum of velocity in x by the Root mean square of velocity in x
t5.30	30	Difference between the Root mean square and the minimum of velocity in y by the Root mean square of velocity in y
t5.31	31	Difference between the Root mean square and the minimum of velocity by the Root mean square of velocity
t5.32	32	Difference between the Root mean square and the minimum of acceleration in x by the Root mean square of acceleration in x
t5.33	33	Difference between the Root mean square and the minimum of acceleration in y by the Root mean square of acceleration in y
t5.34	34	Difference between the mean and the minimum of acceleration by the mean of acceleration
t5.35	35	Ratio of the time of max velocity over the total time
t5.36	36	Number of strokes
t5.37	37-38	Mean of positive and negative velocity in x
t5.38	39-40	Mean of positive and negative velocity in y

6.6.6.1 On the BIOMET Database

840

As we mentioned in Sect. 6.5.4.1, two cases are studied, depending on presence of
time variability or not.

842

With no Time Variability In Table 6.6, the performances of the seven systems are
presented at the Equal Error Rate (EER) point with a confidence interval of 95%.
For more insight, we also report the performance of the two systems based on the
two scores that are fused in Ref1 System, the classical score based on Likelihood
(the associated system is denoted by *Ref1-Lik*) and the score based on the segmen-
tation of the test signature by the target model (the associated system is denoted by
Ref1-Vit).

849

t6.1 **Table 6.6** EERs of the seven systems on the BIOMET database and their Confidence Intervals of
95% in case of no time variability

t6.2	BIOMET—without time variability			
	Skilled forgeries		Random forgeries	
t6.3				
t6.4	System	EER% CI 95%	System	EER% CI 95%
t6.5	Ref1	2.37 ± 0.06	Ref1	1.60 ± 0.07
t6.6	UAM	3.41 ± 0.11	UAM	1.90 ± 0.10
t6.7	Ref1-Vit	3.86 ± 0.08	Globalappr	3.25 ± 0.09
t6.8	Globalappr	4.65 ± 0.10	Ref1-Vit	3.42 ± 0.10
t6.9	Ref1-Lik	4.85 ± 0.11	Ref1-Lik	3.72 ± 0.10
t6.10	GMM	5.13 ± 0.12	GMM	3.77 ± 0.10
t6.11	DTWstd	5.21 ± 0.09	DTWstd	5.25 ± 0.09
t6.12	DTWnorm	5.47 ± 0.11	DTWnorm	5.58 ± 0.10
t6.13	Ref2	8.26 ± 0.15	Ref2	6.83 ± 0.16

It clearly appears in Table 6.6 and Fig. 6.9 that the best approaches are sta-
tistical, particularly those based on Hidden Markov Models (Reference System 1
and UAM's HMM system) for both skilled and random forgeries. Furthermore, at
the EER point, the best HMM-based system is Reference System 1, which performs
the fusion of two sorts of information corresponding to two levels of description of
the signature—the likelihood information that operates at the point level, and the
segmentation information that works on portions of the signature, corresponding to
an intermediate level of description.

850

851

852

853

854

855

856

857

This result is followed at the EER by the performance of UAM's HMM system,
whose output score is the Log-likelihood of the test signature given the model. Nev-
ertheless, when analyzing all functioning points in DET curves (see Fig. 6.9), we
notice that UAM's HMM system is better than Reference System 1 when the False
Acceptance Rate is for skilled forgeries lower than 1% and for random forgeries
lower than 0.3%. Also, UAM's HMM system performs better than Ref1's likeli-
hood score alone and Ref1's Viterbi score alone at the EER, particularly on random

858

859

860

861

862

863

864

Fig. 6.9 DET curves on the BIOMET database of the seven systems in case of no time variability: (a) on skilled forgeries and (b) on random forgeries

forgeries with a relative improvement of around 46%. Indeed, although Ref1 and UAM's system are both based on HMMs, their structural differences are important: Reference System 1 uses a variable number of states (between 2 and 15) according to the client's enrollment signatures length, whereas UAM's system uses only two states, and the number of Gaussian components per state is also different (4 in the case of Reference System,1, and 32 in the case of UAM's System). As already shown in a previous study [13], such differences in the HMM architecture lead to complementary systems, simply because the resulting "compression" of the information of a signature (the model) is of another nature.

