Automatic Online Signature Verification Using HMMs with User-Dependent Structure

J.M. Pascual-Gaspar and V. Cardeñoso-Payo

ECA-SIMM, Dpto. Informática, Universidad de Valladolid,
Campus Miguel Delibes s/n, 47011 Valladolid, Spain
{jmpascual,valen}@infor.uva.es

Abstract. A novel strategy for Automatic online Signature Verification based on hidden Markov models (HMM) with user-dependent structure is presented in this work. Under this approach, the number of states and Gaussians giving the optimal prediction results are independently selected for each user. With this simple strategy just three genuine signatures could be used for training, with an EER under 2.5% obtained for the basic set of raw signature parameters provided by the acquisition device. This results increment by a factor of six the accuracy obtained with the typical approach in which claim-independent structure is used for the HMMs.

1   Introduction

Signature verification is of particular importance within the framework of biometrics, because of its long standing tradition in many identity verification scenarios[1]. In on-line Automatic Signature Verification (ASV) systems, the signer uses special hardware to produce her signature, so that a time sequence of parameters (position, inclination, pressure, . . .) is generated by the device and can be processed by a system in order to characterize spatial and temporal features of the signature. These features are to be useful to build a model or template which could be later used to verify the claimed identity of the same signer with a minimum controlled risk that a forged signature could be taken as genuine. This particular instance of a pattern recognition problem is heavily influenced by intra and inter-user variability affecting the signing process, so that proper selection of a good model for every signature is a key step of any signature verification system.

Hidden Markov Models (HMM) is a widely used probabilistic framework when modelling patterns of temporal sequences. It has been successfully applied to speech recognition tasks [2], on-line handwriting [3] and on-line signature verification [4,5]. A HMM can be roughly described as a graph of interconnected emitting states which topology is defined by means of a transition matrix, made of probability values for every particular state-to-state transition, and where the probability of emission of a given output value is usually modelled by a superposition of Gaussian distributions.

The number of states in the model and the number of Gaussian distributions associated to each state constitute the *structural parameters* of the model. Fixing these structural parameters is a highly specialized problem-dependent task carried out by domain experts, since it is hard to find general efficient algorithms to infer these parameters from the data being modelled.

Given the structure of the model, the classical Baum-Welch algorithm (see [2]) can be successfully applied to estimate the values of the transition probabilities and of the weights and statistical moments of the Gaussians which provide the maximum expected likelihood for the observed time sequence. These values represent what we call the *statistical parameters* of the model.

Table 1 provides a view of the current state of the art in ASV using different modelling alternatives and sometimes combining local and global signature parameters. This gives a reference point of the expected accuracy of present and future systems.

Table 1. Some ASV systems, its error rates and employed techniques

Author	Date	Employed technique	% Error
Nelson et al. [6]	1994	Distance based	EER: 6
Yang et al. [5]	1995	HMM	FR: 1.75, FA: 4.44
Kashi et al. [7]	1997	HMM	EER: 2.5
Nalwa [8]	1997	Own algorithm	EER: between 2 and 5
DiLeece et al. [9]	2000	Multiexpert system	FR: 3.2, FA: 0.55
Jain et al. [10]	2002	String matching	FR: 3.3, FA: 2.7
Igarza et al. [11]	2003	HMM	EER: 9.25
Ortega et al. [12]	2003	HMM	EER: 0.98
Hansheng et al. [13]	2004	Linear regression (ER_2)	EER: 0.2
Fierrez-Aguilar et al. [14]	2005	Local and Global fusion	EER: 0.24

In this work we evaluate the influence of a claim-dependent selection of the structural parameters of the HMM on the accuracy of the automatic ASV. In section 2, a brief discussion of user dependent and user independent structure selection is made. The signature database used in this work is described in section 3 and the set of experiments is presented in section 4. Results and discussion section 5 will show that user dependent selection of the optimal model structure provides average EER below 2.5%, even for skilled forgeries. If we take into account that these experiments were carried out using just 5 raw signature parameters and 3 training signatures per user, we find the final accuracy values really promising and competitive with state of the art ASV systems.

2   User-Dependent Structure of HMMs

HMMs have been successfully applied to high-accuracy ASV systems for the last decade [15,11,5]. Although these systems differ in several important aspects

(e.g. signature pre-processing techniques, signature database characteristics, likelihood and score normalization techniques, ...), they always share a common architectural strategy, since they are based on a global set of structural parameters which, in most cases, are experimentally evaluated using all possible signatures of the database. The structure of the model which provides the best overall performance over all signers is selected. This is what we refer as *ASV approach based on HMM with user-independent structure* (HMM-UIS).

