Open access • Journal Article • DOI:10.1504/IJITM.2012.044062

A new soft biometric approach for keystroke dynamics based on gender recognition

Romain Giot, Christophe Rosenberger

Institutions: University of Caen Lower Normandy

Published on: 01 Dec 2012 - International Journal of Information Technology and Management (Inderscience Publishers Ltd)

Topics: Keystroke dynamics, Keystroke logging and Soft biometrics

A New Soft Biometric Approach For Keystroke Dynamics Based On Gender Recognition

Romain Giot, Christophe Rosenberger

A new soft biometric approach for keystroke dynamics based on gender recognition

Romain Giot* and Christophe Rosenberger

GREYC Research Lab,
ENSICAEN – Université de Caen Basse Normandie – CNRS,
14000 Caen, France
Fax: +33-231538110
E-mail: romain.giot@ensicaen.fr
E-mail: christophe.rosenberger@ensicaen.fr
*Corresponding author

Abstract: Keystroke dynamics allows to authenticate individuals through their way of typing on a computer keyboard. In this study, we are interested in static shared secret keystroke dynamics (all the users type the same password). We present new soft biometrics information which can be extracted from keystroke typing patterns: the gender of the user. This is the first study, to our knowledge, experimenting such kind of information in the field of keystroke dynamics. We present a method for gender recognition through keystroke dynamics with more than 91% of accuracy, on the tested dataset, and we show the improvement on keystroke dynamics authentication method using such kind of information through pattern and score fusion. We obtain a gain of 20% when using gender information against a classical keystroke dynamics method.

Keywords: gender recognition; soft biometrics; keystroke dynamics.

Reference to this paper should be made as follows: Giot, R. and Rosenberger, C. (2012) 'A new soft biometric approach for keystroke dynamics based on gender recognition', *Int. J. Information Technology and Management*, Vol. 11, Nos. 1/2, pp.35–49.

Biographical notes: Romain Giot obtained his Master of Science in 2008 from ENSICAEN. He has been a Research Engineer in the GREYC laboratory for two years. His research interests include biometrics, especially the definition of keystroke dynamics biometric and multibiometrics systems. He is now a PhD student of the University of Caen and works on template update systems for biometric systems.

Christophe Rosenberger is a Full Professor at ENSICAEN, France. He obtained his Master of Science in 1996 and his PhD in 1999 from the University of Rennes I. He works at the GREYC laboratory. His research interests include computer security and biometrics. He is particularly interested in authentication methods for e-transactions applications.

1 Introduction

Biometric systems (Jain et al., 2008) are not 100% accurate, there are always some recognition errors due to imprecisions in the biometric method, various problems during the acquisition or intra-class variability. Performance of these biometric systems can be improved by using various ancillary information. This information can be of the following types: gender, age, height, weight or eye colour (for example) and are collected during the enrolment. It has been shown that such type of information can improve significantly the recognition performance of biometric systems (Jain et al., 2004). Although such kind of information is not sufficient to establish the identity of an individual, it can be used in association of a biometric system in order to strengthen it. Such information is named as *soft biometrics* which can be defined as "*traits as characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals*" (Jain et al., 2004).

We are interested in this paper in a new soft biometric with information that can be used in static keystroke dynamics authentication system (Monrose and Rubin, 1997; Hocquet et al., 2007; Giot et al., 2009b): gender recognition through the typing dynamics. This is the first time, to our knowledge, that gender recognition with keystroke dynamics is experimented (for the moment, this information has only been extracted from facial images or gait (Li et al., 2008). This opens a new path in the soft biometrics field of research. As everybody already uses a keyboard on its computer, keystroke dynamics data are really easy to capture (and the capture is really well accepted by the users). Finding the gender of the user through its typing dynamics has several advantages:

- As every soft biometrics, it is a potential additional parameter allowing to reduce recognition error rates, without having to capture or measure any additional information (the gender recognition is done with the sample used for the verification). It could be used in multi-modal fusion approach.
- It can be used alone in order to verify if the claimed gender of the individual matches the reality (which can be interesting in the web context where users can lie easily).

Our main contributions in this work are:

- We propose a novel framework allowing to guess the gender of an individual through its way of typing a predefined text on a keyboard.
- We show promising results for gender recognition using keystroke dynamics.
- We show promising results for keystroke dynamics authentication using gender information in addition of the keystroke patterns.

This paper is organised as follows. We begin by reviewing related works on soft biometrics in Section 2. We present the proposed method on gender recognition through keystroke dynamics and keystroke dynamics authentication using gender information in Section 3. Section 4 presents the experimental results we obtained. The last section presents the limitations of the proposed approach and its conclusion.

2 Related works

Even if it is an interesting methodology, there are not many works on soft biometrics. This section presents some of the studies which can be found in the state of the art. Such ancillary information can be measured through a defined sensor, guessed by a pattern recognition mechanism or specified by an operator; we present these different types in the state of the art.

