IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 30, 2017, accepted January 7, 2018, date of publication January 15, 2018, date of current version February 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2793966

Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics

RUBEN TOLOSANA, RUBEN VERA-RODRIGUEZ, JULIAN FIERREZ, (Member, IEEE), AND JAVIER ORTEGA-GARCIA, (Fellow, IEEE)

Biometrics and Data Pattern Analytics (BiDA) Lab-ATVS, Universidad Autonoma de Madrid, 28049 Madrid, Spain

Corresponding author: Ruben Tolosana (ruben.tolosana@uam.es)

This work was supported in part by the Project TEC2015-70627-R MINECO/FEDER and in part by the UAM-CecaBank Project. The work of R. Tolosana was supported by the FPU Fellowship from Spanish MECD.

ABSTRACT Systems based on deep neural networks have made a breakthrough in many different pattern recognition tasks. However, the use of these systems with traditional architectures seems not to work properly when the amount of training data is scarce. This is the case of the on-line signature verification task. In this paper, we propose a novel writer-independent on-line signature verification systems based on Recurrent Neural Networks (RNNs) with a Siamese architecture whose goal is to learn a dissimilarity metric from the pairs of signatures. To the best of our knowledge, this is the first time these recurrent Siamese networks are applied to the field of on-line signature verification, which provides our main motivation. We propose both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) systems with a Siamese architecture. In addition, a bidirectional scheme (which is able to access both past and future context) is considered for both LSTM- and GRU-based systems. An exhaustive analysis of the system performance and also the time consumed during the training process for each recurrent Siamese network is carried out in order to compare the advantages and disadvantages for practical applications. For the experimental work, we use the BiosecurID database comprised of 400 users who contributed a total of 11,200 signatures in four separated acquisition sessions. Results achieved using our proposed recurrent Siamese networks have outperformed the state-of-the-art on-line signature verification systems using the same database.

INDEX TERMS Biometrics, deep learning, on-line handwritten signature verification, recurrent neural networks, LSTM, GRU, DTW, BiosecurID.

I. INTRODUCTION

Deep Learning (DL) has become a thriving topic in the last years [1], allowing computers to learn from experience and understand the world in terms of hierarchy of simpler units. DL has enabled significant advances in complex domains such as natural language processing [2] and computer vision [3], among many others. The main reasons to understand the high deployment of DL lie on the increasing amount of available data and also the deeper size of the models thanks to the increased computer resources. However, there are still some tasks in which DL has not achieved state-of-the-art results due to the scarcity of available data and therefore, the inability to train and use those traditional deep learning architectures.

New trends based on the use of Recurrent Neural Networks (RNNs), which is a specific DL architecture, are becoming more and more important nowadays for modelling sequential data with arbitrary length [4]. The range of applications of RNNs can be very varied, from speech recognition [5] to biomedical problems [6]. RNNs are defined as a connectionist model containing a self-connected hidden layer. One benefit of the recurrent connection is that a memory of previous inputs remains in the network internal state, allowing it to make use of past context. Additionally, bidirectional schemes (i.e. BRNNs) have been studied in order to provide access not only to the past context but also to the future [7]. One of the fields in which RNNs has caused more impact in the last years is in handwriting recognition due to the relationship that exists between current inputs and past and future contexts. However, the range of contextual information that standard RNNs can access is very limited due to the well-known vanishing gradient problem [8], [9]. Long Short-Term Memory (LSTM) [10] and Gated Recurrent Unit (GRU) [11]–[13] are RNN architectures that arised with the aim of resolving the

shortcomings of standard RNNs. These architectures have been deployed with success in both on-line and off-line handwriting [8], [14]. Whereas off-line scenarios consider information only related to the image of the handwriting, in on-line scenarios additional information such as X and Y pen coordinates and pressure time functions are also considered providing therefore much better results. Graves *et al.* [8] proposed a system based on the use of Bidirectional LSTM (BLSTM) for recognizing unconstrained handwritten text for both off- and on-line handwriting approaches. The results obtained applying this new approach outperformed a state-of-the-art HMM-based system and also proved the new approach to be more robust to changes in dictionary size. These new approaches have been considered not only for recognizing unconstrained handwriting but also for writer identification. Zhang *et al.* [15] considered a system based on BLSTM for on-line text-independent writer identification. The experiments carried out over both English (133 writers) and Chinese (186 writers) outperformed state-of-the-art systems as well.

Despite the good results obtained in the field of handwriting recognition, very few studies have successfully applied these new RNN architectures to handwritten signature verification. Tiflin and Omlin [16] proposed the use of a system based on LSTM for on-line signature verification. Different configurations based on the use of forget gates and peephole connections were studied considering in the experimental work a small database with only 51 users. The LSTM RNNs proposed in that work seemed to authenticate genuine and impostor cases very well. However, as it was pointed out in [17], the method proposed for training the LSTM RNNs was not feasible for real applications for various reasons. First, the authors considered the same users for both development and evaluation of the system. Moreover, the system should be trained every time a new user was enrolled in the application. In addition, forgeries were required in that approach for training, which may not be feasible to obtain as well. Besides, the results obtained in [16] cannot be compared to any state-of-the-art signature verification system as the traditional measures such as the Equal Error Rate (EER) or calibrated likelihood ratios were not considered. Instead, they just reported the errors of the LSTM-outputs. In order to find some light on the feasibility of LSTM RNNs for signature verification purposes, Otte *et al.* [17] performed in an analysis considering three different real scenarios: 1) training a general network to distinguish forgeries from genuine signatures on a large training set, 2) training a different network for each writer that works perfectly on the training set, and 3) training the network on genuine signatures only. However, all experiments failed obtaining a 23.75% EER for the best configuration, far away from the best state-of-the-art results and concluding that LSTM RNN systems trained with standard mechanisms were not appropriate for the task of signature verification as the amount of available data for this task is scarce compared to other tasks such as handwriting recognition.

The main contributions of the present work are as follows: 1) we propose the use of different RNNs with a Siamese architecture for the task of on-line handwritten signature verification. This Siamese architecture [18] allows getting a close approximation to the verification task learning a dissimilarity metric from pairs of signatures (pairs of signatures from the same user and pairs of genuine-forgery signatures). To the best of our knowledge, to date, recurrent Siamese networks have never been used to model an on-line signature verifier, which provides our main motivation. 2) We propose on-line signature verification systems for a writer-independent scenario. This scenario is preferable over the writer-dependent scenario, as for a real consumer based system, e.g. in banking, the system would not need to be updated (retrained) with every new client who opens an account, avoiding therefore a waste of resources. 3) We propose a strict experimental protocol, in which different users and number of available signatures are considered for the development and evaluation of the systems in order to analyse the true potential of recurrent Siamese networks for signature verification. 4) This work constitutes, to the best of our knowledge, the first analysis of RNNs (i.e. LSTM and GRU) for the two types of forgeries considered in on-line signature verification (i.e. skilled and random or zero-effort forgeries). 5) We perform an exhaustive analysis of the system performance and also the time consumed during the training process for each recurrent Siamese approach in order to compare the advantages and disadvantages of each of them for practical applications. 6) We finally analyse the advantages of considering recurrent Siamese networks with a bidirectional scheme, which is able to access both past and future contexts.

