Fusion et biométrie douce pour la dynamique de frappe au clavier

S Z Syed Idrus, Estelle Cherrier, Christophe Rosenberger

▶ **To cite this version:**

S Z Syed Idrus, Estelle Cherrier, Christophe Rosenberger. Fusion et biométrie douce pour la dynamique de frappe au clavier. colloque COmpression et REprésentation des Signaux Audiovisuels (CORESA), May 2016, Nancy, France. hal-01406711

HAL Id: hal-01406711
https://hal.science/hal-01406711

Submitted on 1 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fusion et biométrie douce pour la dynamique de frappe au clavier

S.Z. Syed Idrus[1] E. Cherrier[2] C. Rosenberger[2]

[1] University Malaisia Perlis, Malaisie
[2] Normandie Univ, France
UNICAEN, GREYC, F-14032 Caen, France
ENSICAEN, GREYC, F-14032 Caen, France
CNRS, UMR 6072, F-14032 Caen, France

syzul@unimap.edu.my, {estelle.cherrier, christophe.rosenberger}@ensicaen.fr

Résumé

La dynamique de frappe au clavier (DDF) est une modalité biométrique comportementale. Moins invasive que les empreintes digitales, cette modalité présente néanmoins de moins bonnes performances que les systèmes biométriques morphologiques. Plusieurs pistes peuvent être envisagées pour augmenter les performances de la DDF : la multibiométrie (la DDF est combinée avec d'autres modalités), l'évaluation de la qualité des données biométriques ou encore la biométrie douce. Cet article présente une approche originale de biométrie douce pour la dynamique de frappe au clavier. Le système proposé est capable d'établir un profil de l'utilisateur, plus précisément de reconnaître : le nombre de mains utilisées pour taper au clavier, le sexe, la catégorie d'âge de l'utilisateur et sa latéralité (droitier/gaucher). Les taux de reconnaissance obtenus sont satisfaisants, et sont améliorés par des processus de fusion des données. Le système est évalué sur une base de données spécialement créée pour cette étude, à la fois en France, et en Norvège.

Mots clefs

Biométrie douce, fusion de données, SVM, profilage

1 Introduction

Un système biométrique est un système d'authentification d'un individu, à l'aide de caractéristiques physiques (visage, empreintes digitales, iris…), physiologiques (ADN, odeur…) ou comportementales (dynamique de frappe au clavier, dynamique de signature, démarche…). La biométrie utilise des techniques de traitement du signal et de reconnaissance de formes. En ce qui concerne le traitement du signal, on peut mentionner : la capture de la donnée (par un appareil-photo, un capteur dédié), le pré-traitement de la donnée (échantillonnage, filtrage pour atténuer le bruit), l'extraction des caractéristiques (temporelles ou spectrales, par exemple dans le cas de la reconnaissance du locuteur). Les techniques de reconnaissance de formes sont largement utilisées pour classifier les données, créer un modèle de l'utilisateur, comparer les données stockées et les données présentées au capteur. Aujourd'hui, il existe de multiples usages des systèmes biométriques : contrôle d'accès physique, contrôle de présence, paiement électronique, etc…Cet article s'intéresse au profilage d'individus à partir de biométrie douce pour la dynamique de frappe au clavier. La notion de *biométrie douce* a été introduite par Jain et al. [1]. Les auteurs définissent les *traits de biométrie douce* de la manière suivante : *un caractère fournissant de l'information sur l'individu, mais manquant d'unicité et de permanence pour différencier suffisamment deux individus*. Ce sont donc des caractéristiques qui ne sont pas suffisantes pour authentifier un individu, mais peuvent aider à la construction d'un profil. L'article [2] cite en example : la couleur de la peau, la couleur des cheveux, la couleur des yeux, la présence d'une barbe, plus généralement des informations extraites à partir du visage, mais aussi le sexe, l'âge, la taille, le poids, la démarche, différentes mesures du corps. De façon naturelle, la biométrie douce permet une recherche plus rapide dans une base de données (par élimination de données non conforme au profil) ainsi qu'une amélioration des performances. La biométrie douce est également considérée comme non invasive (par rapport à des modalités comme les empreintes digitales ou le visage), sans risque d'usurpation d'identité, avec une mise en oeuvre à faible coût.