Jain et al. (2004) presented a generic framework allowing to use a soft biometric information in a recognition system thanks to Bayes probabilities. They obtained an improvement of 5% for the fingerprint recognition system when using information about ethnicity (Asian and non-Asian), gender and height (artificially generated through a Gaussian distribution). The main problem of this approach is the need of a specific materiel to measure the height of the individuals during the authentication process (problem avoided during the experiment by using automatically computed heights).

Ailisto et al. (2006) used the body weight measurement and body fat percentage as additional information for a fingerprint recognition system. They decreased the total error rate from 3.9% to 1.5%. The body weight score consisted in the absolute difference between the reference weight and the measured one which was then normalised (the same procedure is done for the body fat). Their experiment showed that the body weight can be discriminative and used as a soft biometrics (used alone, the total error rate is of 11%), but it is not the case for the body fat percentage. Once again, the choice of the selected soft biometric could be problematic because of the necessity of using a weight-scale which is not necessarily ergonomic.

In the following papers, the information is guessed by a machine learning process, instead of being measured through an adequate sensor. More recently, Jain and Park (2009) used facial marks (freckles, moles and scares) as soft biometrics. The aim of this information is multiple: to add features to the face recognition system, to enable fast search in facial images databases and to facilitate the matching from a partial profile with marks. Rank-1 identification accuracy has been improved for two significant benchmarks.

Soft biometrics can be used to localise people in surveillance databases (Vaquero et al., 2009). In this context, authors are not interested in face recognition, but they want to find people depending on a search query (i.e., show me the bald individuals who entered a given building last Saturday wearing a red shirt and sunglasses.). They used information as visual attributes such as facial hair type, type of eye-wear, bear hair type and clothing colour, some of them can be classified as soft biometrics. One application of such a system would be criminal investigation in order to find a suspect in a scene.

Gender can also be guessed through facial images (Alexandre, 2010). In this paper, the author presented a framework using fusion of different classifiers using shape, texture and colour information of face images at different scales. The framework has been validated on two large datasets and its accuracy was about 90%.

Hwang et al. (2009) proposed a face recognition system using gender information in order to improve performances. In their work, they assumed that the gender recognition system works correctly without any error. They constructed three face models, using LDA (Belhumeur et al., 1997), for the males, the females and both. The recognition process used information from the global model and the gender specific one. The paper presented the fact that female recognition is more difficult than male recognition and

using the gender information improves the verification rate. They validated their work on the FRGC database.

Instead of using a holistic approach to guess the gender of an individual through a facial image, Li et al. (2009) used a patched-based approach by presenting Spatial Gaussian mixture models where information is considered locally and globally. This approach gives better performances (improvement between 40% and 50%) than GMM and SVM-based approaches on the YGA face database.

These previous experiments showed that using soft biometrics can improve the accuracy of biometric systems. The new information can be measured with an appropriate sensor or extracted from various patterns. Most of papers in the state of the art realise gender recognition through facial information. This solution cannot be used in all contexts (computer without a webcam, inappropriate environment…). The aim of this paper is to present a way of extracting gender information through keystroke dynamics samples, and the benefit of using this information to improve the accuracy of a keystroke dynamics authentication system. The next section presents our proposal.

3 Proposed method

The first objective of this paper is to determine the gender of an individual through its way of typing a predefined text. We present in this section the method we use.

3.1 Gender recognition

The first step concerns the feature selection of keystroke patterns useful for gender recognition.

3.1.1 Pattern selection

We use the public GREYC keystroke benchmark database (Giot et al., 2009a) for this work. It is one of the largest databases (in term of number of users and sessions) in keystroke dynamics. To our knowledge, no existing database contains more individuals. In order to reduce the bias due to this high quantity of male information, we only kept the first n male samples (where *n* is the number of female samples). We directly used the extracted features already computed for this benchmark. So, five different features are extracted from each typing sample:

- the *RR latencie*s which are the latencies between each key release
- the *RP latencies* which are the latencies between the release of one key and the pressure of the next one
- the *PP latencies* which are the latencies between each key pressure
- the *PR duration* which is the time of pressure of one key
- the *vector V* which is the concatenation of the four previous timing values (this can also be seen as a template fusion with no reduction dimension).

As these extracted features are already present in the database, we directly use them without computing them again. For each sample, we have at our disposal five different

patterns or timing vectors. These vectors are the different patterns we use all along the study in order to study which one is better than the others for gender recognition.

3.1.2 Machine learning

We use as learning and recognition approach a support vector machine (SVM) (Vapnik, 1998) for the classification of male and female patterns. An SVM has been learnt for each type of pattern (RR, RP, PP, PR, V) in order to compare the performances depending on the selected extracted features.