The remainder of the paper is organized as follows. In Sec. II, our proposed approach based on the use of LSTM and GRU RNNs for signature verification with a Siamese architecture is described together with the bidirectional scheme in order to access to future context as well. Sec. III describes the BiosecurID on-line signature database considered in the experimental work. Sec. IV describes the information used for feeding the LSTM and GRU RNNs. Sec. V describes the experimental protocol and the results achieved with our proposed approach. Finally, Sec. VI draws the final conclusions and points out some lines for future work.

II. PROPOSED METHODS

The methods proposed in this work for improving the performance of on-line signature verification are based on the combination of LSTM and GRU RNNs with a Siamese architecture. A bidirectional scheme is also studied.

A. SIAMESE ARCHITECTURE

The Siamese architecture has been used for recognition or verification applications where the number of categories is very large and not known during training, and where the number of training samples for a single category is small. In our case the main goal of this architecture is to learn a similarity metric from data minimizing a discriminative cost function that drives the dissimilarity metric to be

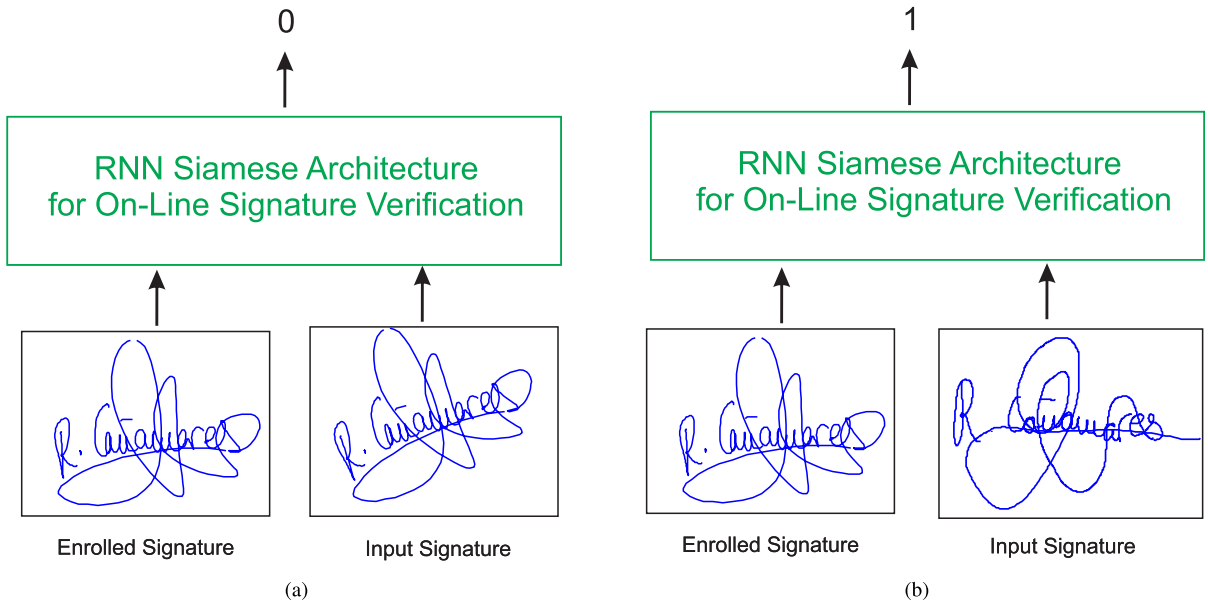

FIGURE 1. Examples of our proposed LSTM and GRU RNN systems based on a Siamese architecture for minimizing a discriminative cost function. (a) Genuine case. (b) Impostor case.

small for pairs of genuine signatures from the same subject, and longer for pairs of signatures coming from different subjects. Fig. 1 shows that idea visually. In previous studies such as [18], the authors proposed the use of Convolutional Neural Networks (CNNs) with a Siamese architecture for face verification. Experiments were performed with several databases obtaining very good results where the number of training samples for a single category was very small. Siamese architectures have also been used in early works for on-line signature verification [19] although not considering RNNs. Bromley *et al.* [19] proposed an on-line signature verification system comprised of two separated sub-networks based on Time Delay Neural Networks (TDNNs) which are one-dimensional convolutional networks applied to time series. Different architectures regarding the number and size of layers were studied. A total of 8 time functions fixed to the same length of 200 points were extracted for X and Y pen coordinates using an old-fashion NCR 5990 Signature Capture Device. The best performance was obtained using two convolutional layers with 12 by 64 units in the first layer and 16 by 19 units in the second one. The threshold was set to detect 80.0% of forgeries and 95.5% of genuine signatures, far away from the results that can be achieved nowadays with state-of-the-art systems [20]–[23].

B. LONG SHORT-TERM MEMORY RNNs

LSTM RNNs [10] have been successfully applied to many different tasks such as language identification considering short utterances [24] or biomedical problems [6] for example. However, the analysis and design of LSTM RNN architectures for new tasks are not straightforward [25].

LSTM RNNs [10] are comprised of memory blocks usually containing one memory cell each of them, a forget gate f, an input gate i, and an output gate o. For a time step t:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$
$$\widetilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (4)$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \quad (5)$$
$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

where W_* and U_* are weight matrices and b_* is the bias vector. The symbol $\odot$ represents a pointwise product whereas σ is a sigmoid layer which outputs values between 0 and 1. The LSTM does have the ability to remove old information from $t-1$ time or add new one from t time. The key is the cell state C_t which is carefully regulated by the gates. The f gate decides the amount of previous information (i.e. h_{t-1}) that passes to the new state of the cell C_t. The i gate indicates the amount of new information (i.e. $\widetilde{C}_t$) to update in the cell state C_t. Finally, the output of the memory block h_t is a filtered version of the cell state C_t, being the o gate in charge of it. Fig. 2 shows a single LSTM memory block at different time steps (i.e. X_{t-1}, X_t and X_{t+1}) for clarification.