La dynamique de frappe au clavier est une modalité biométrique qui consiste à mesurer les rythmes qui se dégagent lorsqu'on tape sur un clavier d'ordinateur [3], [4], [5]. En ce sens, c'est une modalité biométrique *comportementale*, de même que la dynamique de signature, la démarche ou la voix. Parmi les avantages de la dynamique de frappe au clavier par rapport à d'autres modalités, nous pouvons mentionner son faible coût et sa facilité d'usage : en effet, en dehors d'un clavier, aucun capteur ni dispositif supplémentaire n'est nécessaire et les utilisateurs sont habitués à taper un mot de passe. En contrepartie, la dynamique de frappe présente de plus faibles performances que les autres modalités biométriques comme les empreintes digitales, le visage, l'iris. Cela peut s'expliquer par une variabilité intra-

classe élevée. Une façon de gérer cette variabilité est de prendre en compte des informations supplémentaires dans le processus de décision. Cela peut être fait de différentes manières : (i) en combinant la dynamique de frappe au clavier avec une autre modalité biométrique (multibiométrie) ; (ii) en optimisant l'étape d'enrôlement (une donnée biométrique est exploitée pour la génération de la référence seulement si le niveau de qualité est suffisant) ; ou (iii) en combinaison avec des caractéristiques de biométrie douce.

Dans cet article, nous considérons les caractéristiques de biométrie douce suivantes pour la dynamique de frappe (en abrégé DDF) : le nombre de mains (l'utilisateur peut taper au clavier avec une ou deux main(s)), le sexe, l'âge (plus ou moins de 30 ans), la latéralité (droitier ou gaucher). Pour réaliser cette étude, nous avons créé une nouvelle base de données. Deux cas sont considérés : des mots de passe statique et du texte libre. En utilisant des techniques d'apprentissage statistique et des processus de fusion des données, les résultats obtenus sont encourageants.

L'article est organisé comme suit. La partie 2 présente la méthodologie adoptée pour acquérir les données, pour constituer une nouvelle base de données et les méthodes d'apprentissage et de fusion retenues. Les résultats obtenus font l'objet de la partie 3. La partie 4 conclut cet article et présente quelques perspectives.

2 Méthodologie

Cette partie présente la méthodologie adoptée pour collecter la base de données biométrique et pour traiter les données.

2.1 Capture des données de DDF

L'authentification par DDF ne requiert généralement qu'un clavier d'ordinateur et une application dédiée pour récupérer les actions sur ce clavier (appui ou relâchement de touches). Chaque capture est stockée dans une base de données dédiée et comporte les informations suivantes :

- touche concernée
- type d'événement (appui ou relâchement)
- temps d'exécution de l'événement

A partir de ces données brutes, on extrait le modèle de l'utilisateur constitué de différentes données temporelles, comme illustré à la figure 1 :

- pp : intervalle de temps entre deux pressions successives
- rr : intervalle de temps entre deux relâchements successifs
- rp : intervalle de temps entre l'appui sur une touche et le relâchement de la suivante
- pr : intervalle de temps entre un relâchement et l'appui sur la touche suivante

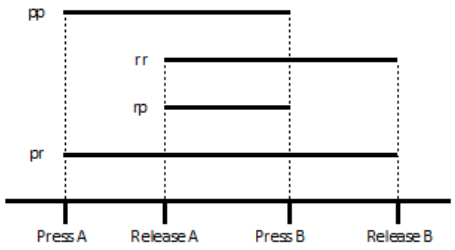

Figure 1 – *Caractéristiques extraites de la DDF*

Ces données sont concaténées dans un vecteur V.