The aim of an SVM is to provide the best linear separator (when using a linear kernel) between two classes by using a system of margin maximisation. As the resolution of such problem uses a scalar product, it is possible to replace it by a kernel function and then obtain a non-linear separation through a transparent changing of feature space (Muller et al., 2001). Several kernels exist and we have chosen to use the radial basis kernel because it is known to work well most of the time. Equation (1) presents the radial basis function.

$$k\left(x_i, x_j\right) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{1}$$

When using an SVM with a radial basis kernel, we need to set two parameters: γ which is the parameter of the kernel and C which is the penalisation coefficient of the SVM. The best SVM classifier is then the one parametrised by the couple (γ, C) having the smallest classification error rate. We have to search which is the best couple (γ, C) for each type of pattern.

Thanks to the python script provided by the libsvm library (Chang and Lin, 2001) library, such a task is easily automated. We launched this script for the five data files (each one representing a different type of pattern). This script works as following (it is an exhaustive search for all the parameters):

- Each data file is normalised in order to have input values in the range [–1; 1].
- Several sets of (γ, C) are tested. For each couple, a five cross validation computation is operated: the data set is splitted in five parts and each part is used as testing set while the other four parts are used as a training set.
- The couple (γ, C) giving the best recognition rate is selected.

Once gender recognition is achieved, we use this information as soft biometrics to improve the performance of keystroke dynamics algorithms.

3.1.3 Gender score

Instead of using the predicted label ({–1, 1}), we compute a score. We compute this score by using the guessed label and its probability. In this method, we obtain a score in the range [0; 1]. Scores closer to 0 represent a male, while scores closer to 1 represent a female. Equation (2) presents the way of computing this score, where predict presents the predicted label and probability its probability:

$$genderscore = \frac{1 - predict * probability}{2} \tag{2}$$

3.2 Keystroke dynamics with gender recognition

As we are able to recognise the gender of users, it is interesting to see if we can use this information in the keystroke dynamics recognition system, in order to improve its performance. We present in the next section the keystroke dynamics recognition we use in this paper.

3.2.1 Keystroke dynamics recognition

In this paper, we used the statistical keystroke distance computation function presented in Hocquet et al. (2007). We used 20 training samples to create the template (which is simply the mean and standard deviation of the enrolled samples) of each user and ten samples for the verification. Equation (3) presents the distance computation method, with v the sample to test against the model, the mean of the enrolled samples and its standard deviation. *Card* (.) is the cardinal of the input sample.

$$biometricscore = 1 - \frac{1}{Card(v)} \sum_{i=0}^{i=Card(v)} \exp\left(\frac{|v_i - \mu_i|}{\theta_i} \right) \quad (3)$$

The resulting score is between 0 and 1 and smaller is better. To define a soft biometrics solution, we need to fuse information extracted from the keystroke dynamics pattern and the gender score. There exist multiple approaches for that in the state of the art.

3.2.2 Information fusion

In this paper, we have tested two different ways of using this new information:

- By doing a *template fusion*: we add to each keystroke sample an additional feature which is the computed gender score (or 0 for a male and 1 for a female when it is manually set). We apply the keystroke recognition on the new template.
- By doing a *score fusion*: for each test sample, we first compute the keystroke recognition score. We have *biometricscore* which represents the distance between the sample and the model, and *genderscore* which represents the guessed gender (or 0 and 1 when manually set). With this information, we can compute a final score as presented in equation (4), where $\mu label$ is the mean value of the gender information of the model and w a weight on the gender recognition result.

$$decisionscore = biometricscore + w * abs\left(\mu_{label} - genderscore\right) \quad (4)$$

In this paper, we have empirically set w to 0.25 in order to give more weight to the keystroke recognition.

In order to test the benefit of this new soft biometrics approach, we identified multiple scenarios.

3.2.3 Tested scenarios

We have tested four different scenarios on the generation of gender information:

- *No gender information:* this is the classical keystroke dynamics authentication scheme, when we only use the comparison score between the reference and a sample.

- *Manual labelling:* each sample is manually labelled depending on the real gender of the user. This can be seen has a supervised approach where an operator informs the keystroke recognition system of the gender of the individual.
- *Automatic labelling:* each sample is automatically labelled by guessing the gender score of each sample (using the previously defined gender recognition method). This is an unsupervised approach, where everything is automatic. Some errors can happened when labelling training and testing samples.
- *Semi-automatic labelling:* the samples used to create the template are manually labelled while the samples used for the validation are automatically labelled. This is a *semi-supervised* approach, where an operator manually labels the enrolment samples while gender information is guessed for the validation samples.

In order to not have samples from the same user both in the training and in the validation sets, for the gender recognition system used during the keystroke dynamics authentication, we have splitted the users in two sets of same size: one SVM is trained for each set and is used to guess the gender of the samples of the other set. By this way, we avoid to recognise user instead of gender.

4 Experimental results

We present in this section some experimental results. First, we show the protocol we use in this paper. Second, the efficiency of gender recognition based on keystroke dynamics is given. Last, the benefit of using this additional information of keystroke dynamics verification is illustrated.