HMM-based systems are followed in performance first by the distance-based approach using 41 global features, and then by the GMM-based system with local features and a personalized number of Gaussians.

On the other hand, it is clearly shown in Fig. 6.9 that the worst systems are elastic distance-based systems. Indeed, we notice that both DTW-based systems that work at the point level reach an Equal Error Rate (EER) of 5.3% on skilled and random forgeries, and Reference System 2 using Levenshtein distance coupled to a coarse segment-based feature extraction reaches an EER of 8.26% on skilled forgeries and 6.82% on random forgeries. This result is unexpected, particularly for the normalized DTW-based system evaluated in this work that uses the same principles in the Dynamic Time Warping (DTW) implementation as the winning system of SVC'2004 [36] (equal weights for all operations of insertion, destruction and substitutions, score normalization by intraclass variance, see Sect. 6.2.2).

This result may be mainly due to two reasons: a) the DTW winning algorithm of SVC'2004 uses only two position derivatives, whereas our DTW-based system uses 25 features, among which we find the same two position derivatives; and b) particularities of the BIOMET database. Indeed, this database has forgeries that are

AQ: Please check the edit of "shtein" to "Levenshtein" is ok.

6 Online Handwritten Signature Verification

in general much longer than genuine signatures (by a factor two or three), as well as highly variable genuine signatures from one instance to another. In this configuration of the data, statistical models are better suited to absorb client intraclass variability and, in particular, we remark that the HMM performs better than the GMM, because it performs a segmentation of the signature, which results in an increased detection power of skilled forgeries with regard to the GMM. This phenomenon is clearly illustrated by the good performance of the system based only on the segmentation score (called *Ref1-Vit*), one of the two scores fused by Reference System 1.

Also, the good ranking of the distance-based approach coupled to a global feature extraction (41 global features), behind the two HMM-based systems, is due to a smoothing effect of the holistic feature extraction that helps to characterize a variable signature, and to detect coarse differences between forgeries and genuine signatures, as signature length among others. Moreover, comparing now the two systems based on Dynamic Time Warping, we notice that the DTW system using the normalization based on intraclass variance is outperformed by the standard DTW in the area of low False Acceptance. This can be explained by the high intraclass variability on BIOMET database, since genuine signatures are in general highly variable as already pointed out. On such clients, the personalized normalization factor distorts the scores obtained on forgeries thus producing false acceptances. To go further, if we suppose that the SVC'2004 Test Set (unfortunately not available to the scientific community) is mainly of the same nature as the SVC development set, we may assert that the data configuration in the BIOMET database is completely opposite of that of SVC, which contains stable genuine signatures and skilled forgeries of approximately the same length as this of the genuine signatures. This enormous difference in the very nature of the data impacts systems in a totally different way and explains that the best approaches in one case are no longer the best in another case. We will pursue our analysis of results with this spirit on the other databases as well.

In Presence of Time Variability Table 6.7 reports the experimental results obtained by each system in presence of time variability. First, we notice an important degradation of performance for all systems; this degradation is of at least 200% for the Global distance-based approach and can even reach 360% for Reference System 1. Indeed, we recall that in this case, time variability corresponds to sessions acquired five months apart, thus to long-term time variability.

