Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords

Syed Zulkarnain Syed Idrus [a,b,c,d,*], Estelle Cherrier [b,c,d], Christophe Rosenberger [b,c,d], Patrick Bours [e]

[a] Universiti Malaysia Perlis, 01000 Kangar, Perlis, Malaysia
[b] Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France
[c] ENSICAEN, UMR 6072 GREYC, F-14032 Caen, France
[d] CNRS, UMR 6072 GREYC, F-14032 Caen, France
[e] NISlab, Gjøvik University College, Gjøvik, Norway

ARTICLE INFO

Article history:
Received 31 October 2013
Received in revised form
6 April 2014
Accepted 25 May 2014
Available online 10 June 2014

Keywords:
Biometrics
Keystroke dynamics
Soft biometrics
Pattern recognition
Data fusion
Computer security

ABSTRACT

This paper presents a new profiling approach of individuals based on soft biometrics for keystroke dynamics. *Soft biometric traits* are unique representation of a person, which can be in a form of physical, behavioural or biological human characteristics that differentiate between him/her into a group people (*e.g.* gender, age, height, colour, race *etc.*). *Keystroke dynamics* is a behavioural biometric modality to recognise how a person types on a keyboard. In this paper, we consider the following soft traits: the hand category (*i.e.* if the user types with one or two hands), the gender category, the age category and the handedness category. For this purpose, we collected a new database. Two cases are studied: static passwords and free text. By combining machine learning and fusion process, the results are promising.

1. Introduction

It is accepted that the way a person types on a keyboard contains timing patterns, which can be used to label him/her and this is called *keystroke dynamics*. Keystroke dynamics is an interesting and a low cost biometric modality (Bours, 2012; Giot et al., 2011a), indeed no additional device is required. Keystroke dynamics belongs to the class of behavioural biometrics, in the sense that the template of a user reflects an aspect of his/her behaviour. Among the behavioural biometric modalities, we can mention signature dynamics analysis, gait recognition, voice recognition, or keystroke dynamics (Impedovo and Pirlo, 2007; Moustakas et al., 2010; Klevans and Rodman, 1997; Monrose and Rubin, 2000). In general, the global performances of behavioural biometric modalities (and especially keystroke dynamics) based authentication systems are lower than the popular morphologic biometric modalities based authentication systems (such as fingerprints, face or iris) (Maio and Jain, 2009; Wildes, 1997). The fact that the

[*] Corresponding author. Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France. Tel.: +33 645246908; fax: +33 231538110.
E-mail addresses: syed-zulkarnain.syed-idrus@ensicaen.fr, syzul@yahoo.com (S.Z. Syed Idrus), estelle.cherrier@ensicaen.fr (E. Cherrier), christophe.rosenberger@ensicaen.fr (C. Rosenberger), patrick.bours@hig.no (P. Bours).

http://dx.doi.org/10.1016/j.cose.2014.05.008
0167-4048/© 2014 Elsevier Ltd. All rights reserved.

performances of keystroke dynamics are lower than other biometric modalities can be explained by the *intra-class* variability of the users behaviour. This intra-class variability pertaining to computer users can be accounted for by a way of typing which is different when they are nervous, or angry, or even sad … (Epp et al., 2011).

One solution to cope with this variability is to study *soft biometrics*, which was first introduced by Jain et al. (2004). In that paper '*soft biometric traits*' are defined as "*characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals*". Jain *et al*. considered gender, ethnicity and height as complementary data for a usual fingerprint based biometric system. Thus, soft biometrics allow a refinement of the search of the genuine user in the database, resulting in a computing time reduction. For example, if the capture corresponds to a male according to a soft biometrics module, then, the standard biometric identification system can confine its search area to male users, without considering female ones.

Since the work of Jain *et al*., several other articles related to soft biometrics can be found in the literature. In the paper (Ailisto et al., 2006), body weight and fat measurements are considered as soft criteria to enhance a standard fingerprint based biometric system. An overview can be found in Dantcheva et al. (2011) about soft biometrics, under a '*Bag of Soft Biometrics*', where Dantcheva *et al*. make a comparison with the pioneering work of Alphonse Bertillon, whose anthropometric criteria gave rise to soft biometrics (Rhodes, 1956). That paper proposes some facial soft biometrics and also body soft biometrics, namely weight and clothes colour detection. In Park and Jain (2010), Park and Jain present how gender or ethnicity and facial marks such as scars, moles and freckles can be used to enhance face recognition. In reference Dong and Woodard (2011), shape based eyebrow features are used for biometric recognition and soft biometric classification. In Denman et al. (2011), the authors use soft biometrics (height and colour model of head, torso and legs) to help identifying people in videos in surveillance networks. Marcialis et al. (2009) use hair colour and ethnicity as soft biometrics combined with face modality.

Regarding keystroke dynamics, Bixler and D'Mello (2013) look into the likelihood of 44 people's behaviour, whether they stay idle, involved or bored when asked to write on a given task. Their results are between 11% and 38% higher than random guessing. In our previous study (Idrus et al., 2013a), the results also show that it is possible to detect users' way of typing by using one/two hand(s) with over 90% recognition rate; gender between 65% and 90%; age between 65% and 82%; and handedness between 70% and 90% correct recognition accuracy with 110 users.