Table 1 Summary of the information provided in the database used in the experiment

Information	*Description*
Users	133 users
Database sample size	7,555 passphrases (60 samples for 100 users)
Data sample length	16 characters ('GREYC laboratory')
Typing error	Not allowed
Controlled acquisition	Yes
Age range	Between 19 and 56
Gender	98 males and 35 females
PC usage frequency	Unknown
User profession	Students, researchers, secretaries, labourers (unknown repartition)
PC usage frequency	Unknown
Keyboard	2 AZERTY keyboards (1 laptop, 1 USB)
Acquisition platform	Windows XP/GREYC keystroke software

4.1 Used benchmark

For this experiment, we have used the benchmark dataset presented in Giot et al. (2009a). A total of 133 users participated to the experiment during several sessions spaced of one

week. During each session, individuals were asked to type six times the password ('GREYC laboratory') on the laptop keyboard and six times on an USB keyboard (by interleaving the input from one keyboard to another one). The individuals did not participated to the same number of sessions, but 100 of them participated to five sessions and provided 60 patterns. The total number of collected patterns is 7,555.

Table 1 presents an overview of the keystroke dynamics database. 35 females participated to the study with 98 males. There are almost three times more males than females which could be a problem during the learning that is why we did not use the whole dataset during the experiments.

In the next section, we quantify the efficiency of gender recognition based on keystroke dynamics.

4.2 Gender recognition accuracy

In this section, we present the results of the gender recognition method. We use all the female patterns and the same number of male ones. Table 2 presents the recognition accuracy for the best configuration of (γ, C) on each kind of pattern with a five folds cross validation. The worst performance we obtain is about 87.31% while the best is about 91.63%. We can say that gender recognition through keystroke dynamics is possible and functional. But, with a five folds cross validation, their is a huge number of patterns used to compute the model (4/5 of the data for the five models), which can be problematic because of the high quantity of data required.

Table 2 Accuracy of gender recognition when using five folds for the cross validation

Extracted features	*RR*	*PP*	*RP*	*PR*	*V*
Accuracy	88.57%	87.31%	88.01%	87.8%	91.63%

Figure 1 Accuracy depending on the quantity of data used to learn the parameters of the model

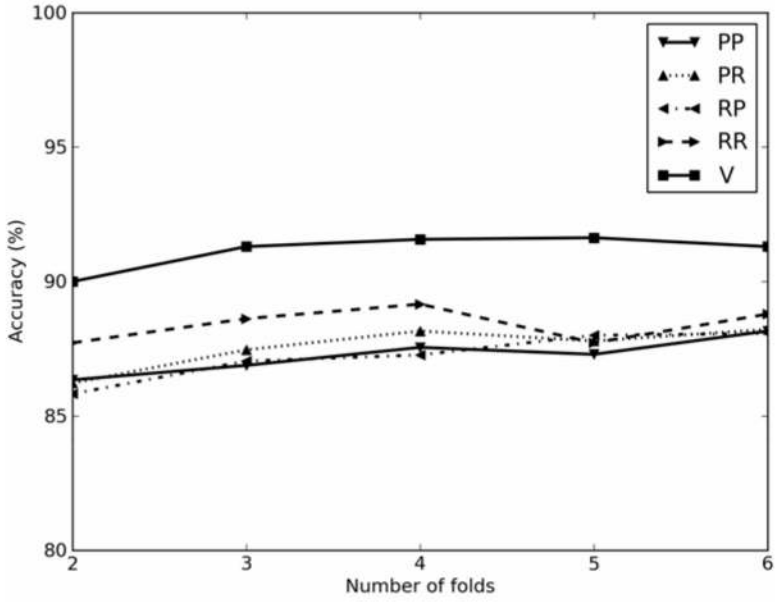


Figure 1 presents the accuracy of the SVM model depending on the chosen extracted feature for different folds in the cross validation. The aim of this figure is to show if results decrease a lot when using less information to create the model.

We can observe that, whatever the chosen number of folds, best performance is always get with the *V* pattern and even if there is a link between quantity to learn the model and accuracy of the gender recognition system, performance difference is not really important.

4.3 Performance of keystroke dynamics authentication using gender recognition

In this section, we present the accuracy improvement of keystroke dynamics recognition systems using the gender information. Considering previous results, we only used the V pattern for this experiment.

We have tested the method proposed in the previous section and another one using the guessed label ({–1;1}) instead of the computed score. In this case, for the score fusion, we use a logical score fusion. When the guessed gender label belongs to the labels of enrolled samples, the final score is the keystroke authentication score, otherwise, it is the maximum distance value of the keystroke authentication method (to simulate an impostor pattern).