As in the previous case of no time variability, results in Table 6.7 and Fig. 6.10, show that the best approaches remain those based on HMMs (Ref1 and UAM's systems). Ref1 is still the best system on skilled forgeries, but on random forgeries UAM system is the best. This can be explained by the fact that the score based on segmentation used in Ref1, as mentioned before, helps in detecting skilled forgeries on BIOMET based mainly on a length criterion. This does not occur on random forgeries of course, and thus there is no longer a substantial difference in segmentation between forgeries and genuine signatures.

t7.1 **Table 6.7** EERs of the seven systems on the BIOMET database and their Confidence Interval (CI) of 95%, in case of time variability

t7.2	BIOMET—with time variability			
t7.3	Skilled forgeries		Random forgeries	
t7.4	System	EER% CI 95%	System	EER% CI 95%
t7.5	Ref1	6.63 ± 0.10	UAM	4.67 ± 0.13
t7.6	UAM	7.25 ± 0.15	Ref1	5.79 ± 0.10
t7.7	Ref1-Vit	7.61 ± 0.12	Globalappr	6.61 ± 0.14
t7.8	Globalappr	8.91 ± 0.14	Ref2	8.69 ± 0.18
t7.9	Ref2	10.58 ± 0.18	Ref1-Vit	8.88 ± 0.17
t7.10	DTWstd	11.36 ± 0.10	GMM	9.01 ± 0.10
t7.11	DTWnorm	11.83 ± 0.51	Ref1-Lik	9.90 ± 0.11
t7.12	Ref1-Lik	12.78 ± 0.15	DTWstd	12.67 ± 0.12
t7.13	GMM	13.15 ± 0.14	DTWnorm	13.85 ± 0.28

Fig. 6.10 DET curves on the BIOMET database of the seven systems in case of time variability: (a) on skilled forgeries and (b) on random forgeries

Taking into consideration time variability, we notice that the system based only on Likelihood score (Ref1-Lik) gives lower results compared to the case without time variability, with degradation around 260% on skilled forgeries (from 4.85% to 12.78%) and random forgeries (from 3.72% to 9.9%). The system based only on the Viterbi score (Ref1-Vit) is however more robust in terms of time variability on skilled forgeries, with a degradation of 190% (from 3.86% to 7.61%). On random forgeries, Ref1-Vit is degraded in the same way as Ref1-Lik (from 3.42% to 8.88%).

6 Online Handwritten Signature Verification

The HMM-based systems are followed in performance by the Global distance-based approach as observed without time variability, then by Reference System2, then by the standard DTW, and finally by the score-normalized-DTW. The GMM-based system, which is the last system on skilled forgeries, behaves differently on random forgeries, but it still remains among the last systems. It is interesting to note that the Global distance-based approach shows the lowest degradation in performance (around 200% on skilled and random forgeries). This approach processes the signature as a whole, and when the signature varies over time, strong local variations appear, but such variations are smoothed by the holistic description of the signature performed by the feature extraction step.

For the same reason, elastic distance works well when coupled to a rough feature extraction detecting “events” and encoding the signature as a string (case of Ref2) while it gives bad results when coupled to a local feature extraction (case of both DTW-based systems).

On the other hand, the same tendencies observed in the previous subsection (without time variability) and due to the characteristics of the BIOMET database, are observed in this case for DTW-based systems and the GMM. The GMM-based system, coupled to a local feature extraction, which gave good results in the previous case, doesn’t resist well to time variability and becomes in this case the last system on skilled forgeries. This result is interesting because it suggests that the piecewise stationarity assumption on the signature, inherent to an HMM, may be helpful with respect to a GMM mostly in presence of long-term time variability. Unfortunately, there has been no online signature verification evaluation yet on long-term time variability data that would permit to assess this analysis. Nevertheless, to have a better insight on the contribution of the segmentation information (the fact of modeling a signature by different states) in the presence of time variability, we compared three statistical approaches on the BIOMET database: Ref1, HMM-based, a GMM with a personalized number of Gaussians, and a GMM with different configurations in the Gaussian mixture (2, 4, 8, 16, 32, 64 and 128 Gaussians). Experimental results obtained by such systems are reported in Table 6.8 (a) and (b).

t8.1 **Table 6.8** EERs of the GMM corresponding to different number of Gaussians on the BIOMET database: (a) on skilled forgeries and (b) on random forgeries

t8.2	<i>N° Gaussians</i>	2	4	8	16	32	64	128	GMM person
t8.3	a) Without variab.	15.88%	12.37%	10.16%	9.16%	8.92%	9.95%	13.53%	5.13%
t8.4	With variab.	26.6%	22.95%	19.9%	18.04%	17.00%	17.6%	21.74%	13.15%
t8.5	b) Without variab.	11.96%	9.47%	7.99%	7.02%	6.71%	13.69%	11.31%	3.77%
t8.6	With variab.	20.02%	16.93%	14.59%	12.99%	12.46%	13.69%	18.03%	9.01%