The objective of this paper is to propose an extended study of soft biometrics for keystroke dynamics from our previous study in Idrus et al. (2013a) on a new biometric benchmark database called 'GREYC-NISLAB Keystroke' (Idrus et al., 2013b) that we have created. We propose in this paper a thorough evaluation of the soft biometrics system and a comparison between static passwords and free text (digraphs). Thus, the novelty (compared to our papers mentioned) is to study to what extent soft biometrics can enhance the recognition performance of keystroke based authentication systems. Furthermore, we show how the performances can be increased significantly by data fusion for passwords. As soft criteria, we propose to test if it is possible to predict if the user:

1. types with one or two hands
2. is a male or a female
3. belongs to a particular age category
4. is right-handed or left-handed

This paper is organised as follows. Section 2 is devoted to the description of the proposed method. In Section 3, we describe the protocol that we applied and present the obtained results on the benchmark database in Section 4. Section 5 presents the conclusions and the different perspectives of this study.

2. Proposed methodology

In general, keystroke dynamics authentication systems involve a keyboard and an application for the capture and processing of the biometric information. Users are required to type on a keyboard running a dedicated application. Each capture is stored in a database within the application in the form of keystroke or timing features for all correct and incorrect entries. These features are composed of several timing values that are extracted, which is the *pattern vector* that is used for the analysis. For each soft criterion, two steps are involved in recognition evaluation: (i) a training step, and (ii) a test step, both relying on a machine learning algorithm. Here we have chosen SVM (Support Vector Machine) (Vapnik, 1998), on account of its efficiency. As a result, we compute the accuracy rate of the prediction of each soft category by the trained SVM. A graphical representation of the overall process is illustrated in Fig. 1. In order to enhance the overall recognition performance, data fusion is then applied.

2.1. Data capture

Different types of features can be extracted from a user while typing on a keyboard (Giot et al., 2011a): "*(i) code of the key; (ii) the type of event (press or release); and (iii) the time of the event*". All this timing information is stored in the form of raw data, which contains (see Fig. 2):

- *ppTime (PP)*: the latency of when the two buttons (keys) are pressed;
- *rrTime (RR)*: the latency of when the two buttons (keys) are released;
- *prTime (PR)*: the duration of when one button (key) is pressed and the other is released;
- *rpTime (RP)*: the latency of when one button (key) is released and the other is pressed.
- *vector (V)*: the concatenation of the previous four timing values.

Subsequently, the keystroke template *V* is utilised for the analysis for each soft category. For keystroke dynamics systems, we apply two approaches, namely: static passwords and free text. Concerning static passwords, we analyse all the typing features previously described. For free text, the

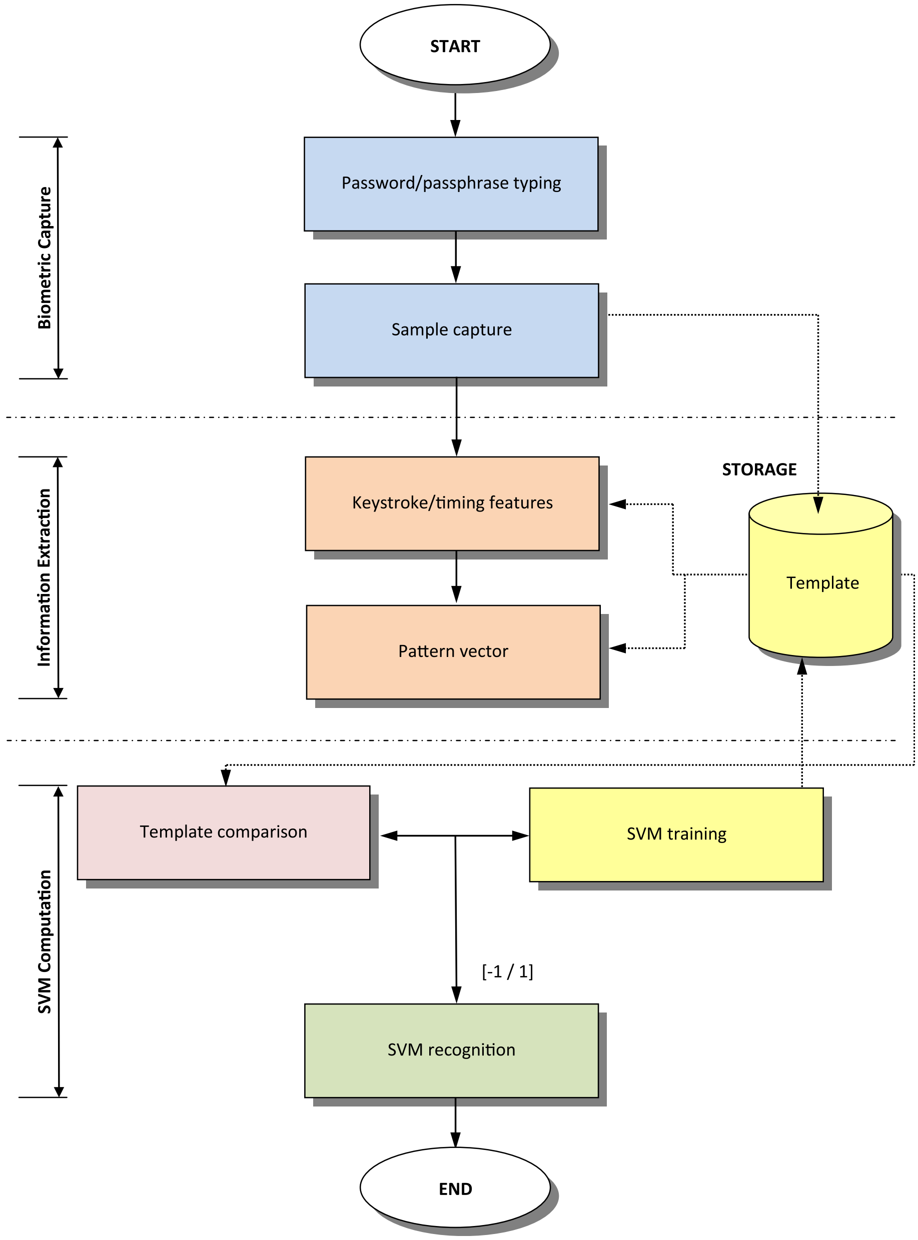


Fig. 1 – The overall process of the proposed system.

analysis is based on digraphs, which are the time latencies between two successive keystrokes *i.e.* digraphs transition time. These typing rhythms are extracted from the users' texts typed without any specific constraint.