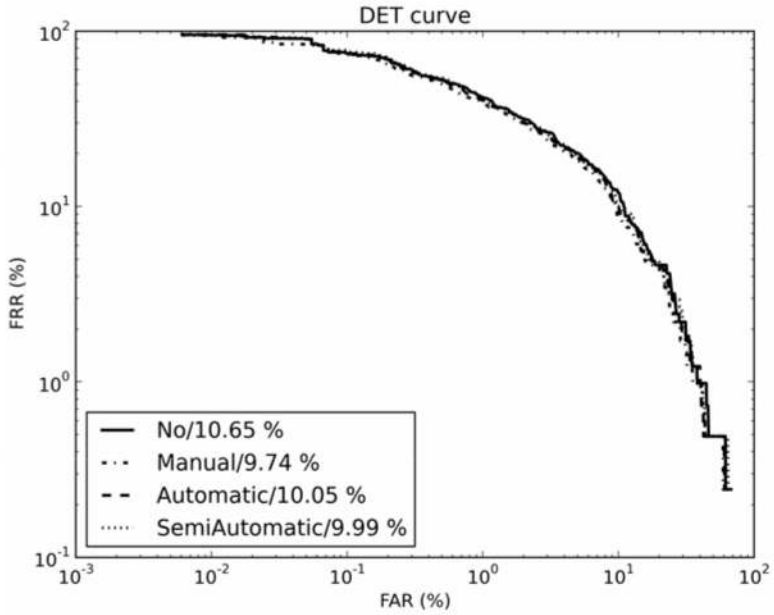
We used a subset of the database having an almost equal repartition of males and females having 30 samples. After removing invalid users (which do not have the right number of samples, or incorrect patterns) we have kept 41 users. For a given method, the quantity of computed intra scores is 410 and the quantity of computed inter scores is 16,400.)

Figure 2 presents the ROC curves of the different variations of the keystroke dynamics authentication method (no use of gender information, manual labelling, automatic labelling and semi-automatic labelling) when using only label –1 and 1 to represent the gender information, while Figure 3 presents the same results when using a score between 0 and 1 to represent the gender information. Table 3 presents the EER value for each variation and the gain of the best method using gender information against the method using no gender information.

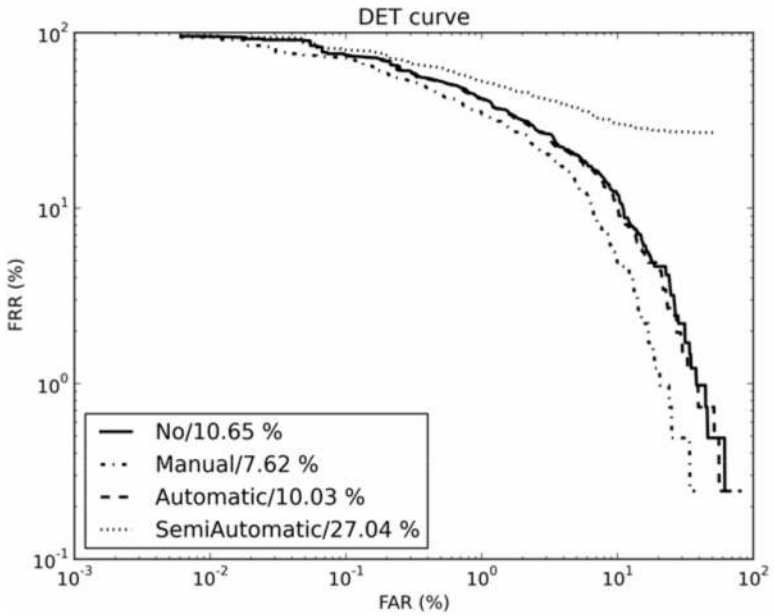
We can see that using the gender information does not always improve the keystroke recognition performance. The methods using a score representing the gender instead of a label seem to be more accurate (which can be easily understood because of the integration of the accuracy of the gender recognition). Most of the time, in the manual or automatic scheme, we can improve the keystroke recognition performance, but (as it was expected), it is not the case for the semi-automatic version. The manual labelling gives better performances than the automatic one, whereas we were expecting the opposite (by expecting the gender recognition system to always do the same error on the different samples of the same individual), but this difference is really light.

Figure 4 presents the score distribution when using a score fusion for the two different schemes. Looking at Figure 4(f), we can understand why the semi-automatic gives such bad results: a lot of genuine samples are wrongly labelled with the automatic procedure (for the test) whereas they were correctly labelled for the manual procedure (for enrolment). With the automatic procedure, these errors are smoothed because the same error can also occur during enrolment and verification.