Fig. 6.11 DET curves on the BIOMET database comparing Ref1 system, a GMM with a personalized number of Gaussians, and a GMM with a different number of Gaussians on skilled forgeries: (a) without time variability and (b) with time variability

Fig. 6.12 DET curves on the BIOMET database comparing Ref1 system, GMM with a personalized number of Gaussians, and GMM with a different number of Gaussians on random forgeries: (a) without time variability and (b) with time variability

Figures 6.11 and 6.12, as well as Tables 6.8 (a) and (b), show that a GMM with a personalized number of Gaussians performs better than a GMM with a common number of Gaussians for all users, even in the best case.

6.6.6.2 On MYCT Database

973

Tables 6.9 and 6.10 report the experimental results obtained by all systems when 974
 tested on the MCYT-100 subset and on the complete MCYT database, with the 975
 evaluation protocols described in Sects. 6.5.4.2 and 6.5.4.3, respectively. 976

t9.1 **Table 6.9** EERs of the seven systems on the MCYT-100 database and their Confidence Interval (CI) of 95%

MCYT-100				
Skilled forgeries			Random forgeries	
System	EER% CI 95%	System	EER% CI 95%	
Ref1	3.41 ± 0.05	Ref1	0.95 ± 0.03	
DTWnorm	3.91 ± 0.07	DTWstd	1.20 ± 0.06	
UAM	5.37 ± 0.08	DTWnorm	1.28 ± 0.04	
Ref1-Vit	5.59 ± 0.07	Ref1-Lik	2.13 ± 0.05	
Ref1-Lik	5.66 ± 0.07	UAM	2.34 ± 0.05	
DTWstd	5.96 ± 0.09	Ref1-Vit	2.44 ± 0.04	
GMM	6.74 ± 0.09	GMM	2.81 ± 0.05	
Globalappr	7.23 ± 0.10	Globalappr	3.15 ± 0.07	
Ref2	10.51 ± 0.13	Ref2	4.95 ± 0.09	

When comparing DET curves on skilled and random forgeries for each database 977
 in Figs. 6.13 and 6.14, it clearly appears that very similar results are obtained on 978
 MCYT-100 and MCYT-330. Therefore, we analyze the experimental results only 979
 on the MCYT-330 database, as it is the complete database. 980

As shown in Fig. 6.14, the best system is Reference System 1, for both skilled 981
 and random forgeries. The fact that Ref1 is the best system on two different data- 982
 bases, BIOMET and MCYT, having different characteristics (size of the population, 983
 sensor resolution, nature of skilled forgeries, stability of genuine signatures, pres- 984
 ence or not of time variability, nature of time variability, etc.) that can strongly in- 985
 fluence systems performance, shows that Reference System 1 holds up well across 986
 databases [34]. 987

Ref1 system is indeed followed at the EER point by the score-normalized DTW 988
 system (an increase of EER from 3.91% to 4.4% on skilled forgeries, and from 989
 1.27% to 1.69% on random forgeries). At other functioning points, the gap between 990
 Ref1 and the normalized DTW increases. This good result of the score-normalized 991
 DTW system can be explained by the fact that this score normalization exploits 992
 information about intraclass variability and the nature of the MCYT database in 993
 this respect. Indeed, as in this database the genuine signatures do not vary as much 994
 as in the BIOMET database, no distortion effects occur on the scores on forgeries 995
 when applying the normalization. This result also confirms the tendency of results 996
 obtained in SVC'2004 [36] concerning the coupling of a normalization based on 997

t10.1 **Table 6.10** EERs of the seven systems on the complete MCYT database (MCYT-330) and their Confidence Interval (CI) of 95%