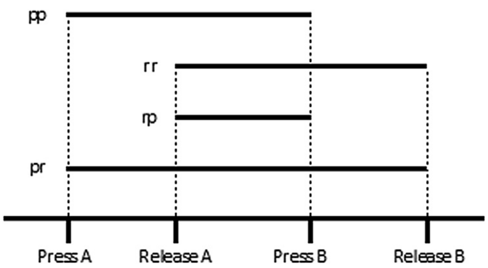


Fig. 2 – Keystroke typing features (Idrus et al., 2013b).

2.2. *Data analysis*

For the data analysis, we recall that we are interested in soft biometrics criteria that can be applied to our biometric database: one or two hand(s); male or female, age < 30 or ≥ 30 years old, right- or left-handed. This subsection presents the methodology in which we followed to analyse keystroke data. Classification is performed by training, for each soft criterion (hand, gender, age, and handedness categories), using a Support Vector Machine. We use LibSVM (Chang and Lin, 2011) with the Radial Basis Function (RBF) kernel (Hsu et al., 2003; Hearst et al., 1998). In order to maximise the performance, we have to determine which is the best couple for our computation. We set the following values for the parameters: $C = 128$ is the penalisation coefficient of the SVM; $\gamma = 0.125$ is the parameter of the kernel, as introduced by Hsu et al. (2003).

The computation of the SVM process is repeated for 100 iterations for each percentage of the training ratio, to produce an averaged recognition rate.

2.3. Data fusion process

For the data fusion, we apply two techniques based on *majority voting* and *score fusion* with binary classifications. For the sake of clarity, we take the example of gender category. There are more men than women in the database (*i.e.* 78 males; 32 females). We select data to have the same number within each category, so here, we randomly remove 46 males. We keep the same users sub-sample for each password, and we train one SVM per soft category. To avoid the influence of sample extraction, the whole process (from the extra men removal to the fusion) is repeated 100 times, with a different random draw of 32 males each time. The presented results are the average of these 100 classifications. Now, we present the chosen fusion processes.

First fusion process: *majority voting*. The predicted label (+1 or −1) is exploited in the first fusion method: the majority voting. Since there are 5 passwords, the majority is easily obtained.

Second fusion process: *score fusion*. We compute this score by using the predicted label and its associated probability. This method, we obtain a score in the range [0; 1], then we compute the average of 5 probabilities to decide the final class. If the average is above 0.5 then the label 1 is assigned, otherwise label 0.

Once this process has been completed, we can compute the *confusion matrix* (Stehman, 1997) to obtain the correct recognition rate for each class. To compute the recognition rate (for gender category), we apply formula (1), where *M_correct* and *F_correct* are respectively the total number of correctly predicted Males and Females. A large value of *r* guarantees a large correct recognition rate for the considered category. Subsequently, we will be able to evaluate to what extent both fusion processes can enhance the performance.

$$r = \frac{M\_correct + F\_correct}{total\_data} \times 100\% \qquad (1)$$

3. Experimental protocol

3.1. Static passwords

In this section, we briefly describe the protocol that we applied. We refer the interested reader to our previous paper (Idrus et al., 2013b) for more details. As mentioned earlier, we created a biometric benchmark database. The database can facilitate and accelerate reproducible and comparable research. In Idrus et al. (2013a), an experiment has been performed in two locations: France and Norway, and a total of 110 individuals had volunteered to participate. We used two desktop keyboards (French keyboard for users in France and Norwegian keyboard for users in Norway) *i.e.* AZERTY and QWERTY (this is not a classical QWERTY keyboard, however, we do not use specific Norwegian keys), respectively. Giot *et al.* work show that the keyboard does not influence the performance (Giot et al., 2011b).

During the data acquisition, some metadata such as gender, age and handedness were collected. We have chosen passphrases that are well-known in both countries, which are between 17 and 24 characters long including spaces (see Table 1). All the participants were asked to type the 5 different passphrases 20 times (10 times with one hand and 10 times with two hands).

We have used GREYC Keystroke software developed at GREYC Laboratory (downloadable from the following address: http://www.ecole.ensicaen.fr/~rosenber/keystroke.html), to capture the biometric data. At the end of the data collection, a total of 11,000 data samples are in the proposed biometric benchmark database. For each user, 7 out of 10 samples are used for training and testing data. The first three entries for each user are not taken into account because a leeway was given to the users to allow them to train themselves for each of the given passphrases. We justify why three entries have been discarded by operational reasons: (i) noticed that we made 10 captures for each password entry because we want to avoid the volunteers from being annoyed, having to type (the same text) too many times; and (ii) by removing more than 3 captures will definitely lead to a smaller database. So it is more an operational justification than a statistical one, which could have made sense with much more data.