2.2 Description de la base de données

Une base de données conséquente (110 utilisateurs) a été créée pour étudier la biométrie douce pour la dynamique de frappe au clavier. La collecte des données a eu lieu à la fois en France et en Norvège, auprès de volontaires (étudiants ou enseignants, chercheurs, personnels de laboratoires de recherche, grand public...), provenant de 24 pays différents. Au total, 70 français et 40 norvégiens ont participé à l'expérimentation. En plus des données de DDF, nous avons des informations sur les volontaires : leur sexe, leur âge, leur latéralité. . .Plus de détails sur cette base peuvent être trouvés dans l'article [6]. On peut mentionner d'autres études du même genre, comme les références [7], [8]. La figure 2 donne quelques statistiques sur le profil des utilisateurs.

Participant	70 (France) & 40 (Norway)
Gender	78 men & 32 women
Age Range	< 30 & ≥ 30 years old
Handedness	98 right-handed (70 men / 28 women); 12 left-handed (8 men / 4 women)

Figure 2 – *Statistiques sur le profil des utilisateurs*

D'après les travaux de Giot *et al.* [9], les meilleures performances pour l'authentification à base de DDF sont obtenues avec des mots de passe connus. Puisque notre expérimentation se passe en France et en Norvège, nous avons choisi des mots de passe connus dans les deux pays, notés $P1$ à $P5$ (Password 1 à Password 5) : *leonardo di caprio*, *the rolling stones*, *michael schumacher*, *red hot chili peppers* et *united states of america*. Ces mots de passe comportent entre 17 et 24 caractères (espaces inclus). Les volontaires doivent taper chaque mot de passe 10 fois, successivement avec une main, puis avec deux mains. Pour capturer les données, nous avons utilisé le logiciel développé au Laboratoire GREYC, que l'on peut librement télécharger à l'adresse suivante : `http://www.ecole.ensicaen.fr/~rosenber/keystroke.html`. Ce

logiciel est décrit plus en détails dans l'article [5]. La base de données collectées comporte ainsi 11000 données (5 mots de passe $\times$ 10 captures $\times$ 2 façons de taper $\times$ 110 utilisateurs).
A partir des données collectées, nous définissons 4 traits de biométrie douce, répartis chacun en deux classes C_1 et C_2 :

— La façon de taper
C_1 = une main (droite ou gauche en fonction de la latéralité), C_2 = deux mains
— Le sexe
C_1 = homme, C_2 = femme
— L'âge
C_1 = moins de 30 ans, C_2 = plus de 30 ans
— La latéralité
C_1 = droitier, C_2 = gaucher

A partir de cette base de données biométriques de dynamique de frappe au clavier, nous présentons les méthodes que nous allons avons choisies pour prédire le profil de chaque utilisateur.

2.3 Apprentissage et fusion

Pour chaque utilisateur, nous choisissons de ne pas tenir compte des trois premières captures, pour chaque mot de passe : cela correspond au temps nécessaire pour apprendre à taper correctement chaque mot de passe et permet de supprimer des hésitations qui ne révèlent pas la véritable façon de taper.
Pour chaque caractéristique de biométrie douce, pour chaque mot de passe, un Séparateur à Vastes Marges (SVM) [10] est entraîné à reconnaître les deux classes C_1 et C_2. On utilise la bibliothèque LibSVM [11] avec un noyau gaussien radial (RBF : Radial Basis Function), et les valeurs de paramètres suivantes : $\gamma = 0.125$ le coefficient de pénalisation $C = 128$. Pour l'apprentissage, une partie de la base de données est conservée (entre 1% et 90%), le reste de la base sert aux tests et à l'évaluation de la performance. Plus il y a de données d'apprentissage, meilleures sera le taux de reconnaissance de chaque classe. Pour les classes déséquilibrées (hommes-femmes, droitiers-gauchers, voir la figure 2), on sélectionne au hasard dans la classe surnuméraire un nombre de données correspondant au nombre total de données de l'autre classe.