MCYT-330				
	Skilled forgeries		Random forgeries	
	System	EER% CI 95%	System	EER% CI 95%
t10.5	Ref1	3.91 ± 0.09	Ref1	1.27 ± 0.04
t10.6	DTWnorm	4.40 ± 0.08	DTWnorm	1.69 ± 0.05
t10.7	Ref1-Vit	5.81 ± 0.12	DTWstd	1.75 ± 0.08
t10.8	UAM	6.31 ± 0.11	UAM	1.97 ± 0.05
t10.9	Ref1-Lik	6.57 ± 0.10	Ref1-Vit	2.91 ± 0.06
t10.10	DTWstd	7.04 ± 0.09	Ref1-Lik	2.93 ± 0.06
t10.11	Globalappr	7.45 ± 0.09	Globalappr	3.22 ± 0.06
t10.12	GMM	7.45 ± 0.09	GMM	3.83 ± 0.07
t10.13	Ref2	12.05 ± 0.10	Ref2	6.35 ± 0.10

Fig. 6.13 DET curves of the seven systems on the MCYT-100 database: (a) on skilled forgeries and (b) on random forgeries

intraclass variance and a DTW. To go further, MCYT is more similar in two main 998 characteristics to the SVC Development Set, than it is to the BIOMET database: 999 on one hand more stability in genuine signatures, and on the other hand skilled 1000 forgeries not having an important difference in length with respect to genuine sig- 1001 natures. Finally, we notice that on the complete MCYT database, the statistical ap- 1002 proach based on the fusion of two sorts of information, Reference System 1, that 1003 participated in the SVC'2004 and was then outperformed by this normalized DTW 1004 approach, gives the best results in this case. Of course, some other differences in na- 1005 ture between MCYT and SVC databases, like for instance the influence of cultural 1006

6 Online Handwritten Signature Verification

Fig. 6.14 DET curves of the seven systems on the MCYT-330 database: (a) on skilled forgeries and (b) on random forgeries

types of signature on approaches (MCYT contains only Western signatures while 1007
SVC mixes Western and Asian styles), may be responsible of this result. 1008

Nevertheless, we still notice on MCYT a phenomenon already observed on BIO- 1009
MET: that the normalization based on intraclass variance coupled to DTW increases 1010
considerably the False Rejection Rate in the zone of low False Acceptance Rate. The 1011
standard DTW gives lower results but does not show this effect in this zone. This 1012
phenomenon also increases significantly on random forgeries. 1013

Although UAM's HMM system was favored on MCYT since it was optimized 1014
on the first 50 writers and the test is performed on the totality of the 330 writers, it 1015
is ranked behind Ref1 and the DTW with intraclass variance normalization. 1016

On the other hand, we observe that on MCYT the worst system is the Global 1017
distance-based approach, followed by Reference System 2, while the GMM with 1018
local features remains in between the best model-based approaches (HMM-based) 1019
and the worst systems. Indeed, as the database shows less variability in the genu- 1020
ine signatures and less differences in the dynamics between the forgeries and the 1021
genuine signatures than BIOMET, a global feature extraction coupled with a simple 1022
distance measure, or alternatively an elastic distance with a coarse segment-based 1023
feature extraction, are not precise enough to discriminate genuine signatures from 1024
forgeries in this case. 1025

6.7 Conclusions

1026

We have made in the present chapter a complete overview of the online signa- 1027
ture verification field, by analyzing the existing literature relatively to the main ap- 1028
proaches, by recalling the main results of the international evaluations performed, 1029

by describing the existing public databases at the disposal of the researcher in the field, and by discussing the main challenges nowadays.

Our experimental work on seven online signature verification systems and three different public databases with different protocols completes the view of the field. We have indeed compared in this work model-based approaches (two HMM-based systems, and a GMM-based system) coupled to local features, and distance-based approaches (two elastic distances coupled to local features, an elastic distance coupled to a segment-based feature extraction, and a standard distance coupled to a holistic feature extraction).