C. GATED RECURRENT UNIT RNNs

GRU [11], [12] is a relatively new type of RNNs which has been inspired by the LSTM unit but is much simpler to compute and implement. In addition, the results obtained using this novel RNN system seems to be very similar to the LSTM RNN system [26]. The main difference between GRU and LSTM RNNs resides in the number of gates used to control the flow of information. Whereas the LSTM unit contains three different gates (i.e. forget f, input i and output o gate), the GRU unit only owns two gates (i.e. reset gate r

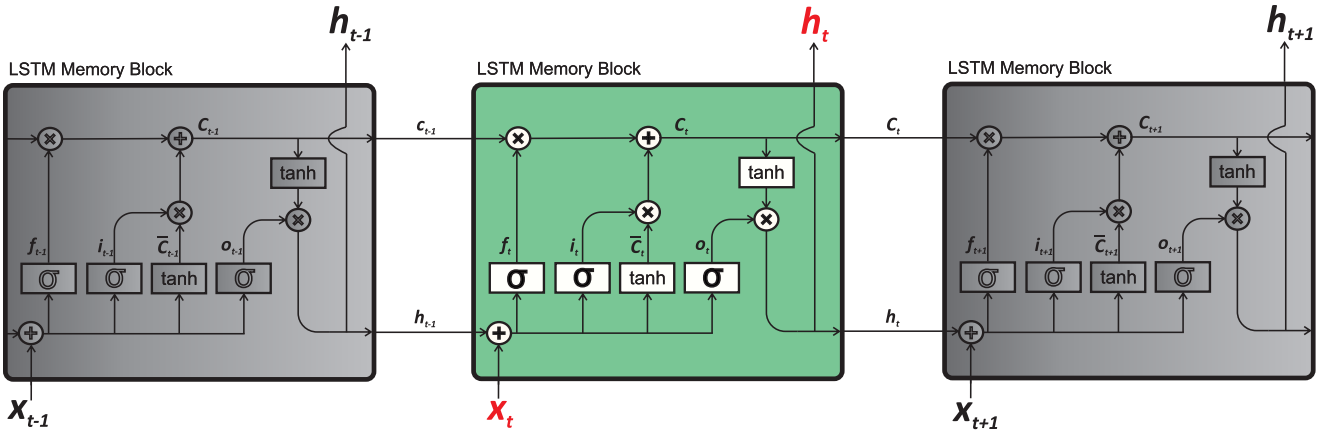

FIGURE 2. Scheme of a single LSTM memory block at different time steps (i.e. X_{t-1}, X_t and X_{t+1}).

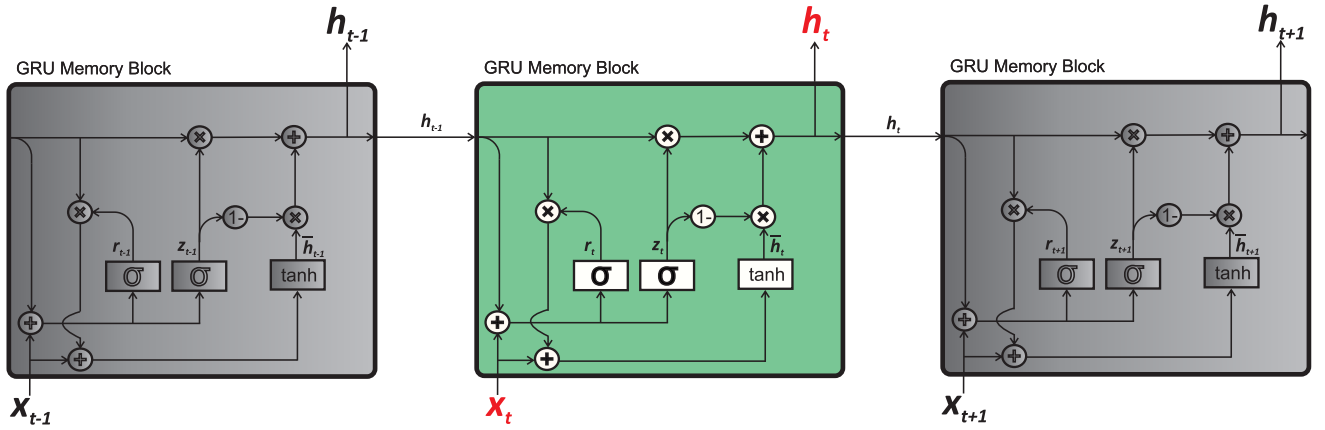

FIGURE 3. Scheme of a single GRU memory block at different time steps (i.e. X_{t-1}, X_t and X_{t+1}).

and update gate z). For a time step t:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{7}$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{8}$$

$$\widetilde{h_t} = \tanh(W_h x_t + U_h(h_{t-1} \odot r_t) + b_h) \tag{9}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h_t} \tag{10}$$

where W_* and U_* are the weight matrices and b_* is the bias vector. The symbol $\odot$ represents a pointwise product whereas σ is a sigmoid layer which outputs values between 0 and 1. The GRU does have the ability to remove old information from $t-1$ time or add new one from t time. The reset gate r_t is in charge of keeping in the current cell state (i.e. $\widetilde{h_t}$) the information of the previous time step (i.e. h_{t-1}) or reset it with the information of only the current input (i.e. x_t). Finally, the update gate z_t filters how much information from the previous time step and current cell state will flow to the current output of the memory block (i.e. h_t). Fig. 3 shows a single GRU memory block at different time steps (i.e. X_{t-1}, X_t and X_{t+1}) for clarification.

D. BIDIRECTIONAL RNNs

The RNN schemes explained before in Sec. II-B and II-C are the original ones. These schemes have access only to the past and present contexts. However, for some applications such as handwriting or speech recognition the chance of having access to the future context can further improve the system performance [5], [8]. Schemes which also allow access to the future context are known as Bidirectional RNNs (BRNNs) [7]. BRNNs combine a RNN that moves forward through time beginning from the start of the sequence with another RNN that moves backward through time beginning from the end of the sequence [1]. Fig. 4 shows a typical scheme of a BRNN system at different time steps (i.e. X_{t-1}, X_t and X_{t+1}) for clarification. The bottom part of the scheme propagates the information forward in time (towards the right) while the top part of the scheme propagates the information backward in time (towards the left). Thus at each point t, the output units O_t can benefit from a relevant summary of the past in its h_t^f input and from a relevant summary of the future in its h_t^b input [1].