On répète 100 fois l'entraînement d'un SVM pour chaque pourcentage fixé de données d'apprentissage, et la performance de la classification est calculée comme le taux moyen de reconnaissance sur ces 100 essais.
Pour valider les résultats obtenus, des intervalles de confiance sont calculés (voir les références [12] et [13] pour plus de détails). L'équation (1) donne la formule de l'indice de confiance à 95% :

$$CI = m(taux) \pm 1.96 \frac{\sigma(taux)}{\sqrt{N}} \qquad (1)$$

où $m(taux)$ et $\sigma(taux)$ sont respectivement la moyenne et l'écart-type du taux de reconnaissance sur $N = 100$ itérations.

Pour améliorer les performances du système de biométrie douce, c'est-à-dire la reconnaissance de chaque trait de biométrie douce, un processus de fusion peut être appliqué. Les techniques de fusion des données sont des techniques classiques de reconnaissance de formes, très souvent utilisées en biométrie pour définir des systèmes multibiométriques [14], [15], [16]. Le principe est le suivant : pour chaque critère de biométrie douce, au lieu de considérer séparément le résultat de chaque classifieur (un par mot de passe $P1$ à $P5$), ces cinq résultats vont être fusionnés. Il existe plusieurs façons de fusionner les données issues de plusieurs systèmes biométriques : on peut appliquer une fusion de captures, de caractéristiques, de score, de rang, de décision (pour plus de détails, voir [15], [16]). Plus la fusion a lieu proche de la capture des données, plus la quantité d'information à fusionner est importante : de quelques Mo pour la fusion de captures (par exemple fusion de deux images d'empreintes digitales), à 1 bit pour la fusion de décision (décision finale = oui ou non, 1 ou 0...). Dans cet article, nous nous intéressons au vote majoritaire et à la fusion de scores, à partir de la décision prise par les 5 classifieurs et de la probabilité associée. Les deux processus de fusion sont illustrés à la figure 3.

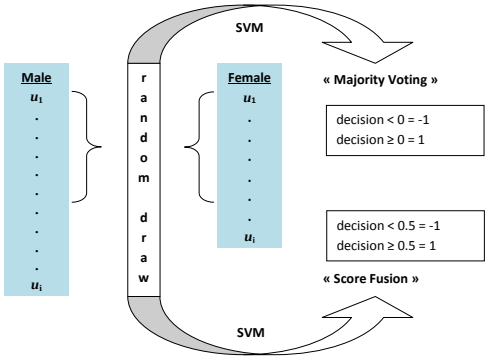

Figure 3 – *Fusion des données pour le critère* sexe

La partie suivante présente les résultats obtenus pour les différents processus d'apprentissage.

3 Résultats

3.1 Apprentissage séparé pour chaque mot de passe

On s'intéresse dans un premier temps à l'évolution du taux moyen (la moyenne est calculée sur 100 itérations) de reconnaissance de chaque critère de biométrie douce, pour chaque mot de passe. Les figures 4 à 7 montrent les résultats obtenus.