Our experiments lead to the following conclusions. We had three databases but one is a subset of the other, so mainly we are in the presence of two data configurations in terms of acquisition protocol, population, sensor, nature of forgeries, etc. Also, one of the databases allowed to evaluate the seven systems in presence of long-term (several months) time variability, for the first time in the literature.

Moreover, two out of the seven systems considered in our experimental study had already been evaluated at the First International Signature Verification Competition in 2004 [36], SVC'2004, and one of the distance-based approaches compared is based on the same principles as those of the winning algorithm [3]; these facts gave us many elements of comparison in our analysis.

We noticed an important variation of systems' ranking from one database to another and from one protocol to another (concerning the presence of long-term time variability). This ranking also varied with respect to SVC test set [36]. This fact rises many important questions; indeed, when the data configuration in a database is totally at the opposite of that of another, regarding for example stability of genuine signatures and resemblance of skilled forgeries to genuine signatures in terms of dynamics, this will impact the systems and the best approaches in one case will no longer be the best in another case.

Nevertheless, across databases, some tendencies have emerged.

First, HMM-based systems outperformed Dynamic Time Warping systems and GMM-based system; in particular Reference System 1, based on the fusion of two information levels from the signature, has the same behavior across databases and protocols. This system is classified as the first in all cases, and no tuning was performed on a development set in any of such cases, at the opposite of the other HMM-based system. A sophisticated double-stage and personalized normalization scheme is at the origin of these nice properties [34].

Second, the approach that won SVC'2004, Dynamic Time Warping with the intraclass variance normalization, is not always ranked in the same way across databases and protocols. We have shown that the intraclass variability has a distortion effect on forgeries' scores when the normalization factor is high because of high variance in genuine scores (the distance between the forgery and the genuine signature is normalized by the intraclass variance distance). Therefore, this approach is very sensitive to the nature of the data, in terms of genuine signature variability. We have remarked that on MCYT data this approach is ranked second behind Ref1, while it is ranked sixth out of seven on BIOMET without time variability and

6 Online Handwritten Signature Verification

seventh in the presence of time variability (the Global distance-based approach and Ref2 based on coarse feature extraction and elastic distance perform better in this case, as we explain later).

Third, the distance-based approach using City Block distance coupled to a holistic feature extraction is well ranked on data with high intraclass variability (when the genuine signatures are highly variable from one instance to another) and also when long-term time variability is present in the data.

The Gaussian Mixture Model coupled with the same local feature extraction used by Reference System 1 gives lower results than HMM-based approaches in general and the distance-based approach using City Block distance with a holistic feature extraction. We also obtained an interesting result: the GMM coupled to local features resists poorly to long-term time variability (five months). One open question, given the good results obtained by the GMM-based systems at the BioSecure Evaluation Campaign (BMEC'2007) [2] in the presence of short-term time variability (2-3 weeks), is whether a GMM-based system coupled with a holistic feature extraction would perform better in the same long-term time variability conditions?

More generally, we observed the terrible impact of time variability on systems, causing degradation of at least 200% for the distance-based approach coupled with a holistic feature extraction, and can reach 360% for the best system, Reference System 1. Indeed, we recall that in this case, time variability corresponds to sessions acquired five months apart, thus to long-term time variability. More studies on this are necessary. A remaining challenge in research is certainly the study of the possibility of performing an update of the writer templates (references) in the case of distance-based approaches, or an adaptation of the writer model in the case of model-based approaches. Also, personalized feature selection should be explored by the scientific community since it may help to cope with intraclass variability, which is the main problem in signature and even with time variability. Indeed, one may better characterize a writer by those features that show more stability for him/her.

Finally, the scientific community may find in this work the access to a permanent evaluation framework, composed of publicly available databases, associated protocols and baseline reference systems in order to be able to compare their systems to the state of the art.

Acknowledgments

This work was funded by the IST-FP6 BioSecure Network of Excellence. J. Fierrez-Aguilar and F. Alonso are supported by a FPI Scholarship from Consejería de Educación de la Comunidad de Madrid and Fondo Social Europeo (Regional Government of Madrid and European Union).