As an alternative to the HMM-UIS approach, the selection of the best structural parameters of the model could be carried out independently for each user, exploiting specific characteristics of the signature. This is what we call *ASV based on HMM with user-dependent structure* (HMM-UDS). Under this approach, the model would be adapted to the user or, at least, to a given class of users sharing a set of statistically similar features. Although some heuristics have been evaluated to guide this selection, in the present paper we will concentrate in an exhaustive search procedure of the optimal model structure. Although this approach is not directly usable in practical systems, we can use the results as an upper bound for the best accuracy which could be obtained when switching to a UDS.
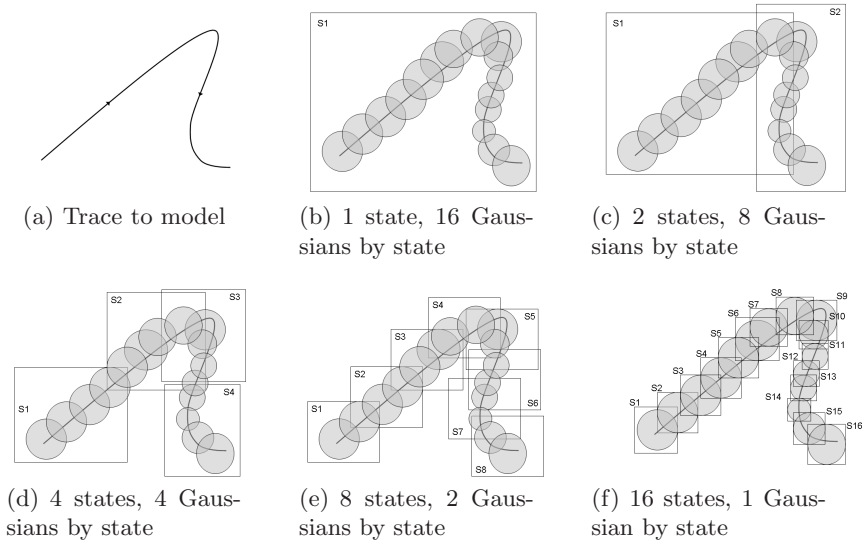
To further support our view of the importance of a user-dependent selection of the structural parameters of the model, we include Figure 1 to illustrate the correlation between form and structural parameters of the HMM in a given fictitious stroke of a signature. The five subfigures illustrate five different ways of covering a sample signature trace with states, each with the same fixed number of Gaussians. The same number of degrees of freedom ($N_S \times N_G = 16$ has been used in all cases, starting with the coarse-grained solution of figure 1-b), in which there is no dynamic modelling provided by state change, to the one in figure 1-f), where 16 states are used to model the stroke dynamics. For a given number of degrees of freedom, it should be expected that the model with lower number of Gaussians would perform better when using raw features without time derivatives, since it will better resemble time evolution of the stroke.

At the same time, the number of states will be highly influenced by the amount of statistically different samples available to initialize the model. As a consequence, the compromise between N_S and N_G is to be highly influenced both by the geometric and temporal characteristics of the signature and by the observable variability of features along it over the set of training samples. This is what motivated the experimental study we will present in the following sections.

3   Data Acquisition and Pre-processing

All the experiments have been carried out using the MCYT signature database [16], which contains on-line signatures acquired with a WACOM Intuos A6 USB digital tablet. A total of 333[1] different users contributed to the database, each of them producing 25 genuine signatures across five time-spaced sessions in which

---

[1] While the delivered and filtered version of MCYT database contains just 330 signatures, the raw complete original version was used in this study.

(a) Trace to model

(b) 1 state, 16 Gaussians by state

(c) 2 states, 8 Gaussians by state

(d) 4 states, 4 Gaussians by state

(e) 8 states, 2 Gaussians by state

(f) 16 states, 1 Gaussian by state

Fig. 1. Samples of how to model a signature's stroke

each user also produced 25 'over-the-shoulder' skilled forgeries of 5 different users. With this procedure, a total of $333 \times (25 + 25) = 16650$ signatures were produced, 8325 genuine and 8325 skilled forgeries. The input tablet provided a series of 100 vector samples per second, each vector including the raw parameters we used in this experiment: pen coordinates X and Y, pressure level P and pen orientation in terms of azimuth and elevation (see [16] for further details).