Figure 2 Global performance of keystroke dynamics recognition when using gender information as a label ({–1, 1}), (a) template fusion (b) score fusion

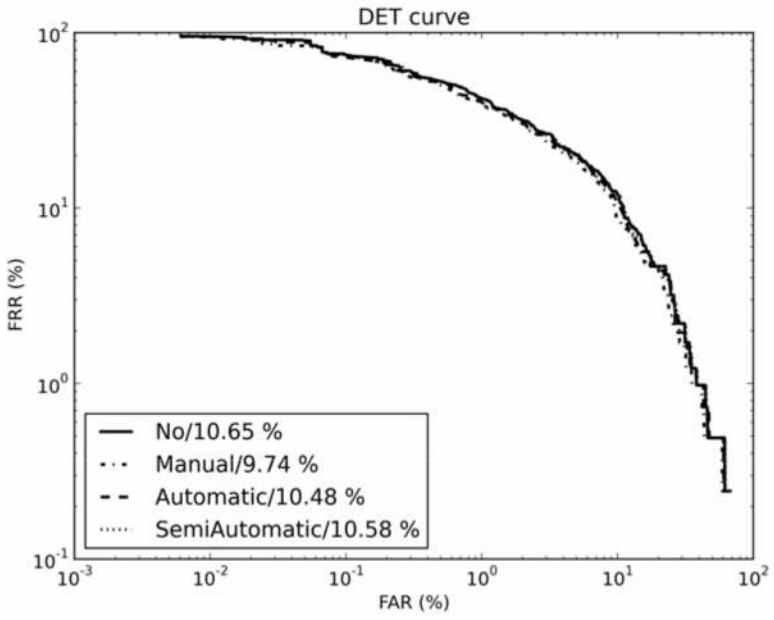


(a)

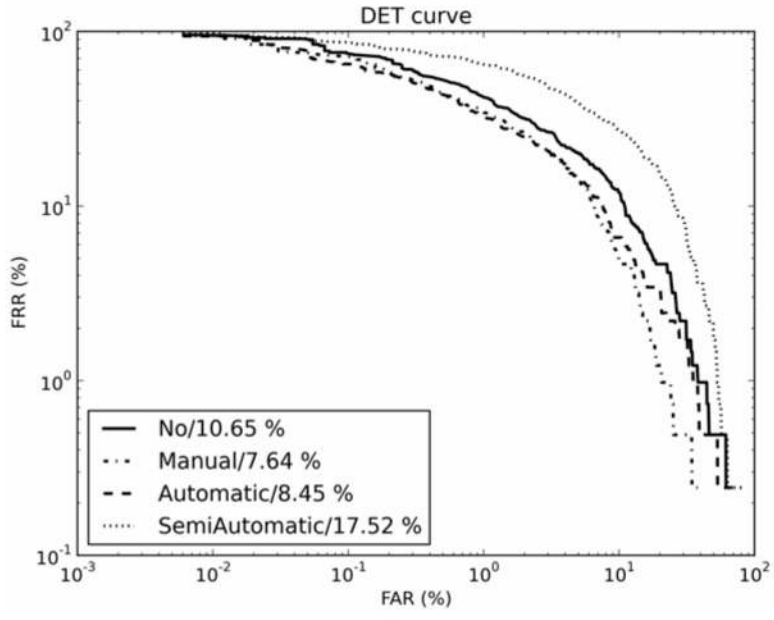


(b)

Figure 3 Global performance of keystroke dynamics recognition when using gender information as a score ([0; 1]), (a) template fusion (b) score fusion



(a)



(b)

Table 3 EER value of the keystroke dynamics authentication method when using gender information in the pattern

Method	*No*	*Manual*	*Automatic*	*Semi-automatic*
EER (label/template)	10.65%	*9.74%*	*10.05%*	*9.99%*
EER (label/score)	10.65%	10.65%	10.74%	27.64%
EER (score/template)	10.65%	*9.74%*	*10.48%*	*10.58%*
EER (score/score)	10.65%	*7.64%*	*8.45%*	17.52%
Best	10.65%	*7.64%*	*8.45%*	*9.99%*
Gain	x	28.26%	20.67%	6.20%

Notes: EER (a/b) represents the EER value where a is the representation of the gender and b is the kind of fusion. The best line represents the best value for each type of gender recognition, and the gain line represents the gain against the method without any gender information. Italic values are results better than no gender use and the underline values are the best scenario for each scheme.

Figure 4 Score distribution when using the score fusion of keystroke authentication result and gender authentication result (score and label), (a) manual score (b) manual label (c) automatic score (d) automatic label (e) semi-automatic score (f) semi-automatic label (g) no gender