We define two classes C_1 and C_2 for each category as follows:

- *Hand category*: C_1 = One Hand: only one hand is used (right/left depending on the handedness of the user); C_2 = Two Hands: both hands are used.
- *Gender category*: C_1 = Male; C_2 = Female.
- *Age category*: C_1 = <30 years old; C_2 = ≥30 years old.
- *Handedness category*: C_1 = Right-handed; C_2 = Left-handed.

Here, for hand category, we use all the data. Whereas for the other soft biometrics information, we only use data corresponding to the usual way of typing, that is 2 hands.

To validate the proposed recognition system, we compute Confidence Intervals (CI). A CI is necessary when it is associated with the recognition rate of the soft biometric trait to reinforce the confidence in the obtained results. It represents a measure of confidence on the estimated error rate. It is based on a re-sampling, which consists of a random draw with a replacement of new values of example from the test base. For each draw, the data are randomly selected. This is done $N = 100$ times in order to calculate the CI, where we perform the computation of the recognition rate for each of the N tries. The CI can be determined based on the percentiles of the normal distribution. Here, the CI at 95% is defined by Equation (2), where m(rate) is the mean of the recognition rates over N

Table 1 – Passphrases.

Label	Description	Character size
Password 1 (P_1)	leonardo dicaprio	17
Password 2 (P_2)	the rolling stones	18
Password 3 (P_3)	michael schumacher	18
Password 4 (P_4)	red hot chilli peppers	22
Password 5 (P_5)	united states of america	24

iterations, and $\sigma(\text{rate})$ is the corresponding standard deviation.

$$CI = m(\text{rate}) \pm 1.96 \frac{\sigma(\text{rate})}{\sqrt{N}} \tag{2}$$

3.2. *Free text*

Subsequently, we perform a distance measure to consider the different timing information between two-character sequences known as *digraphs*. The types passwords are considered as a whole, and only digraphs are kept. Digraphs are the latency times between two successive keystrokes. We extract the keystroke features using the mean and variance of digraphs. Here, we consider free text as the collection of the 5 passwords. Therefore, the digraphs appear with an occurrence between one and four. To obtain significative results, we restrict to digraphs with an occurrence equal or larger than 2. Thus, we consider three categories of digraph: (i) 11 with two occurrences; (ii) 2 with three occurrences; and (iii) 1 with four occurrences. Consequently, there are a total of 14 occurrences as shown in Fig. 3.

In some instances, the digraphs appear numerous times and because of that the size of the timing vector may differs from one digraph instance to another (Davoudi and Kabir, 2009). In a long text, there may be more than one instance of a digraph. However, the mean of all these instances is used as a corresponding latency time. It was shown in Sim and Janakiraman (2007) that the typing pattern of a letter sequence may change when it is part of a larger word. For example, digraph ‘IS’ has different timing information in typing the word ‘**IS**’ and the word ‘FUTUR**IS**TIC’.

Finally, we compute the confusion matrix in order to obtain the correct recognition rate. For each class, it presents the percentage of correctly classified users. We define our soft biometrics information as shown in Table 2.

Table 2 – Soft biometrics information class label.

One hand = 1	Two hands = −1
Male = 1	**Female = −1**
<30 years old = 1	≥30 years old = −1
Right-handed = 1	Left-handed = −1

4. Experimental results

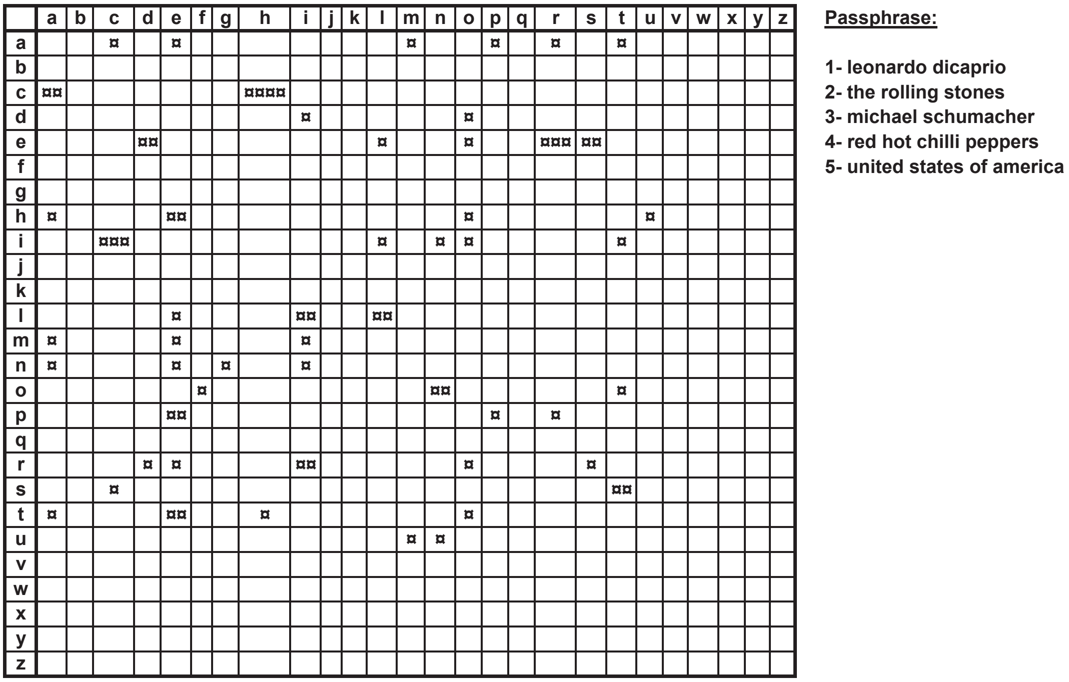
In the first subsection, we quantify the performance results of soft biometrics for keystroke dynamics with static passwords. As mentioned, distance measure are calculated for different timing information between two-character sequences, and hence we show that with any combinations of two-key characters (digraphs), significant results are obtained with free text illustrated in the following subsection.