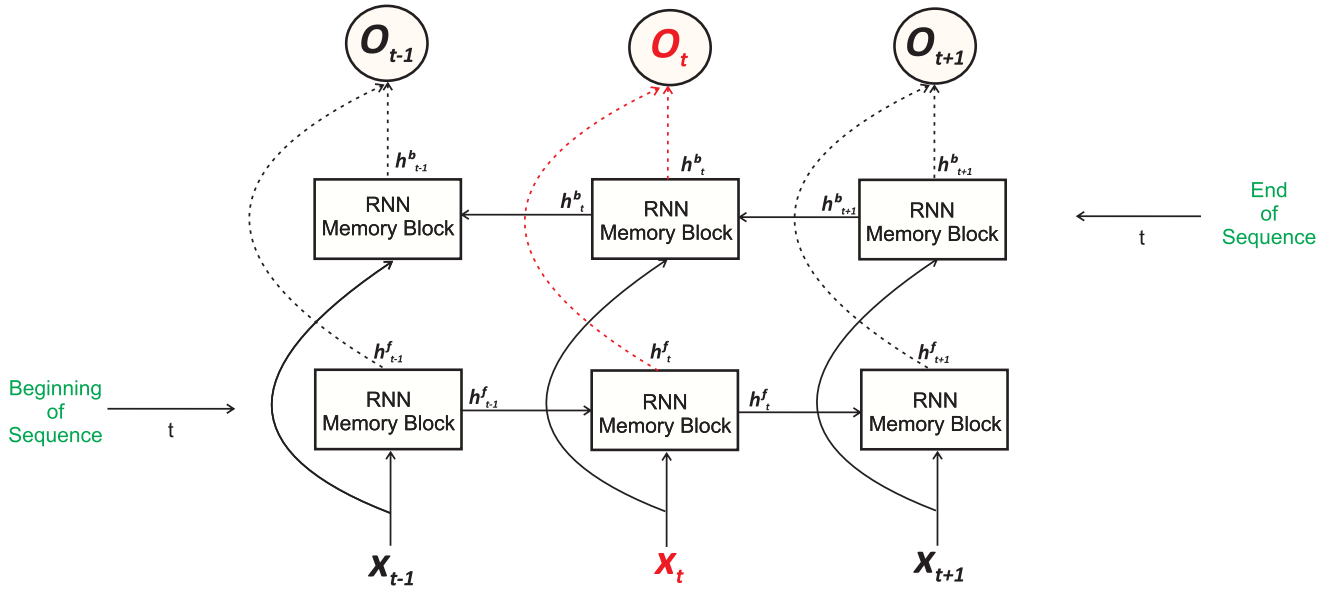

FIGURE 4. Scheme of a typical Bidirectional RNN system at different time steps (i.e. X_{t-1}, X_t and X_{t+1}). The bottom part of the scheme propagates the information forward in time (towards the right) while the top part of the scheme propagates the information backward in time (towards the left). Thus at each point t, the output units O_t can benefit from a relevant summary of the past in its h^f_t input and from a relevant summary of the future in its h^b_t input. Adapted from [1].

III. ON-LINE SIGNATURE DATABASE

The BiosecurID database [27] is considered in the experimental work of this paper. This database is comprised of 16 original signatures and 12 skilled forgeries per user, captured in 4 separate acquisition sessions leaving a two-month interval between them. There are a total of 400 users and signatures were acquired considering a controlled and supervised office-like scenario. Users were asked to sign on a piece of paper, inside a grid that marked the valid signing space, using an inking pen. The paper was placed on a Wacom Intuos 3 pen tablet that captured the following time signals of each signature: *X* and *Y* pen coordinates (resolution of 0.25 mm), pressure (1024 levels) and timestamp (100 Hz). In addition, pen-ups trajectories are available. All the dynamic information is stored in separate text files following the format used in the first Signature Verification Competition (SVC) [28], [29], where one of our previous signature verification systems was ranked first against skilled forgeries. The acquisition process was supervised by a human operator whose task was to ensure that the collection protocol was strictly followed and that the captured samples were of sufficient quality (e.g. no part of the signature outside the designated space), otherwise the subjects were asked to repeat the signature.

IV. TIME FUNCTIONS REPRESENTATION

The on-line signature verification system proposed in this work is based on time functions (a.k.a. local system) [30], [31]. For each signature acquired, signals related to *X* and *Y* pen coordinates and pressure are used to extract a set of 23 time functions, similar to [32]. All time functions are included in Table 1.

TABLE 1. Set of time functions considered in this work.

#	Feature
1	x-coordinate: x_n
2	y-coordinate: y_n
3	Pen-pressure: z_n
4	Path-tangent angle: θ_n
5	Path velocity magnitude: v_n
6	Log curvature radius: ρ_n
7	Total acceleration magnitude: a_n
8-14	First-order derivate of features 1-7: $\dot{x_n}, \dot{y_n}, \dot{z_n}, \dot{\theta_n}, \dot{v_n}, \dot{\rho_n}, \dot{a_n}$
15-16	Second-order derivate of features 1-2: $\ddot{x_n}, \ddot{y_n}$
17	Ratio of the minimum over the maximum speed over a 5-samples window: v^r_n
18-19	Angle of consecutive samples and first order difference: α_n, $\dot{\alpha_n}$
20	Sine: s_n
21	Cosine: c_n
22	Stroke length to width ratio over a 5-samples window: r^5_n
23	Stroke length to width ratio over a 7-samples window: r^7_n

V. EXPERIMENTAL WORK

A. EXPERIMENTAL PROTOCOL

The experimental protocol considered in this work has been designed in order to analyse and prove the feasibility of both LSTM and GRU RNNs for on-line signature verification in practical scenarios. Therefore, different users and signatures are considered for the two main stages, i.e., development of the RNNs system (Sec. V-B1) and the final evaluation (Sec. V-B2). Additionally, the two most common types of forgeries are considered here: skilled, the case when a forger tries to imitate the signature of another user of the system, and random or zero-effort, the case when a forger uses his own signature claiming to be another user of the system.

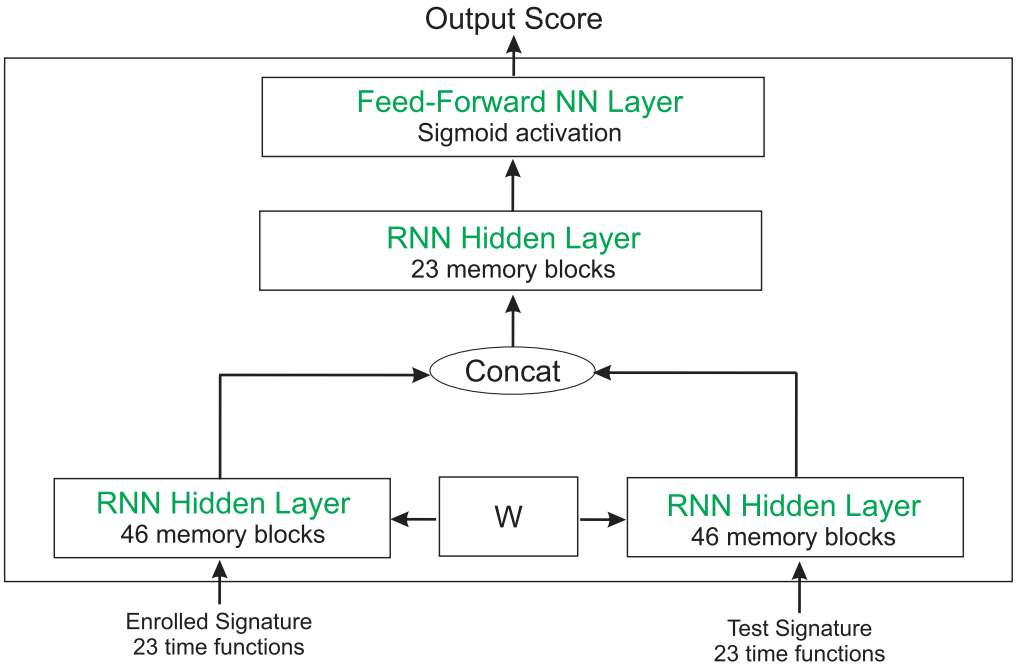

FIGURE 5. End-to-end on-line signature verification system proposed in this work and based on the use of LSTM and GRU RNNs with a Siamese architecture.