Table 2. MCYT global statistics summary

Gen (For)	Mean	σ (%)	Max.	Min.
Length (cm)	23.98 (24.25)	18% (38%)	47.51 (71.28)	6.12 (3.44)
Duration (s)	5.79 (7.15)	41% (74%)	20.20 (53.06)	0.57 (0.50)
Speed (cm/s)	5.71 (4.59)	28% (56%)	19.46 (23.96)	1.29 (0.51)

The basic statistics of relevant global signature parameters for this database are shown in Table 2, both for genuine instances and forgeries. The signature length results to be the most stable feature and the easiest to reproduce, while duration and speed show higher deviation both intra and, specially, inter-user. A geometric normalization is performed to remove the absolute position offsets, since a grid of cells was used during the acquisition process to capture several signatures per sheet.

4   Experimental Setup

A single experiment can be defined as a function $E(N_U, N_S, N_G, I_K) \rightarrow EER$ depending on the user identity N_U, the number of states N_S, the number of

Gaussians per state N_G, and the kind of forgery considered in the evaluation phase I_K ($K = S, R$, S:skilled, R:random). In each experiment the same three multi-session signatures were used for training (the first ones from each of the 3 central sessions).

Two collections of experiments have been carried out in order to compare HMM-UIS and HMM-UDS strategies. In both of them a Bakis Left-To-Right-noskip topology was chosen for the HMMs and EER was used as the objective function for the optimization in HMM-UDS strategy.

In the first collection of experiments, the HMM-UIS strategy was evaluated using 49 $M_{UIS} = (N_S, N_G)_{7x7}$ different structure configurations, both N_S and N_G ranging from 1 to 64 in steps of power of 2. The average EER obtained with each model was calculated and also the number of users for whom a model structure could be trained was annotated. This is a relevant parameter, since not for every user is always possible to initialize a model with any number of states and Gaussians. In these experiments, we evaluated EER using only random forgeries.

In the second collection of experiments, we evaluated the influence of the selection of optimal values for N_S and N_G independently for each user (HMM-UDS). Here, the models were trained using the same experimental conditions used in the first collection, to allow comparison between both approaches. A set of 555 possible structures were evaluated for each user $M_{UDS} = (N_S, N_G)_{111x5}$, the number of states ranging from 1 to 111 and the number of Gaussians from 1 to 5. From these models, the optimal model (M_{opt}) was selected as the one that brought the lowest EER or the lower number of states for equal EER values. The average EER over all the M_{opt} models was taken as the global average accuracy. Evaluation was carried out using both random and skilled forgeries in this case.

5   Results and Discussion

Table 3 shows error rates using the 49 models of the M_{UIS} test matrix. The number in parenthesis in each cell represents the number of users N_{VU} for whom a valid model structure was trainable. As expected, N_{VU} decreases as the number of degrees of freedom N_{FD} increases, since for some users there are not enough initialization data. Six of these configurations ($N_{FD} = 1024, 2048$ and 4096) were not trainable for any user and are shown as empty cells in the table.

Since any ASV system should balance accuracy and good generalization capabilities, the model structure with lowest EER and for which all subjects in the database can be trained is chosen: the configuration composed of 32 states and one Gaussian by state. This HMM-UIS configuration produced an EER of 16.29% using just three training samples. Of course, better EER exist in the table, but they come at the cost of smaller generalization capabilities, since the number of valid trainable users is really small.

Table 4 shows error rates obtained using the HMM-UDS approach, which are clearly lower than the ones in the HMM-UIS approach. The adaptation of the number of states individually for each user drastically improved the accuracy of