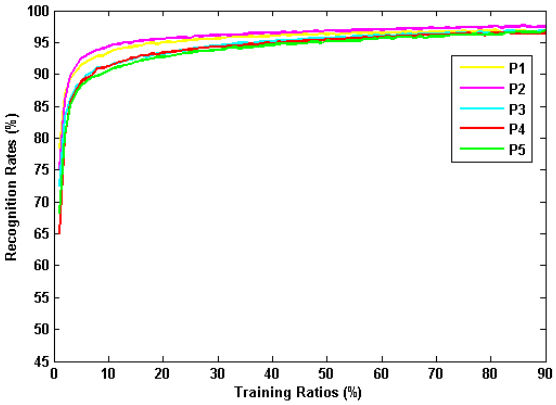

Figure 4 – *Taux de reconnaissance du nombre de mains utilisées*

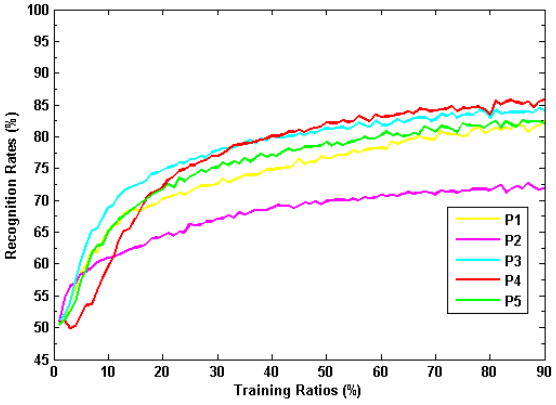

Figure 5 – *Taux de reconnaissance du sexe de l'utilisateur*

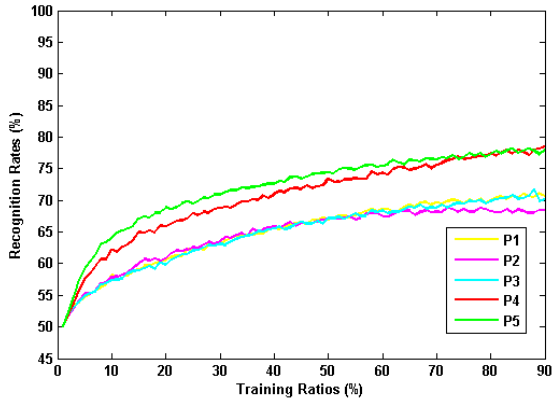

Figure 6 – *Taux de reconnaissance de la classe d'âge*

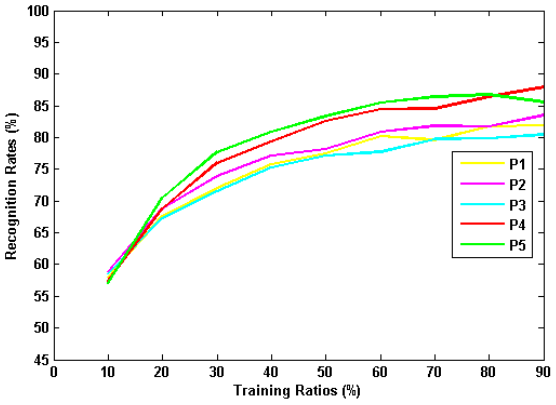

Figure 7 – *Taux de reconnaissance de la latéralité*

Les résultats obtenus sont également présentés dans le tableau 1 avec les intervalles de confiance correspondants, dans le cas où 50% des données disponibles sont utilisées pour la phase d'apprentissage. On constate que les performances sont comprises entre 63% et 96%.

3.2 Amélioration des performances par fusion

Dans cette partie, nous étudions l'influence de la fusion des 5 mots de passe sur le taux de reconnaissance de chaque critère de biométrie douce. Le tableau 2 présente les résultats obtenus pour les deux processus de fusion décrits dans la partie 2.3.
On constate que la fusion de score est plus efficace que la fusion de décision par vote majoritaire ce qui est tout à fait normal : les performances atteignent au maximum 100% de reconnaissance pour le nombre de mains utilisées, et au minimum 86% pour la catégorie d'âge.

4 Conclusion et perspectives

Cet article s'intéresse à la biométrie douce pour la dynamique de frappe au clavier. Les critères de biométrie douce ne sont pas suffisants pour authentifier un utilisateur, mais permettent d'en établir un profil. Ce profil peut ensuite être utilisé à de nombreuses fins : amélioration des performances de recherche dans une grande base de données, amélioration des performances d'un système biométrique classique : le profil vient renforcer la confiance dans le score de décision. En ce qui concerne la dynamique de frappe au clavier, les critères étudiés sont les suivants : nombre de mains utilisées, sexe, catégorie d'âge et latéralité. Les taux de reconnaissance obtenus en sortie des SVM entraînés pour chaque mot de passe sont compris entre 63% et 96%. La fusion des données disponibles pour les 5 mots de passe permet d'atteindre des performances au-delà de 86%. Nos travaux futurs concerneront l'exploitation des informations de profilage pour améliorer les performances d'un système d'authentification par DDF.