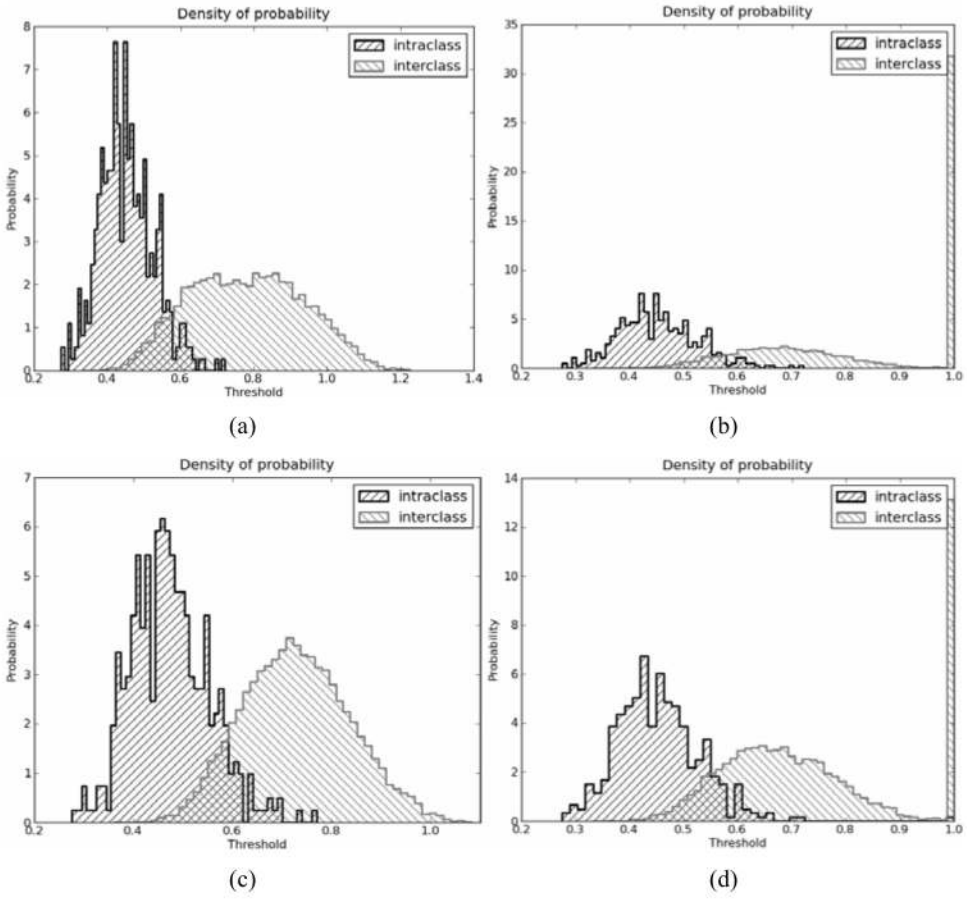
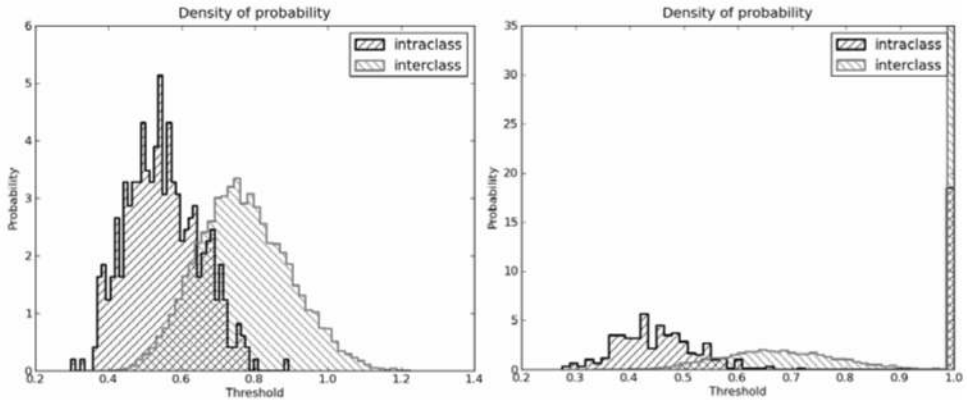
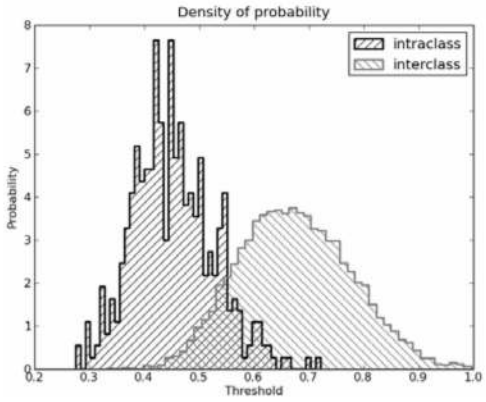


Figure 4 Score distribution when using the score fusion of keystroke authentication result and gender authentication result (score and label), (a) manual score (b) manual label (c) automatic score (d) automatic label (e) semi-automatic score (f) semi-automatic label (g) no gender (continued)



(e)

(f)



(g)

Briefly stated the manual labelling (using the score fusion) gives the best result but cannot be used in a real context. The automatic approach provides a less good result (but not far from the manual one) but can completely be used in an operational context in a static shared password keystroke dynamics context. The benefit of this last approach compared to the classic one is about 20%.