4.1. *Passwords: static*

We performed several computations by using SVM. We recall that we present the evolution of the average (over 100 computations) recognition rate, while varying the percentage of the training data (from 1% to 90%), for each of the four soft category.

- Hand category recognition

Fig. 4 illustrates the results of the recognition rates for different training ratios with one hand (C_1) and two hands (C_2) for five passphrases P_1 to P_5. To compute these results, an equal amount of data is used for both classes, in particular 770 data samples for each class. In this experiment, the results are promising, since from a ratio of training data over 50% of the total data of the 110 users, the recognition rate is over 90%. This means that if the database contains more than 55 users,



	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a			¤		¤								¤			¤		¤		¤						
b																										
c	¤¤							¤¤¤¤																		
d									¤						¤											
e				¤¤								¤			¤			¤¤¤	¤¤							
f																										
g																										
h	¤				¤¤										¤						¤					
i			¤¤¤									¤		¤	¤					¤						
j																										
k																										
l					¤				¤¤			¤¤														
m	¤				¤				¤																	
n	¤				¤		¤		¤																	
o						¤								¤¤						¤						
p					¤¤											¤		¤								
q																										
r				¤	¤				¤¤						¤				¤							
s			¤																	¤¤						
t	¤				¤¤			¤							¤											
u													¤	¤												
v																										
w																										
x																										
y																										
z																										

Passphrase:

1- leonardo dicaprio
2- the rolling stones
3- michael schumacher
4- red hot chilli peppers
5- united states of america

Fig. 3 – Digraphs and its number of occurrences.

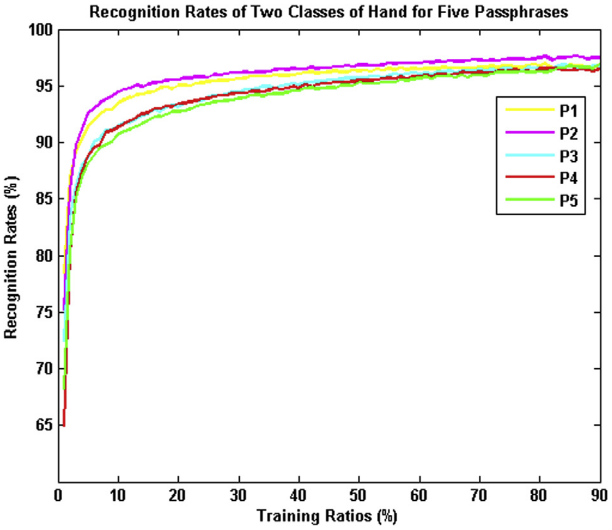


Fig. 4 – Average values for 100 iterations of recognition rates at 1–90% training ratios with two classes of hand for five passphrases (Idrus et al., 2013a).

the soft biometric system is able to determine if the user types with one or two hands.

- Gender category recognition

Fig. 5 illustrates the results of the recognition rates for different training ratios with males (C_1) and females (C_2) for passphrases P_1 to P_5. Only 30% of the data samples of male users are used (but all samples belong to female) in order to have equilibrated classes (*i.e.* 224 data samples related to male participants and 224 data samples related to female participants). The recognition rate depends on the particular passphrase and ranges from 70% to 86%.

- Age category recognition

Fig. 6 illustrates the results of the recognition rates for different training ratios with <30 years old (C_1) and ≥30 years old (C_2) for passphrases P_1 to P_5. We remove 46% of the data samples of class C_1 to have equal size data classes each with the data of 51 users. The recognition rate for a ratio over 50% is slightly less than the other soft criteria, namely between 67% and 78%.

- Handedness category recognition

Fig. 7 illustrates the results of the recognition rates for different training ratios with right-handed (C_1) and left-handed (C_2) for passphrases P_1 to P_5. We keep only 12% of the right-handed class and all the left-handed class to have equal size classes. The obtained recognition rate tends to vary more than the other soft categories, but stays between 76% and 88%, which are nevertheless quite good results. However, as mentioned, the selected database for this category contains only 12 users in each class, therefore the performances are decreased and the confidence intervals are wider compared to other soft criteria with 110 users in each class.

- Confidence interval

Table 3 shows the CI computed with a training dataset containing 50% of the whole database, for different categories (*i.e.* hand, gender, age, handedness).

4.2. *Free text: digraphs*

We performed similar analysis with SVM as mentioned in Section 4.1. The first results deal with averaging of recognition rates (100 iterations) on all four soft categories for different percentage of training data ranging from 1% to 90%.