The first 300 users of the BiosecurID database are used for the development of the system, while the remaining 100 users are considered for the evaluation. For both stages, the 4 genuine signatures of the first session are used as training signatures, whereas the 12 genuine signatures of the remaining sessions are left for testing. Therefore, inter-session variability is considered in our experiments. Skilled forgery scores are obtained by comparing training signatures against the 12 available skilled forgery signatures for each user whereas random forgery scores are obtained by comparing the training signatures with one genuine signature of 12 other random users.

Finally, three different scenarios are analysed regarding the type of forgery considered for training the RNN systems: 1) **"skilled"**, the case which considers only pairs of genuine and skilled forgery signatures, 2) **"random"**, the case which considers only pairs of genuine and random forgery signatures, and 3) **"skilled + random"**, the case which considers pairs of both genuine/skilled and also genuine/random signatures in order to train just one system for both types of forgeries.

B. RESULTS

1) DEVELOPMENT RESULTS

This section describes the development and training of our proposed LSTM and GRU RNN systems with a Siamese architecture considering the 300 users of the development dataset. Three kinds of pairs of signatures can be used as inputs of the RNN systems: 1) two genuine signatures performed by the same user, 2) one genuine signature from the claimed user and one skilled forgery signature performed by an impostor, and 3) one genuine signature from the claimed user and one random forgery signature. For each of these three cases there are a total of $4 \times 12 \times 300 = 14,400$ comparisons, having the same number of genuine and impostor signatures for testing. Our RNN systems are implemented under Theano [33] with a NVIDIA GeForce GTX 1080 GPU.

In order to find the most suitable RNN system architecture we explored different configurations regarding the number of time functions used as inputs and the complexity level of the RNN system (i.e. number of hidden layers and memory blocks per hidden layer). In all cases, we considered our proposed Siamese architecture in order to learn a dissimilarity from pair of signatures. Our first attempt was based on the use of some of the 11 most commonly used time functions from a total of 23 (i.e., x_n, y_n, z_n, θ_n, v_n, ρ_n, a_n, $\dot{x}_n$, $\dot{y}_n$, $\ddot{x}_n$, $\ddot{y}_n$ from Sec. IV) and a RNN system based on two RNN hidden layers (with 22 and 11 memory blocks, respectively) and finally, a feed-forward neural network layer. Both input-to-hidden and hidden-to-hidden layers are fully-connected. The system performance obtained over the evaluation dataset was 8.25% EER. Then, we decided to increase the complexity of the RNN system in order to achieve better results over the evaluation dataset. First, we added a new RNN layer comprised of 6 memory blocks on top of the second RNN layer providing a 20.00% EER over the evaluation dataset, so this configuration was discarded. Another approach was based on the use of the original configuration based on two RNN hidden layers but increasing the number of memory blocks (44 and 22 per RNN hidden layer, respectively) achieving

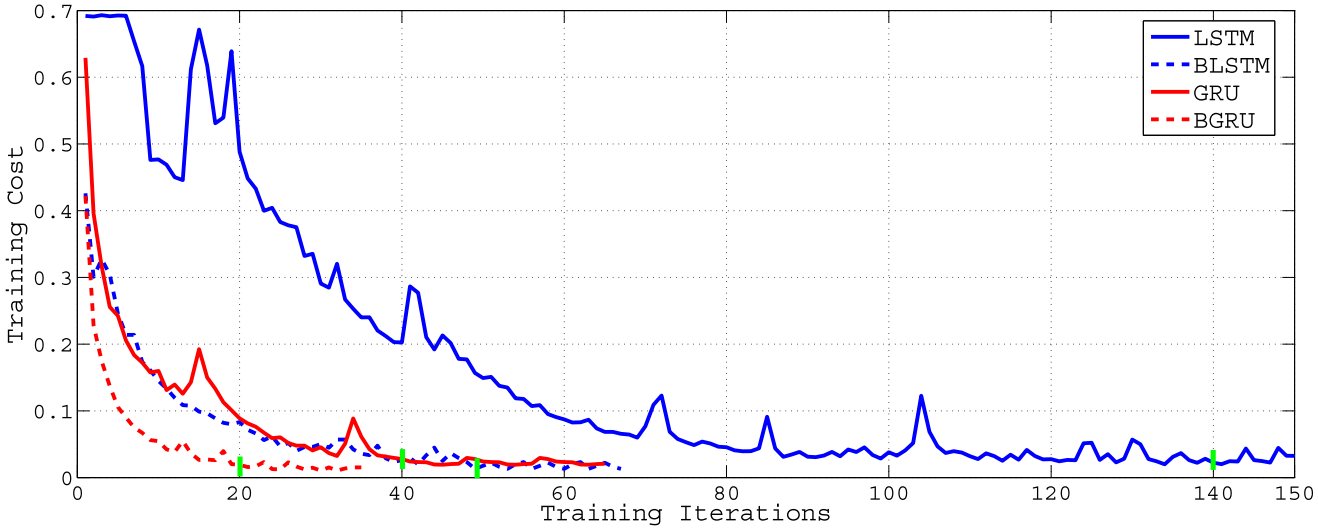

FIGURE 6. Considered RNNs cost during training for the "skilled" scenario. A small green vertical line indicates for each proposed RNN system the training iteration which provides the best system performance over the evaluation dataset.

a 10.00% EER, being this result worse compared to the 8.25% EER of the original configuration. We concluded that increasing the complexity of the RNN system always ended up with a worse generalization over the evaluation dataset (i.e. overfitting). Then we decided to feed the RNN system with as much information as possible, i.e., all 23 time functions.

After repeating the same previous exploration, the best topology obtained for both LSTM and GRU proposed RNNs is based on the use of two RNN hidden layers and finally, a feed-forward neural network layer. Fig. 5 shows our proposed end-to-end on-line signature verification system. The first layer is composed of two LSTM/GRU hidden layers with 46 memory blocks each and sharing the weights between them. The outputs provided for each LSTM/GRU hidden layer of the first layer are then concatenated and serve as input to the second layer which corresponds to a LSTM/GRU hidden layer with 23 memory blocks. Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score for each pair of signatures.