5 Perspectives and conclusions

In this paper, we have presented, for the first time in the keystroke dynamics research field, a study on the gender recognition through the keystroke dynamics of an individual when typing a predefined text. The next step is to be able to get this information with free text which will grow the quantity of potential use of such a mechanism.

The quantity of information available on the database is not really huge (although most of keystroke dynamics studies use even less users and captures in their studies)

which gives results slightly better than it would be with a bigger dataset. But creating such a database stays quite difficult because of the time it takes (look at the quantity of public keystroke databases available).

The protocol could be improved: shuffling data before doing the cross validation, using different male patterns and averaging the results, being sure to not have a user with patterns in training set and other patterns in validation site (to not learn user instead of gender in the first experiment).

In a non-supervised scenario, the best results are obtained when using a template fusion of all the extracted features (keystroke typing timings and gender score). There is less than 9% of recognition error. In a supervised scenario, we obtain the best results with the score fusion procedure (fusion of the keystroke recognition score and the gender recognition score). Using gender recognition score instead of gender label gives better results.

Now, that we know we are able to discriminate males from females in a keystroke dynamics system, it is possible to use this information as a soft biometric in order to reduce the error rate of classical keystroke dynamics methods. It is also time to check this kind of information regardless of the used password. Using such information would be useful for automatically personalising web applications depending on the gender of the user, fighting against paedophilia or social network monitoring.

We showed that using this information as ancillary information for keystroke dynamics authentication, we improve the EER value of more than 20%.

Acknowledgements

The authors would like to thank the Société Française des Statistiques and the Région Basse-Normandie for their financial support.

References

Ailisto, H., Vildjiounaite, E., Lindholm, M., Mäkelä, S-M. and Peltola, J. (2006) 'Soft biometrics – combining body weight and fat measurements with fingerprint biometrics', *Pattern Recognition Letters*, Vol. 27, No. 5, pp.325–334.

Alexandre, L.A. (2010) 'Gender recognition: a multiscale decision fusion approach', *Pattern Recognition Letters*, February.

Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) 'Eigenface vs. fisherfaces: recognition using class specific linear projection', *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.711–720.

Chang, C.C. and Lin, C-J. (2001) *Libsvm: A Library for Support Vector Machines*.

Giot, R., El-Abed, M. and Christophe, R. (2009a) 'GREYC keystroke: a benchmark for keystroke dynamics biometric systems', *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, September, pp.1–6, District of Columbia, IEEE Computer Society, doi: 10.1109/BTAS.2009.5339051, Washington, USA.

Giot, R., El-Abed, M. and Christophe, R. (2009b) 'Keystroke dynamics with low constraints SVM based passphrase enrollment', *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, September, District of Columbia, IEEE Computer Society. doi: 10.1109/BTAS.2009.5339028, Washington, USA.

Hocquet, S., Ramel, J-Y. and Cardot, H. (2007) 'User classification for keystroke dynamics authentication', *The Sixth International Conference on Biometrics (ICB2007)*, pp.531–539.

Hwang, W., Ren, H., Kim, H., Kee, S-C. and Kim, J. (2009) 'Face recognition using gender information', *IEEE 16th International Conference on Image Processing (ICIP 2009)*, November, IEEE Signal Society, Cairo, Egypt.

Jain, A.K. and Park, U. (2009) 'Facial marks: soft biometric for face recognition', *IEEE International Conference on Image Processing (ICIP)*.

Jain, A.K., Dass, S.C. and Nandakumar, K. (2004) 'Soft biometric traits for personal recognition systems', *Proceedings of International Conference on Biometric Authentication*.

Jain, A.K., Flynn, P. and Ross, A.A. (Eds.) (2008) *Handbook of Biometrics*, Springer US.

Li, X., Maybank, S.J., Yan, S., Tao, D. and Xu, D. (2008) 'Gait components and their application to gender recognition', *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews*, Vol. 38, No. 2, p.145.

Li, Z., Zhou, X. and Huang, T.S. (2009) 'Spatial Gaussian mixture model for gender recognition', *IEEE 16th International Conference on Image Processing (ICIP 2009)*, November, IEEE Signal Society, Cairo, Egypt.

Monrose, F. and Rubin (1997) 'Authentication via keystroke dynamics', *Proceedings of the 4th ACM Conference on Computer and Communications Security*, pp.48–56, ACM Press New York, NY, USA,.

Muller, K.R., Mika, S., Ratsch, G., Tsuda, K. and Scholkopf, B. (2001) 'An introduction to kernel-based learning algorithms', *IEEE Transactions on Neural Networks*, Vol. 12, No. 2, pp.181–201.

Vapnik, V. (1998) *Statistical Learning Theory*, Wiley New York.

Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A. and Turk, M. (2009) 'Attribute-based people search in surveillance environments', *IEEE Workshop on Applications of Computer Vision (WACV)*.