Fig. 8 illustrates the evolution of the recognition rates on different training ratios with C_1: one hand, male, age < 30

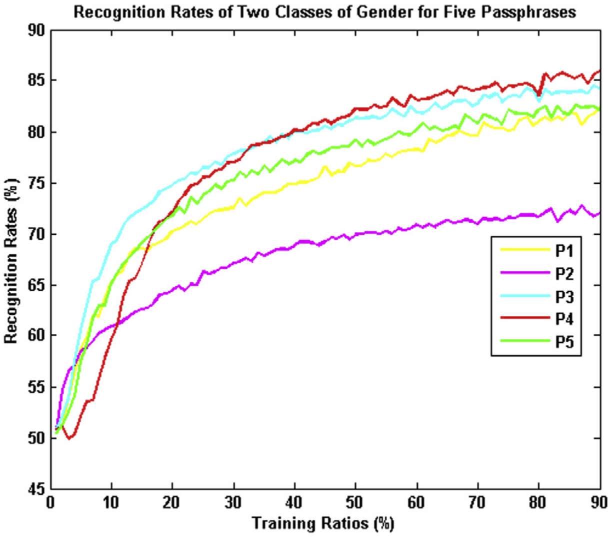


Fig. 5 – Average values for 100 iterations of recognition rates at 1–90% training ratios with two classes of gender for five passphrases (Idrus et al., 2013a).

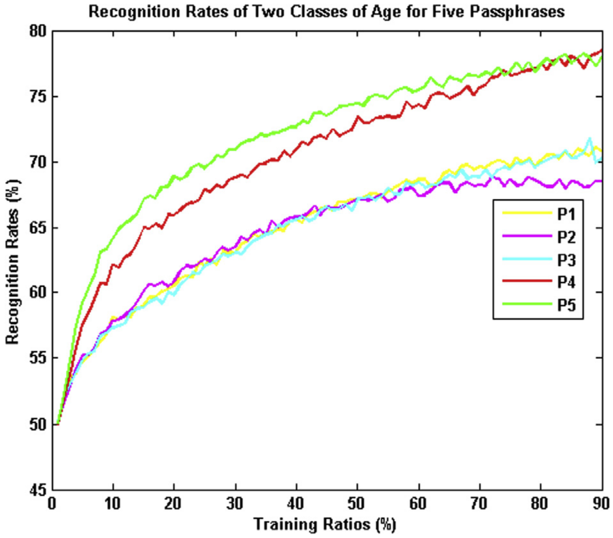


Fig. 6 – Average values for 100 iterations of recognition rates at 1–90% training ratios with two classes of age for five passphrases (Idrus et al., 2013a).

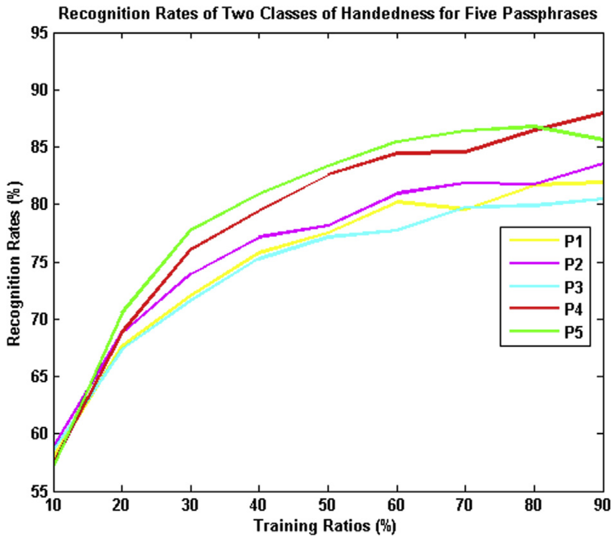


Fig. 7 – Average values for 100 iterations of recognition rates at 10–90% training ratios with two classes of handedness for five passphrases (Idrus et al., 2013a).

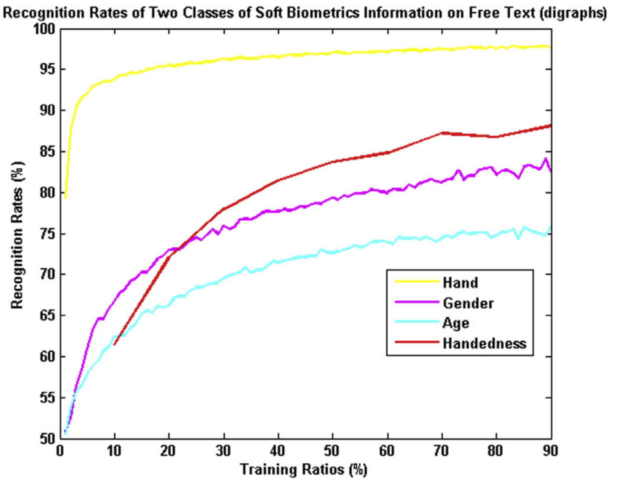


Fig. 8 – Average values for 100 iterations of recognition rates at 1–90% training ratios with two classes of soft biometrics information with 14 digraphs (occurrences ≥ 2) on free text.

years old, right-handed; and C_2: two hands, female, age ≥ 30 years old for all four different soft biometrics information. In this experiment, the results are promising, from the ratio of 50% of total data used for SVM training, the recognition rate for Hand Class Recognition is over 90%; Gender Class Recognition is between 79% and 84%; Age Class Recognition is between 72% and 75%; and Handedness Class Recognition is between 83% and 88%. Table 4 summarises the performance comparison of recognition rate between passwords and free text at 50–90% training ratio.

4.3. *Confusion matrix: majority voting and score fusion for passwords*

In order to further enhance the performance, we perform data fusion, where we show that there is a great increase in the recognition accuracy rate results. The results of the obtained confusion matrix have improved significantly by fusing the data on all soft biometrics information at 50% training ratio based on *static passwords*. We apply the same equation and ratio on free text, and here the results of the corresponding confusion matrix are based on *digraphs*. Then, the obtained performances are compared with three SVM computations: (i) before fusion for static passwords and free text; (ii) fusion based on *majority voting*; and (iii) fusion based on *score*; and only for static passwords in (ii) and (iii). Here, the fusion does not involves free text because all of the digraphs data are in the passwords. Table 5 summarises this information.