		Taux de reconnaissance et intervalle de confiance				
Critère	**Nombre de données**	P_1	P_2	P_3	P_4	P_5
Nb mains	770 par classe	$96\% \pm 0.1\%$	$96\% \pm 0.1\%$	$95\% \pm 0.1\%$	$94\% \pm 0.1\%$	$94\% \pm 0.1\%$
Sexe	224 par classe	$74\% \pm 0.3\%$	$69\% \pm 0.3\%$	$70\% \pm 0.2\%$	$78\% \pm 0.2\%$	$76\% \pm 0.2\%$
Age	357 par classe	$64\% \pm 0.2\%$	$64\% \pm 0.2\%$	$63\% \pm 0.2\%$	$69\% \pm 0.2\%$	$69\% \pm 0.2\%$
Latéralité	84 par classe	$72\% \pm 1.2\%$	$73\% \pm 1.2\%$	$72\% \pm 1.2\%$	$72\% \pm 1.3\%$	$73\% \pm 1.2\%$

Tableau 1 – *Taux de reconnaissance avec intervalles de confiance pour un apprentissage avec* 50% *des données*

		By fusing	
Critère	**Sans fusion**	**Vote majoritaire**	**Fusion de score**
Nb mains	94%	100%	100%
Sexe	63%	86%	92%
Age	55%	87%	86%
Latéralité	62%	85%	92%

Tableau 2 – *Comparaison des performances sans et avec fusion pour un apprentissage avec* 50% *des données*

Références

[1] A. Jain, S. Dass, et K. Nandakumar. Soft biometric traits for personal recognition systems. *Biometric Authentication*, pages 1–40, 2004.

[2] Antitza Dantcheva, Carmelo Velardo, Angela D'angelo, et Jean-Luc Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2) :739–777, 2011.

[3] L.C.F. Araujo, Jr. Sucupira, L.H.R., M.G. Lizarraga, L.L. Ling, et J.B.T. Yabu-Uti. User authentication through typing biometrics features. *Signal Processing, IEEE Transactions on*, 53(2) :851–855, 2005.

[4] Daniele Gunetti et Claudia Picardi. Keystroke analysis of free text. *ACM Trans. Inf. Syst. Secur.*, 8(3) :312–347, 2005.

[5] R. Giot, M. El-Abed, et C. Rosenberger. Greyc keystroke : A benchmark for keystroke dynamics biometric systems. Dans *Biometrics : Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, 2009.

[6] S.Z.S. Idrus, E. Cherrier, C. Rosenberger, et P. Bours. Soft biometrics database : A benchmark for keystroke dynamics biometric systems. Dans *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*, 2013.

[7] M. Bertacchini, C. Benitez, et P. Fierens. User clustering based on keystroke dynamics. Dans *In XVI Congreso Argentino de Ciencias de la Computacion (CACIC 2010)*, 2010.

[8] Romain Giot et Christophe Rosenberger. A new soft biometric approach for keystroke dynamics based on gender recognition. *International Journal of Information Technology and Management*, 11(1) :35–49, 2012.

[9] Romain Giot, Alexandre Ninassi, Mohamad El-Abed, et Christophe Rosenberger. Analysis of the acquisition process for keystroke dynamics. Dans *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–6. IEEE, 2012.

[10] V. Vapnik. *Statistical learning theory*. Wiley New York, 1998.

[11] C.-C. Chang et C.-J. Lin. Libsvm : A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3), 2011.

[12] A. Mayoue. Biosecure tool - performance evaluation of a biometric verification system. `http://svnext.it-sudparis.eu/svnview2-eph/ref_syst/Tools/PerformanceEvaluation/doc/howTo.pdf`, 2007.

[13] R.M. Bolle, N.K. Ratha, et S. Pankanti. Evaluation techniques for biometrics-based authentication systems. Dans *15th Internat. Conf. Pattern Recogn.*, 2000.

[14] A. Ross et A. Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13), 2003.

[15] A. Ross, K. Nandakumar, et A. Jain. *Handbook of multibiometrics*, volume 6. Springer, 2006.

[16] Romain Giot, Baptiste Hemery, Estelle Cherrier, et Christophe Rosenberger. La multibiométrie. Dans *Traitement du signal et de l'image pour la biométrie*, page Chapitre 9. Hermès, 2012.