Fig. 6 shows the training cost of the considered RNNs with the number of training iterations for the "skilled" scenario. Four different RNN-based systems are considered, i.e., LSTM, GRU and their bidirectional schemes (i.e. BLSTM and BGRU). A small green vertical line is included in the figure for each proposed RNN system indicating the training iteration which provides the best system performance over the evaluation dataset, with a training cost value very close to zero. Similar results were obtained for both "random" and "skilled + random" scenarios as well. It is important to remark two different aspects of the figure. First, the difference in the number of training iterations needed between normal (i.e. LSTM and GRU) and bidirectional schemes (i.e. BLSTM and BGRU). For example, the best LSTM configuration is obtained after 140 training iterations whereas only around 50 iterations are needed for the BLSTM RNN system. This shows the importance of considering both past and future contexts in order to train RNNs faster and also with a lower value of training cost. Additionally, it is important to highlight the difference in the number of training iterations between both LSTM and GRU RNN systems. As the GRU memory block is a simplified version of the LSTM memory block (see Sec. II-C) the number of parameters to train are lower and therefore, we are able to get similar and even better values of training cost with fewer number of training iterations compared to the LSTM RNN system.

2) EVALUATION RESULTS

This section analyses the performance of the proposed RNN systems trained in the previous section for the three different training scenarios considered (i.e. "skilled", "random" and "skilled + random"). The remaining 100 users (not used for development) are used here. Regarding the system performance, two different cases are considered. First, the evaluation of the system performance considering scores directly from all pairs of signatures (i.e. 1vs1) and second, the case of performing the average score of the four one-to-one comparisons (i.e. 4vs1) as there are four genuine training signatures per user. In order to make comparable our approach to related works, we have considered a highly competitive system based on the popular DTW approach [23] with a total of 9 out of 27 different time functions selected using the Sequential Forward Feature Selection (SFFS) algorithm.

Tables 2 and 3 show the system performance in terms of EER(%) for our Proposed RNN-based Systems for both 1vs1 and 4vs1 cases, respectively. In addition, Table 4 shows the system performance in terms of EER(%) for the DTW-based System [23] for both 1vs1 and 4vs1 cases, over the same evaluation set of Tables 2 and 3. We now analyse the results obtained for each of the three different training scenarios considered.

TABLE 2. 1vs1 Evaluation Results: System performance in terms of EER(%) for the three different training scenarios considered, i.e., "skilled", "random" and "skilled + random".

	Train: "skilled"		Train: "random"		Train: "skilled + random"	
	Skilled	Random	Skilled	Random	Skilled	Random
LSTM	6.44	24.48	13.31	5.38	7.94	6.22
GRU	7.69	29.42	15.63	6.92	7.67	5.98
BLSTM	**5.60**	24.48	15.31	**5.28**	**6.83**	**5.38**
BGRU	6.31	19.14	12.56	5.33	7.88	5.52

TABLE 3. 4vs1 Evaluation Results: System performance in terms of EER(%) for the three different training scenarios considered, i.e., "skilled", "random" and "skilled + random".

	Train: "skilled"		Train: "random"		Train: "skilled + random"	
	Skilled	Random	Skilled	Random	Skilled	Random
LSTM	5.58	24.03	15.17	4.08	6.17	3.67
GRU	6.25	28.69	13.92	4.25	5.58	3.63
BLSTM	**4.75**	24.03	15.58	3.89	**5.50**	3.00
BGRU	4.92	19.69	12.33	**3.25**	5.92	**2.92**

TABLE 4. 1vs1 and 4vs1 DTW-based Evaluation Results: System performance in terms of EER(%).

	1vs1	4vs1
Skilled	10.17	7.75
Random	0.94	0.50

a: SKILLED TRAINING SCENARIO

First, we analyse in Tables 2 and 3 the case in which only pairs of genuine and skilled forgery signatures are used for developing the systems (i.e. "skilled"). Overall, very good results have been obtained for all Proposed Systems when skilled forgeries are considered. Bidirectional schemes (i.e. BLSTM and BGRU) have outperformed normal schemes, highlighting the importance of considering both past and future contexts. In addition, both LSTM and GRU RNN systems have achieved very similar results proving their feasibility for handwritten signature verification. Analysing the results obtained in Tables 2 and 4 for the 1vs1 case, our Proposed BLSTM System has achieved the best results with a 5.60% EER, which corresponds to an absolute improvement of 4.57% EER compared to the 10.17% EER achieved for the DTW-based System. This result (i.e. 5.60% EER) outperforms related state-of-the-art results for the case of considering just one signature for training [20]. Analysing the results obtained in Tables 3 and 4 for the 4vs1 case, our Proposed BLSTM System achieves a 4.75% EER, which corresponds to an absolute improvement of 3.00% EER compared to the 7.75% EER achieved for the DTW-based System. Moreover, it is important to highlight that the result obtained with our Proposed BLSTM System for the case of using just one training signature (1vs1) outperforms the result obtained with the DTW-based System (i.e. 5.60% vs 7.75% EER) for the 4vs1 case. Additionally, our Proposed BLSTM system outperforms other state-of-the-art signature verification systems such as the one proposed in [34] and based on fusion of a function-based system based on DTW and a feature-based system based on Mahalanobis distance (i.e. 4.75% vs 4.91% EER) for the case of considering 4 training signatures. These results show the high ability of our proposed approach for learning even with small amounts of signatures. However, the results obtained in Tables 2 and 3 for our Proposed RNN Systems when random forgeries are considered are far away from the state-of-the-art results. The best result has been obtained using our Proposed BGRU System with a value of 19.14% EER whereas a 0.50% EER is obtained in Table 4 for the DTW-based System. These bad results obtained for random forgeries make sense in this case as only skilled and not random forgeries were used for training the RNNs.

b: RANDOM TRAINING SCENARIO

In order to see the ability of the RNN systems to detect different types of forgeries, Tables 2 and 3 also show the system performance in terms of EER(%) for the scenario in which our Proposed RNN Systems are trained using only pairs of genuine and random forgery signatures (i.e. "random"). Overall, a high improvement of the system performance is achieved for the case of random forgeries compared to the results previously analysed in the "skilled" train scenario. The best result corresponds to our Proposed BGRU System with a 3.25% EER. However, as happened for the "skilled" train scenario previously commented, bad results are achieved for the task in which the RNN system is not trained (i.e. skilled forgeries in this "random" train scenario).