4.4. *Discussions*

From the previous results, we are able to see that the performances differ from one soft category to another. For static passwords, fusion processes namely *majority voting* and *score fusion* techniques have significantly increased the recognition performance rate on all soft biometrics characteristics from the initial results. *Score fusion*, however, gives better results, where we can see the results of this performance have increased significantly.

The results of free text are slightly superior to those of static passwords as illustrated in Table 5. As mentioned, we consider free text as the collection of the 5 passwords. With a total of only 14 occurrences consisting in three categories of digraph namely (i) 11 with two occurrences; (ii) 2 with three occurrences; and (iii) 1 with four occurrences, nevertheless, the results are quite promising.

5. Conclusions and perspectives

In this paper, we propose a new soft biometric approach for keystroke dynamics. It consists of predicting the users' way of typing by defining the hand category *i.e.* number of hands used to type (one/two); gender category; age category; and handedness category, where the results are promising for both static passwords and free text. Moreover, we are able to enhance the soft biometrics recognition rate for static

Table 3 – Confidence interval computation at 50% training ratio for 5 passphrases and the data distribution (number of users) in each class.

	P_1	P_2	P_3	P_4	P_5
Hand (770 data samples: C_1 & C_2)	96% ± 0.1%	96% ± 0.1%	95% ± 0.1%	94% ± 0.1%	94% ± 0.1%
Gender (224 data samples: C_1 & C_2)	74% ± 0.3%	69% ± 0.3%	70% ± 0.2%	78% ± 0.2%	76% ± 0.2%
Age (357 data samples: C_1 & C_2)	64% ± 0.2%	64% ± 0.2%	63% ± 0.2%	69% ± 0.2%	69% ± 0.2%
Handedness (84 data samples: C_1 & C_2)	72% ± 1.2%	73% ± 1.2%	72% ± 1.2%	72% ± 1.3%	73% ± 1.2%

Table 4 – Summary of performance comparison of recognition rates for passwords and free text at 50–90% training ratios.

	Passwords	Free text
Hand (770 data samples: C_1 & C_2)	>90%	>90%
Gender (224 data samples: C_1 & C_2)	Between 70% and 86%	Between 79% and 84%
Age (357 data samples: C_1 & C_2)	Between 67% and 78%	Between 72% and 75%
Handedness (84 data samples: C_1 & C_2)	Between 78% and 88%	Between 83% and 88%

passwords significantly by data fusion and achieve higher performance accuracy. Another part of this work is a comparative study between passwords and free text, where both approaches give good results. For passwords, it is based on static texts where the users are obliged to type with some constraints *i.e.* users type specific pre-defined strings. Free text, on the other hand, with any combinations of two-key characters (digraphs) is also able to provide good recognition rates without any specific constraints on the users when typing.

For free text, it may be considered as suitable recognition if a set of password is created by the user himself/herself, and hence having his/her own freedom of texts choice. Nonetheless, the effectiveness of digraphs is discriminative only when they are word-specific *i.e.* digraph features depend on the word context they are occurred in Zhong et al. (2012). Therefore, the obtained results could be used as a reference model to assist the biometric system to better recognise a user by a way he/she types on a keyboard. This will not only strengthen the authentication process by hindering an impostor trying to enter into the system, but also cut down on the computation time.

The results presented in this paper can be used to improve user authentication based on keystroke dynamics by combining two pieces of information: (i) 'scores' provided by the biometric authentication system when comparing the reference to a stored template; and (ii) a 'reliability index' by verifying the concordance between one extracted soft biometric information (such as gender) and the known information. The results in this work could also be applied, for example, in securing social networks, where the soft biometric characteristics of a person in a chat room can be checked against his/her claimed profile.

Table 5 – Performance comparison before and after fusion for passwords, and free text at 50% training ratio.

Technique	Soft biometrics information	Before fusion	By fusing	
			Majority voting	Score fusion
Password	Hand category	93.66%	100%	100%
	Gender category	62.5%	85.71%	92.14%
	Age category	55.49%	86.67%	85.71%
	Handedness category	61.65%	84.52%	91.67%
Free text	Hand category	96.57%		
	Gender category	80%		
	Age category	65.71%		
	Handedness category	83.33%		

REFERENCES

Ailisto H, Vildjiounaite E, Lindholm M, Mkel S-M, Peltola J. Soft biometrics – combining body weight and fat measurements with fingerprint biometrics. Pattern Recognit Lett 2006;27(5):325–34.

Bixler R, D'Mello S. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In: Proceedings of the 2013 international conference on intelligent user interfaces. ACM; 2013. pp. 225–34.

Bours P. Continuous keystroke dynamics: a different perspective towards biometric evaluation. Inf Secur Tech Rep 2012;17(1–2):36–43. http://dx.doi.org/10.1016/j.istr.2012.02.001.

Chang C-C, Lin C-J. Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol TIST 2011;2(3):27. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Dantcheva A, Velardo C, Dangelo A, Dugelay J-L. Bag of soft biometrics for person identification. Multimed Tools Appl 2011;51(2):739–77.

Davoudi H, Kabir E. A new distance measure for free text keystroke authentication. In: 14th International CSI computer conference (CSICC) 2009. IEEE; 2009. pp. 570–5.