References

1111

1. <http://www.biosecure.info/>. 1112
2. www.int-evry.fr/biometrics/bmec2007/. 1113
3. A. Kholmatov and B.A. Yanikoglu. Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408, 2005. 1114
4. W. D. Chang and J. Shin. Modified dynamic time warping for stroke-based on-line signature verification. In *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 724–728, Brazil, 2007. 1116
5. J. G. A. Dolfig. *Handwriting recognition and verification, a Hidden Markov approach*. PhD thesis, Philips Electronics N.V., 1998. 1119
6. J. G. A. Dolfig, E. H. L. Aarts, and J. J. G. M. Van Oosterhout. On-line signature verification with hidden markov models. In *Proc. of the International Conference on Pattern Recognition*, pages 1309–1312, Brisbane, Australia, 1998. 1121
7. J. Fierrez, D. Ramos-Castro, J. Ortega-Garcia, and J. Gonzales-Rodriguez. Hmm-based on-line signature verification: feature extraction and signature modelling. *Pattern Recognition Letters*, 28(16):2325–2334, December 2007. 1122
8. J. Fierrez and J. Ortega-Garcia. On-line signature verification. In A. K. Jain, A. Ross, and P. Flynn, editors, *Handbook of Biometrics*, pages 189–209. Springer, 2008. 1124
9. J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni. An on-line signature verification system based on fusion of local and global information. In *Proc. of 5th IAPR Intl. Conf. on Audio- and Video -based Biometric Person Authentication, AVBPA*, Springer LNCS, New York, USA, July 2005. 1125
10. J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE Transactions on Systems, Man and Cybernetics, part C*, 35(3):418–425, 2005. 1126
11. BioSecure Benchmarking Framework. <http://share.int-evry.fr/svnview-eph/>. 1127
12. S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux-Les Jardins, J. Lanter, Y. Ni, and D. Petrovska-Delacretaz. Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *Proc. of 4th International Conference on Audio and Video-Based Biometric Person Authentication*, pages 845–853, Guildford, UK, 2003. 1128
13. S. Garcia-Salicetti, J. Fierrez-Aguilar, F. Alonso-Fernandez, C. Vielhauer, R. Guest, L. Allano, T. Doan Trung, T. Scheidat, B. Ly Van, J. Dittmann, B. Dorizzi, J. Ortega-Garcia, J. Gonzalez-Rodriguez, M. Bacile di Castiglione, and M. Fairhurst. *Biosecure Reference Systems for On-Line Signature Verification: A Study of Complementarity*, pages 36–61. Annals of Telecommunications, Special Issue on Multimodal Biometrics, France, 2007. 1129
14. R. Guest, M. Fairhurst, and C. Vielhauer. Towards a flexible framework for open source software for handwritten signature analysis. In M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nuernberger, and W. Gaul, editors, *From Data and Information Analysis to Knowledge Engineering, Proceedings of the 29 Annual Conference of the German Classification Society GfKI 2005*, pages 620–629. Springer, Berlin, 2006. ISBN 1431-8814. 1130
15. S. Hangai, S. Yamanaka, and T. Hamamoto. Writer verification using altitude and direction of pen movement. In *International Conference on Pattern Recognition*, pages 3483–3486, Barcelona, September 2000. 1131
16. A. K. Jain, F. D. Griess, and S. D. Connell. On-line signature verification. *Pattern Recognition*, 35(12):2963–2972, December 2002. 1132
17. R. Kashi, J. Hu, W. L. Nelson, and W. Turin. A hidden markov model approach to online handwriting signature verification. *International Journal on Document Analysis and Recognition*, 1(2):102–109, Jul 1998. 1133
18. Y. Komiya and T. Matsumoto. On-line pen input signature verification ppi (pen-position/pen-pressure/pen-inclination). *Proc. IEEE International Conference on SMC*, pages 41–46, 1999. 1134