c: SKILLED+RANDOM TRAINING SCENARIO

Finally, Tables 2 and 3 show the system performance in terms of EER(%) for the case in which our Proposed RNN Systems are trained using pairs of genuine and skilled forgery signatures and also pairs of genuine and random forgery signatures (i.e. "skilled + random"). Analysing the results obtained for skilled forgeries, the best system performance has been obtained using our Proposed BLSTM System with a value of 5.50% EER. Moreover, the result obtained with our Proposed BLTM System for the case of using just one training signature (1vs1) still outperforms the result obtained

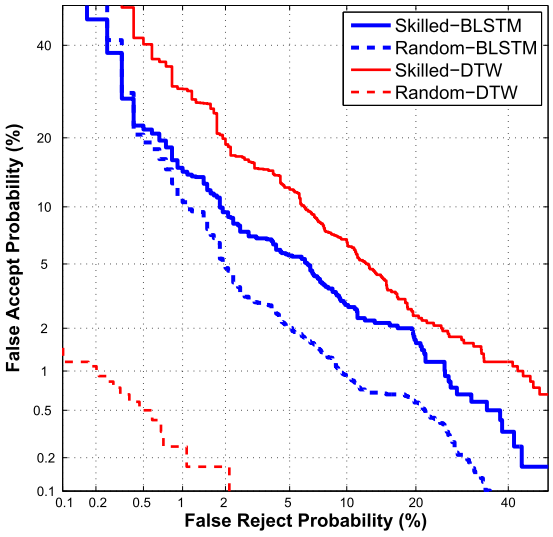

FIGURE 7. System performance results obtained using our Proposed BLSTM System for the 4vs1 case and "skilled + random" train scenario over the BiosecurID evaluation dataset.

with the DTW-based System for the 4vs1 case (i.e. 6.83% vs 7.75% EER), showing the high ability of our proposed approach for learning even with small amounts of signatures. Analysing the results obtained for random forgeries, our Proposed BLSTM System has achieved a 3.00% EER. These results prove the ability of RNN-based systems to detect two different types of forgeries using just one system. Despite of the high improvements achieved when both skilled and random forgeries are used for training the RNNs, the 3.00% EER obtained using our Proposed BLSTM System can not outperform the 0.5% EER obtained using the DTW-based System against random forgeries. Fig. 7 shows the DET curve of both Proposed BLSTM and DTW-based Systems for the 4vs1 case and "skilled + random" train scenario for completeness. In order to achieve state-of-the-art results for both skilled and random forgeries, a possible solution is to perform two consecutive stages similar to [23]: 1) first stage based on DTW optimized for rejecting random forgeries, and 2) our Proposed RNNs Systems in order to reject the remaining skilled forgeries. Another recent example of multiple classifier contribution for signature is [35].

VI. CONCLUSIONS

The main contribution of this work is to assess the feasibility of different RNNs systems in combination with a Siamese architecture [18] for the task of on-line handwritten signature verification. As far as we know, this work provides the first complete and successful framework on the use of multiple RNN systems (i.e. LSTM and GRU) for on-line handwritten signature verification considering both skilled and random types of forgeries. The BiosecurID database comprised of 400 users and 4 separated acquisition sessions has been considered in the experimental work, using the first 300 users for development and the remaining 100 users for evaluation. Three different scenarios regarding the type of forgery considered for training the RNN system is proposed (i.e. "skilled", "random", "skilled + random"). Additionally, two different cases have been considered. First, the evaluation of the system performance considering scores directly from all pairs of signatures (i.e. 1vs1) and second, the case of performing the average of scores of the four one-to-one comparisons (i.e. 4vs1) as there are 4 genuine training signatures per user (from the first session).

Regarding the development of our Proposed RNN Systems, it is important to remark the difference in the number of training iterations needed between normal (i.e. LSTM and GRU) and bidirectional schemes (i.e. BLSTM and BGRU). This shows the importance of considering both past and future contexts in order to train RNNs faster and also with a lower value of training cost. Additionally, it is important to highlight the difference in the number of training iterations between both LSTM and GRU RNNs as the GRU memory block is a simplified version of the LSTM memory block with fewer parameters to train.