Table 3. Errors as %EER with HMM-UIS approach tested using random forgeries

$N_S \setminus N_G$	1	2	4	8	16	32	64
1	$36.43_{(333)}$	$35.19_{(333)}$	$33.08_{(333)}$	$31.88_{(333)}$	$30.19_{(333)}$	$28.64_{(333)}$	$28.29_{(324)}$
2	$35.55_{(333)}$	$33.93_{(333)}$	$31.77_{(332)}$	$30.11_{(324)}$	$28.21_{(306)}$	$27.61_{(287)}$	$29.92_{(260)}$
4	$34.80_{(333)}$	$32.40_{(327)}$	$29.84_{(319)}$	$27.88_{(310)}$	$26.76_{(277)}$	$29.53_{(210)}$	$37.47_{(90)}$
8	$31.11_{(333)}$	$29.71_{(330)}$	$26.80_{(312)}$	$25.83_{(287)}$	$27.28_{(241)}$	$35.91_{(104)}$	$42.59_{(7)}$
16	$24.20_{(333)}$	$23.74_{(321)}$	$22.38_{(306)}$	$22.90_{(254)}$	$30.47_{(99)}$	$55.88_{(9)}$	
32	$\mathbf{16.29}_{(333)}$	$16.27_{(309)}$	$15.85_{(259)}$	$16.36_{(107)}$	$21.04_{(8)}$		
64	$11.82_{(324)}$	$11.48_{(262)}$	$8.94_{(107)}$	$8.56_{(7)}$			

Table 4. Results with HMM-UDS models

| N_G	% EER ($I_K = R$)	% EER ($I_K = S$)
1	3.83	3.29
2	3.46	3.29
3	4.08	3.42
4	4.63	3.71
5	4.96	3.59
$\mathbf{N_{Gopt.}}$	**2.33**	**2.06**

the system, although the impact of the number of Gaussians by state was not so relevant as the influence of the optimization of the number of states.

Average EER using both random and skilled forgeries is shown in this table and, again, just 3 signatures were always used for training. The first five rows at the table represent the error when N_G was fixed and the optimum number of states was selected. As expected, the best results arise for models with low N_G values (1 and 2). The last row shows the average error obtained when both the number of states and the number of Gaussians by state are selected to optimize EER. With this 'two-dimension' optimization the error rate is reduced by 33% and 37% for random and skilled forgeries respectively.

Figure 2 shows the histograms of N_S and N_G for the experiments with random (a,b)and skilled forgeries (c,d). From figures fig. 2-a and fig.2-c it can be seen that the upper limit of 111 states per model can be increased in future works expecting better results from it because many models reached their best performance with the highest values of N_S. With respect the number of Gaussians distributions it seems that for random forgeries (fig. 2-b) a low number of Gaussians per state performs better, however in the case of skilled forgeries (fig. 2-d) a higher number of Gaussians discriminate better this kind of impostor.

To illustrate the relationship between signature complexity and number of states, signatures with different visual complexities are plotted in figure 3 besides their optimal number of states (one Gaussian per state was used in these signatures).

After all these experiments, we came to the conclusion that a reliable ASV-HMM system must be accurate for the majority of users, reporting a low mean error rate, and also that is very important for the real application acceptance

(a) N_S (random forgeries)

(b) N_G (random forgeries)

(c) N_S (skilled forgeries)

(d) N_G (skilled forgeries)
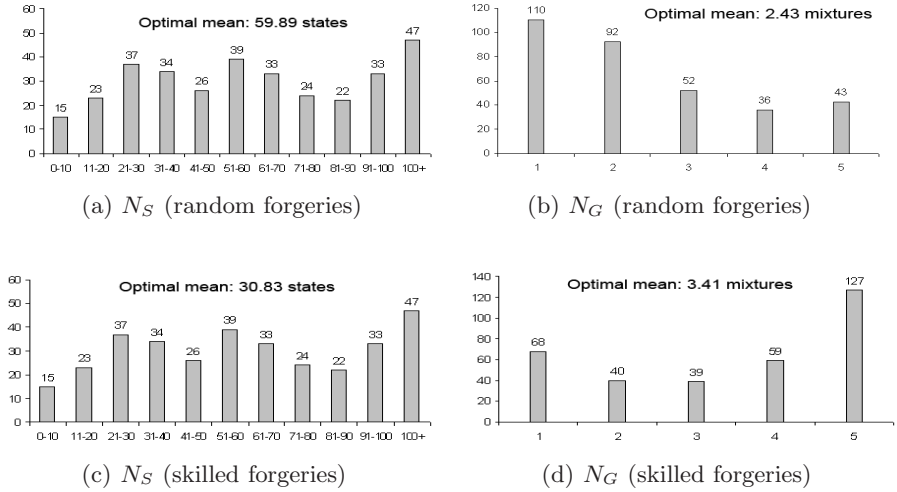
Fig. 2. N_S and N_G histograms for random forgeries (a,b) and skilled forgeries (c,d) tests

(a) 12 states   (b) 55 states   (c) 93 states
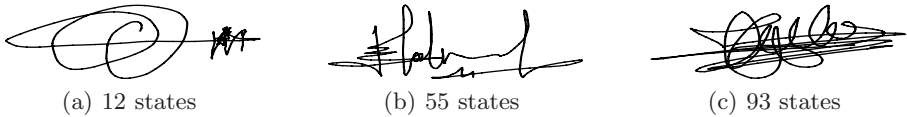
Fig. 3. Samples of signatures modelled using the HMM user-dependent structure approach

that the system works properly for very different types of signatures, not being desirable the existence of users getting high errors rates because their signatures are simplistic or inconsistent.