6 Online Handwritten Signature Verification

19. V. I. Levenshtein. *Binary codes capable of correcting deletions, insertions and reversals*, 1163
volume 10 of *Soviet Physics*. 1966. 1164
20. A. Martin, G. Doddington, T. Kamm, and M. Przybocki M. Ordowski. The det curve 1165
in assessment of detection task performance. In *Proc. Eurospeech '97*, volume 4, pages 1166
1895–1898, Rhodes, Greece, 1997. 1167
21. M. Martinez-Diaz, J. Fierrez, and J. Ortega-Garcia. Universal background models for dynamic 1168
signature verification. *Proceedings of the First IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, September 27-29th 2007. 1169
22. J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, 1171
A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. Mcyt baseline 1172
corpus: A bimodal biometric database. *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6):395–401, December 2003. 1173
23. R. Plamondon, W. Guerfali, and M. Lalonde. Automatic signature verification: a report on 1175
a large-scale public experiment. In *Proceedings of the International Graphonomics Society*, 1176
pages 9–13, Singapore, June 25 – July 2 1999. 1177
24. R. Plamondon and G. Lorette. Automatic signature verification and writer identification – the 1178
state of the art. *Pattern Recognition*, 22(2):107–131, 1989. 1179
25. R. Plamondon and S. Srihari. On-line and off-line handwriting recognition: A comprehensive 1180
survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84, 2000. 1181
26. L. Rabiner and B.H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall Signal 1182
Processing Series. 1993. 1183
27. D. A. Reynolds and R. C. Rose. Robust text-independent speaker identification using gaussian 1184
mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, 1185
Jan 1995. 1186
28. J. Richiardi and A. Drygajlo. Gaussian mixture models for on-line signature verification. In 1187
International Multimedia Conference, Proceedings 2003 ACM SIGMM workshop on Biometrics methods and applications, pages 115–122, Berkeley, USA, Nov 2003. 1188
29. G. Rigoll and A. Kosmala. A systematic comparison of on-line and off-line methods for 1190
signature verification with hidden markov models. In *Proc. of 14th International Conference on Pattern Recognition*, pages 1755–1757, Brisbane, Australia, 1998. 1191
30. S. Schimke, C. Vielhauer, and J. Dittmann. Using adapted levenshtein distance for on-line signature 1193
authentication. In *Proceedings of the ICPR 2004, IEEE 17th International Conference on Pattern Recognition*, 2004. ISBN 0-7695-2128-2. 1194
31. O. Uréche and R. Plamondon. Document transport, transfer and exchange: Security and commercial 1196
aspects. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition*, pages 585–588, 1999. 1197
32. O. Uréche and R. Plamondon. Systèmes numériques de paiement sur internet. *Pour la science*, 1199
260:45–49, 1999. 1200
33. O. Uréche and R. Plamondon. Digital payment systems for internet commerce: the state of 1201
the art. *World Wide Web*, 3(1):1–11, 2000. 1202
34. B. Ly Van, S. Garcia-Salicetti, and B. Dorizzi. On using the viterbi path along with hmm 1203
likelihood information for online signature verification. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Special Issue on Recent Advances in Biometric Systems*, 37(5):1237–1247, October 2007. 1204
35. M. Wirocius, J.-Y. Ramel, and N. Vincent. Selection of points for on-line signature comparison. 1205
In *Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR)*, Tokyo, Japan, September 2004. 1206
36. Dit-Yan Yeung, Hong Chang, Yimin Xiong, Susan George, Ramanujan Kashi, Takashi 1210
Matsumoto, and Gerhard Rigoll. Svc2004: First international signature verification 1211
competition. In *International Conference on Biometric Authentication (ICBA)*, volume 3072 of *Springer LNCS*, pages 16 – 22, Hong Kong, China, July 15-17 2004. 1212
37. T. G. Zimmerman, G. F. Russell, A. Heilper, B. A. Smith, J. Hu, D. Markman, J. E. Graham, 1214
and C. Drews. Retail applications of signature verification. In A. K. Jain and K. Ratha, editors, 1215
Biometric Technology for Human Identification, volume 5404, pages 206–214. Proceedings of 1216
SPIE, Bellingham, 2004. 1217