Analysing the results obtained using the 100 users of the evaluation dataset, our Proposed BLSTM System has achieved for the "skilled + random" train scenario and 4vs1 case values of 5.50% and 3.00% EER for skilled and random forgeries, respectively. Moreover, the result obtained with our Proposed BLSTM System for the case of using just one training signature (1vs1) still outperforms the result obtained with the highly competitive system based on the popular DTW approach for the 4vs1 case (i.e. 6.83% vs 7.75% EER), showing the high ability of our proposed approach for learning even with small amounts of signatures. Finally, it is important to highlight the results obtained in this work compared to the ones obtained by Otte *et al.* in [17] where all experiments failed obtaining for the best case a 23.75% EER. In that work, standard LSTM architectures seemed not to be appropriate for the task of signature verification. For future work we will address two important current challenges in on-line signature verification: 1) input device interoperability, i.e., signatures for training and testing the system are acquired using different devices, and 2) mixed writing-tool, i.e., signatures for training and testing the system are acquired using different writing tools (stylus or finger). For this we will make use of larger and novel databases [36] in combination to the recurrent Siamese networks described in this work.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[2] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3104–3112.
[3] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. 29th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.
[4] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[5] A. Graves, A. R. Mohamed, and G. Hinton, "Towards end-to-end speech recognition with recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 14. 2014, pp. 1764–1772.
[6] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, no. 1, pp. 201–218, 2000.
[7] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[8] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.
[9] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Networks*, S. C. Kremer and J. F. Kolen, Eds. 2001.
[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[11] K. Cho *et al.* (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." [Online]. Available: https://arxiv.org/abs/1406.1078
[12] K. Cho, B. V. Merrienboer, D. Bahdanau, and Y. Bengio. (2014). "On the properties of neural machine translation: Encoder-decoder approaches." [Online]. Available: https://arxiv.org/abs/1409.1259
[13] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. (2014). "Empirical evaluation of gated recurrent neural networks on sequence modeling." [Online]. Available: https://arxiv.org/abs/1412.3555
[14] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 545–552.
[15] X.-Y. Zhang, G.-S. Xie, C.-L. Liu, and Y. Bengio, "End-to-end online writer identification with recurrent neural network," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 285–292, Apr. 2016.
[16] C. Tiflin and C. Omlin, "LSTM recurrent neural networks for signature verification," in *Proc. Southern African Telecommun. Netw. Appl. Conf.*, 2003, pp. 1–5.
[17] S. Otte, M. Liwicki, and D. Krechel, "Investigating long short-term memory networks for various pattern recognition problems," in *Proc. Mach. Learn. Data Mining Pattern Recognit.*, 2014, pp. 484–497.
[18] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. Comput. Vis. Pattern Recognit.*, 2005, pp. 539–546.
[19] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1993, pp. 737–744.
[20] M. Diaz, A. Fischer, M. A. Ferrer, and R. Plamondon, "Dynamic signature verification system based on one real signature," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 228–239, Jan. 2018.
[21] Y. Liu, Z. Yang, and L. Yang, "Online signature verification based on DCT and sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2498–2511, Nov. 2014.
[22] M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally, "Mobile signature verification: Feature robustness and performance comparison," *IET Biometrics*, vol. 3, no. 4, pp. 267–277, 2014.
[23] M. Gomez-Barrero, J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Plamondon, "Enhanced on-line signature verification based on skilled forgery detection using sigma-lognormal features," in *Proc. IEEE/IAPR Int. Conf. Biometrics (ICB)*, May 2015, pp. 501–506.
[24] R. Zazo, A. Lozano-Diez, J. Gonzalez-Dominguez, D. T. Toledano, and J. Gonzalez-Rodriguez, "Language identification in short utterances using long short-term memory (LSTM) recurrent neural networks," *PLoS ONE*, vol. 11, no. 1, p. e0146917, 2016.
[25] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio. (2014). "How to construct deep recurrent neutal networks." [Online]. Available: https://arxiv.org/abs/1312.6026
[26] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," *J. Mach. Learn. Res.*, vol. 37, pp. 2342–2350, Jul. 2015.
[27] J. Fierrez *et al.*, "BiosecurID: A multimodal biometric database," *Pattern Anal. Appl.*, vol. 13, no. 2, pp. 235–246, May 2010.
[28] D.-Y. Yeung *et al.*, "SVC2004: First international signature verification competition," in *Proc. IAPR Int. Conf. Biometric Authentication*, 2004, pp. 16–22.
[29] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target dependent score normalization techniques and their application to signature verification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 418–425, Aug. 2005.
[30] M. Martinez-Diaz, J. Fierrez, and S. Hangai, "Signature features," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds. Springer, 2015, pp. 1375–1382.
[31] J. Fierrez and J. Ortega-Garcia, *On-Line Signature Verification*. Springer, 2008, pp. 189–209.
[32] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez, "Update strategies for HMM-based dynamic signature biometric systems," in *Proc. 7th IEEE Int. Workshop Inf. Forensics Secur.*, Nov. 2015, pp. 1–6.
[33] F. Bastien *et al.*, "Theano: New features and speed improvements," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1–9.
[34] J. Galbally, M. Diaz-Cabrera, M. A. Ferrer, M. Gomez-Barrero, A. Morales, and J. Fierrez, "On-line signature recognition through the combination of real dynamic data and synthetically generated static data," *Pattern Recognit.*, vol. 48, no. 9, pp. 2921–2934, 2015.
[35] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez, "Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification," *IEEE Access*, vol. 3, pp. 478–489, May 2015.
[36] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Benchmarking desktop and mobile handwriting across COTS devices: The e-BioSign biometric database," *PLoS ONE*, vol. 12, no. 5, p. e0176792, 2017.

RUBEN TOLOSANA received the M.Sc. degree in Electrical Engineering in 2014 from Universidad Autonoma de Madrid. In April 2014, he joined the Biometrics and Data Pattern Analytics (BiDA) Lab-ATVS at the Universidad Autonoma de Madrid, where he is currently collaborating as an assistant researcher pursuing the Ph.D. degree. In 2015 he was awarded with a FPU research fellowship from Spanish MECD. His research interests are mainly focused on signal and image processing, pattern recognition, deep learning, and biometrics, particularly in the areas of handwriting and handwritten signature. He is author of several publications and also collaborates as a reviewer in many different international conferences (e.g. ICDAR and ICB) and high-impact journals (e.g. IEEE TRANSACTIONS ON CYBERNETICS and *International Journal of Pattern Recognition and Artificial Intelligence*). Finally, he has participated in several National and European projects focused on the deployment of biometric security through the world.

RUBEN VERA-RODRIGUEZ received the M.Sc. degree in telecommunications engineering from the Universidad de Sevilla, Spain, in 2006, and the Ph.D. degree in electrical and electronic engineering from Swansea University, U.K., in 2010. Since 2010, he has been with the Biometric Recognition Group, Universidad Autónoma de Madrid, Spain, first as a recipient of a Juan de la Cierva Post-Doctoral Fellowship from the Spanish Ministry of Innovation and Sciences, and as an Assistant Professor since 2013. His current research interests include signal and image processing, pattern recognition, and biometrics, with emphasis on signature, face and gait verification and forensic applications of biometrics. He is actively involved in several National and European projects focused on biometrics.

JULIAN FIERREZ received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Universidad Politecnica de Madrid, Spain, in 2001 and 2006, respectively. Since 2002, he has been affiliated with the Biometric Recognition Group, first with the Universidad Politecnica de Madrid, and since 2004 with the Universidad Autonoma de Madrid, where he is currently an Associate Professor. From 2007 to 2009, he was a Visiting Researcher with Michigan State University, USA, under the Marie Curie Fellowship. His current research interests include general signal and image processing, pattern recognition, and biometrics, with emphasis on signature and fingerprint verification, multibiometrics, biometric databases, system security, and forensic applications of biometrics. He has been actively involved in multiple EU projects focused on biometrics (e.g. TABULA RASA and BEAT), has attracted notable impact for his research, and is a recipient of a number of distinctions, including: EBF European Biometric Industry Award 2006, EURASIP Best PhD Award 2012, Medal in the Young Researcher Awards 2015 by the Spanish Royal Academy of Engineering, the Miguel Catalan Award to the Best Researcher under 40 in the Community of Madrid in the general area of Science and Technology, and the 2017 IAPR Young Biometrics Investigator Award. Since 2016 he is Associate Editor for IEEE TRANS. ON INFORMATION FORENSICS AND SECURITY and the IEEE BIOMETRICS COUNCIL NEWSLETTER.

JAVIER ORTEGA-GARCIA received the M.Sc. degree in electrical engineering and the Ph.D. degree (*cum laude*) in electrical engineering from Universidad Politécnica de Madrid, Spain, in 1989 and 1996, respectively. He is currently a Full Professor at the Signal Processing Chair in Universidad Autónoma de Madrid-Spain, where he holds courses on biometric recognition and digital signal processing. He is a founder and Director of the BiDA-Lab, Biometrics and Data Pattern Analytics Group. He has authored over 300 international contributions, including book chapters, refereed journal, and conference papers. His research interests are focused on biometric pattern recognition (online signature verification, speaker recognition, human-device interaction) for security, e-health and user profiling applications. He chaired Odyssey-04, The Speaker Recognition Workshop, ICB-2013, the 6th IAPR International Conference on Biometrics, and ICCST-2017, the 51st IEEE International Carnahan Conference on Security Technology.