The histogram in figure 4-a) illustrates the distribution of the number of users sharing a same EER interval when random forgeries were used. We emphasize the following three results: a) a high number of users yield no verification errors (28% of the models give 0% EER); b) 86% of the models have an EER lower than 5%; c) only three models report EER over 15%, 15.81% being the worst EER result of our system.

Table 4 shows that no significant differences are found with respect to the random case when skilled forgeries were used. In fact, only a slight improvement can be depicted, which could be attributed to the fact that forged signatures were produced without information on the signature dynamics, which is difficult to infer for complex signatures. In spite of these similar average EER results, the appearance of the EER histogram in figure 4-b) is completely different to the one in random forgeries. Many of the genuine signatures of the users resulted to be difficult to forge when a different optimal number of states and Gaussians is chosen for each user, since the dynamics are hidden in these optimal number
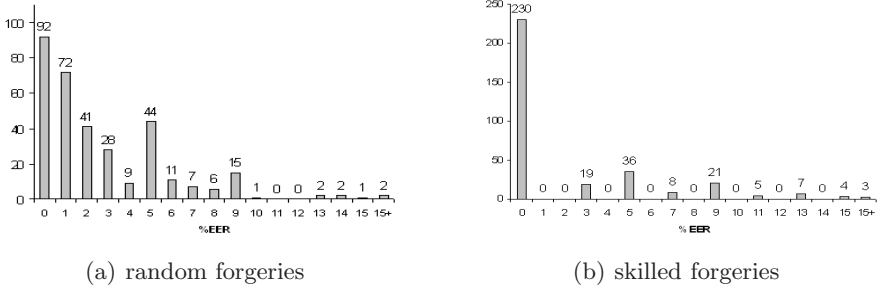
(a) random forgeries   (b) skilled forgeries

Fig. 4. EER histograms (in %)

(a) Duration vs. N_s (b) Duration vs. $N_s \times N_g$ (c) # pen-ups vs. $N_s \times N_g$
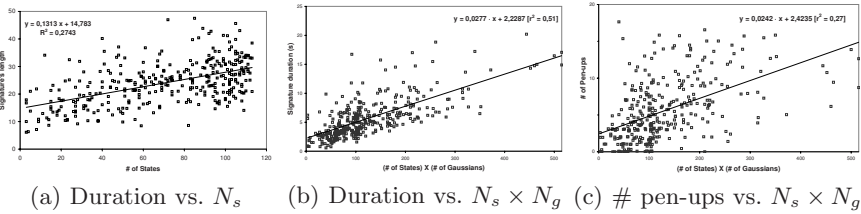
Fig. 5. Relevant global parameters in terms of optimal claim-dependent N_s and N_g

of degrees of freedom. A higher number of models with low EER show (230 producing 0% EER), although there also exists a group of users which produced simplistic or non-consistent signatures easier to forge, leading to a final average EER similar to the one found in random forgeries. This might have been clearly improved with a higher control over the acquisition process.

Recent studies on the combination of local and global features for ASV have provided reference rankings on the relevance of several global features [15]. The total length of a signature and the number of strokes (or the basically equivalent number of pen-ups) are shown to be the most relevant of these global parameters.

In order to test the correlation between the optimum number of degrees of freedom (N_S, N_G) of the user dependent models and the value of these global parameters for a given user, we have carried out a linear correlation analysis shown in figure 5.

In these figures, we prove that there is a reasonably good linear correspondence $(r^2 = 0.51)$ between the length of a signature and the number $N_S \times N_G$, which could provide a basic guideline on a more efficient model selection strategy. As for the number of pen-ups, there is not such a clear correspondence. This could be related to the fact that independent strokes could be better modelled after separate HMM models and then the results merged, compared to a single HMM based model as the one we are using here. Even less information can be extracted about the correspondence between signature length and optimum number of states. In any case, we conclude that further research is worthy on

the correspondence between global signature parameters and optimum structural parameters of the models.