Denman S, Bialkowski A, Fookes C, Sridharan S. Determining operational measures from multi-camera surveillance systems using soft biometrics. In: 8th IEEE international conference on advanced video and signal-based surveillance (AVSS) 2011. IEEE; 2011. pp. 462–7.

Dong Y, Woodard DL. Eyebrow shape-based features for biometric recognition and gender classification: a feasibility study. In: International joint conference on biometrics (IJCB) 2011. IEEE; 2011. pp. 1–8.

Epp C, Lippold M, Mandryk R. Identifying emotional states using keystroke dynamics. In: Proceedings of the 2011 annual conference on human factors in computing systems; 2011. pp. 715–24.

Giot R, El-Abed M, Rosenberger C. Keystroke dynamics overview. In: Yang J, editor. Biometrics/book 1, vol. 1. InTech; 2011a. pp. 157–82. URL, http://www.intechopen.com/articles/show/title/keystroke-dynamics-overview.

Giot R, El-Abed M, Hemery B, Rosenberger C. Unconstrained keystroke dynamics authentication with shared secret. Comput Secur 2011b;30(67):427–45. http://dx.doi.org/10.1016/j.cose.2011.03.004.

Hearst M, Dumais S, Osman E, Platt J, Scholkopf B. Support vector machines. Intell Syst Appl IEEE 1998;13(4):18–28.

Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification; 2003.

Idrus S, Cherrier E, Rosenberger C, Bours P. Soft biometrics for keystroke dynamics. In: Kamel M, Campilho A, editors. Image analysis and recognition. Lecture notes in computer science, vol. 7950. Springer Berlin Heidelberg; 2013a. pp. 11–8.

Idrus S, Cherrier E, Rosenberger C, Bours P. Soft biometrics database: a benchmark for keystroke dynamics biometric systems. In: 2013 International conference of the biometrics special interest group (BIOSIG); 2013. pp. 281–8.

Impedovo S, Pirlo G. Verification of handwritten signatures: an overview. In: 14th International conference on image analysis and processing (ICIAP) 2007. IEEE; 2007. pp. 191–6.

Jain AK, Dass SC, Nandakumar K. Soft biometric traits for personal recognition systems. In: Proceedings of international conference on biometric authentication. Springer; 2004. pp. 731–8.

Klevans RL, Rodman RD. Voice recognition. Artech House, Inc.; 1997.

Maio D, Jain AK. Handbook of fingerprint recognition. Springer; 2009.

Marcialis GL, Roli F, Muntoni D. Group-specific face verification using soft biometrics. J Vis Lang Comput 2009;20(2):101–9.

Monrose F, Rubin AD. Keystroke dynamics as a biometric for authentication. Future Gener Comput Syst 2000;16(4):351–9.

Moustakas K, Tzovaras D, Stavropoulos G. Gait recognition using geometric features and soft biometrics. Signal Process Lett IEEE 2010;17(4):367–70.

Park U, Jain A. Face matching and retrieval using soft biometrics. IEEE Trans Inf Forensics Secur 2010;5(3):406–15.

Rhodes HTF. Alphonse Bertillon, father of scientific detection. Abelard-Schuman; 1956.

Sim T, Janakiraman R. Are digraphs good for free-text keystroke dynamics?. In: Conference on computer vision and pattern recognition (CVPR) 2007. IEEE; 2007. pp. 1–6.

Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 1997;62(1):77–89.

Vapnik V. Statistical learning theory. Wiley; 1998.

Wildes R. Iris recognition: an emerging biometric technology. Proc IEEE 1997;85(9):1348–63.

Zhong Y, Deng Y, Jain AK. Keystroke dynamics for user authentication. In: Conference on computer vision and pattern recognition workshops (CVPRW) 2012. IEEE; 2012. pp. 117–23.

Syed Zulkarnain Syed Idrus received the B.Sc. degree in Information Systems Engineering from University of Manchester Institute of Science and Technology (UMIST), United Kingdom and M.Sc. degree in Computer Engineering from Universiti Malaysia Perlis (UniMAP), Malaysia in 2001 and 2008, respectively. He is a Senior Lecturer at UniMAP (from 2009) and currently pursuing a Ph.D. degree in Computer Science and Applications at University of Caen Lower-Normandy, France specialising in biometrics. His research interest includes biometrics, pattern recognition, encryption, and information security.

Estelle Cherrier is an Associate Professor at ENSICAEN, France. She obtained her Ph.D. degree from the Collegium ingénieur de l'Université de Lorraine in 2006. She works at the GREYC Laboratory where she is a permanent member of the research group in E-payment & Biometrics. Her research interests include biometrics, signal processing and chaos system.

Christophe Rosenberger is a Full Professor at ENSICAEN, France. He obtained his Ph.D. degree from the University of Rennes I in 1999. He works at the GREYC Laboratory where he leads the research group in E-payment & Biometrics. His research interests include biometrics (definition of biometric systems and privacy issues). He is involved in developing authentication solutions for e-transactions applications.

Patrick Bours studied mathematics at the Eindhoven University of Technology in the Netherlands. He got his M.Sc. and Ph.D. with a specialisation in coding theory. After his studies, he worked at the Netherlands National Communication Security Agency (NLNCSA) in the area of cryptology. In July 2005, he moved to Norway where he joined the Norwegian Information Security Laboratory (NISlab), which is a part of Gjøvik University College. Since September 2009, he holds a Full Professor position. He specialised in behavioural biometrics. His current research focus is on gait recognition, and static and continuous keystroke dynamics.