6   Conclusions

In this work, we provided experimental evidences of the fact that data driven user dependent structure optimization of the HMM models could bring lower EER in ASV systems. The influence of two relevant structural parameters, the number of states N_S in the model and the number of Gaussians by state N_G, was evaluated and N_S was the parameter which provided better observable improvement. HMM-UDS strategies lead to more accurate and reliable ASV systems using a smaller number of training signatures, which always represents an advantage for practical use cases. User adaptation shows to cope well with intra-user variability while providing good inter-user discrimination.

An EER of 2.33% for random forgeries and 2.06% for skilled forgeries has been obtained. This represents a factor of 6 gain over the HMM-UIS strategy for the random forgery scenario within the same experimental conditions. Since optimization was carried out using an exhaustive search, it might not be useful in practical systems. Nevertheless, the results provide an lower bound for the best obtainable EER which encourages for further experimentation on data driven model selection strategies. Also, a better parameterization including time-dependent features will of course provide an overall increase of accuracy, according to the results found in other works [15].

Acknowledgements

References

1. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. Transactions on pattern analysis and machine intelligence 22(1), 63–84 (2000)
2. Rabiner, L.R.: A tutorial on hidden markov models and selected application in speech recognition. Proceedings of IEEE 77(2), 257–286 (1989)
3. Hu, J., Brown, M.K., Turin, W.: Hmm based on-line handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. 18(10), 1039–1045 (1996)
4. Fierrez-Aguilar, J.: Adapted Fusion Schemes for Multimodal Biometric Authentication. PhD thesis, Esc. Técnica Superior de Ing. de Telecomunicación (2006)
5. Yang, L., Widjaja, B.K., Prasad, R.: Application of hidden markov models for signature verification. Pattern Recognition 28(2), 161–170 (1995)
6. Nelson, W., Turin, W., Hastie, T.: Statistical methods for on-line signature verification. IJPRAI 8(3), 749–770 (1994)

7. Kashi, R.S., Hu, J., Nelson, W.L., Turin, W.: On-line handwritten signature verification using hidden markov model features. Document Analysis and Recognition 2 (1997)
8. Nalwa, V.S.: Automatic on-line signature verification. Proceedings of the IEEE 85(2), 215–239 (1997)
9. Lecce, V.D., Dimauro, G., Guerriero, A., Impedovo, S., Pirlo, G., Salzo, A.: A multi-expert system for dynamic signature verification. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 320–329. Springer, Heidelberg (2000)
10. Jain, A., Griess, F., Connell, S.: On-line signature verification. Pattern Recognition 35(12), 2963–2972 (2002)
11. Igarza, J.J., Goirizelaia, I., Espinosa, K., Hernáez, I., Méndez, R., Sánchez, J.: Online handwritten signature verification using hidden markov models. In: Sanfeliu, A., Ruiz-Shulcloper, J. (eds.) CIARP 2003. LNCS, vol. 2905, pp. 391–399. Springer, Heidelberg (2003)
12. Ortega-Garcia, J., Fierrez-Aguilar, J., Martin-Reillo, J., Gonzalez-Rodriguez, J.: Complete signal modeling and score normalization for function-based dynamic signature verification. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 658–667. Springer, Heidelberg (2003)
13. Lei, H., Palla, S., Govindaraju, V.: Er2: An intuitive similarity measure for online signature verification. In: IWFHR '04: Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04), Washington, DC, USA, pp. 191–195. IEEE Computer Society Press, Los Alamitos (2004)
14. Fierrez-Aguilar, J., Nanni, L., Lopez-Peñalba, J., Ortega-Garcia, J., Maltoni, D.: An on-line signature verification system based on fusion of local and global information. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 523–532. Springer, Heidelberg (2005)
15. Fierrez-Aguilar, J., Krawczyk, S., Ortega-Garcia, J., Jain, A.K.: Fusion of local and regional approaches for on-line signature verification. In: Li, S.Z., Sun, Z., Tan, T., Pankanti, S., Chollet, G., Zhang, D. (eds.) IWBRS 2005. LNCS, vol. 3781, pp. 188–196. Springer, Heidelberg (2005)
16. Ortega, J., Fierrez, J., Simon, D., Gonzalez, J., Hernaez, I., Igarza, J.J., Vivaracho, C., Escudero, D., Moro, Q.: Mcyt baseline corpus: a bimodal biometric database. IEEE Proc. Visual Image Signal Processing 150(6), 395–401 (2003)